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.

Biasing switching outcomes in the *Escherichia coli fim* site-specific recombination system through DNA supercoiling and nucleoid-associated proteins

By

Colin P. Corcoran

A dissertation presented for the degree of Doctor of Philosophy, in the faculty of Science, University of Dublin, Trinity College

2010

Declaration

I, Colin Peter Corcoran, am the sole author of this thesis. The work presented herein represents my own work, except where duly acknowledged in the text, and has not been previously presented for a higher degree at this or any other University.

I agree that this thesis may be lent or copied at the discretion of the Librarian, Trinity College Dublin.

Colin Peter Corcoran

Summary

Type 1 fimbriae are typically expressed in nutrient poor environments and facilitate colonization through attachment when the bacterium can no longer support a motile lifestyle that demands a high metabolic flux. They play a key role in the colonization of various host tissues by *E. coli* and in biofilm formation on abiotic surfaces such as catheters. The expression of type 1 fimbriae is phase variable and controlled by the site-specific inversion of a DNA segment (*fimS*), which contains the promoter for the structural components of type 1 fimbriae. In bacteria growing in mid-exponential phase at 37°C in LB broth, the FimB protein inverts the *fimS* element from the on phase to the off phase and back again at approximately equal rates. However, when novobiocin is used to inhibit DNA gyrase activity, allowing DNA to become more relaxed, the FimB-catalyzed *fimS* inversion reaction adopts a pronounced bias in favour of the on orientation. Previously, the leucine-responsive regulatory protein (LRP) was identified as playing a role in determining the directionality in this reaction. Specifically, the LRP protein is required to maintain the *fimS* element in the on orientation and it does this by binding to the LRP-1 and LRP-2 sites within *fimS*. In the present study, the IHF protein bound to the IHF-1 site was identified as a second recombination directionality factor (RDF) that is required to maintain *fimS* in the on orientation when gyrase activity is inhibited. The location of the IHF-1 binding site in the non-inverting chromosomal DNA immediately adjacent to IRL explains the differential interaction of IHF with LRP bound within *fimS* because all three sites are only juxtaposed when the switch is in the phase-on orientation.

Previous work has suggested a role for the H-NS protein in influencing the FimB-mediated inversion of *fimS*. Careful examination of H-NS interactions with *fimS* and its flanking regions showed that H-NS interaction with the switch was contingent on *fimS* orientation. This study showed that the H-NS binding site in the *fimS* region is distributed across three incomplete sites; one of these is in the invertible *fimS* element and is adjacent to the P_{fimA} promoter and the others are located in the non-inverting parts of the chromosome immediately outside the inverted repeats (IRs) that flank *fimS*. The absence of H-NS causes a dramatic increase in the percentage of phase-on cells in a population indicating that H-NS binding overlapping the left IR (IRL) in the phase-off orientation prevents switching from phase-off to phase-on.

A major finding of this study was that the ubiquitous reporter gene *gfp*, which encodes the Green Fluorescent Protein, was bound by the global repressor protein H-NS. H-NS binding in *gfp* caused a repression of transcription *in vivo* and thus previous studies utilizing *gfp* may be complicated by added H-NS regulation. In this study, a highly fluorescent, fast-folding variant of *gfp* (*gfp+*) was reverse-engineered to reduce H-NS binding affinity without altering the amino acid sequence of the protein. This new *gfp*, gfp^{TCD}, is bound with lower affinity by H-NS and does not repress local transcription.

The reverse engineering method for reducing H-NS affinity for a region of DNA was developed as part of this study and was the subject of a patent application. This novel method can be applied to any DNA, including other fluorescent proteins

Acknowledgments

I would like to thank Charlie for providing me with the opportunity to do this PhD and for all of his helpful advice and guidance during its course. I would also like to thank my committee members Tim, Jay, Stephen and Cyril for all their constructive comments and advice. A special thanks to all the past and present members of the Dorman lab who have been so generous with their time and from whom I've learned so much and to Joan for all her helpful advice on writing a patent.

To my fellow PhD students (Fi, Fitzy, Kel, Heather, Alex, Finian…); thanks for providing such a solid platform of alcoholic support…just what's needed when things aren't working!! I won't forget the look of resignation in the managers eyes when they'd see us walking into the Porterhouse…they knew straight away it was gonna be a busy night. I'll miss all the crazy dancing and the 'lifts' but I'm sure the bouncers won't! I think the SGM in Cork has to get a special mention…the Moyne hit that like a ton of bricks. Only Moyners could take a nice quiet evening and turn it into such drunken fun…up singing on the stage, irish/dirty-dancing in front of the whole conference. Worked out well though-Fi's dancing impressed the organisers so much they gave her a prize! :) Thanks to Fitzy for my recent addiction to chilli. To Shano for being my rock. Well, not really…but what a trip to Madison. Think that's all I should say about that.

Thanks to the lads from home (Granty, Stevo, Cod, Dermo, Behan and Mono) for always makin sure there was a session planned in Portmarnock; a great way to get away from the stresses of college life and to Shea for always havin some random last-minute adventure ready to go. Thanks to all the climbin crowd (Andy, Andy Locheadan, Alex, Heather and Paul) for all the long relaxin hours at the climbin wall and the unbelieveable trip to the Burren; climbin from dawn to dusk in such a spectacular place…I won't forget it anytime soon. Thanks to Andy and Alex for a great trip to Spain; great climbin and even better mojitos…the perfect way to unwind when finishing up your thesis! Thanks to my Hurling managers Jackie, Frank and Tony for getting me back into the game in the last year; hopefully we'll be celebrating next week.

Thanks to my parents and brothers for their constant support and to my Granda who passed away recently.

Most importantly I have to mention my long-suffering girlfriend Katie, who apart from the usual effort of putting up with me has had the added inconvenience of putting up with me writing a thesis. I couldn't have done it without her.

Publications

Full papers

1. **Corcoran CP, CJ Dorman. 2009.** DNA relaxation dependent phase biasing of the *fim* genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP. *Molecular Microbiology* **74:** 1071-1082.

2. **Dorman CJ, CP Corcoran. 2009.** Bacterial DNA topology and infectious disease. *Nucleic Acids Research* **37:** 672-678.

3. **Corcoran CP, ADS Cameron, CJ Dorman. 2010.** H-NS silences *gfp,* the Green Fluorescent Protein gene: gfp^{TCD} is a genetically remastered *gfp* gene with reduced susceptibility to H-NS-mediated transcription silencing and with enhanced translation. *Journal of Bacteriology* **192:** 4790-4793.

Published conference abstracts

1. **Corcoran CP, CJ Dorman. 2008.** Biasing switching outcomes in the *Escherichia coli fim* site-specific recombination system through DNA supercoiling and nucleoid-associated proteins (NAP). 163rd meeting of the Society for General Microbiology, Trinity College Dublin, Ireland.

2. **Corcoran CP, Dorman CJ. 2009.** DNA supercoiling and nucleoid-associated proteins (NAPS) bias DNA inversion directionality in the *Escherichia coli fim* site-specific recombination system. Phage Conference, Madison, Wisconsin, USA.

3. **Dorman CJ, Guadarrama S, Corcoran CP. 2009.** Global and local gene regulation in Gram-negative bacteria. Current Trends in Biomedicine Workshop on *Bacterial Regulatory Networks*, Univerisidad Internacional de Andalucia, Baeza, Spain.

Patent application

1. **Dorman CJ and Corcoran CP. 2010.** Patent application PCT/EP2010/057856 *A Method for Improving Gene Expression.* European Patent Office.

Table of Contents

List of Figures

List of Tables

Chapter 1.

General Introduction.

1.1 Overview

Pathogenic bacteria, including pathovars of the Gram-negative bacterium *Escherichia coli*, display an intricate array of surface adhesins that allow colonization of a range of niches. Surface adhesins are displayed temporally and their expression often determines the tropism of the bacteria. The role of type 1 fimbriae in virulence of uropathogenic *E. coli* (UPEC) is well established. Adhesion by type 1 fimbriae to mono-mannose residues in the bladder epithelium is essential for the internalization of UPEC, which results in formation of intracellular bacterial communities that aid bacterial persistence and recurrent infection. Type 1 fimbriae are therefore important determinants of virulence. The expression of type 1 fimbriae is phase variable and controlled by the site-specific inversion of a DNA segment (*fimS*), which contains the promoter for the structural components of type 1 fimbriae. Phase variation of type 1 fimbriae often results in the bacterial population containing a mixture of fimbriate and afimbriate cells. If the fimbriation is advantageous to the particular subset of bacteria they will proliferate in preference to the non-expressing subset. However, if the fimbriation is disadvantageous that particular subset will not proliferate or may be eradicated, while the non-expressing subset of the population may survive. Heterogeneity in the population is therefore an important determinant of bacterial persistence. Biochemical processes that are dependent on infrequent molecular events involving small numbers of molecules, such as site-specific recombination, are inherently stochastic (random) events. Intrinsic factors control the rate and direction of *fimS* inversion but the concentrations and activities of these intrinsic factors can be altered in response to environmental changes to increase the probability of *fimS* adopting a specific orientation.

The aim of this chapter is to provide a review of the literature relevant to this study. The structure of type 1 fimbriae and the organisation of the *fim* locus are outlined. The phase variation of type 1 fimbriae is predominantly under the control of two site-specific recombinases, FimB and FimE. These two proteins and their role in inversion of the *fimS* are detailed. Inversion of *fimS* requires and is regulated by a number of nucleoid-associated proteins (NAPs) including Integration Host Factor (IHF), the Leucine Responsive Regulatory Protein (LRP) and the nucleoid associated protein H-NS. These proteins and how they govern the regulation of type 1 fimbrial expression form an important part of this study and are thus described in detail. DNA supercoiling and its role in regulation of *fimS* is

also a key element of this study and are described in detail. Other regulatory inputs that are important for the understanding of type 1 fimbrial gene regulation are described.

1.2 Type 1 fimbriae.

Type 1 fimbriae are adhesive organelles and were the first fimbriae to be described (Brinton, 1959). Type 1 fimbriae, also referred to as type 1 pili, are important virulence determinants expressed in approximately 70% of *E. coli* strains as well as in most members of the *Enterobacteriaciae* family (Kuehn *et al.*, 1994; Soto and Hultgren, 1999). They also contribute to virulence in pathogenic Gram-negative bacteria (Kaper *et al.*, 2004). Interaction between type 1 fimbriae and receptor structures has been shown in a number of studies to play a key role in the colonization of various host tissues by *E. coli* (Yamamoto *et al.*, 1990; Bloch *et al.*, 1992) and in biofilm formation on abiotic surfaces such as catheters (Pratt and Kolter, 1998; Schembri and Klemm, 2001a). They play a particularly important role in colonization of the bladder by uropathogenic *E. coli* (UPEC) (Iwahi *et al.*, 1983; Hultgren *et al.*, 1986; Connell *et al.*, 1996). Bacteria in the urinary tract experience severe hydrodynamic shear forces that, without attachment to host cells, would result in rapid clearance (Schembri and Klemm, 2001a). Type 1 fimbriae, like many surface exposed proteins, are highly immunogenic and thus phase variation is important to prevent clearance through strong induction of the innate immune response (Kaper *et al.*, 2004). Type 1 fimbriae are typically expressed in nutrient poor environments and facilitate colonization through attachment when the bacterium can no longer support a motile lifestyle that demands a high metabolic flux (Müller *et al.,* 2009). Expression of adhesins such as type 1 fimbriae are inversely regulated with expression of flagella, which are required for motility (Pesavento and Hengge, 2009). Binding of type 1 fimbriae to host cells contribute to biofilm formation, which allows for efficient nutrient scavenging, provides increased resistance to host immune responses and antibiotics, and provides a reservoir for persistent infections (Costerton *et al.*, 1999; Mulvey *et al.*, 2001; Wright *et al.*, 2007). Type 1 fimbriae are involved in adherence to mucosal and phagocytic cells (Connell *et al.*, 1996). A typical type 1 fimbriated bacterium has 200–500 peritrichously arranged fimbriae on its surface (Klemm and Christiansen, 1990). Each fimbria consists of two distinct components; the major component is a rod containing ~3000 units of the main fimbrial subunit FimA. The second component thin fibrillium contains the adhesin FimH. The structure, transport and assembly of type 1 fimbriae are illustrated in Fig. 1.1.

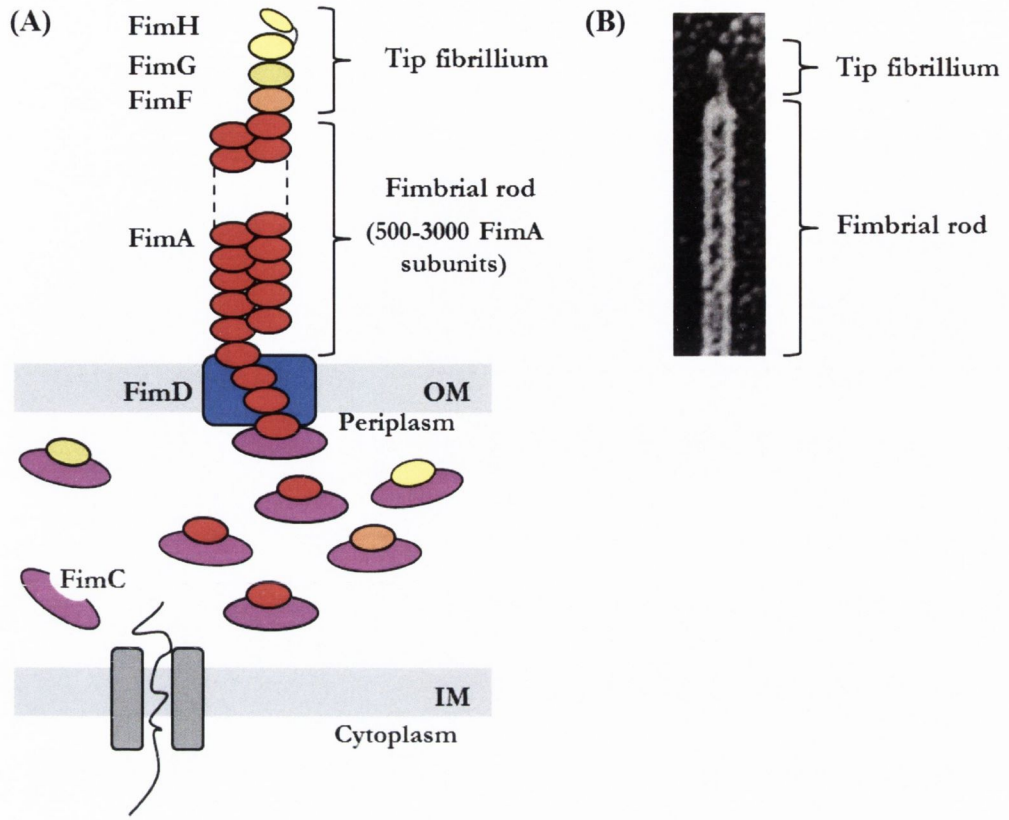


Fig. 1.1 Structure and assembly of type 1 fimbriae. (A) Schematic representation of type 1 pilus assembly according to the chaperone–usher pathway. Structural subunits (FimA, FimF, FimG, FimH) enter the periplasm in an unfolded conformation. They subsequently form 1:1 complexes with the chaperone FimC, which folds the structural subunits and delivers them to the usher FimD. The structural subunits pass through the FimD pore and are incorporated into the growing pilus. (B) High-resolution electron micrograph revealing the composite structure of a type 1 fimbria (Jones *et al.*, 2005). IM, inner membrane; OM, outer membrane

1.2.1 Main fimbrial subunit (FimA)

FimA is a 17-kDa protein that forms a right-handed helix with an average of 3.125 FimA subunits per turn. The rod has a length of 1-2μM, diameter of ~70 Å (7 nm), a central pore of about 20 to 25 Å, and a pitch of about 24 Å (Brinton, 1965; Russell and Orndorff, 1992; Hahn *et al.*, 2002).

1.2.2 Determinant of binding specificity (FimH)

The minor component of the fimbriae is a thin fibrillum that extends ~ 16 Å from the tip of the rod and contains the binding specificity determinant for type 1 fimbria, FimH (Jones *et al.*, 1995). FimH is also interspersed along the fimbrial shaft, although only FimH proteins exposed on the tip are involved in receptor binding (Ponniah *et al.*, 1991). FimH contains a lectin-like domain (Choudhury *et al.*, 1999) that confers binding specificity of type 1 fimbriae to mannosylated glycoproteins (Krogfelt *et al.*, 1990). FimH adhesion to mannose residues dramatically increases under shear stress conditions by a catch-bond mechanism (Tchesnokova *et al.*, 2008; Yakovenko *et al.*, 2008), whereby under increasing shear conditions, a mannose binding pocket of the lectin domain of FimH adopts a new conformation, relative to the pilin anchored domain of FimH, that increases FimH binding affinity. This switches binding from rolling bacterial adhesion to strong binding with cessation of bacterial movement and prevents clearance by urination (Thomas *et al.*, 2002). While FimH is highly conserved among members of the *Enterobactereciae* (Abraham *et al.*, 1988), minor mutations in the protein can lead to significant functional heterogeneity (Sokurenko *et al.*, 1994; Schembri *et al.*, 2000; Sokurenko *et al.*, 2004). The affinity of FimH variants toward mannose targets can vary due to changes in the primary structure of FimH.

E. coli isolated from faeces typically express FimH that is capable of binding only to tri-mannose residues receptors. In contrast, the FimH adhesins present in urinary tract isolates carry minor mutations (compared to the faecal isolates) that enhance their ability to recognize mono-mannose receptors, which are abundant in the urinary tract (Sokurenko *et al.*, 1995). Strains with FimH alleles that allow binding to mono-mannose residues have a significantly higher tropism for the uroepithelium and are correspondingly more virulent (Sokurenko *et al.*, 1998; Sokurenko *et al.*, 2004). FimH variants (e.g. FimH-Arg66) that have unusually strong mono-mannose binding affinity are cleared from the urinary track

more rapidly than the congenic strain, due to reduced resistance to shearing conditions caused by the alteration in FimH (Weissman *et al.*, 2007).

Other mutations in the FimH adhesion can lead to production of fimbriae with a twisted appearance that promote clumping of bacterial cells (Schembri *et al.*, 2001) in a molecular handshake mechanism similar to Ag43 mediated auto-aggregation (Danese *et al.*, 2000; Kjaergaard *et al.*, 2000).

The agglutination assay is a standard diagnostic test for the expression of type 1 fimbriae. It involves incubation of bacterial cells with eukaryotic cells that express mannosylated glycoproteins, such as erythrocytes (Salit and Gotschlich, 1977) or yeast (*Saccharomyces cerevisae*) cells (Eshdat *et al.*, 1981). Binding of the type 1 fimbriae to receptors on the eukaryotic cells causes agglutination. This can be inhibited by the addition of exogenous mannose.

1.2.3 Minor components (FimF and FimG)

Other components of type 1 fimbriae include FimF and FimG, which form part of the fibrillar tip (Fig. 1.1) (Hanson and Brinton, 1988; Choudhury *et al.*, 1999). FimF (16 kDa) and FimG (14 kDa) are not essential for the production of functional type 1 fimbriae but mutants in these loci produce fimbriae with altered length and abundance (Russell and Orndorff, 1992). Strains containing *fimG*, *fimH* and double knockout mutants are capable of agglutinating guinea pig erythrocytes (Russell and Orndorff, 1992). The absence of FimF reduces the number of fimbriae per cell, whereas the absence of FimG caused causes production of extra-long fimbriae (Russell and Orndorff, 1992). The double mutant displayed the characteristics of both single mutants. On the basis of the phenotypes exhibited by the single and double mutants, it was proposed that FimF aids the initiation of fimbrial assembly and that FimG inhibits fimbrial polymerization (Russell and Orndorff, 1992).

1.2.4 Fimbrial assembly proteins (FimC and FimD).

The assembly machinery is comprised of two specialized classes of proteins, the periplasmic chaperone FimC, which contains two immunoglobulin-like domains separated by a deep cleft, and the usher protein FimD that forms a pore in the outer membrane (Thanassi *et al.*, 1998) required for type 1 pilus assembly (Klemm and Christiansen, 1990).

The structural subunits (FimA, FimF, FimG and FimH) aggregate as monomers in solution and cannot fold into a native three-dimensional structure. The steric information required

for folding is donated to the hydrophobic core of the subunit by the FimC (Choudhury *et al.*, 1999). Once transported through the outer membrane the fimbrial components self-assemble. The FimCH chaperone–adhesin complex binds to the FimD usher (Dodson *et al.*, 1993) faster and with higher affinity than the other chaperone-subunit complexes (Saulino *et al.*, 1998). The kinetic dissociation rates of all of the chaperone complexes are relatively slow. These kinetics combined result in the early presentation of FimCH to FimD, which is a critical factor positioning the FimH adhesin at the tip of the pilus (Saulino *et al.*, 1998).

1.3 The *fim* operon

In *E. coli*, the genes involved in the regulation, synthesis and assembly of type 1 fimbriae are arranged in a nine gene, tri-cistronic operon (Swaney *et al.*, 1977) located between minutes 97.81 and 97.98 of the linkage map of *E. coli* K-12 (Fig. 1.2) (Brinton *et al.*, 1961; Burland *et al.*, 1995; Blattner *et al.*, 1997). Fimbrial gene expression is controlled through inversion of a 314-bp invertible element (*fimS*) containing the promoter for the major fimbrial subunit *fimA* (Abraham *et al.*, 1985). Type 1 fimbriae are encoded by the genes *fimA, fimI, fimC, fimD, fimF, fimG and fimH*, which are transcribed from the *fimA* promoter (P_{fimA}) (Abraham *et al.*, 1985; Burland *et al.*, 1995). The control of fimbrial expression is mediated by two site-specific recombinases FimB and FimE, encoded by genes located immediately upstream of the invertible element (Klemm, 1986).

1.4 Site-specific recombination.

Site-specific recombination is required for many important biological processes including bacterial genome replication, conjugative transposition of plasmid DNA, resolution of catenated DNA circles, integration and excision of phage DNA and virulence, through phase or antigenic variation of surface components (Hallet and Sherratt, 1997). Site-specific recombination involves DNA breakage and reunion of DNA and requires no DNA synthesis or high-energy cofactor (Grindley *et al.*, 2006). Unlike general recombination that requires large segments of homologous DNA, site-specific recombinases catalyse recombination between substrates that share limited sequence identity (Grindley *et al.*, 2006).

The minimal requirements for tyrosine recombinase-catalysed recombination are two DNA segments termed "core sites", which contain binding sites for the recombinases separated

by repeat regions where the strand exchange takes place (often referred to as a spacer). This arrangement allows the binding of four recombinase molecules that when brought in close proximity, can catalyse the recombination reaction (Fig. 1.3). The sequence identity usually extends over the short strand exchange region and flanking recombinase binding sites and is critical to the recombination reaction in most cases studied (Rajeev *et al.*, 2009). The outcome of a recombination event depends on the relative orientation and location of the two spacers. Intra-molecular recombination between inverted sites, as is the case with *fim*, causes the inversion of the intervening DNA. Intra-molecular recombination between directly repeated sites, as is the case with lambda phage, causes the excision of the intervening DNA. Recombination between sites on separate DNA molecules will integrate one molecule into the other (Hallet and Sherratt, 1997). Bacteriophages, such as phage lambda, use this cut-and-paste mechanism to insert their own genomes into the genome of a bacterial host (Nash, 1981). Some reactions, including Fim recombinase mediated DNA inversion, require accessory proteins to promote juxtaposition of the spacer regions (termed synapsis; see sections 1.8.2 and 1.8.4).

1.5 The serine and tyrosine site-specific recombinases

Site-specific recombinases can be divided into two families according to the nucleophile that is used for cutting and rejoining the defined DNA segments: the tyrosine recombinase family (also known as the lambda integrase (Int) family) and the serine recombinase family (also known as the resolvase/invertase family) (Stark *et al.*, 1992; Grindley *et al.*, 2006). The lamda-integrase family, to which the Fim recombinases belong, can be identified by the conserved Arg-His-Arg-Tyr catalytic tetrad, of which the tyrosine residue performs the important catalytic role (Esposito and Scocca, 1997; Grainge and Jayaram, 1999). The family of tyrosine site-specific recombinases contains over 150 members from all domains of life (Nunes-Duby *et al.*, 1998). Some of the best-studied members of the family are phage lambda integrase (Int) and the Flp, Cre, and Xer recombinases.

While the enzymes from both families show significant functional overlap, the recombination mechanisms of the two families of enzymes are distinctly different. Tyrosine recombinases break and re-join single strands in pairs to form a Holliday junction intermediate Fig. 1.3. By contrast, serine recombinases cut all strands in advance of strand exchange and religation (Grindley *et al.*, 2006). Despite the difference in mechanism, enzymes from either group perform similar functions. For example, the Hin serine

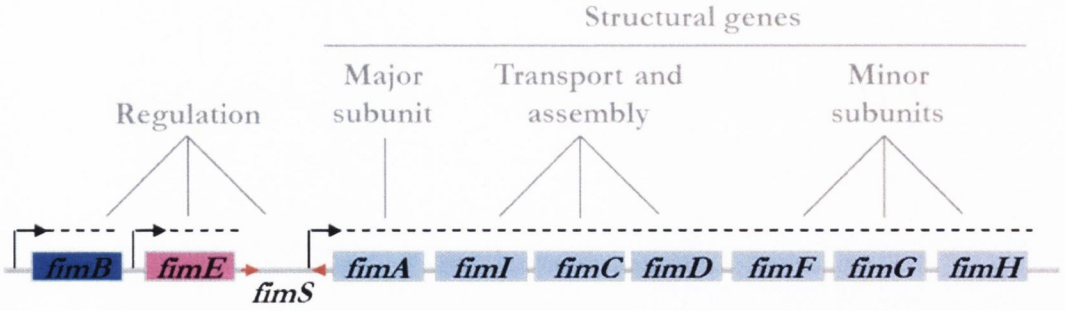


Fig. 1.2 Arrangement of the *fim* operon. The genes encoding type 1 fimbriae are arranged in an operon. The 7 structural genes are located downstream of the P_{fimA}, which is located on an invertible element *fimS*. Two recombinases, encoded by *fimB* and *fimE*, control *fimS* inversion of *fimS* by binding to sites flanking the 9-bp inverted repeats (red triangles) and catalysing strand exchange.

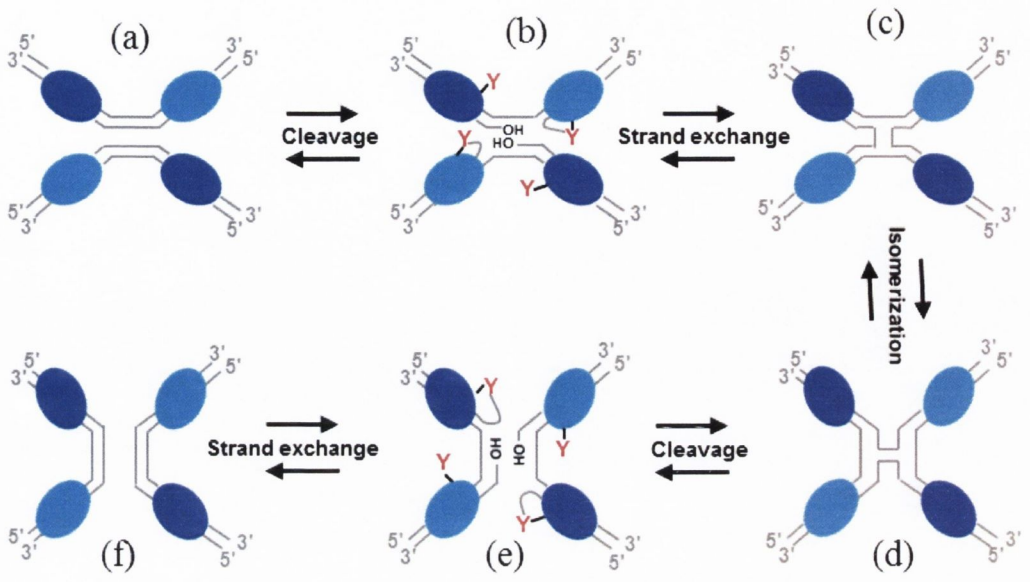


Fig. 1.3 The Cre–*loxP* site-specific recombination pathway. Current model of the Cre–*loxP* site-specific recombination mechanism. Two Cre monomers (blue or light blue ovals) bind to recombination sites flanking the strand exchange region and promote synapsis of the DNA substrates through protein–protein interactions (a). One monomer bound on each DNA duplex cleaves a phosphodiester bond on the adjacent DNA (*cis* cleavage) using a conserved tyrosine side-chain (indicated by Y) to form a covalent 3' phosphotyrosine linkage (b). The liberated 5' hydroxyl groups (OH) can either attack the adjacent phosphotyrosine linkage to return to the starting configuration or attack the partner substrates to exchange one pair of strands and form a Holliday junction intermediate (c). Isomerization of the Holliday intermediate (d) alters the position of the catalytic tyrosines (Y) deactivating the first pair of monomers and activating the second pair, allowing cleavage (e) and exchange (f) of the second pair of DNA strands to give recombinant products. This model is believed to be a general mechanism for tyrosine site-specific recombinases (van Duyne *et al.*, 2005).

recombinase and the Fim recombinases both control surface protein expression by site-specific DNA inversion (Simon *et al.*, 1980; Klemm, 1986).

1.5.1 Mechanism for tyrosine site-specific recombination.

The availability of crystal structures for a number of tyrosine site-specific recombinases in complex with their substrates has provided an understanding of how tyrosine site-specific recombinases use a common reaction pathway to mediate recombination in a range of physiological processes (Grainge and Jayaram, 1999). The interaction of the Cre recombinase from bacteriophage P1 with its target sites (*loxP*) is the best characterized recombination event in the Int family. Several structures of intermediates in the Cre-*loxP* pathway have been solved in recent years, which provided a detailed mechanism of action (Guo *et al.*, 1997; Gopaul *et al.*, 1998; Guo *et al.*, 1999) (Figure). Structural insights gained from the crystal structures of Int (Aihara *et al.*, 2003; Biswas *et al.*, 2005), XerD (Subramanya *et al.*, 1997) and FLP (Chen *et al.*, 2000) indicate that the Cre-*loxP* pathway is a general recombination pathway for tyrosine site-specific recombinases (Van Duyne, 2001).

In the Cre-*loxP* model, strand exchange proceeds by a strand-swapping mechanism similar to that proposed by Nunes-Duby (1995) (Fig. 1.3) (Nunes-Duby *et al.*, 1995). Cre binds to its recognition element by forming a 'C-shaped clamp' around the DNA duplex. The 13 bp recombinase-binding element (RBE) in the Cre–DNA structure is bent only slightly as a result of Cre–DNA contacts, but the full recombination site (34 base pairs) is bent sharply in the strand exchange (spacer) region between the half-sites (Van Duyne, 2001). The spacer regions are aligned (synapsed) in an antiparallel fashion. The recombination reaction is then performed by a tetramer of the recombinase protein in two spatially and temporally distinct steps (Fig. 1.3). In the first step, two strands in different duplexes at one end of the exchange region (spacer) are cleaved by the catalytic tyrosine, which forms a hydrogen bonded to the adjacent phosphate, that can be attacked by either the free hydroxyl group on the same or opposing DNA duplex (Fig. 1.3b). Attack by the *trans* hydroxyl group leads to strand exchange and formation of a Holliday junction (Fig. 1.3c). Cleavage of a phosphodiester bond of the DNA backbone by the recombinase monomer bound adjacent to it is termed *cis* cleavage. Conversely, cleavage by the monomer bound on the opposing DNA duplex is called *trans* cleavage. Int, Cre, XerD and Flp all cleave DNA in *cis*, although some tyrsosine-integrase recombinases do act in *trans* (Van Duyne, 2002).

The second step involves isomerization of the recombination complex to move the tyrosine of the previously inactive recombinase monomers into an ideal position for cleavage (Fig. 1.3d). This allosteric change moves the tyrosine of the monomers that catalysed the initial strand cleavage step into an unreactive position. The second strand cleavage and exchange step is performed in this new conformation (Fig. 1.3e), after which both strands of the DNA duplex have been exchanged and the recombination reaction is complete (Fig. 1.3f). Some recombination reactions, including the recombination of *fimS*, require the binding of accessory factors that alter DNA topology to promote interactions between the two recombinase dimers and thus promote formation of the recombination complex (see section 1.8).

1.6 Site-specific recombinases FimB and FimE

Fimbrial gene expression in *E. coli* is controlled through inversion of a 314-bp invertible element (*fimS*) containing the promoter for the major fimbrial subunit *fimA* (Eisenstein, 1981; Abraham *et al.*, 1985). Inversion of *fimS* leads to either fimbriate (phase on) or afimbriate cells (phase off) cells. Inversion of *fimS* is catalysed by two integrase-like tyrosine site-specific recombinases, FimB and FimE, which are encoded by genes located immediately upstream of *fimS* (Fig. 1.2) (Klemm, 1986). These proteins are small (FimB 25-kDA; FimE 23.5-kDa), basic, and are highly homologous at the level of amino acid sequence (48% identical) (Klemm, 1986), but have distinct activities. As suggested by the family name, the catalysis of DNA inversion by a member of the integrase family is an unusual feature of the Fim recombinases (Nunes-Duby *et al.*, 1998).

The FimB protein inverts the switch in the ON-to-OFF (resulting in afimbriate cells) and the OFF-to-ON direction (resulting in fimbriate cells) with approximately equal efficiencies at a rate of $\sim 10^{-3}$ inversions per cell per generation, whereas FimE inverts it predominantly in the ON-to-OFF direction at the much higher rate of up to 0.7 inversions per cell per generation (Blomfield *et al.*, 1991a; Gally *et al.*, 1993). As a result, when FimB and FimE are co-expressed in bacteria growing in Luria–Bertani at 37°C, the switching preference of FimE predominates and most bacteria in the population are in the OFF phase for fimbrial expression (McClain *et al.*, 1991). Both recombinases act by binding at a half-site internal to *fimS* at the border of the switch, and a second half site in proximal DNA external to *fimS*. Inversion of *fimS* involves DNA cleavage, strand exchange and religation reactions within the 9-bp left and right inverted repeats (IRL and IRR) that flank *fimS* (Dove and Dorman,

1996; Gally *et al.*, 1996; Kulasekara and Blomfield, 1999; Burns *et al.*, 2000; Holden *et al.*, 2007; McCusker *et al.*, 2008). Since only the 9-bp repeats, and not the flanking half-sites are identical both recombinases have varying binding affinities for the 4 half-sites. For example, FimE binds poorly to the combination of half sites that flank the IRL in the phase off orientation, which contributes to the low rate of inversion by FimE from phase off to phase on (Gally *et al.*, 1996; Kulasekara and Blomfield, 1999).

The Fim recombinases are remarkably tolerant of sequence changes in *fimS*. Recent studies have shown that extensive rearrangements of the external and internal half sites (Holden *et al.*, 2007; McCusker *et al.*, 2008) and complete replacement of the *fimS* region with unrelated DNA (Holden *et al.*, 2007) are tolerated by the recombinases, and in some cases even increases the inversion frequency.

Mutants deficient in FimE arise frequently and in those cases, FimB-alone catalyses *fimS* inversion (Blomfield *et al.*, 1991b).

1.7 Control of directionality in lambda Int-mediated integration and excision.

Bacteriophage lambda integrase (Int) is a tyrosine site-specific recombinase that mediates lambda phage integration into and excision out of the bacterial chromosome (Landy, 1989). This involves site-specific recombination between phage and chromosome encoded core sites (Nash, 1981; Richet *et al.*, 1988; Kim and Landy, 1992; Segall and Nash, 1993).

The product of integrative recombination between the viral core site (*attP*) and chromosomal core site (*attB*) is an integrated prophage bordered by hybrid *att* sites (*attL* and *attR*), which are themselves substrates for excisive recombination (Van Duyne, 2005).

The Int-system has a highly evolved mechanism for controlling the recombination directionality, which includes a phage encoded excision enhancing protein (Xis) and host encoded proteins IHF and the factor for inversion stimulation (FIS). Xis binds site-specifically as a homodimer to DNA and introduces a bend of >140° in DNA (Thompson and Landy, 1988; de Vargas and Landy, 1991). Binding of Xis alters DNA topology and the occupancy of Int and IHF binding sites, which promotes excision and inhibits integration (Better *et al.*, 1983; de Vargas and Landy, 1991). FIS can substitute for Xis and promote excision by binding to a site overlapping the site for one Xis protomer (Thompson *et al.*, 1987; Ball and Johnson, 1991a; b). FIS can introduce a bend of 50° to 90° depending on flanking sequences (Pan *et al.*, 1994; Pan *et al.*, 1996). This DNA bending activity may facilitate formation of the excissive nucleoprotein complex in the absence of Xis. FIS also

promotes excision by aiding Xis binding (Papagiannis *et al.*, 2007). IHF binds at 3 sites (H1, H2 and H') in phage lambda and can promote Int interaction with low affinity sites flanking the core regions (De Vargas *et al.*, 1989a). IHF binding to H1 is required for integrative but not excissive recombination, whereas H2 and H' are required for both reactions (Bushman *et al.*, 1985; Thompson *et al.*, 1986). IHF binding to H1 inhibits excision. Since IHF levels increase upon entry into stationary phase (see section 1.8.1) the H1 site is more likely to be occupied during poor growth conditions. This growth phase control is further supported by the role for FIS in stimulating excision at low Xis concentrations. FIS levels dramatically increase upon nutrient upshift (Ball *et al.*, 1992), which results in an increase in excision of phage lambda (Ball and Johnson, 1991b).

The regulatory inputs of growth phase regulated host proteins (IHF and FIS) appear to promote excision in good growth conditions and repress excision in poor growth conditions. Nutrient availability supports the production of high levels of phage and also ensures the presence of a viable bacterial population to infect upon transition to lytic growth. It therefore appears that the lambda phage has incorporated host regulatory proteins to promote its survival.

The integrative and excissive recombination events can also by distinguished by their requirement for negative supercoiling. Negative DNA supercoiling is required for lambda phage integration but not excision due to the different DNA topology formed in the nucleoprotein complexes formed during integration and excission (Crisona *et al.*, 1999).

1.7.1 Regulation of *fimB* and *fimE* expression.

Differential expression of the *fimB* and *fimE* genes in response to changing environmental signals allows for variation in recombinase levels (Schwan *et al.*, 1992; Gally *et al.*, 1993; Schwan *et al.*, 2002; Sohanpal *et al.*, 2004; Lahooti *et al.*, 2005; Sohanpal *et al.*, 2007). Since FimE-mediated switching is dominant to FimB-mediated switching, altering the relative levels of the two recombinases is an important component of controlling fimbrial gene expression (McClain *et al.*, 1991).

Feedback mechanisms play an important role in maintaining heterogeneous populations that are pre-adapted for changes in the environment (Balaban *et al.*, 2004). For example, *fimS* contains a Rho-dependent transcription terminator that, when in the phase-off orientation, causes shortening of the *fimE* transcript. The shortening of this transcript decreases the half-life of *fimE* mRNA and hence reduces the level of FimE protein in the

cell (Sohanpal *et al.*, 2001; Joyce and Dorman, 2002; Hinde *et al.*, 2005). A reduction in FimE levels allows FimB-mediated switching to phase on, which in turn allows for the accumulation of FimE. Accumulation of FimE causes the rapid switching of *fimS* back to phase off. The cyclic control of *fimS* orientation through FimE is referred to as orientational control (Sohanpal *et al.*, 2001; Chu and Blomfield, 2007). Orientational control is believed to provide a form of memory to fimbrial expression and is modelled to be responsible for maintaining *fimS* in phase on for sufficient time to allow full fimbriation of the cell (Wolf and Arkin, 2002; Chu and Blomfield, 2007).

Another method of ensuring population heterogeneity involves *fimS* in the ON orientation. It has been reported that cells undergo phase variation independently of *fimS* orientation or P_{fimA} activity resulting in 10% of phase ON cells remaining afimbriate (McClain *et al.*, 1993). The effects of nucleoid-associated proteins and environmental changes on transcription of *fimB* and *fimE* are discussed in sections 1.8 and 1.10.

1.8 Trans acting factors regulate phase variation of type 1 fimbriae

1.8.1 Integration host factor (IHF)

Integration host factor (IHF) is a small, basic protein that was originally identified as the host factor required for integration of phage lambda (Miller and Nash, 1981). IHF is a heterodimeric protein whose subunits are encoded by *ihfA* (formerly known as *himA*) and *ihfB* (formerly known as *himD* or *hip*) (Weisberg *et al.*, 1996). These genes are found at distinct locations in the chromosome and are subject to independent regulatory influences (Aviv *et al.*, 1994). Each subunit is ~11kDa and they share ~30% sequence identity. Although IHF is usually considered as a heterodimer, its subunits are capable of forming functional homodimers that are less stable *in vitro* (Zulianello *et al.*, 1994) but functional *in vivo* (Werner *et al.*, 1994; Zablewska and Kur, 1995; Hiszczynska-Sawicka and Kur, 1997; Mangan *et al.*, 2006). IHF binds tightly to a well-defined consensus sequence that consists of a 13-bp element WATCAANNNNTTR (W is dA or dT, R is dA or dG, and N is any nucleotide) (Yang and Nash, 1989; Goodrich *et al.*, 1990; Hales *et al.*, 1994) and an upstream A+T rich region (Hales *et al.*, 1996). While sites with a very good match to the consensus provide good targets for IHF, sites that match less perfectly are often still bound by IHF, albeit with lower affinity (Goodrich *et al.*, 1990; Yang and Nash, 1995a). IHF

dimers interact with the minor groove of DNA (Yang and Nash, 1989) interchelating hydrophobic residues, located on two extended arms, between the DNA base pairs causing severe wrapping of the DNA around the protein (Rice *et al.*, 1996) and creating the characteristic U-turn in DNA of up to 180° (Fig. 1.4) (Robertson and Nash, 1988; Thompson and Landy, 1988; Rice *et al.*, 1996).

The intracellular concentration of IHF is growth phase dependent and increases with the onset of the stationary phase (Bushman *et al.*, 1985; Aviv *et al.*, 1994; Ditto *et al.*, 1994; Weglenska *et al.*, 1996; Murtin *et al.*, 1998; Azam and Ishihama, 1999). Thus, IHF plays an important role in integrating stationary phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium (Mangan *et al.*, 2006). IHF binds specifically to high-affinity (nM Kd) and non-specifically to low-affinity (μM Kd) sites and is therefore considered to have an important role in DNA compaction and genome organization (Oppenheim *et al.*, 1993; Yang and Nash, 1995a; Ali *et al.*, 2001).

IHF is directly involved in regulating a diverse range of processes (Freundlich *et al.*, 1992) including transcriptional activation (Giladi *et al.*, 1990; Sheridan *et al.*, 1998), chromosome replication (Ryan *et al.*, 2002), phage integration (Bushman *et al.*, 1985), transposition (Crellin *et al.*, 2004) and DNA inversion (Dorman and Higgins, 1987). One example where IHF facilitates long-range molecular interactions is found in the lambda integration system. The role of IHF in lambda integration was effectively replaced with other DNA bending proteins (CRP and HU) or through the introduction of DNA that forms a stable bend in the DNA (Goodman and Nash, 1989; Goodman *et al.*, 1992). Similar studies showed the importance of helical phasing in the juxtaposition of distal sites by IHF binding (Surette *et al.*, 1989). IHF binding to DNA can activate transcription of certain promoters through a mechanism not involving stabilizing long-range protein-protein interaction. For example, IHF activates transcription of the *ilvG* promoter by binding to an upstream site located in a highly A+T-rich region (~88% A+T). IHF binding prevents the A+T-rich region from becoming single stranded, transferring the superhelical energy downstream to the *ilvG* promoter and increasing transcription by promoting the formation of an open transcription complex (Parekh and Hatfield, 1996; Parekh *et al.*, 1996; Sheridan *et al.*, 1998).

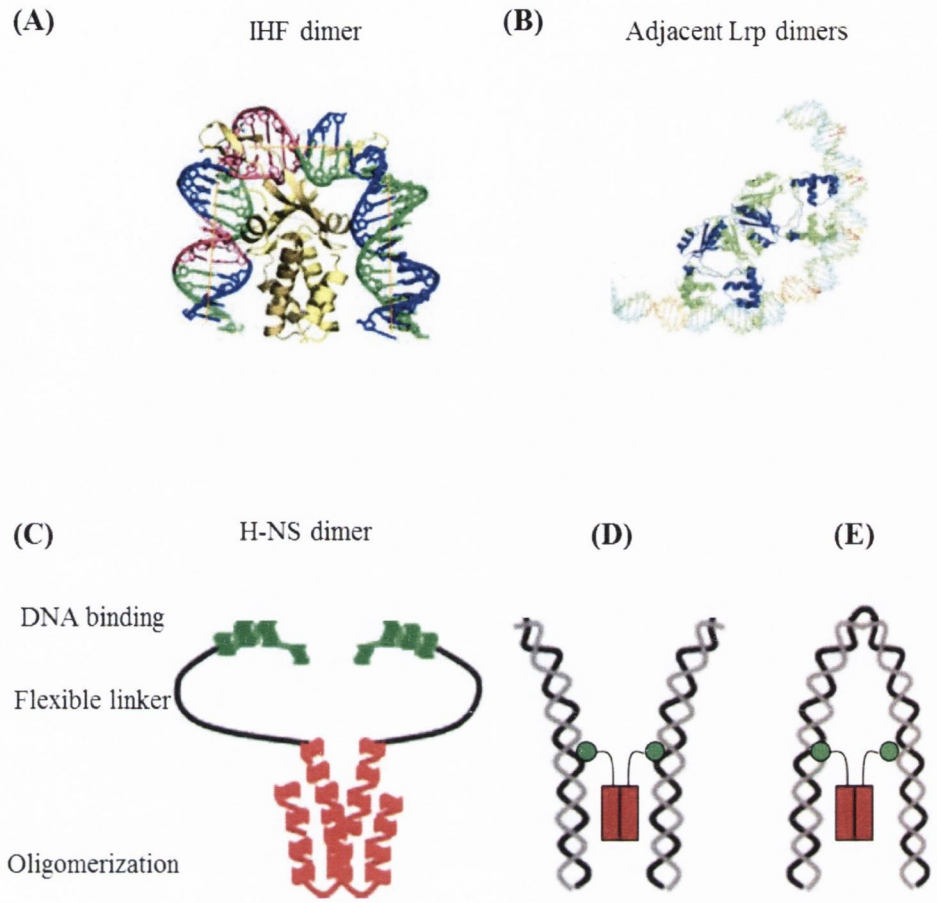


Fig. 1.4 The nucleoid associated proteins IHF, LRP and H-NS alter DNA topology. (A) Binding of an IHF dimer to DNA introduces a bend of >160° in the DNA. IHF; yellow ribbon strucuture (Yang, 2008). (B). Binding of LRP dimers to adjacent LRP sites introduces a bend of up to 135° in the DNA. LRP dimers; blue and green ribbon structures (de los Rios and Perona, 2007). (C) Ribbon structure of a H-NS dimer showing the DNA binding domains (green) and oligomerization domains (red) joined by flexible liner domains (black lines). H-NS binding to DNA can cause (D) inter- and (E) intra-molecular bridging of DNA. For clarity the interaction of a single H-NS dimer with DNA is shown as a simplified version of the ribbon structure (Rimsky, 2004).

1.8.2 Integration host factor and *fim* gene expression

A knockout mutation in either of the genes encoding the IHF subunits completely arrests both FimB- (Dorman and Higgins, 1987; Eisenstein *et al.*, 1987; Blomfield *et al.*, 1997) and FimE- (Blomfield *et al.*, 1997) mediated inversion of *fimS*. Two perfect matches to the IHF consensus were identified in the *fim* sequence (Fig. 1.5). The first site (IHF-1) is located outside *fimS*, centred ~30-bp upstream of the left inverted repeat (Dorman and Higgins, 1987; Eisenstein *et al.*, 1987). The second site (IHF-2) is located asymmetrically within *fimS* (Dorman and Higgins, 1987). Each of these sites was confirmed by gel retardation assay and DNA protection studies, the results of which are consistent with binding of a single dimer of IHF to each site (Blomfield *et al.*, 1997). By analogy with the lambda integrase system (de Vargas *et al.*, 1989b), it is thought that the IHF protein plays an architectural role in promoting efficient recombination through the appropriate juxtapositioning of the 9-bp inverted repeats (Fig. 1.6). Analysis of the switching frequencies of *fimS* containing mutated IHF binding sites confirmed that IHF had a direct effect on *fimS* switching for both FimB- and FimE- mediated recombination. The combined binding site mutations reduced FimB and FimE mediated switching by 100 fold. While this fully accounted for the reduction in FimB mediated switching in an IHF deficient cell, it was markedly lower than the 15,000 fold reduction seen in FimE mediated switching in an IHF deficient cell (Blomfield *et al.*, 1997). This suggests a role for IHF in transcriptional activation of *fimE*. IHF is also required for full activation of the *fimA* (Dorman and Higgins, 1987) and *fimB* promoters (see section 1.10.1) (Sohanpal *et al.*, 2007).

1.8.3 Leucine responsive regulatory protein (LRP)

The leucine responsive protein (LRP) is a major global regulator of the Asn/LRP family that is present in *E. coli* growing in minimal medium at 37°C at a level of ~3000 dimers per cell (Willins *et al.*, 1991). It is thought that the primary function of LRP is to monitor the nutritional state of the cell to adjust its metabolism to changing nutritional conditions and to coordinate these changes with the physical environment of the cell (Cho *et al.*, 2008). LRP levels increase in response to nutrient stress (Landgraf *et al.*, 1996). Depending on the architecture of the individual promoter, LRP can act as an activator and/or a repressor and its affects can be antagonized, enhanced or not affected by the presence of branched-chain amino acids such as leucine. LRP regulates >10% of the *E. coli* genome (Hung *et al.*, 2002; Tani *et al.*, 2002; Cho *et al.*, 2008).

Among the genes regulated by LRP in *E. coli* are operons involved in amino acid biosynthesis and catabolism (Calvo and Matthews, 1994), nutrient transport (Haney *et al.*, 1992) and P and type 1 fimbria production (Braaten *et al.*, 1992). In general, among LRP-regulated operons, those which function in biosynthetic pathways are stimulated by LRP while those involved in catabolic pathways are repressed (Calvo and Matthews, 1994). LRP plays a key role in regulation of genes upon nutrient depletion and entry into stationary phase (Tani *et al.*, 2002). Mutations that confered a growth advantage during prolonged carbon starvation were linked to the *lrp* locus and found to abolish the DNA binding ability of LRP, thus dramatically altering global gene expression and allowing adaptation to the new environment (Zinser and Kolter, 2000).

LRP binds cooperatively (Chen *et al.*, 2005b) to a consensus sequence YAGHAWATTWTDCTR, where Y = C or T, H = not G, W = A or T, D = not C, and R = A or G (Wang *et al.*, 1993; Cui *et al.*, 1995). LRP is an 18.8 kDa protein that forms higher order structures. Upon binding of a LRP dimer to a single site in DNA, LRP bends DNA by ~52°. Binding of a second LRP dimer to an adjacent site increases the angle of bending to at least 135° (Fig. 1.4) (Wang and Calvo, 1993). Further binding of LRP to adjacent sites can allow the formation of an LRP octamer that wraps the DNA duplex around its outside surface by interacting with the DNA binding N-termini of successive monomers (de los Rios and Perona, 2007). At nM concentrations LRP exists predominantly as a dimer (Cui *et al.*, 1996; Chen *et al.*, 2001c). It can self-associate to form an octamer or an hexadecamer, with the hexadecamer being the predominant species at high cellular concentrations (μM range) (Chen *et al.*, 2001c; Chen *et al.*, 2005b; de los Rios and Perona, 2007). The presence of leucine causes the dissociation of LRP hexadecamers to octamers, which can alter gene expression at sensitive promoters and thus links gene expression with environmental signals (Chen *et al.*, 2001c; Chen and Calvo, 2002).

While exogenous leucine does not affect the total amount of LRP within *E. coli* cells, it dramatically affects the quaternary structure and DNA-binding characteristics of LRP. It promotes the dissociation of hexadecamer to the leucine-bound octamer (Chen *et al.*, 2001c; Chen and Calvo, 2002), increases the binding of LRP to non-specific sites on DNA (Chen *et al.*, 2001b), decreases the intrinsic affinity of LRP for binding to a single site (Cui

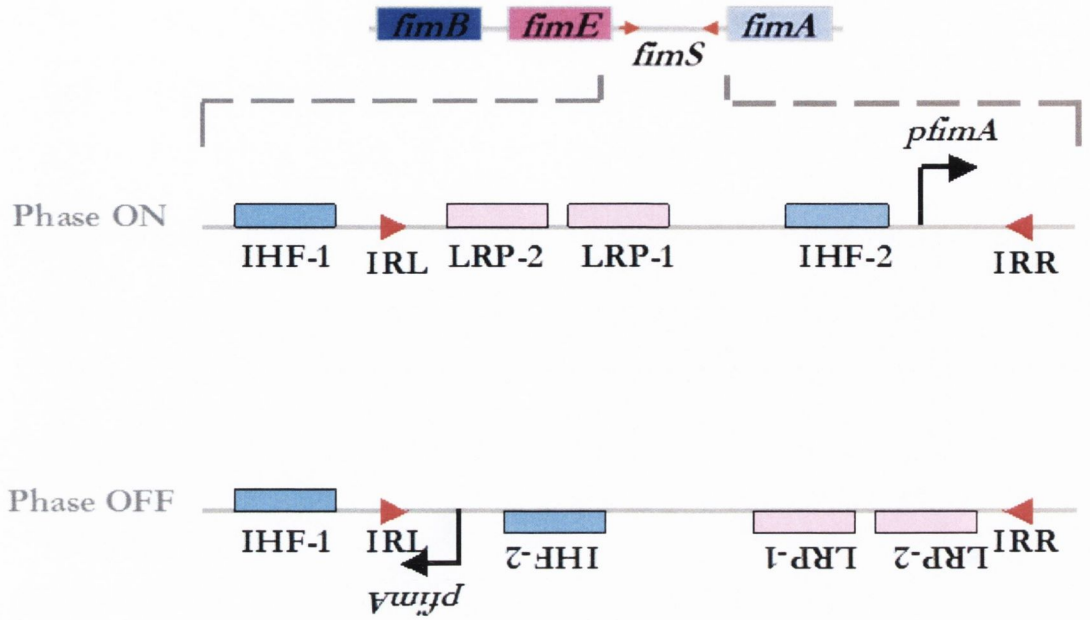


Fig. 1.5 Arrangement of nucleoid-associated protein binding sites in *fimS*. LRP binds to two sites asymmetrically within *fimS* (LRP-1 and LRP-2). IHF binds to two sites (IHF-1 and IHF-2). IHF-2 is located asymmetrically within *fimS* proximal to the *fimA* promoter (*pfimA*). IHF-1 is located in a static (non-inverting) region upstream of the IRL.

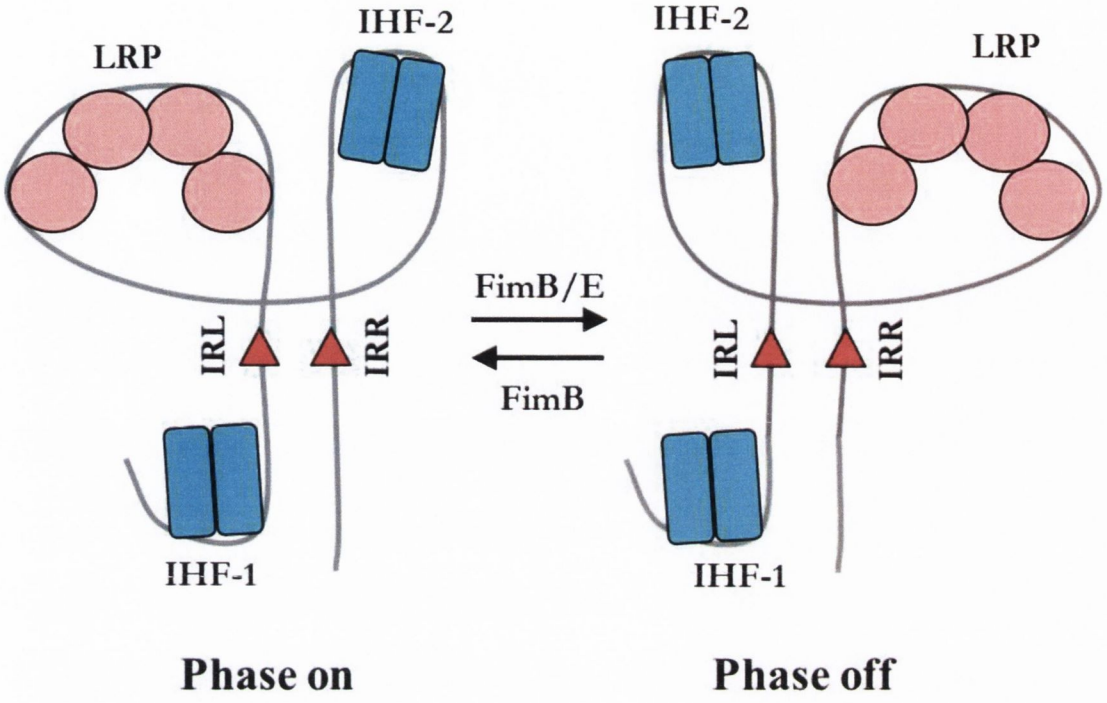


Fig. 1.6 The nucleoid associated proteins IHF and LRP alter DNA topology at *fimS* (A) The DNA bending activity of IHF and the DNA wrapping ability of LRP may facilitate juxtaposition of the inverted repeats (IRL and IRR). Site juxtaposition is a pre-requisite for strand exchange catalyzed by FimB and FimE. This model also highlights the unique topologies of the phase on and phase off invertasome complexes with respect to the fixed (non-inverting) reference point IHF-1.

et al., 1995) and it increases the cooperativity with which dimers bind to adjacent sites (Chen *et al.*, 2005b).

LRP has only a 20–400-fold binding discrimination between specific and non-specific DNA sequences (Peterson *et al.*, 2007). It binds with high affinity to non-specific sites. Non-specific binding of LRP shows significant cooperativity and can lead to an octameric solenoid-like structure that is similar to that formed by high-affinity, specific, binding of LRP to multiple adjacent sites (Chen *et al.*, 2005b; Peterson *et al.*, 2007). Therefore high affinity, non-specific binding of LRP may play a key role in genome organization under certain conditions (D'Ari *et al.*, 1993; Peterson *et al.*, 2007).

1.8.4 Leucine responsive regulatory protein and *fim* gene expression

LRP has only minor effects on transcription of *fimB* and *fimE* (Blomfield *et al.*, 1993) but plays a major role in regulation of *fimS* inversion. It binds at two discrete sites located asymmetrically within *fimS* (Fig. 1.5) (Gally *et al.*, 1994) and is required for efficient inversion of *fimS*. In LRP deficient cells, or when LRP binding to both sites within *fimS* is abolished, FimE-mediated switching is 100-fold lower and FimB-mediated switching is 50-fold lower than in the wild type (Blomfield *et al.*, 1993; Gally *et al.*, 1994).

The *fim* switch is subject to environmental control by temperature and independently by the amino acids alanine, isoleucine, leucine, and valine (Gally *et al.*, 1993). Both *fimB*-promoted switching and *fimE*-promoted switching are stimulated by these amino acids, and this stimulation requires LRP. While the addition of leucine stimulates *fimS* inversion, it reduces LRP binding in *fimS in vitro* (Roesch and Blomfield, 1998). Leucine therefore affects inversion of *fimS* by modulation of the occupancy of a third, low-affinity, LRP binding site in *fimS* that, when occupied, inhibits inversion of *fimS* (Roesch and Blomfield, 1998; Lahooti *et al.*, 2005).

LRP binding in *fimS* is believed to be required for efficient synapsis (Fig. 1.6) and is required for the maintenance of the DNA relaxation induced phase-on bias (see section 1.9.2) (Kelly *et al.*, 2006; Corcoran and Dorman, 2009; Müller *et al.*, 2009).

1.8.5 Histone-like nucleoid structuring protein (H-NS).

The histone-like nucleoid structuring protein (H-NS) is an abundant nucleoid-associated protein. H-NS protein levels are maintained through growth at 10,000-20,000 copies per cell (Spassky *et al.*, 1984). H-NS contains an amino terminal DNA binding domain that is connected by a flexible linker to the carboxyl terminal oligomerization domain (Fig. 1.4) (Shindo *et al.*, 1995; Badaut *et al.*, 2002; Bloch *et al.*, 2003; Rimsky, 2004). This structural arrangement allows H-NS dimers to create DNA-protein-DNA bridges between separate DNA strands, which play a major role in DNA compaction (Spassky *et al.*, 1984; Spurio *et al.*, 1992; Spurio *et al.*, 1997; Stella *et al.*, 2005; Dame *et al.*, 2006; Noom *et al.*, 2007; Maurer *et al.*, 2009). The abundance of H-NS and its ability to form DNA protein-DNA bridges provides support for the hypothesis that this protein can act as a topological insulator on DNA loops in the folded chromosome of the nucleoid (Hardy and Cozzarelli, 2005; Noom *et al.*, 2007). H-NS molecules can form bridges on different portions of the same DNA molecule leading to trapping of RNA polymerase at a promoter with concomitant repression of transcription (Prosseda *et al.*, 2004; Nagarajavel *et al.*, 2007; Liu *et al.*, 2010; Stoebel *et al.*, 2008). In addition, H-NS can spread laterally from nucleation sites on DNA to silence transcription without a need for bridging (Amit *et al.*, 2003; Chen *et al.*, 2005a; Liu *et al.*, 2010). H-NS silences ~15% of the *E. coli* and *S. enterica* genomes and the genes affected contribute to virulence, motility, metabolism and cell adhesion (Noom *et al.*, 2007). Deletion of the *hns* gene results in pleiotropic effects on gene expression (Higgins *et al.*, 1988). H-NS binds with low sequence specificity to regions of high A+T content and strong intrinsic DNA curvature (Yamada *et al.*, 1990; Owen-Hughes *et al.*, 1992; Lucchini *et al.*, 2006; Navarre *et al.*, 2006; Oshima *et al.*, 2006).

Other intrinsic characteristics, such as DNA flexibility, may also contribute to the specificity of H-NS binding. DNA that has become single stranded (melted) is considerably more flexible than double stranded DNA (Forties *et al.*, 2009), which may aid DNA bridging by H-NS. While a consensus binding site has been proposed for H-NS (Lang *et al.*, 2007b), it was later shown that the A+T content of the 10-bp motif (70% A+T) and not the order of the nucleotides was important (Dillon *et al.*, 2010). H-NS has a high off-rate from DNA indicating that there is substantial breathing of bridged regions (Dame *et al.*, 2006). This allows effective competition for DNA binding by other proteins *in vivo*. Since an active RNA polymerase can generate 75 pN of force and as a force of only 7 pN is

required to disrupt a DNA-HNS-DNA bridge, a transcribing RNA polymerase would be sufficient to displace H-NS from a downstream region (Dame *et al.*, 2006). This allows H-NS to stabilize DNA loops while still allowing for dynamic reorganization (Dame *et al.*, 2006; Noom *et al.*, 2007).

1.8.6 Histone-like nucleoid structuring protein and *fim* gene expression.

H-NS affects inversion of *fimS* both directly and indirectly. H-NS deficient cells have increased rates of FimB-mediated inversion (Higgins *et al.*, 1988; Kawula and Orndorff, 1991; Donato *et al.*, 1997; Schembri *et al.*, 1998). Although transcription of *fimB* increases in the absence of H-NS (Donato *et al.*, 1997) the associated increase in FimB protein levels is not the primary cause of the rapid-switching phenotype that is characteristic of *hns* null alleles (Dove and Dorman, 1996; Donato *et al.*, 1997). H-NS binding in *fimS* is believed to repress inversion directly by an unknown mechanism (Kawula and Orndorff, 1991; O' Gara and Dorman, 2000). Overexpression of H-NS results in decreased inversion of *fimS* (May *et al.*, 1990), most likely through repression of *fimB* expression. The effects of H-NS are typically masked in FimE+ backgrounds due to the rapid ON to OFF switching, however H-NS is essential for maintaining the FimB-mediated inversion bias to phase off that is induced by transcription from within *fimE* (O' Gara and Dorman, 2000).

H-NS plays an important role in regulating *fimS* inversion and P_{fimA} activity in response to changing environmental temperature (Dorman and Ní Bhriain, 1992). Since changes in transcription and changes in temperature cause changes in DNA supercoiling (Dorman, 2006), H-NS can be considered an important mediator of the sensitivity of fimbrial gene expression to changes in DNA supercoiling.

1.8.7 Recombinase overexpression and accessory protein redundancy

Increasing the amount of FimB in the cell by supplying *fimB* cloned in a multicopy plasmid increases the rate of bidirectional inversion of *fimS* (Klemm, 1986; Dove and Dorman, 1996) and allows for efficient switching to phase on in a FimE+ cell (Klemm, 1986). Over-expression of FimB also restores DNA inversion in cells that are deficient in IHF and LRP, which were previously thought to be essential for inversion of *fimS* (Dove and Dorman, 1996). Presumably the over-expression of FimB leads to increased binding of FimB to *fimS*, allowing the DNA bending activity of FimB to replace the predicted architectural roles of IHF and LRP in aligning the IRs for efficient inversion (Dove and Dorman, 1996).

When overexpressed, FimB binding at the right inverted repeat (IRR) is responsible for repressing transcription of the *fimA* promoter by approximately 10-fold (Dove and Dorman, 1996). Thus, overexpression of FimB in a FimE+ background leads to an increased number of cells with a phase on switch that have reduced levels of fimbriation, compared to a normal fimbriae expressing cell (Klemm, 1986; Dove and Dorman, 1996).

Overexpression of FimE also represses transcription of the *fimA* promoter by over 10-fold in a phase locked-on cell (Smith, 1999). Unusually, increasing levels of FimE also causes an increase in phase on cells in the population (Pallesen *et al.*, 1989; Smith, 1999).

1.9 DNA supercoiling

The bacterial chromosome is a single, circular DNA molecule (Cairns, 1963). Supercoiling represents a state of relatively greater free energy in a closed circular DNA molecule that arises when the molecule has either an excess (positive supercoiling) or a deficiency (negative supercoiling) of duplex turns relative to the same molecule that has been nicked on a single strand and is thus energy free or relaxed (Dorman, 2006). Bacterial DNA almost always exists in a state of negative supercoiling (Cozzarelli, 1980). DNA gyrase introduces negative supercoils into relaxed DNA by a double-strand breakage, passage and religation mechanism that uses ATP as a source of energy; it removes positive supercoils by the same mechanism (Gellert *et al.*, 1976a; Cozzarelli, 1980; Champoux, 2001). DNA topoisomerase I removes negative supercoils by a DNA single-strand cleavage, swivelase and religation mechanism that relies on the energy stored in the negatively supercoiled DNA (Brown *et al.*, 1979; Champoux, 2001).

DNA supercoiling levels vary with the stage of growth of the bacterial culture (Dorman *et al.*, 1988), with DNA becoming more relaxed at the onset of stationary phase when the metabolic flux in the cell and the concentration of ATP are diminished (Balke and Gralla, 1987; van Workum *et al.*, 1996). These physiological conditions correlate with a reduction in the negative supercoiling activity of DNA gyrase. ATP hydrolysis by gyrase can be inhibited by treatment with the drug novobiocin (Gellert *et al.*, 1976b). As the concentration of novobiocin increases, the level of gyrase activity in the cell is reduced, allowing DNA topoisomerase I to relax the DNA (Gellert *et al.*, 1976b; Heddle *et al.*, 2000). DNA supercoiling levels are responsive to changes in other environmental signals such as temperature (Goldstein and Drlica, 1984; Dorman *et al.*, 1990), anaerobiosis (Yamamoto and Droffner, 1985; Dorman *et al.*, 1988; Hsieh *et al.*, 1991a) and osmolarity (Higgins *et al.*, 1988; Hsieh *et al.*, 1991b). Homeostasis is maintained since transcription of

DNA gyrase and topoisomerase 1 is alternatively regulated in response to changes in DNA supercoiling levels (Menzel and Gellert, 1983; Tse-Dinh, 1985).

The movement of RNA polymerase along DNA during transcription alters the local DNA topology such that positive supercoils accumulate ahead of the transcribing polymerase and negative supercoils accumulate behind the transcribing polymerase (Liu and Wang, 1987; Samul and Leng, 2007). This transcription-induced alteration in DNA supercoiling can affect local gene expression (Fang and Wu, 1998b; a).

Negative supercoiling can be further defined as the deficit in linking number (L_k) where L_k = Twist (T_w) + Writhe (W_r). While writhe illustrates the shape of the DNA, T_w is best illustrated as the number of base pairs in a single helical turn, and hence, the number of helical turns in the DNA (Fuller, 1971). Both of these parameters change in response to changing supercoiling levels (ΔL_k) to adopt the lowest free energy conformation. A change in twist can either involve the smooth torsional deformation (change in the number of base pairs per helical turn) or alternatively, can involve a portion of the DNA molecule becoming single stranded (Benham, 1979). The constraint of this helical twist (e.g. through protein binding the DNA) can cause a corresponding alteration in writhe as the lowest energy state is achieved.

Stress induced duplex destabilisation (SIDD) regions are typically A+T-rich regions in DNA and hence are prone to strand separation when placed under a superhelical stress (ΔT_w). When nucleoid associated proteins (e.g. LRP and IHF) bind within a region which is prone to strand separation such as a SIDD region, the bound protein constrains the DNA and prevents strand separation (Sheridan *et. al.*, 1998). Since the strand separation is necessary to adopt the lowest free energy state, the altered DNA twist (ΔT_w) is transferred to another region (often A+T-rich but sometimes containing G+C-rich stretches), which then becomes single-stranded (Sheridan *et. al.*, 1998; Opel *et. al.*, 2004). By this mechanism, nucleoid associated proteins can facilitate increased promoter activity through increased RNA polymerase accessibility. Negative supercoiling may play a similar role in promoting site-specific recombination by promoting single-strand formation at inverted repeats.

1.9.1 DNA supercoiling and site-specific recombination

Recombination by tyrosine site-specific recombinases can be affected by changes in negative DNA supercoiling through a number of different mechanisms; (i) an increase in the efficiency of formation of the synaptic complex (by an increase in the rate of the forward reaction or a decrease in the rate of the backward reaction); (ii) an increase in the catalytic recombinase-mediated reaction (Holliday junction formation); or (iii) an increase in the efficiency of resolution of the Holliday junction, possibly produced by modifications in the geometry of the Holliday junction brought about by changes in DNA supercoiling density (Mikheikin *et al.*, 2006).

Negative supercoiling of DNA increases the juxtaposition of distal sites and thus aids synapsis (Vologodskii and Cozzarelli, 1996). In the Int-system DNA supercoiling facilitates formation of a higher-order nucleoprotein complex at *att*P, that catalyses integrative recombination with the chromosomal site (*att*B) (Richet *et al.*, 1986, 1988). DNA supercoiling is also essential for synapsis in serine-recombinase mediated recombination systems such as circular molecule resolution by the Tn*3* resolvase (Benjamin and Cozzarelli, 1988), the Hin-mediated inversion of the *hin* segment that controls antigenic variation of flagellar proteins in *Salmonella* (Moskowitz *et al.*, 1991; Sanders and Johnson, 2004; Dhar *et al.*, 2009) and the Gin mediated inversion of the G segment in bacteriophage Mu (Plasterk *et al.*, 1984).

DNA supercoiling is also known to influence some recombination reactions after synapsis by lowering the activation energy required to complete the recombination reaction and thus increasing the rate of recombination (Castell and Halford, 1989).

1.9.2 DNA supercoiling and *fim* gene expression

FimB-mediated inversion of *fimS* is sensitive to changes in DNA supercoiling. Previous work has shown that when DNA gyrase is progressively inhibited by novobiocin treatment, the *fimS* element shows a novobiocin-dose-dependent bias towards the on phase (Dove and Dorman, 1994; Kelly *et al.*, 2006; Müller *et al.*, 2009). In addition, it has been shown in earlier work that this off-to-on bias is contingent on binding by the LRP protein to two adjacent sites, LRP-1 and LRP-2, within *fimS* (Kelly *et al.*, 2006) (Fig. 1.7). The two LRP binding sites are located asymmetrically within *fimS,* raising the possibility that they may contribute to a structure that distinguishes the phase-on switch from the phase-off switch. This hypothesis postulates the existence of a reference point that somehow communicates

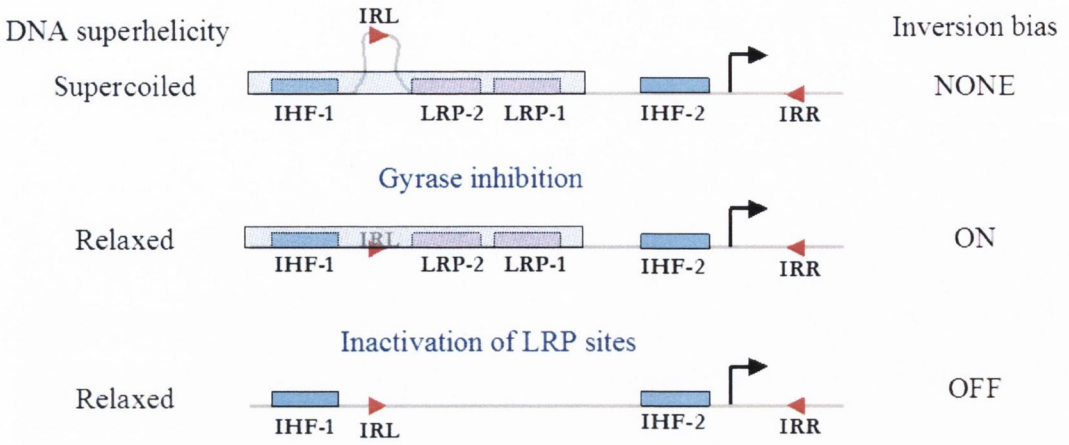


Fig. 1.7 A model for the FimB-mediated, LRP-dependent, phase on bias of *fimS* in response to DNA relaxation. The model involves the formation of a nucleoprotein complex between LRP and an unidentified reference point, potentially IHF bound at IHF-1, at the IRL in the phase on orientation where, upon DNA relaxation, the IRL is sequestered. This prevents efficient switching from phase on to phase off, which results in an increase in the amount of on cells in a population. This bias is reversed in the absence of LRP binding in *fimS*.

with the LRP-occupied LRP-1 and LRP-2 sites in the phase-on switch. An inspection of the DNA adjacent to *fimS* revealed that, in the on phase, the two LRP binding sites lie close to a well characterized binding site for IHF (Fig. 1.7).

This novobiocin-induced phase-on bias is independent of transcription from within *fimA* (Kelly *et al.*, 2006). However, transcription from within *fimS* is essential for the establishment of a phase bias caused by transcription from within *fimE* (O' Gara and Dorman, 2000). CSH50 contains an *IS*1 insertion in *fimE* that prevents expression of FimE (Blomfield *et al.*, 1991b) and also prevents transcriptional read through into *fimS* (O' Gara and Dorman, 2000). Strains that contain an inactivated *fimE* gene but which allow transcription into *fimS* are strongly biased to phase off. This phase-off bias was removed in the absence of transcription from within *fimS* and also dependent on H-NS (O' Gara and Dorman, 2000).

The action of topoisomerase 1 is required for inversion of *fimS* (Dove and Dorman, 1994). Inactivation of topoisomerase 1 causes an increase in negative supercoiling levels. Reduction in negative supercoiling levels using novobiocin failed to restore switching in the Δ*topA* strain (Dove and Dorman, 1994) possibly indicating that TopA is required to reduce transcription-induced hyper-negative supercoils that are inhibitory to DNA inversion.

DNA supercoiling facilitates synapsis in many recombination reactions and may have a similar role in synapsis of the *fimS* IRs. DNA supercoiling may also facilitate recombination by facilitating single-stand formation at the IRs (See section 1.9). The *fimS* region is highly A+T-rich. For example, the 296-bp *fimS* region between the IRs is 70% A+T while the 100-bp 5' to the IRL (where IHF-1 is centrally located) is 79% A+T. The high A+T content of these regions makes them highly unstable and prone to single-strand formation (Bi and Benham, 2004). In contrast, the 9-bp IRs are highly G+C rich (67%). Since recombination requires single-strand formation at the IRs, transfer of superhelical energy from the A+T-rich regions in *fimS* by IHF and LRP binding may contribute to melting of the IRs and may thus promote recombination. This model is also compatible with the proposed structural roles for IHF and LRP in promoting synapsis of the IRs that allows for *fimS* inversion (Fig. 1.6). A reduction in DNA supercoiling could therefore affect both synapsis and strand separation, preventing recombination under certain conditions. Considering the asymmetric nature of *fimS* topology due to the asymmetric binding of IHF and LRP binding in *fimS* (Fig. 1.5), a reduction in DNA supercoiling may affect switching

from ON to OFF and OFF to ON differently. This directionality would be reminiscent of the differential requirement for DNA supercoiling in Int-mediated recombination by phage lambda (section 1.7).

1.10 Control of fimbriation by signal transduction.

Multiple signal transduction pathways alter expression of type 1 fimbriae in response to changes in the environment. One effective mechanism of controlling fimbriation is through altering the levels of the recombinases, particularly levels of FimB. A number of pathways for this are summarized below.

1.10.1 Sialic acid response

During inflammation, host defences enhance the release of the common sialic acids N-acetylneuraminic acid (Neu_5Ac) and N-acetylglucosaminic acid (GlcNAc) (Kriat *et al.*, 1991), which are known to inhibit transcription of *fimB* and thus inhibit type 1 fimbrial gene expression (El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004; Condemine *et al.*, 2005; Sohanpal *et al.*, 2007). The promoters for *fimB* and the divergently transcribed gene *nanC* are located within the large (~1.4 kb) intergenic region and are co-regulated by a number of proteins. Two proteins, NanR and NagC, are responsive to cellular concentrations of Neu_5Ac and GlcNac and control transcription of *nanC,* which encodes a voltage-dependent outer membrane channel that transports Neu5Ac (Condemine *et al.*, 2005).

NanR and NagC binding near the *nanC* promoter (P_{nanC}) represses transcription of *nanC* and activates transcription of *fimB*. IHF binding in between two NagC sites promotes their interaction and thus promotes activation of the *fimB* promoter (Sohanpal *et al.*, 2007). NanR, which binds directly overlapping the *nanC* promoter, does not require IHF to enhance transcription of *fimB* (Sohanpal *et al.*, 2007).

The regulation of *fimB* by NanR and NagC provides a mechanism for modulation of type 1 fimbrial gene expression according to the host immune response (Kriat *et al.*, 1991; El-Labany *et al.*, 2003; Sohanpal *et al.*, 2004; Condemine *et al.*, 2005; Sohanpal *et al.*, 2007). The inhibition of fimbriation in response to inflammation may allow *E. coli* to limit the inflammatory response and prevent clearance by the immune system. However, a moderate inflammatory response may be advantageous since sialic acid can be converted by *E. coli* into the high-energy metabolite GlcNAc-6-P (Plumbridge and Vimr, 1999).

1.10.2 Stringent response

The small nucleotide guanosine tetraphosphate (ppGpp) is produced in response to multiple stresses including amino acid limitation, nutrient limitations and circumstances that cause growth arrest (Magnusson *et al.*, 2005). Aided by DksA, ppGpp binds in the secondary channel of RNA polymerase (RNAP) near the active site and alters transcription (Artsimovitch *et al.*, 2004; Paul *et al.*, 2004; Perederina *et al.*, 2004). ppGpp affects multiple genes in response to poor growth conditions, including the positive regulation of amino acid biosynthesis promoters (Traxler *et al.*, 2008). It is also a key component of the stringent response where in response to amino acid starvation uncharged tRNAs bind to the ribosomal A site, signalling ribosome-associated RelA to synthesize ppGpp (Wendrich *et al.*, 2002) which leads to the transcription of stable RNAs (ribosomal and transfer RNAs) being shut-off (Chatterji and Ojha, 2001).

ppGpp enhances transcription of *fimB,* which leads to increased fimbriation (Aberg *et al.*, 2006, 2008). In contrast to its usual role facilitating ppGpp, DksA deficient strains are hyper-fimbriate indicating a negative role for DksA in controlling fimbrial gene expression (Magnusson *et al.*, 2007; Aberg *et al.*, 2008). DksA does promote transcription of *fimB in vitro*. A model for this discrepancy between *in vivo* and *in vitro* data predicts an increased role *in vivo* for the DksA homologues GreA and GreB in induction of *fimB* transcription in the absence of DksA (Aberg *et al.*, 2008). Enhanced transcription of *fimB* in the absence of DksA is independent of previously known regulators of *fimB* transcription, such as RpoS, H-NS, NanR or NagC, (Aberg *et al.*, 2008). The role of ppGpp in promoting fimbriation is consistent with the involvement of type 1 fimbriae in promoting a sessile lifestyle and biofilm formation in nutrient poor conditions.

1.11 Regulatory cross-talk between surface adhesions

The expression of surface adhesions in uropathogenic *E. coli* are intrinsically linked (Holden and Gally, 2004; Kaper *et al.*, 2004). Many uropathogenic strains contain multiple fimbrial gene clusters with different binding specificities, which determine tropism of the organism (Kaper *et al.*, 2004). While some pathogenic strains contain over 16 fimbrial gene clusters (Low *et al.*, 2006), this section will provide a brief overview of well characterized examples of adhesin gene cross talk in uropathogenic *E. coli*.

1.11.1 Pyelonephritis associated pili

In strains that cause cystitis, type 1 fimbriae are continually expressed and the infection is confined to the bladder (Connell *et al.*, 1996). Pyelonephritis associated pili (P-pili) are expressed in the kidney where they bind specifically to digalactoside receptors (Korhonen *et al.*, 1986). Induction of P-pili leads to repression of type 1 fimbrial genes in the same bacterial cell (Xia *et al.*, 2000; Gunther *et al.*, 2001; Holden *et al.*, 2001; Holden *et al.*, 2006; Holden *et al.*, 2007). PapB, which can both activate and repress P-pili gene expression, directly inhibits FimB activity and promotes FimE activity, resulting in rapid switching of *fimS* from the ON to OFF orientation (Xia *et al.*, 2000). This cross-regulation may facilitate nephritic infection by releasing bacteria from the bladder epithelium, allowing movement through the ureters to the kidneys.

1.11.2 Antigen 43

The cross-talk between adhesins extends further than co-ordinate regulation of P-pili and type 1 fimbriae. An impressive hierarchical cascade exists where expression of P-pili is dominant to expression of type 1 fimbriae, which in turn is dominant to expression of antigen 43 (Ag43) that is responsible for auto-aggregation of *E. coli* through an intercellular handshake mechanism. Ag43 contributes to biofilm formation (Danese *et al.*, 2000; Kjaergaard *et al.*, 2000) and long-term persistence of uropathogenic *E. coli* in the urinary tract (Ulett *et al.*, 2007). Expression of fimbriae inhibits the interaction of Ag43 and prevents auto-aggregation by preventing intimate association with neighbouring bacterial cells (Hasman *et al.*, 1999). Since fimbriae are phenotypically dominant to Ag43 the temporal expression of fimbriae is important for bacterial persistence. Expression of Ag43, which is encoded by the *flu* gene, is dramatically increased when the genes encoding type 1 fimbriae are deleted (Schembri and Klemm, 2001b). The mechanism for this cross-regulation is unclear but the composition of the fimbriae, which contain disulphide residues, may alter the cellular thiol-disulphide status of the cell leading to reduction of redox-sensitive DNA binding protein OxyR, which is a negative regulator of *flu* (Wallecha *et al.*, 2002; Wallecha *et al.*, 2003).

1.12 Aims of this study

FimB-mediated inversion of *fimS* is sensitive to changes in DNA supercoiling. Previous work has shown that when DNA gyrase is progressively inhibited by novobiocin treatment, the *fimS* element shows a novobiocin-dose-dependent bias towards phase on (Dove and Dorman, 1994; Kelly *et al.*, 2006; Müller *et al.*, 2009). It has been shown in earlier work that this off-to-on bias is contingent on binding by the LRP protein to two adjacent sites (LRP-1 and LRP-2) within *fimS* (Kelly *et al.*, 2006). The two LRP binding sites are located asymmetrically within *fimS,* raising the possibility that they may contribute to a structure that distinguishes the phase on switch from the phase off switch (Fig. 1.7). This hypothesis postulates the existence of a reference point that somehow communicates with the LRP-occupied LRP-1 and LRP-2 sites in the phase-on switch. An inspection of the DNA adjacent to *fimS* revealed that, in the on phase, the two LRP binding sites lie close to a well characterized binding site for IHF (IHF-1). The initial aim of this study was to determine the role, if any, for IHF bound at IHF- 1 in the DNA relaxation-induced phase-on bias (Chapter 3). A role for H-NS in phase variation of type 1 fimbriae has been well established (see section 1.8.6). H-NS was previously shown to be essential for FimB-mediated inversion biasing to phase off caused by head-to-head transcription from within *fimE* and *fimS*. Since transcription induces positive supercoils (similar in effect to relaxation of negatively supercoiled DNA), a key aspect of this study was to determine whether H-NS played a role in the DNA relaxation-induced phase-on bias. This was investigated by genetic analysis of *fimS* inversion in the presence and absence of H-NS. The mechanism of H-NS affecting *fimS* was investigated *in vitro* and provided novel insights into *fimS* regulation and the role for H-NS influencing a site-specific recombination event (Chapter 3).

In order to aid the rapid simultaneous analysis of *fimS* orientation and P_{fimA} activity, a transcriptional fusion of *fimA* to the gene encoding the green fluorescent protein (*gfp*) was created. Analysis of this fusion led to new insights into the molecular biology of the *gfp* gene and subsequently to the development of a new *gfp* gene, which is a more accurate reporter of transcriptional activity (Chapter 4).

Chapter 2.

Materials and Methods.

2.1 Chemicals and growth media.

All chemicals and reagents used in this study were purchased from BDH Chemicals Ltd, Bio-Rad, Clontech, Invitrogen™, Novagen, PALL, Pfizer, Pierce, Promega, Qiagen, RBC Biosciences, Roche Molecular Biochemicals and Sigma Chemical Company. DNA restriction and modifying enzymes were obtained from Roche, New England Biolabs (NEB), Promega, Stratagene or Fermentas. In addition, several molecular biology 'kits' were used during this study. The basic principle of each kit is described in brief below, without giving a complete protocol.

2.1.1 Growth conditions and media

Materials for preparing growth media were obtained from Difco (BD Diagnostic systems). All media were sterilised by autoclaving at 120°C for 20 min prior to use. Aqueous solutions, which would be damaged by autoclaving, e.g. amino acid or antibiotic solutions, were sterilised by filtration through 0.2 μm SupaTop Syringe Filters (Anachem). Media were supplemented with the appropriate antibiotics as required.

All quantities listed below are for the preparation of 1 litre of medium in distilled, deionised water (ddH_2O). Solid media were prepared by the addition of 15 g of agar per litre of medium prior to autoclaving.

L Broth and L agar plates

Bacterial strains were routinely cultured in L broth unless otherwise stated. L agar plates were used commonly throughout this study for culturing of strains, and selection of transformants and transductants.

L 10 g bacto-tryptone, 5 g yeast extract, 5 g NaCl, pH7

Luria Bertani (LB) and LB agar plates

LB was used in *fimS* inversion assays.

LB 10 g bacto-tryptone, 5 g yeast extract, 10 g NaCl, pH7

Some experiments required variation in NaCl concentration. For this, NaCl-free LB was prepared to the above recipe but only made up to 900 ml before autoclaving. Aliquots of this broth could later be used at any NaCl concentration by adding $^{1}/_{10}^{th}$ volume of concentrated sterile NaCl solution or ddH_20.

MacConkey Lactose Agar and X-gal agar

MacConkey Lactose agar was used to indicate the β-galactosidase activity of bacterial colonies. The MacConkey base contains phenol red, a pH indicator that results in a red colour under acidic conditions. β-galactosidase-producing colonies are a red colour due to the production of acidic end products during the fermentation of lactose. In contrast colonies producing no, or very low levels, of β-galactosidase are a white or a pale pink colour.

MacConkey agar base (Difco): 40 g MacConkey agar base, 10 g lactose, 15 g agar

Alternatively, X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside) was added to LB agar to allow blue/white screening for β-galactosidase activity. This was preferred with Δ*hns* strains, as these strains had reduced viability after growth on MacConkey agar. X-Gal was prepared as a 20 mg/ml stock solution in dimethyl formamide and stored in the dark at -20°C. X-Gal was added after autoclaving and used in solid media at a final concentration of 20 μg ml^{-1}.

2.1.2 Antibiotics

Antibiotics were used at the following concentrations: carbenicillin, 100 μg ml^{-1}; kanamycin, 20 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; tetracycline, 12.5 μg ml^{-1}.Antibiotics were dissolved in ddH_20 with the exception of chloramphenicol and tetracycline, which were dissolved in ethanol. Tetracycline was stored in the dark. All antibiotics were prepared as 1000X stocks, filter sterilised and stored at -20°C. Novobiocin was prepared fresh before use in ddH_20 to a stock concentration of 25 or 100 mg ml^{-1}.

2.2 Culture conditions

Bacterial cultures were grown aerobically in liquid medium at 37°C unless otherwise indicated in the text. Cultures were typically grown by inoculating single colonies into 2 ml

of LB broth in a sterile test tube and incubating at the required temperature overnight. When larger volumes were required cultures were generally grown by inoculating 1% of an overnight culture into a volume of fresh medium appropriate to the experimental procedure, in a suitably sized conical flask. Bacterial cultures grown on agar plates were typically incubated at 37°C for 18 hr.

Assays involving the DNA gyrase inhibitor novobiocin were performed predominantly as follows: bacteria containing the *fimA–lacZ* transcriptional fusion were screened for their Lac phenotype on MacConkey lactose indicator medium as previously described (Dove and Dorman, 1994). Distinct phase-on (red) or phase-off (white) colonies were used to inoculate overnight cultures that were in turn used to inoculate 250 ml flasks containing 25 ml of LB and these were grown to an OD_{600} of ~0.1 at which point novobiocin (stock solution 100 mg ml^{-1}) was added to a final concentration of 25, 50, 75 or 100 μg ml^{-1}. Cultures were further incubated overnight (20 h) before sampling for determination of *fim* switch orientation.

The majority of novobiocin-mediated gyrase inhibition assays were performed using *fimB* cloned on a multi-copy plasmid (pSLD203), which dramatically increases bidirectional inversion of *fimS* (Dove and Dorman, 1996). Rapid switching allows for rapid response to environmental changes and thus, rapid equilibration of *fimS* orientation in a population.

CSH50 (wild type *fimB* at the native chromosomal location) has lower switch rates than strains containing pSLD203 and those experiments that involved CSH50 were performed as follows. Phase on or phase off colonies were distinguished by colony morphology; phase-on colonies are small and smooth while phase off colonies are larger and have a rough appearance (Hasman *et al.*, 2000). Phase on or phase off colonies were resuspended in 100 μl LB broth and used to innoculate 3 ml of LB in a 15 ml test tube. Cultures were grown until exponential phase at which point they were diluted and used to innoculate pre-warmed 3-ml broths containing increasing concentrations of novobiocin (0-25μg/ml). Dilutions were calculated to allow 15 generations before cessation of growth (OD_{600} of ~3), assuming each generation resulted in a doubling of the OD_{600} of the culture. Using 3-ml cultures in 15-ml flasks allows for higher throughput of samples than using 25-ml flasks. 25-ml cultures are required to monitor *fimS* orientation by southern blotting (Dove and Dorman, 1994) but are not required when monitoring *fimS* orientation by PCR based assay (Smith, 1999).

2.3 Bacterial Strains, plasmids and bacteriophage

2.3.1 Bacterial strains

All bacterial strains used in this study were derivatives of *Escherichia coli* K-12. Both the genotype and the source of the strains used, together with details relevant to their construction are listed in Table 2.1. Bacterial strains were maintained as permanent stocks in 20% glycerol in LB broth with appropriate antibiotic selection and stored at -70°C.

2.3.2 Plasmids

Plasmid constructs used in this study are listed in Table 2.2 together with relevant details and source.

2.3.3 Bacteriophage

The bacteriophage used in this study was bacteriophage P1 *vir*.

2.4 Oligonucleotides

The sequences of all oligonucleotides used in this study are listed in Table 2.3. Oligonucleotides were purchased from MWG-Biotech or Integrated DNA Technologies (IDT). Primers were delivered as lyophilised DNA and were resuspended in ddH_20 to a final concentration of 100pmol/µl before use.

2.5 Genetic techniques

2.5.1 Use of the generalised transducing phage P1 *vir*

Throughout this study bacteriophage P1 *vir* was routinely used to transduce mutant alleles between *E. coli* strains. Phage P1 can encapsidate up to ~70-kb of bacterial DNA during lytic growth of the phage. Extraction and incubation of these phage particles allows transfer of genetic material from the donor strain into a recipient where homologous recombination with the host DNA may result in allelic exchange. The *vir* mutation prevents the phage from generating P1 lysogens (circular phage molecules that are self-replicating) among transductants (Sternberg and Hoess, 1983).

Table 2.1. Bacterial strains used during this study.

Strain	Relevant details[a]	Reference/source
CSH50	λ pL(209) *fimE*::IS*1*	Miller (1972)
VL386	CSH50*fimA-lacZ*	Freitag *et al.*, (1985)
CJD 808	VL386 *fimB*::*kan*. *fimS* phase locked-on	Dove and Dorman (1994)
CJD 957	VL386 *fimB*::*kan*. *fimS* phase locked-off	Dove and Dorman (1994)
VL386 IHF 1	VL386 *recD* IHF 1 x VL386	Corcoran and Dorman (2009)
VL386 Lrp 1/2	VL386 *recD* LRP1/2 x VL386	Corcoran and Dorman (2009)
VL386 IHF 1, LRP 1/2	VL386 *recD* IHF 1, LRP1/2 x VL386	Corcoran and Dorman (2009)
VL386 *hns*$^-$	*tetRA* element from Tn*10* inserted 205 bp into the *hns* gene	Corcoran and Dorman (2009)
VL386 *hns*$^-$, IHF 1	VL386 *hns*$^-$ x VL386 IHF 1	Corcoran and Dorman (2009)
VL386 *hns*$^-$, LRP 1/2	VL386 *hns*$^-$ x VL386 LRP1/2	Corcoran and Dorman (2009)
VL386 *hns*$^-$, IHF 1, LRP1/2	VL386 *hns*$^-$ x VL386 IHF 1, LRP1/2	Corcoran and Dorman (2009)
VL386 *recD*	VL386 *recD*::Tn*10*	Smith and Dorman, (1999)
VL386 *recD* IHF 1	Δ*fimB*::*kan*, 4bp mutation in the *fimS* IHF 1 site.	Corcoran and Dorman (2009)
VL386 *recD* IHF 2	Δ*fimB*::*kan*, multiple mutations that in the *fimS* IHF 2 site .	This study
VL386 *recD* LRP1/2	Δ*fimB*::*kan*, binding site mutations in *fimS* LRP sites 1 and 2	Kelly *et al.*, (2006)

VL386 *recD* IHF 1, LRP1/2	Δ*fimB*::*kan*, binding site mutations in IHF site 1 and LRP sites 1 and 2.	Corcoran and Dorman (2009)
CSH50*fimA-gfp-cat*	*fimA* transcriptional fusion to *gfp*. *gfp-cat* from pZep08 inserted was inserted at the PstI site of *fimA*.	Corcoran *et al.*, (2010)
CSH50*fimA-gfp-cat.rv*	*fimA* transcriptional fusion to *gfp*. The *cat* gene is reversed compared to CHS50*fimA-gfp-cat* so that transcription is directed towards *fimS*.	This study
CSH50*fimA-gfp+kan*.1	Kan gene from pKD4 integrated to replace the *cat* gene in CHS50*fimA-gfp-cat*	This study
CSH50*fimA-gfp+kan*.2	Kan gene from pKD4 integrated to replace the *cat* gene in CHS50*fimA-gfp-cat.rv*.	This study
CSH50*fimA-gfp*+.1	Kan excised from CSH50*fimA-gfp+kan*.1 using pCP20	This study
CSH50*fimA-gfp*+.2	Kan excised from CSH50*fimA-gfp+kan*.2 using pCP20	This study
CSH50*fimA-gfp*^{TCD}*-kan*	*fimA* transcriptional fusion to *gfp*^{TCD}. *gfp*^{TCD}*-kan* inserted at the PstI site of *fimA*.	Corcoran *et al.*, (2010)
CSH50*proU*+98-*gfp*+	*gfp*+-*kan* inserted at +98 bp from the transcriptional start site of *proU*	Corcoran *et al.*, (2010)
CSH50*proU*+98-*gfp*^{TCD}	*gfp*^{TCD}*-kan* inserted at +98 bp from the transcriptional start site of *proU*	Corcoran *et al.*, (2010)
CSH50*proU*+98-*lacZ*	*lacZY-cat* inserted at +98 bp from the *proU* transcriptional	Corcoran *et al.*, (2010)

	start site	
CSH50*proU*+936- *gfp*+	*gfp*+-*kan* inserted at +936 bp from the transcriptional start site of *proU*	Corcoran *et al.*, (2010)
CSH50*proU*+936- gfp^{TCD}	gfp^{TCD}-*kan* inserted at +936 bp from the transcriptional start site of *proU*	Corcoran *et al.*, (2010)
CSH50*proU*+936- *lacZ*	*lacZY-cat* inserted at +936 bp from the transcriptional start site of *proU*	Corcoran *et al.*, (2010)
DMS1976	MG1655 *lacA*::*cat*	Stoebel *et al.*, (2009)

a. Transductional crosses are represented as A x B, where A is the donor and B is the recipient.

Table 2.2. Plasmids

Plasmid name	Relevant details	Reference/source
pMMC106	pMCL210 cut with XhoI, NheI and relegated to delete the *lac* promoter. Amp^R.	McCusker *et al.*, (2008)
pMMC108	pMMC106 containing a 550-bp region of *fim* (*fimS-fimA*) inserted at the PstI site (phase off).	McCusker *et al.*, (2008)
pMMC108IHF-2	pMMC106 containing binding site mutations in IHF-2	This study
pCPC2	700 bp (from *fimE*::IS*1* to *fimA-lacZ)* inserted in PstI site of pMMC106 (*fimS* phase off).	Corcoran and Dorman (2009)
pCPC2 IHF-1	pCPC2 containing a 4-bp mutation in the IHF-1 site	Corcoran and Dorman (2009)
pSGS501	*fimB*::*kan*, *fimE*::IS*1*, *fimA-lacZ* cloned in pACYC184, *fimS* phase off. Kan^R, Cm^R	Smith and Dorman, (1999)
pSGS501 IHF-1	pSGS501 containing binding site mutations in site IHF-1	Corcoran and Dorman (2009)
pSGS501 IHF-2	pSGS501 containing binding site mutations in IHF-2	
pSGS501 LRP-1, LRP- 2	pSGS501; binding site mutations in sites LRP -1 and LRP-2	Kelly *et al.*, (2006)
pSGS501 IHF-1, LRP-1, LRP-2	pSGS501containing binding site mutations in sites IHF- 1 and LRP-1 and LRP-2	Corcoran and Dorman (2009)
pSLD203	*fimB* cloned in pUC18	Dove and Dorman, (1994)
pCoC1	gfp^{TCD} provided by DNA2.0 in the custom vector pJ204. pUC origin, Amp^R	Corcoran *et al.*, (2010)

pCoC2	*kan* cassette from pKD4 cloned downstream of gfp^{TCD} in pCoC1 (Amp^R, Kan^R).	Corcoran *et al.*, (2010)
pCoC2-BFP	gfp^{TCD} chromophore mutated (Y66H) to convert GFP to BFP.	This study
pPro	Propionate inducible cloning vector, Amp^R	Lee and Keasling (2005)
pPro-*gfp*+	*gfp*+ under the control of the propionate promoter	Corcoran *et al.*, (2010)
pPro-gfp^{TCD}	gfp^{TCD} under the control of the propionate promoter	Corcoran *et al.*, (2010)
pPro-bfp^{TCD}	bfp^{TCD} under the control of the propionate inducible promoter	Corcoran *et al.*, (2010)
pKD46	Lambda genes *gam*, *bet* and *exo* under the control of the arabinose inducible pBad promoter (Amp^R).	Datsenko and Wanner (2000)
pKD4	*kan* gene flanked by FRT sites that allows its excision by the *Flp* recombinase. Not competent for replication in standard *E. coli* K-12 strains. (Kan^R).	Datsenko and Wanner (2000)
pCP20	Temperature sensitive origin of replication. Expresses the *Flp* recombinase.	Datsenko and Wanner (2000)
pZep08	Transcriptional fusion plasmid containing the green fluorescent protein gene (*gfp*+). (Kan^R, Cm^R, Amp^R)	Hautefort *et al.*, (2003)

Table 2.3. Oligonucleotides

Primer name [a]	Primer sequence (5' - '3) [b]
OL 4	CCG TAA CGC AGA CTC ATC CTC
OL 20	GAG TTT TAA TTT TCA TGC TGC TTT CC
Ihf1.int.fw	GCA CCT CAA AAA CAC CAT CA
fimS.rv	GAC ATG GGC AGT CGT TCT GTA C
fimE.chip.fw	GAT TAT CTC GGG CAT CGA AA
fimA.chip.rv	AGA GCC GAC AGA ACA ACG AT
hns.fw	CTC AAC AAA CCA CCC CAA TA
hns.rv	TGG CGG GAT TTT AAG CAA GT
hns.mut.fw	TAA ACT GCA GCA ATA TCG CGA AAT GCT GAT CGC TGA CTA ATT AAG ACC CAC TTT CAC ATT
hns.mut.rv	TTG CTT GAT CAG GAA ATC GTC GAG GGA TTT ACC TTG CTC ACT AAG CAC TTG TCT CCT G
Ihf1.fw.pstI	TTG GCG GCT GCA GGC AGG CCT ATC GCA TTA TTC GCG AT
fimS.rv.pstI	GGC GGC TGC AGG ACA TGG GCA GTC GTT CTG TAC
Ihf1.mut.fw	GAG AAG AGG TTT GAT TTG GCT TAT CGA TAA TAA AGT TAA AAC AAC AAA TAA ATA CAA GAC
Ihf1.mut.rv	GTC TTG TAT TTA TTT GTT GTT TTA ACT TTA TTA TCG ATA AGC CAA ATC AAA CCT CTT CTC
Ihf1.bs.fw.BIO	CTC ATA AAC GAA AAA TTA AAA AGA GAA G
Ihf1.bs.rv	ACG TAA CTT ATT TAT GAT ATG GAC AG
IHF2.mut.fw	ATA GAA ATA ATT TAC AGC GAC CGT CTC CGG ATG CTT TTT GTC G
IHF2.mut.rev	CGA CAA AAA GCA TCG CCA CAC GCT GCC TGT AAA TTA TTT CTA T
IHF2.bs.fw.BIO	GAA GAA TGA GGA AGA TGC G
IHF2.bs.rv	CCA AAA GAT GAA ACA TTT GGG
+98proU-gfp+.fw	GGC AAT TAA ATT AGA AAT TAA AAA TCT TTA TAA AAT ATA ATG AGC GTT CTA GAT TTA AGA
+936proU-gfp+.fw	CCG CCG GAC ACC GAA TGG CTT AAT TCG TAA

	AAC CCC TTA ATG AGC GTT CTA GAT TTA AGA
+98proU-gfp^{TCD}.fw	GGC AAT TAA ATT AGA AAT TAA AAA TCT TTA TAA AAT ATG ATT GAT TAA GAA GGA GAT
+936proU-gfp^{TCD}.fw	CCG CCG GAC ACC GAA TGG CTT AAT TCG TAA AAC CCC TTG ATT GAT TAA GAA GGA GAT
proU.(stop).kan.rv	GTC CGC CGC TGG CGT GGT ATC CCA CGG ATT ATT TTG ATC ACA TAT GAA TAT CCT CCT TA
+98proU-lacZ.fw	GCA ATT AAA TTA GAA ATT AAA AAT CTT TAT AAA ATA TAA TGA GCG GAT AAC AAT TTC ACA
+936proU-lacZ.fw	CGC CGG ACA CCG AAT GGC TTA ATT CGT AAA ACC CCT TAA TGA GCG GAT AAC AAT TTC ACA
proU.(stop).cat.rv	GTC CGC CGC TGG CGT GGT ATC CCA CGG ATT ATT TTG ATC ACA TAT GAA TAT CCT CCT TAG
proU.fw	AGG GGT TGC CCT CAG ATT CTC
proU.rv	GTC AGT CGG TGC AGT CGT C
gfp+.bs.fw	ATG AGT AAA GGA GAA GAA CTT TTC
gfp+.bs.bio.rv	TTA TTT GTA GAG CTC ATC CAT G
gfp+.bs.rv	TTA TTT GTA GAG CTC ATC CAT G
gfpTCD.bs.fw	ATG AGC AAA GGC GAA GAG CT
gfpTCD.bs.bio.rv	TTA CTT ATA CAG TTC ATC CAT ACC G
gfpTCD.bs.rv	TTA CTT ATA CAG TTC ATC CAT ACC G
proV.bs.fw	AGG GTG TTA TTT TCA AAA ATA TCA C
proV.bs.bio.rv	CAT ATG CGG CAT TAA GGC AA
lacZ.bs.fw	ATG ACC ATG ATT ACG GAT TCA CTG G
lacZ.bs.bio.rv	AGC GCG GCT GAA ATC ATC AT
fimA.fw	GGA AAG CAG CAT GAA AAT TAA AAC TC
fimA.rv	GGT TAT TGA TAC TGA ACC TTG AAG G
kan.XmaI.fw	CAT AAC GAG CCC GGG TGT AGG CTG GAG CTG CTT C
kan.XmaI.rv	CAT AAC GAG CCC GGG CAT ATG AAT ATC CTC CTT A
fimA.gfpTCD.kan.fw	TCT GCA CAC CAA CGT TTG TTG CGC TAC CCG CAG CTG AAC TCA TAT GAA TAT CCT CCT TA

fimA.gfpTCD.kan.rv	GAT TGA TGC GGG TCA TAC CAA CGT TCT GGC TCT GCA GGA TTG ATT AAG AAG GAG AT
fimE.HindIII.fw	CCC AAG CTT GCA TTA TTC GCG ATG CCG GTA TTG
fimD.HindIII.rv	CCC AAG CTT CGA GTT GCT GAG TAA CGG TGA G
gfp+.fw.pstI	CCC CCT GCA GTA AGA GCG TTC TAG ATT TAA GAA GGA G
cat.rv.pstI	CCC CCT GCA GCG TCA TTT CTG CCA TTC ATC C
cat.swap.fw	TAG CAG CAC GCC ATA GTG AC
cat.swap.rv	CGG TCG CTA CCA TTA CCA GT
fimA.gfp+.cat.fw	GAT TGA TGC GGG TCA TAC CAA CGT TCT GGC TCT GCA GTA ATG AGC GTT CTA GAT TTA AGA
fimA.gfp+.cat.rv	TCT GCA CAC CAA CGT TTG TTG CGC TAC CCG CAG CTG AAC TCT ACG AGA CAG CAC ATT AAC
kan.int.fw	AAT GTC ATG ATA ATA ATG GTT TCT TAG ACG TCA GGT GGC GTG TAG GCT GGA GCT GCT TCG
kan.int.rv	AGT TCA GCT GCG GGT AGC GCA ACA AAC GTT GGT GTG CAG ACA TAT GAA TAT CCT CCT TA
gfpTCD Y66H.fw	CAA CTT TGG TGA CCA CCC TGA CCC ATG TGT TCC AGT GTT TTA GCC GTT AC
gfpTCD Y66H.rv	GTA ACG GCT AAA ACA CTG GAC ACC ATG GGT CAG GGT GGT CAC CAA AGT TG
gfp.ChIP.1.fw	GGG TCA TAC CAA CGT TCT GG
gfp.ChIP.1.fw	TTG TGC CCA TTA ACA TCA CCA TC
gfp.ChIP.2.fw	CTA CAA GAC GCG TGC TGA AGT C
gfp.ChIP.2.rv	GTT TGT GTC CGA GAA TGT TTC CAT C
gfp.ChIP.3.fw	CAT GGC ATG GAT GAG CTC TAC AAA
gfp.ChIP.3.rv	TTT CCT TAC GCG AAA TAC GGG C
gfp.ChIP.4.fw	TCA CTA CCG GGC GTA TTT TT
gfp.ChIP.4.rv	TTG AGC AAC TGA CTG AAA TGC
gfp.ChIP.5.fw	TGT CGG CAG AAT GCT TAAA TG
gfp.ChIP.5.rv	CTG CCA TTC ATC CGC TTA TT

a. BIO identifies primers that have a 5' biotin tag.

b. Restriction enzyme sites used for cloning are underlined.

2.5.2 Preparation of a high titre P1 *vir* lysate

A high titre lysate of the donor strain (strain containing the allele for transduction) was made as outlined previously (Thomason *et al.*, 2007). Donor strains were grown in 3ml of L broth containing 5mM $CaCl_2$ and 0.2% glucose at 37°C, 200rpm, to an OD_{600} 0.4-0.6. An overnight culture (1/100 dilution) or single colony was used as an inoculum. 100 µl of high titre P1*vir* (~10^9 phage ml^{-1}) was added to one of the cultures. Initial lysis was visible typically 1 hr after addition of the phage with further lysis progressing rapidly. This resulted in clearing of the culture and accumulation of clumps of cell debris. Several drops of chloroform were added to the lysate, which was then vortexed and centrifuged at 13,200 rpm for 3 min to pellet cell debris and any remaining whole cells. The supernatant was then passed through a 0.45 µM filter into a 2ml Eppendorf tube and if not used immediately, stored at 4°C.

2.5.3 Transduction with P1 *vir*

Recipient strains were grown in 2-ml cultures to turbidity (typically overnight). Cells were pelleted at 6,000 rpm for 5 min and re-suspended in 2 ml L broth containing 10ml $MgSO_4$ and 5mM $CaCl_2$.

The transduction reactions were set up as follows in 2ml Eppendorf tubes;

	A	B	C	D	E
Phage (µl)	-	1	10	100	100
Cells (µl)	100	100	100	100	-

Reaction tubes were mixed gently and incubated at 37°C for 30 min in a static waterbath, after which 200 µl of 1M sodium citrate (NaCi) was added to each tube using an Eppendorf multi-pippetor. NaCi chelates the Ca^{2+} ions that are required for phage adhesion and thus stops further infection. ~1 ml of L broth was added to each tube and incubated for 1 hr to allow expression of the antibiotic resistance gene. Cells were then centrifuged, all but ~100 µl of supernatant removed, resuspended and plated onto L agar containing 10mM NaCi and with the appropriate antibiotic selection. Reactions A and E are negative controls. Transductants were streaked to single-colony purity twice and screened by PCR for the presence of the allele or alleles of interest. Permanent stocks of the new strain were then made.

If no transductants were obtained the following points were considered. One reason for transductions using P1*vir* not working was a low phage titre. Some protocols suggest 1 hr growth or growth to an OD_{600} of ~0.2 before the addition of phage but lysis of a more turbid culture produced higher phage titres, which often considerably increased the yield of transformants. If no high-titre phage stock was available when preparing the donor lysogen, a high titre stock was prepared as follows; an L agar plate containing 5 mM $CaCl_2$ was lawned with a strain of interest, 100µl of turbid culture, and allowed to dry. 10 µl spots of a neat P1*vir* stock were placed in each quadrant of the plate, which was incubated upright overnight at 37°C. Zones of clearing were visible where the phage has replicated and lysed the cells. Using a sterile spatula, the layer of phage/cell debris was scraped into a 1.5 ml Eppendorf tube, to which ~200 µl of L broth containing 5 mM $CaCL_2$, followed by 2-3 drops of chloroform, were added. The phage/cell debris mixture was centrifuged, filter sterilized and stored as describe for liquid phage/cell debris mixtures. This method produces consistently high P1*vir* titres.

Secondly, when transducing alleles that caused a reduction in growth rate it was found that colonies from successful transductions were often not visible after overnight growth. In such cases, incubation was continued for 2 days to allow transductants colonies to become visible.

2.6 Purification and precipitation of DNA

2.6.1 Isolation of plasmid DNA

Plasmid DNA was isolated from bacterial cells using the HiYield plasmid mini-kit (RBC Biosciences) according to the manufacturers' instructions. Plasmid DNA was eluted in 50-100 µl ddH_20. The kit is based on the alkaline-lysis method of plasmid extraction. This involves lysis of the bacterial cells, denaturation of the DNA into single stranded molecules followed by a short renaturation step. Plasmid DNA is a small closed circular molecule it can easily reform under renaturing conditions. Only double-stranded DNA binds to the silica column provided in the kit, thus separating the plasmid DNA from denatured chromosomal DNA. Plasmid DNA was routinely eluted in 50µl of ddH_20 although larger plasmids >5 kb were eluted using elution buffer pre-warmed to 60°C, as described in the manufacturers' instructions.

2.6.2 Purification of linear DNA

Linear DNA, either from PCR amplification or restriction digestion of plasmid DNA was purified after electrophoresis through from TAE (40 mM Tris, 1mM EDTA, 0.114% (v/v) glacial acetic acid)/agarose gels containing 1 μg/ml of ethidium bromide. The DNA fragment of interest was identified by a brief exposure to short-wave (254 nm) UV light to minimize DNA damage caused by UV. The gel containing the DNA was cut out using a surgical blade and the DNA extracted using the HiYield gel/PCR DNA fragments extraction kit (RBC Biosciences).

2.6.3 Precipitation of DNA

Ethanol precipitation of linear DNA was routinely performed using Pellet Paint co-precipitant (Novagen) according to the manufacturers' guidelines. Pellet Paint is a visible dye labelled nucleic acid carrier that aids the precipitation and visualization of nucleic acids during alcohol precipitation. Precipitation with Pellet Paint negates the need for prolonged low-temperature incubation and thus dramatically reduces the time required for precipitation. Initially, the following ethanol precipitation protocol was used. $^{1}/_{10}{}^{th}$ volume of 3M sodium acetate, and 3 volumes 100% ethanol were added to the DNA sample. The contents of the tube were mixed by inversion and incubated at -70°C for at least 1hr before centrifugation at 13400 g for 10 min. The pellet was resuspended by vortexing in 200 μl of 70% ethanol, re-centrifuged and the ethanol aspirated away from the pellet. The pellet was then air-dried and resuspended in a suitable volume of ddH_20 or the desired buffer.

2.7 Manipulation of DNA *in vitro*

2.7.1 Restriction endonuclease cleavage of DNA

For ligation reactions, typically 0.5-2.0 μg of plasmid and linear DNA were cut with 1-2 U of restriction enzyme in a 30 μl volumes containing the appropriate reaction buffer. 1 U is defined as the amount of enzyme required to fully digest 1 μg of DNA in 1 hr. For digests involving simultaneous cleavage of DNA by two endonucleases, a buffer was chosen in which both enzymes had adequate activity (>75%) according to the manufacturer's guidelines. For double digests where no compatible buffer was available, sequential digests separated by DNA purification were performed. Reactions were typically incubated at the recommended temperature for 1-2 hr.

2.7.2 Phosphatase treatment of restriction endonuclease-cleaved DNA

Cleavage of DNA by restriction enzymes leaves phosphoryl groups on the DNA. These phosphoryl termini are required for formation of phosphodiester bonds between adjacent ends of DNA. Shrimp alkaline phosphatase (SAP; Roche) catalyses the dephosphorylation of 5' phosphates from DNA and RNA. Removal of the 5' phosphoryl group of digested plasmid DNA molecules prior to use in cloning reactions prevents self-ligation and decreases the empty-vector background in cloning experiments. SAP was used for blunt-ended cloning or when cloning into a single restriction site.

Typically DNA sufficient for a number of ligation reactions was incubated with 2 U SAP in a 20-µl volume in 1 x dephosphorylation buffer (supplied with the enzyme) and incubated for 30 min at 37°C. 5' blunt ended fragments were incubated for 1-2 hr at 37°C. SAP was irreversibly inactivated by heat treatment for 15 min at 65°C. DNA was recovered as described in section 2.6.2.

2.7.3 Ligation of DNA molecules

Bacteriophage T4 DNA ligase (Roche) was used to clone DNA inserts into plasmid vectors. T4 DNA ligase, in the presence of ATP and Mg^{2+} ions, catalyzes the formation of phosphodiester bonds between neighbouring 3'-hydroxyl and 5'-phosphate ends in double-stranded DNA. Single-stranded nicks in double-stranded DNA are also closed by T4 DNA ligase. A 1:3 molar ratio of vector DNA to insert DNA was used for cloning of sticky ends while a 1:5 molar ratio was used for cloning of blunt-ended fragments. Also, while both types of reaction were incubated for 16 hr, sticky-end ligations were incubated at 4°C and blunt-end ligations were incubated at 16°C.

Ligations were performed in 30 µl volumes to which 2 µl of T4 DNA ligase was added and the mixture incubated at room temperature for 5-10 min. Typically a standard *E. coli* cloning strain (either Dh5α or XL1-Blue) was transformed with 10 µl of the ligation reaction by the calcium chloride method (section 2.8.2).

2.7.4 DNA amplification by Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) was used for the *in vitro* amplification of DNA products. The PCR method is based on a technique in which double-stranded template DNA is thermally denatured, and then annealed to specific oligonucleotide primers flanking the region to be amplified. A thermostable DNA polymerase in the presence of

dNTPs carries out second-strand synthesis from the annealed primers. The denaturation and annealing steps are then repeated for 20-35 cycles such that subsequent rounds of strand synthesis are initiated using both the parent DNA strands and the newly-synthesised daughter strands resulting in exponential amplification of the DNA region of interest from only a few molecules of original template (Saiki *et al.*, 1988).

2.7.5 Amplification of DNA using Phusion Polymerase

Phusion DNA polymerase was used as standard for PCR reactions as it contains all the desired characteristics that were previously specific to individual DNA polymerases (André, 2009). For example, the DNA polymerase from *Thermus aquaticus* (Taq) has high processivity but does not contain proof-reading activity and is thus considered to be a low-fidelity enzyme. The DNA polymerase from *Pyrococcus furiosus* (Pfu) does have proof-reading activity and can amplify DNA with high-fidelity, but it has reduced processivity and requires longer extension times than Taq. Phusion DNA Polymerase (Phu) is a *Pyrococcus*-like enzyme that has been fused to a domain that stabilizes the association of the polymerase with the DNA, reducing the dissociation constant and allowing faster DNA amplification (Wang *et al.*, 2004). Phu has approximately 50-fold and 6-fold higher fidelity than the *Thermus aquaticus* (Taq) and *Pyrococcus furiosus* (Pfu) DNA polymerases, respectively. Calculation of annealing temperature for Phusion PCR is done differently to other commonly used polymerases and was determined using the calculator provided by Finnzymes (https://www.finnzymes.fi/tm_determination.html).

PCR reactions were carried out according to the polymerase manufacturers' instructions; briefly 50 µl PCR reactions contained 0.5 U Phusion polymerase, 200 µM dNTPs, 0.5 µM of each primer and 10-50 ng DNA template. Reactions were typically performed in the high-fidelity buffer provided with the enzyme. Reactions were performed in 500-µl thin-walled PCR tube (Sarstedt) using a Peltier Thermal Cycler, model PTC-200 (MJ Research). DNA was amplified using the following PCR cycle: Initial denaturation, 98°C for 3 min, followed by 34 cycles (98°C for 10 sec, annealing temperature for 30 sec and 72°C for 15-30 sec/kb) and a final extension of 10 min at 72°C.

2.7.6 Direct amplification of DNA from colonies or cultures

Direct amplification of DNA from colonies or cultures using PCR was carried out as described in 2.7.5. except instead of using purified DNA as template single colonies were resuspended in 20 µl ddH_2O, boiled for 5 min and 2 µl used for each PCR reaction. 50 µl of overnight cultures were also boiled for 5 min and 2 µl used for template in each PCR reaction.

2.7.7 Determination of *fimS* orientation by PCR

The orientation of the invertible *fimS* DNA element was determined by a PCR-based assay as previously described (Smith, 1999). This method exploits a restriction fragment length dimorphism, which arises due to the orientation-dependent location of a unique BstUI restriction site within the amplified DNA. The switch region was typically amplified with primers OL4 and OL20 (Table 2.3) to generate a 726-bp DNA product. DNA was amplified with Phusion polymerase (Finnzymes) using the following PCR cycle: denature 98°C for 3 min followed by 34 cycles (98°C for 10 sec , 66°C for 30 sec and 72°C for 22 sec) and a final extension of 10 min at 72°C. 10 µl of each reaction was digested with 1 µl BstUI (2U/µl) in a 20µl reaction and incubated for 1-2hrs at 60°C. Digested PCR products were resolved on 2% agarose (w/v) gels run at 100 V in TAE buffer. The size and relative concentrations of the BstUI restriction products are dependent on the orientation of the invertible promoter. Phase on cells were identified by two DNA bands on an agarose gel of 433 bp and 293 bp. Phase off cells were identified by two bands of 539 bp and 187 bp. Band intensities were quantified by densitometry using ImageJ analysis software.

2.7.8 Site-directed mutagenesis

Site-directed mutagenesis (SDM) was performed using the QuikChange II SDM kit according to the manufacturer's recommendations (Stratagene). The QuikChange II kit can be used to make single mutations or multiple point mutations, to exchange DNA sequences and to delete or insert single or multiple nucleotide base pairs into a DNA target contained on a small plasmid (<4 kb). Primers, complementary to the region of the vector to be mutated, contained 15-20 bp of homology on each side of the non-homologous nucleotides. 10-50 ng of plasmid DNA was used as a substrate with 125 ng of each oligonucleotide. A standard annealing temperature of 55°C degrees was used for all primers; if SDM was unsuccessful this was

lowered to 53°C. Following amplification, the product was digested for 2 hr with the restriction enzyme DpnI to remove the methylated parental DNA template. Only newly synthesized DNA containing the required mutation should remain, which is then transformed into an appropriate *E. coli* strain. Typically 1 µl of the 30-µl reaction was transformed by the calcium chloride method (section 2.8.2). If no colonies were obtained, the remainder of the SDM reaction was transformed and cells were allowed recover overnight before plating on appropriate antibiotic selection.

2.7.9 Quantitative PCR

Quantitative PCR (qPCR) was used to determine the enrichment of DNA after chromatin immunoprecipitation (ChIP) and was performed as described previously (Dillon *et al.*, 2010). qPCR was performed on enriched DNA and unenriched (input) DNA using the Rotor-Gene 3000 real-time PCR machine (Corbett Research) and SYBR green (QuantiTect) as per manufacturers instructions. Each PCR product was between 100- and 150 bp in length and generated only a single specific product (analyzed by agarose gel electrophoresis after a 40 cycle PCR). ChIP DNA samples were diluted 1 in 10 in AnalaR water (BDH). Each reaction was also performed using 20 ng of input DNA (4 ng/µl). The SYBR green PCR was set up in a 25 µl reaction containing 2.5 µl AnalaR water, 5 µl 1.5 µM forward and reverse primer mix, 12.5 µl SYBR green PCR mix and 5 µl of the DNA sample. The following thermocycle was used: 1. 95°C for 15sec, 2. 52°C for 60sec, 3. 72°C for 15 sec. This was repeated for 40 cycles. Cycle threshold (C_T) values were extracted and the change in cycle threshold (ΔC_T) for each probe determined as follows; $\Delta C_T = C_T$ input sample - C_T ChIP sample.

The ΔC_T of the ChIP probe with the highest C_T value (i.e. the lowest amount of amplified DNA), was normalized to 0 by adding that ΔC_T to the ΔC_T of all of the ChIP probes. This ensures a positive value for the fold enrichment calculation.

Fold enrichment was calculated as follows; Fold enrichment= $(1+\text{PCRyield})^{\Delta CT}$. The PCR yield refers to the efficiency of the amplification. Since PCR amplification is exponential, the amount of DNA is doubled in every cycle. Therefore, when the C_T values of 10-fold dilutions of input DNA are plotted against the log10 of the amount of input DNA per

reaction, the slope should be -3.32 [$2^{3.32}$=10]. When the slope approaches this number the PCR yield=1.

2.8 Transformation of *E. coli* with DNA

2.8.1 Transformation of *E. coli* with plasmid DNA

E. coli was transformed with plasmid DNA using calcium chloride coupled to heat-shock (Sambrook and Russell, 2001). This method was routinely used for transformation with purified plasmid DNA or ligation reaction products.

2.8.2 Preparation of competent cells by the calcium chloride method

1 ml of an overnight culture of the strain to be transformed was used to inoculate 100 ml of L broth and the culture was grown to an OD_{600} ~ 0.2-0.4. The cells were pelleted by centrifugation at 4000 g for 10 min and the bacterial pellet resuspended in 20 ml of ice-cold $CaCl_2$ solution (100 mM $CaCl_2$, 10% glycerol v/v). This step was repeated, after which the cells were resuspended in 2 ml of cold $CaCl_2$ solution. Cells were subsequently incubated on ice for between 1-4 hr, then aliquoted and stored at -70°C.

2.8.3 Transformation of calcium chloride competent cells

DNA to be used in transformation (10 ng-100 ng in a volume not exceeding 10 µl) was added to a sterile 2-ml Eppendorf tube containing competent cells (100- 200 µl) and incubated on ice for 30 min. The tubes were then placed with rapid warming at 42°C for 1 min and then returned to ice for 2 min. L broth was added to a final volume of 1 ml and the culture was incubated at 37°C for 1hr. Subsequently, 100 µl of the transformation mix was plated with an appropriate selection. The remainder of the transformation mixture was centrifuged for 3 min at 6,000 g. Pelleted cells were resuspended in 100 µl of L and spread onto an appropriate selective plate. Cells to which no DNA had been added but had otherwise been treated in the same way served as a control for contamination. Following overnight incubation, single colony transformants were streaked to single-colony purity on fresh selective plates.

2.8.4 Transformation of *E. coli* with linear DNA

E. coli was transformed with linear DNA, generated by PCR or by digestion of plasmid DNA, using a high-voltage electroshock treatment termed electroporation. This was used for techniques requiring high-efficiency uptake of linear DNA, primarily the creation of gene knockouts or gene fusions by allelic replacement (see section 2.16).

2.8.5 Preparation of electrocompetent cells

1 ml of an overnight culture of the recipient strain was used to inoculate 100 ml of LB broth and grown with aeration at 37°C to an OD_{600} ~0.2-0.4. The bacterial culture was chilled on ice for 15 min after which the cells were pelleted by centrifugation at 6000 g for 10 min. Cells were then resuspended in 10 ml of ice-cold sterile ddH_2O, the centrifugation step repeated and the cells resuspended in 5 ml of ice-cold sterile ddH_2O. After another centrifugation step the cells were finally resuspended in 2ml of cold 10% (v/v) glycerol. The electrocompetent cells were stored at -70°C in 50 µl aliquots.

2.8.6 Electroporation of electrocompetent cells

DNA (100ng -2 µg) to be electroporated, in a volume not exceeding 5 µl, was added to the 50 µl aliquot of electrocompetent cells and incubated on ice for 2 min. The mixture was then transferred to a pre-chilled electroporation cuvette (EquiBio) with a gap-width of 2 mm. The cuvette was then left on ice for 5 min. It was wiped with a tissue to remove any moisture and placed in the Gene Pulser chamber (Bio-Rad). A pulse, typically of 4-5 msec duration, was delivered with the Gene Pulser set to deliver 2.5 kV from the 25 µF capacitor and a Pulse Controller resistance setting of 200 Ω. 1 ml of pre-warmed L broth was added to the cuvette and the cells transferred to a sterile Eppendorf tube, which was incubated at 37°C with agitation for at least 1 hr. Cells were pelleted by centrifugation at 9,700 g for 2 min, resuspended in 100 µl of broth and plated onto LB agar plates supplemented with the appropriate selective antibiotic. Cells to which no DNA had been added were treated in the same way and thus served as a control for contamination. After incubation at 37°C overnight transformants were single-colony purified on fresh LB agar plates with the appropriate antibiotic selection.

2.9 Assays based on spectrophotometry

2.9.1 Monitoring the growth of bacterial cultures

The growth of bacterial cultures was monitored by measuring the optical density of the culture at a wavelength of 600 nm (OD_{600}) in a spectrophotometer (Genesys 10uv; ThermoSpectronic). Bacterial cultures were assayed as 1-ml fractions. Neat sample was used for dilute cultures (OD_{600} <0.4). Cultures of OD_{600} >0.4 were diluted in sterile culture medium to maintain the absorbance reading between 0.1 and 0.4, where the relationship between OD_{600} and cell count is linear. Sterile culture medium was used as a blank to account for background absorbance.

2.9.2 Assay of β-galactosidase

β-galactosidase activity of strains harbouring *lacZ* fusions was measured as described previously (Miller, 1992) but with some minor differences. Bacterial culture samples were chilled on ice for 20 min. To permeabilize cells, 500 µl Z buffer, 100 µl chloroform and 50 µl 0.1% SDS was added to each 500 µl sample in a 2 ml Eppendorf tube. The sample was mixed by votexing and placed on ice until used. β-galactosidase activity was monitored kinetically using a multiscan ascent plate reader (Thermo labsystems). Reactions were performed in 96-well flat-bottomed microtiter plates. The total volume of each reaction was maintained at 200 µl. Typically 20 µl of permeabilized cells were assayed using 180 µl of the chromogenic β-galactosidase substrate o-nitrophenyl-β-D-galactopyranoside (ONPG; 4 mg/ml in Z buffer)

The kinetics of substrate hydrolysis at 37°C was measured for at least ten samples, at 30 second intervals after an initial 3 minute lag period. The measurements were plotted (OD_{414} vs Time) and the slope of the line used to determine β-galactosidase activity according to the following equation:

Slope (OD_{414}/time) / (OD_{600} x volume (ml) of cells used)

Z-Buffer: 60 mM $Na_2HPO_4.2H_20$

40 mM $NaH_2PO_4.2H_20$

50 mM β-mercaptoethanol

10 mM KCl

1 mM $MgSO_4.7H_2O$

pH 7.0; made up to 1 litre with ddH_2O

2.9.3 Determination of nucleic acid concentrations

The concentration of DNA samples was determined using a NanoDrop device (Thermo Scientific). DNA concentration is measured by the absorbance at 260 nm (A_{260}) and reported in ng/µl. ddH_20 was used to calibrate the machine which then also measures the absorbance of the buffer to calibrate a base line.

2.10 Flow cytometry

Green Fluorescent Protein (GFP) levels in single cells were measured by flow cytometry. 20 µl of overnight cultures were added to 1 ml of 2% (vol/vol) formaldehyde/phosphate buffer saline (PBS). Samples were left at 4°C under tin-foil until measuring fluorescence using a flow cytometer (Beckan Coulter). The following detection settings were used;

	Volts	Gain
Forward scatter	55	100
Side scatter	400	500
F1	650	7.5

2.11 Fluorescence microscopy

Green Fluorescent Protein (GFP) levels in single cells were visualized using fluorescence microscopy as described previously (Cabrera and Jin, 2003). Briefly, strains to be analysed were cultured overnight in appropriate conditions. 100 µl of bacterial culture was removed, centrifuged at 4,000 *g* and the pelleted cells resuspended in 100 µl of PBS. Formaldehyde was added to a final concentration of 3%, cells were mixed and incubated at room temperature for 30 min. The fluorescent DNA stain DAPI (4',6-diamidino-2-phenylindole; 2µl of 10µg/ml solution) was added and cells were incubated for a further 15 min. DAPI stained cells were diluted 1/10 in PBS and 10 µl of the diluted cell solution pipetted onto a clean microscope slide. Cells were mounted using either ~10 µl of Vectashield (Vector Labs) or 1% TAE/agarose. Mounted cells were immediately covered with a coverslip.

Congenic strains (*gfp*+ or gfp^{TCD} inserted at the *proU* locus) cultured using identical conditions were imaged on a single slide to limit experimental error between samples.

Microscopy was performed using a Nikon Eclipse E400 microscope equiped with a 100X objective lens (Plan Fluor; Nikon) and epifluorescence filters. Images were captured using a DXM 1200 digital camera (Nikon).

Excitation of cells mounted using TAE/agarose resulted in rapid photobleaching; therefore images were acquired within 1 sec of illumination. Vectashield dramatically reduced photobleaching of GFP but were still imaged within 10 sec of illumination. No differences in initial fluorescence levels were observed between samples mounted using TAE/agarose or Vectashield.

2.12 Chromatin immunoprecipitation

Chromatin immunoprecipitation (ChIP) using a monoclonal antibody allows identification of DNA that was bound *in vivo* by a protein of interest . By purifying the antibody bound DNA-protein complexes, the DNA bound by the protein is enriched relative to DNA that is not bound by the same protein. The fold enrichment is an indication of the affinity of the protein for the DNA. Its affinity for sequences in a whole genome can be measured using ChIP, fluorescent labelling of the DNA followed by hybridization to a microarray slide (ChIP-on-chip). Alternatively, the enrichment of DNA of interest in the immunoprecipated DNA can be quantified using quantitative PCR (qPCR). ChIP using a monoclonal H-NS-specific antibody has been described previously (Dillon *et al.*, 2010) and is further described below.

2.12.1 Day1. Fixing cells, fragmenting the DNA and adding the antibody.

This protocol is for cell volumes of exponentially growing cultures in 25 ml of broth. The culture used in this study was sampled in late exponential growth phase (OD_{600} 0.7) in standard growth conditions (200rpm, 37°C). Cells were harvested by centrifugation at room temperature, at 4000 rpm for 5 min. The cell pellet washed once in PBS pre-heated to 37 °C and transferred to a glass flask. 1.355 ml of formaldehyde solution (38% stock; final concentration 1%) was added to cross-link the DNA-protein interactions. Formaldehyde was added drop-wise and left with gentle agitation for 30 min at room temperature. To stop

the cross linking reaction, 3.425 ml of 2 M glycine was added drop-wise (final concentration 0.125 M) while maintaining agitation, and left for a further 5 min. Cells were chilled on ice and centrifuged at 4°C for 10 min.

The supernatant was removed, the cell pellet resuspended in 600 µl of lysis buffer (LY) and chilled on ice. After 10 min, 1.4 ml of immuno-precipitation dilution buffer (IPDB) was added and the mixture transferred to a 5 ml falcon tube for sonication (Sanyo/MES Soniprep sonicator) using the following settings.

Amplitude (microns): 10
Number of bursts: 20
Length of bursts: 30 sec.

The sonication probe was cleaned between samples using 0.1 % SDS followed by ddH_2O (10 microns, 15 sec each) and dried before use. Samples were kept on ice at all times during sonication and then sonication probe was kept ~ 1 cm below the surface. Using these conditions, DNA was sheared to fragments ranging between 300 bp and 1000 bp.
The following steps were carried out at 4°C unless stated otherwise.
The sonicated mixture was transferred to a 2-ml Eppendorf tube and centrifuged at 13,000 rpm for 5 min. This pellets all unlysed cells (cream) and cell debris (black). The supernatant was transferred to a 15-ml falcon tube containing 1 ml IPDB. To this 50 µl of normal rabbit IgG (Upstate Biotechnology) was added and incubated for 1 hr on a rotating wheel. After this, 100 µl of homogenous (after vortexing) protein G-agarose suspension (Roche) was added using a p1000. The mixture was left to rotate for 3 hr (or overnight).
The previous two steps were performed to remove any DNA-protein complexes that non-specifically interacted with rabbit IgG. The mixture was centrifuged at 3,000 rpm for 2 min and left on ice for a further 2 min to allow the beads settle. The supernatant was removed and 1.35 ml was put into two 2-ml Eppendorf tubes.
5 µl of an H-NS specific monoclonal antibody (provided by Professor Jay Hinton) was added to 1.35 ml of chromatin. Unused chromatin was stored at -70°C. The antibody containing mixture was left rotating for 3 hr (or overnight).

2.12.2 Day 2. Binding the antibody bound DNA to, and washing the antibody bound DNA from, protein G agarose.

The chromatin-antibody mixture was centrifuged at 13,000 rpm for 5 min at 4°C. This was pelleted any precipitated SDS that would otherwise interfere with the protein G agarose. Samples were transferred to a new 1.5 ml Eppendorf tube, to which 50 µl of homogenous protein G agarose suspension was added. This mixture was incubated for 3 hr at 4°C with rotation at which point it was centrifuged at 13,000 rpm for 1 min at 4°C. (After each step involving the centrifuging of protein G agarose, the sample was left on ice for 1 min after each spin to allow the resin settle further.)

The supernatant was removed and the resin pellet washed twice with 750 µl of IP wash buffer (IPWB) 1. The pellet was then washed **once** in 750 µl IPWB 2 (a harsh detergent that removed non-specifically bound proteins), followed by 2 washes with TE, pH 8.0 (diluted IPWB 2). The sample was centrifuged at 7,500 rpm for 2 min after each wash step, with the tube being left undisturbed for 2 min on ice after each centrifugation step.

The immune complexes (DNA-protein-antibody) were eluted at room temperature from the resin by adding 225 µl of IP elution buffer (IPEB). The sample was vortexed briefly after the elution step and centrifuged at 7,500 rpm for 2 min at room temperature. This elution step was repeated and the supernatants combined in a new 2-ml Eppendorf tube.

0.2 µl of RNase A (10mg/ml stock) and 27 µl NaCl (final concentration 0.3 M) were added to the sample, which was then mixed thoroughly by pipetting and incubated at 65°C for 6 hr (or overnight). This heat step also reversed the formaldehyde cross-linking of the protein to the DNA.

After cooling to room temperature, 9 µl of proteinase K (10 mg/ml, 20 U/mg, GibcoBRL) was added to the sample and incubated at 45 °C overnight (or at least 3 hr)

2.12.3 Day 3. Phenol chloroform extraction of the enriched DNA.

Two microlitres of tRNA (5mg/ml, Invitrogen) was added to the sample immediately before the addition of 250 µl of phenol (Tris saturated, pH 8.0, Sigma), followed immediately by the addition of chloroform (do one sample completely before moving to the next sample.)

The mixture was vortexed and then centrifuged at 13,200 rpm for 5 min at room temperature. The aqueous layer was transferred to a new 2ml epindorf. 500 µl chloroform

was added (to remove residual phenol), the sample vortexed and the centrifugation and transfer steps repeated.

One microliter (5mg/ml stock) of glycogen (DNA carrier, Roche), 1 µl tRNA (5mg/ml stock, aids visualization of the DNA pellet, Invitrogen) and 50 µl (1/10 total volume) of 3 M sodium acetate (pH 5.2) was added to the recovered DNA (mixed well) before the addition of 1.35 ml 100% ethanol. The DNA was precipitated at -70°C for at least 30 min. Samples were centrifuged at 13,000 rpm for 20 min at 4°C. The DNA pellets appeared glassy or opaque and were firm. The supernatant was removed and the pellet washed with 500 µl ice cold 70 % ethanol (resuspended by inversion) before being centrifuged for 5 min at 13,000 rpm. The supernatant was removed and the DNA left to air dry at room temperature for 10 min to evaporate residual ethanol. The DNA pellets were resuspended in 50 µl of ddH_2O (BDH). 5 microlitres of the immunoprecipitated DNA was run on a 2 % TAE/agarose gel. Samples were stored at -20°C and diluted 1/5 before use in quantitative PCR reactions.

2.12.4 Buffers

Lysis buffer (LY)

- 50 mM Tris-HCl pH 8.1
- 10 mM EDTA
- 1% SDS
- Sigma protease inhibitor tablet

IP dilution buffer (IPDB)

- 20 mM Tris-HCl pH 8.1
- 150 mM NaCl
- 2 mM EDTA
- 1% Triton X-100
- 0.01% SDS
- Sigma protease inhibitor tablet

IP wash buffer 1 (IPWB1)

- 20 mM Tris-HCl pH 8.1
- 50 mM NaCl
- 2 mM EDTA
- 1% Triton X-100
- 0.1% SDS

IP wash buffer 2 (IPWB2)

- 10 mM Tris-HCl pH 8.1
- 250 mM LiCl
- 1 mM EDTA
- 1% NP-40
- 1% deoxycholic acid

IP elution buffer (IPEB)

- 100 mM $NaHCO_3$
- 1% SDS

TE (pH 8.0)

- 10 mM Tris base (pH 8.0)
- 1 mM EDTA

2.13 Proteomic analysis

2.13.1 Preparation of total cellular extracts

Crude cellular extracts for SDS-PAGE and western immunoblot analysis were prepared as described below. A sufficient volume of cells was harvested to provide a 50-µl sample after equalization of sample density to an OD_{600} of 0.2. Fifty microliters of 2x Laemmli buffer (Sigma) was added to each sample, mixed and boiled for 10 min. Samples were stored at -20°C or 10µl loaded immediately onto a 12% SDS-PAGE gel.

Laemmli 2x contains 4% SDS, 20% glycerol (v/v), 10% 2-mercaptoethanol (v/v), 0.004% bromophenol blue (w/v) and 0.125 M Tris-HCl, pH approx. 6.8, made up to 100 ml with ddH2O. The presence of SDS and 2-mercaptoethanol makes Laemmli buffer unsuitable for non-denaturing electrophoresis (section 2.15).

2.13.2 Separation of proteins by SDS-PAGE

The SDS-PAGE gels used consisted of a stacking gel containing a low percentage of acrylamide for focusing of the proteins before migration through the high percentage acrylamide matrix which separates the proteins by molecular mass. The electrophoresis apparatus used was a Mini-Protean II® vertical electrophoresis cell supplied by Bio-Rad, assembled in accordance with the manufacturers' instructions.

Briefly, a 12% (w/v) separating gel mix was made by mixing 2 ml Protogel (National Diagnostics), 1.25 ml of 1.5 M Tris-HCl (pH 8.8), 50 µl 10% (w/v) SDS and 1.648 ml of ddH_2O. Immediately prior to pouring, 50 µl of 10% (w/v) ammonium persulphate (freshly made) and 2 µl of N,N,N',N'-tetramethyl-ethylenediamine (TEMED) were added to the gel mix to promote polymerisation. The gel mix was applied between the glass plates of the electrophoresis rig using a clean 10 ml pipette. The separating gel was then overlaid with 300 µl isopropanol (100%) and allowed to polymerise for 20 min. Upon polymerization, the isopropanol was removed by inverting the gel and using absorbent paper (Protran). A stacking gel was then poured on top of the separating gel. The stacking gel was made by the mixing of 0.833 ml Protogel (National Diagnostics), 1.25 ml 0.5 M Tris-HCl (pH 6.8), 50 µl 10%(w/v) SDS and 2.667 ml of ddH_2O. Prior to pouring, polymerisation was promoted by the addition of 50 µl 10% (w/v) ammonium persulphate and 5 µl of TEMED to the gel mix. Following the insertion of a well-forming comb, gels were allowed to polymerise for 15 min.

The polymerised gel cassette was assembled in the electrophoresis chamber as directed by the manufacturer. The apparatus was filled with 1 x Tris-glycine running buffer [25 mM Tris-HCl, 250 mM glycine (electrophoresis grade) pH 8.3, 0.1% (w/v) SDS]. The comb was removed from each gel and the wells were flushed out with running buffer. 10 µl of prepared samples were electrophoretically separated on 12% SDS-polyacrylamide gels. Five microliters of Laemmli buffer was added to any lanes not containing a sample. Pre-

stained protein marker (NEB) was used to estimate the molecular mass of proteins of interest. Electrophoreses was at 90 V for 90min.

2.13.3 Visualisation of proteins in SDS-PAGE gels

Gels were removed from between the glass electrophoresis plates and a small portion of the upper left-hand corner of the gel was removed to facilitate subsequent orientation. Gels were stained using Coomasie Brilliant Blue (0.1% (w/v) Coomassie R250, 10% (v/v) glacial acetic acid, 40% (v/v) methanol). Gels were covered in Coomasie stain and microwaved at high power for 30 sec. After 10min with agitation the Coomasie stain was replaced with destain solution (1l destain; 200 ml methanol, 100 ml glacial acetic acid, 700 ml ddH_20). The gel/solution was microwaved again for 30sec. Destaining typically took 30-60min which agitation and replacing destain when it became blue. Paper tissue was usually added in the corner of the container to absorb the stain and thus increase the effectiveness of the destain.

2.13.4 Transfer of proteins to nitrocellulose membrane

Following electrophoresis, gels were removed from the glass electrophoresis plates and were covered on one side with 0.2 mM Protran nitrocellulose membrane (Schleicher and Schuell). The gel/nitrocellulose was sandwiched between absorbant paper (Protran) in a Mini Trans-blot electrophoretic transfer cell (BioRad). Transfer was at 90V for 60min in ice-cold buffer buffer (25 mM Tris, 192 mM glycine, 20% methanol) ensuring that the gel was between the cathode (-) and the nitrocellulose membrane.

2.13.5 Western immunoblot analysis

Equal loading and consistent transfer to nitrocellulose was confirmed by staining the membrane with Ponceau S for 1 min followed by extensive washing with distilled water. The membrane was then blocked for 2 hr using 20mls 5% Marvel low-fat milk powder in phosphate buffered saline (PBS). The membrane was then, incubated with shaking for 1hr at room tempterature with a primary antibody typically diluted 1/2000 in 20mls of 5% (w/v) Marvel/(PBS). The primary antibody/marvel solution was then decanted and stored at -20°C for re-use. The nitrocellulose membrane was washed repeatedly with PBS for 1hr followed by incubation with a HRP-conjugated secondary antibody, which was incubated

for 1 hr at room temperature. The 1 hr wash step was repeated before incubation for 5min using SuperSignal substrate (Pierce). Detection of emitted light was performed using developer and fixer solutions (Kodak) and Hyperfilm (Amersham Biosciences), typically requiring a 30 sec exposure.

2.14 Protein purification.

IHF purification was attempted during this study but is not referred to in the results chapters. Information provided by Professor Stephen Goodman is presented here for posterity. IHF should be purified from *E. coli* strain DH5α pHN1084 (ampicillin resistant) which contains *ihf*A and *ihf*B under the control of the IPTG inducible P*tac* promoter. Induction with 1 mM IPTG for 1 hr during mid-log phase is recommended. The purification process has been described in detail elsewhere (Nash *et al.*, 1987). A cleared lyaste with a single pass over a P-11 phosphocellulose column should result in 50% pure IHF. IHF antibody, provided by Prof. Goodman, is 20% cross-reactive with HU. Therefore, an *ihfA ihfB* double mutant [K3374: *E. coli* N99 *himD*::*cat* and Δ*himA* (no marker)] was provided as a negative control for western blots to detect IHF in eluted fractions. These strains were also kindly provided by Prof. Goodman. It is also possible to purify IHF homodimers as described previously (Werner *et al.*, 1994).

2.15 Electrophoretic mobility shift assays (EMSA)

EMSA analysis is based on the reduced electrophoretic mobility of DNA bound by protein compared with identical DNA unbound by protein. A number of protocols were used during the course of this work and they are detailed below.

2.15.1 EMSA analysis using IHF

Purified IHF (a kind gift from Prof. Steven Goodman) was stored in 0.6M potassium chloride (KCl), TrisCl pH 7.4 in 10% glycerol at -80°C. The final concentration of KCl for IHF binding to DNA should be less than 100 mM. Once the salt was dialyzed away, unless DNA was present, the IHF aggregated irreversibly. The protein would then have to be denatured and renatured to restore activity. IHF was therefore diluted in storage buffer (0.6M KCL TrisCl pH 7.4 in 10% glycerol) to the desired concentration and used as a 10X stock for EMSA.

The binding of purified IHF protein to the wild type and mutant binding sites in *fimS* DNA was tested by electrophoretic mobility shift assay (EMSA) as described previously (Yang and Nash, 1995b). Appropriate primers, one of which contained a 5' Biotin molecule, were used to ampify ~120-bp regions of DNA containing either the wild type or mutant IHF sites after which the DNA probes were gel purified. 0.5 ng of DNA were incubated in 0.5x TBE (45 mM Tris-borate [pH 8.3] containing 1.25 mM disodium EDTA), 5 μg ml^{-1} bovine serum albumin (BSA), 10% (v/v) glycerol, 60 mM KCl and 1 μg poly.di.dc with increasing concentrations (typically 0 to 30 nM) of purified IHF in a final reaction volume of 20 μl. Samples were incubated for 30 min at room temperature and loaded onto gels without the addition of loading dye. Protein-DNA complexes were resolved by electrophoresis through a 6% polyacrylamide gel (Invitrogen) for 1 hr at room temperature. The gel was then transferred to Biodyne B membrane (Pall), UV crosslinked (GS GeneLinker, BioRad) and developed using the electrophoretic mobility shift assay kit (Pierce) as per manufacturer's recommendations. While 6% polyacrylamide gels (Invitrogen) were used for IHF EMSAs, issues with supply and the short shelf-life of the gels lead to a cessation of their use in future EMSA. A recipe for non-denaturing polyacrylamide gels is provided in section 2.15.4.

2.15.2 EMSA analysis using H-NS.

H-NS was kindly provided by Professor Jay Hinton. A detailed protocol for purification of native over-expressed H-NS has been described previously (Tupper *et al.*, 1994). Purified H-NS was snap frozen in liquid nitrogen, stored at -70°C and thawed as follows; the H-NS containing vial was thawed rapidly in warm running water and once thawed, immediately placed on ice. Thawed H-NS was not re-frozen but instead maintained as a concentrated solution (1.3 mg/ml) at 4°C. Similar to IHF, H-NS aggregates in the absence of salt (NaCl) and thus was maintained in a high salt storage buffer (400 mM NaCl) until use in the EMSA. The final concentration of NaCl must be <100mM to allow efficient interaction of H-NS with the DNA.

No change in H-NS DNA binding activity was detected after 6 months at 4°C. Diluted H-NS solutions (100 mM H-NS) were typically used for 1-2 weeks before being discarded. More dilute solutions were discarded immediately after use. Two different methods were used for H-NS EMSA, which are detailed in sections 2.15.3 and 2.15.4.

H-NS storage buffer: 20mM Tris-HCL, 1 mM EDTA, 100 µg/ml BSA, 1 mM DTT, 10% glycerol and 400 mM NaCl.

H-NS reaction buffer: 20mM Tris-HCL, 1 mM EDTA, 100 µg/ml BSA, 1 mM DTT, 10% glycerol and 80 mM NaCl.

2.15.3 Competative H-NS bandshifts.

To assess the interaction of H-NS with *fimS*, regions encompassing *fimS* were PCR amplified from phase-locked substrates. Multiple fragments (~1 µg total DNA per reaction) were incubated in 20 µl reaction mixtures containing increasing concentrations of purified H-NS protein (final concentrations; 0, 7, 35, 70, 100, 135 and 165 nM). The reactions were incubated at room temperature for 30 min before loading (without the addition of loading dye) onto a 3% molecular screening agarose/TAE gel and electrophoresed at 90 V for 1 h. While low percentage agarose solutions can be easily dissolved by heating in a microwave, 3% gels molecular screening agarose is very viscous and should be dissolved using a steam oven before pouring. After electrophoresis, gels were stained for 20 min in ethidium bromide (1 µg/ml) and visualized under UV light using an AlphaImager 2200 (AlphaInnotech) gel documentation system.

2.15.4 Determining H-NS binding affinity for individual probes.

The electrophoresis apparatus used was as described for SDS-PAGE. The components that were exposed to buffers were thoroughly washed to remove residual SDS. Appropriate primers, one of which contained a 5' Biotin molecule, were used to amplify the DNA of interest. Probes used in this study were 717 bp (the equivalent to the length of the full *gfp* coding sequence).

H-NS binding to each probe (0.4 ng DNA per reaction, 50 pM) was carried out in 20 µl reaction mixtures containing increasing concentrations of purified H-NS protein. Reactions were incubated at 4°C for 30 min during which time the 5% poly-acrylamide gel [5% acrylamide/bisacrylamide (30:1) (National Diagnostics), 2% glycerol, 0.5X TBE. TEMED and APS were also added to promote polymerization as described for SDS-PAGE] was pre-run at 90V. This removes free-radicals produced during the polymerization process that could damage the DNA. 10 µl of each reaction was loaded (without the addition of loading

dye) and electrophoresed at 90 V for 2 h (4°C) followed by electrophoretic transfer (30V for 1 hr) to Biodyne B 0.45 μM membrane (Pall). 0.5X TBE (45 mM Tris-borate [pH 8.3] containing 1.25 mM disodium EDTA), was used as both running and transfer buffers. The wet membrane was UV treated twice at 150 mJoule in a GS Gene Linker UV chamber (Biorad). The Chemiluminescent Nucleic Acid Detection Module (Pierce) was used as per manufacturers' instructions followed by signal detection using developer and fixer solutions (Kodak) and Hyperfilm (Amersham Biosciences). Since H-NS binds co-operatively, DNA typically shifted over a narrow range of protein concentrations. Also since H-NS does not bind site-specifically, at high concentrations H-NS bound DNA non-specifically (Tupper *et al.*, 1994). Densitometric analysis was performed using Image J software (U. S. National Institutes of Health; http://rsb.info.nih.gov/ij/). The data were plotted as fraction of free DNA vs protein concentration to determine the apparent dissociation constants (K_{app}), which is approximately equal to the protein concentration at which half the free DNA has become bound. K_{app} were calculated by Dr. Andrew Cameron using Prism 5.0 (GraphPad Software) as described in Corcoran *et al.*, (2010). Cooperative DNA binding by H-NS was modeled using nonlinear regression analysis for site specific binding with a Hill slope. The K_{app} values reported include ± standard deviation.

2.16 Allelic replacement using linear DNA.

Unlike some bacteria, *E. coli* is not readily transformable by linear DNA (Smith *et al.*, 1981). This is due in part to degradation of incoming DNA by intracellular exonucleases (Hoekstra *et al.*, 1980). The *recB*, *recC* and *recD* genes of *E. coli* encode the subunits of the exonuclease V (ExoV) protein that can perform several enzymatic activities (Kowalczykowski *et al.*, 1994). The RecBCD complex promotes homologous recombination and mutations in either *recB* or *recC* result in decreased recombination (Willetts and Mount, 1969; Lloyd and Buckman, 1985). ExoV activity is also missing in these mutants. ExoV activity can be removed by mutations in *recD* but this mutation does not affect recombination (Chaudhury and Smith, 1984; Amundsen *et al.*, 1986; Biek and Cohen, 1986). This discovery allowed *recD* derivatives of *E. coli* to be used in gene disruption experiments using linear DNA (Russell *et al.*, 1989).

2.16.1 Allelic replacement using *recD* strains

recD deficient strain (VL386*recD*) has previously been used for incorporation of mutations into *fimS* (Smith, 1999; Kelly *et al.*, 2006). In this study, mutations were initially introduced by site-directed mutagenesis into *fimS* cloned on a small plasmid (pCPC2 or pMMC108) before being cloned into pSGS501 (Smith, 1999) or one of its derivatives (Kelly *et al.*, 2006). These plasmids were subsequently digested with EcoRV releasing an 8.5-kb fragment including the *fimB* gene interrupted by the *kan* gene and the *fimS* mutations. The digestion products were electrophoresed through 0.7 % (w/v) agarose/ TAE, to allow clear separation of the *fim* fragment from the linearized plasmid backbone (4.25-kb). The *fim* fragment was then gel extracted and precipitated using Pellet Paint co-precipitant (Novagen). Roughly two micrograms of this fragment was electroporated into strain VL386*recD* in a volume not exceeding 5 µl. Transformants were selected on L agar containing kanamycin and after single colony purification on MacConkey-lactose agar, phase locked-off colonies (white) were screened for the presence of the mutations by PCR and digestion at novel sites. Presumptive integrants were then sequenced (GATC Biotech) to ensure the correct series of mutations were present.

2.16.2 Allelic replacement using lambda-Red.

The ability of the λ phage to recombine in recombination-deficient *rec* strains indicated that the phage must encode proteins for its own recombination (Smith, 1988). The λ phage encoded genes γ, β and *exo* were identified in screens for recombination deficient λ phage (λ Red) (Smith, 1988) and can be used to promote integration of linear DNA in *E. coli* (Murphy, 1998; Datsenko and Wanner, 2000). The λ Red genes enhance recombination of linear DNA compared to *rec* mutants (Murphy, 1998) and their expression in a host cell such as *E. coli* allows integration of PCR products with limited (~40-bp) homology (Datsenko and Wanner, 2000).

Allelic replacement using linear DNA was as described previously (Datsenko and Wanner, 2000). Briefly, strains for integration were transformed with the pKD46, which contains the λ Red genes under the control of an arabinose inducible promoter. pKD46 has a temperature sensitive origin of replication and was thus maintained at 30°C to allow it to be propagated. Strains were grown to exponential phase (OD_{600} ~0.2) and arabinose was added to a final concentration of 0.2%. Cells were harvested after 1hr as described for

electroporation. Primers suitable to amplify the DNA to be integrated were designed with ~40-bp flanking regions homologous to the insertion site. DNA was gel purified and ~200 ng was used for electroporation as described in section 2.8.4. All integrations were confirmed by PCR amplification and DNA sequencing (GATC biotech). All genetic recombinants made using this method were P1 transduced into an isogenic background as induction of the λ Red genes could have led to unrecognized genomic re-arrangements in the original strain.

Chapter 3.

DNA relaxation-dependent phase biasing of the *fim* genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP.

3.1 Introduction

Type 1 fimbriae contribute to virulence in pathogenic Gram-negative bacteria and are associated with the establishment of an attached, as opposed to a planktonic, lifestyle (Wright *et al.,* 2007). Type 1 fimbriae promote bacterial adherence and persistence by binding to glycosylated mannose residues in host epithelial cells and are important components in the initial stages of biofilm formation (Wright *et al.*, 2007). Uropathogenic *E. coli* that cannot express type 1 fimbriae are attenuated for virulence (Connell *et al.*, 1996; Gunther *et al.*, 2002). Type 1 fimbriae are highly immunogenic (Brinton, 1959; Eisenstein *et al.*, 1983) and their expression represents a major metabolic burden accounting for up to 8% of total cell protein in fimbriate cells (Schembri and Klemm, 2001b). The phase-variable expression of type 1 fimbriae is an important determinant of virulence in uropathogenic *E. coli* (Gunther *et al.*, 2001). Fimbrial gene expression in *E. coli* is controlled through inversion of a 314-bp invertible DNA element (*fimS*) containing the promoter for the major fimbrial subunit gene *fimA* (Abraham *et al.*, 1985), leading to either fimbriate (phase-on) or afimbriate (phase-off) cells (Fig. 1.5). In $FimE^{+}$ cells the higher unidirectional switching rate of FimE compared to FimB results in *fimS* remaining predominantly in phase off (McClain *et al.*, 1991). *fimE* expression is post-transcriptionally repressed by *fimS* in the phase-off orientation (Sohanpal *et al.*, 2001; Joyce and Dorman, 2002; Hinde *et al.*, 2005). Once FimE levels decrease FimB-mediated switching to phase on can occur. Switching of *fimS* to phase on relieves the repression of *fimE* expression and allows FimE to accumulate. This cyclic mechanism of fimbrial gene expression is termed orientational control and provides an efficient mechanism for controlling phase variable expression of type 1 fimbriae (Wolf and Arkin, 2002; Chu and Blomfield, 2007). However mutations in *fimE* occur frequently (Blomfield *et al.*, 1991b) and in the absence of the FimE recombinase, inversion of *fimS* is catalysed exclusively by FimB. Under standard growth conditions FimB catalyses inversion of *fimS* from phase-on to phase-off and phase-off to phase-on with approximately equal efficiencies. There are, however, circumstances when FimB-mediated DNA inversion displays a marked orientational bias. For example, inversion of *fimS* becomes biased towards phase-on in response to DNA relaxation and this is contingent on LRP binding within *fimS* (Dove and Dorman, 1994; Kelly *et al.*, 2006; Corcoran and Dorman, 2009; Müller *et al.*, 2009). FimB-mediated switching can thus be considered programmable with multiple regulatory signals combining to bias the switching outcome. The roles of DNA relaxation, IHF, LRP and H-

NS in biasing the switching outcome of FimB-mediated inversion of *fimS* were investigated.

3.2 Results

3.2.1 Allelic replacement of *fimS*

To assess accurately the role of the site-specific binding proteins IHF and LRP in the biasing of FimB-mediated *fimS* inversion it was necessary to inactivate the binding sites of these proteins in their native locations within the invertible element. The strategy for allelic replacement of NAP binding sites in *fimS* (Smith, 1999; Kelly *et al.*, 2006) is outlined in section 2.16.1 and Fig. 3.1. Mutations were initially created by SDM on a suitable small plasmid containing the *fimS* region (phase off).

While an existing plasmid (pMMC108) was suitable for mutation and cloning of sites within *fimS* it did not contain a unique restriction site upstream of IHF-1to allow cloning of the mutated site into pSGS501. The *fimS* region was amplified from a phase off substrate (pSGS501) using primers IHF1.fw.PstI and fimS.rv.PstI to amplify a ~700 bp region including the unique StuI site located upstream of IHF-1. This PCR product contained flanking PstI restriction sites that were used to clone the *fim* fragment into PstI cut pMMC106 to create pCPC2. This plasmid was suitable for mutagenesis of the IHF-1 site. The mutated sites were cloned into the *fim* locus of a larger plasmid (pSGS501) that contained an 8.5-Kb region of the *fim* operon. Digestion of this plasmid with EcoRV released the *fim* fragment which was integrated into the chromosome of a *recD* deficient strain. Homologous recombination of the linear DNA replaced the native *fimB* gene with an insertionally-inactive copy containing a gene encoding kanamycin resistance (*kan*) and, depending on the location of the downstream cross-over event, replaced wild-type *fimS* with the version containing mutated protein binding sites. Switching was restored by supplying *fimB* in *trans* using pSLD203 (Dove and Dorman, 1996). pSLD203 is essential to restore efficient switching in strains lacking IHF and LRP (Dove and Dorman, 1996). Inactivation of IHF and LRP binding sites causes a dramatic (>50 fold) reduction in FimB and FimE mediated inversion of *fimS* (Blomfield *et al.*, 1993; Gally *et al.*, 1994; Blomfield *et al.*, 1997). IHF and LRP are believed to play a key role in alignment of the inverted repeats, which is a prerequisite for site-specific recombination (Fig. 1.6). Efficient

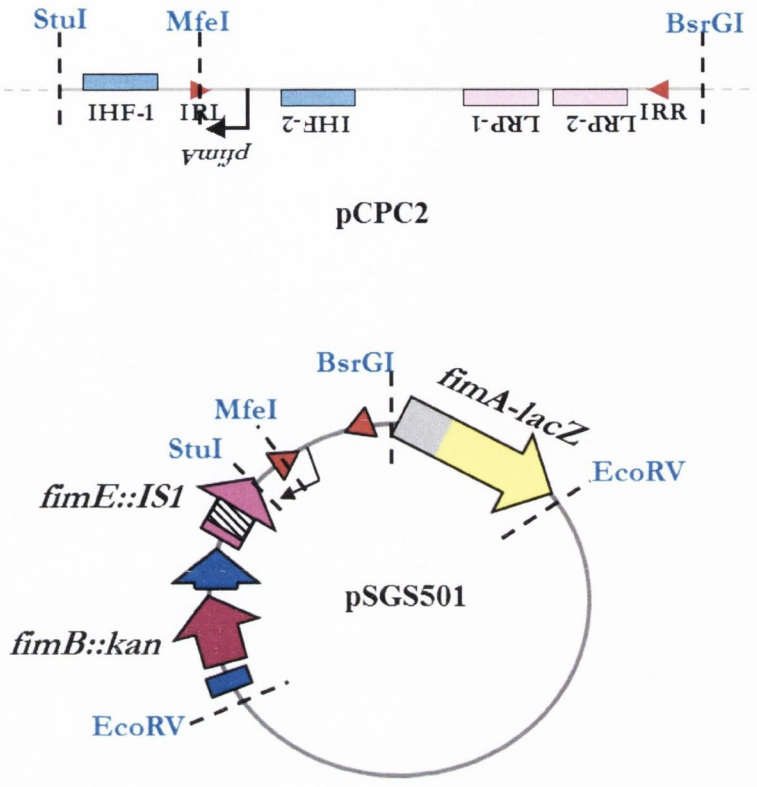


Fig. 3.1 Strategy for allelic replacement of nucleoid-associated protein binding sites in *fimS*. Mutations are introduced into *fimS* located on a small plasmid (pCPC2) by site-directed mutagenesis. Restriction sites flanking the mutations (StuI, MfeI, BsrgI) were used to clone the mutated site(s) into the identical location on pSGS501. For clarity *fimS* details are only shown on the pCPC2 fragment. pMMC108 is similar to pCPC2 but does not contain the StuI site and was thus used only to mutate sites within *fimS*. Digestion of pSGS501 with EcoRV released an 8.5-kb fragment containing *fimB*::*kan*, *fimE*::*IS1*, *fimS* and *fimA-lacZ*. The linear DNA was integrated into the chromosome (section 2.16.1). Integration of mutated *fimS* sites was confirmed by PCR amplification and digestion using unique restriction sites and sequencing.

switching can be restored in IHF and LRP deficient cells by over-expression of FimB in *trans* using pSLD203 (Dove and Dorman, 1996). This may involve increased co-operativity between FimB molecules stabilizing FimB interactions with sites flanking the IRs and facilitating the replacement of IHF- and LRP-induced bends with FimB-induced bends that allow for efficient synapsis (Dove and Dorman, 1996).

The increase in FimB levels also causes a dramatic increase in switching frequency in WT cells. Importantly, the novobiocin induced phase-on bias, which was first demonstrated in strains expressing wild type levels of FimB, is not affected by over-expression of FimB (Dove and Dorman, 1994; Kelly *et al.*, 2006).

3.2.2 IHF binding to site IHF-1 is required to bias *fimS* inversion to the on phase.

Previous work has shown that when DNA gyrase is progressively inhibited by novobiocin treatment, the *fimS* element shows a novobiocin-dose-dependent bias towards the on phase (Dove and Dorman, 1994; Kelly *et al.*, 2006; Müller *et al.*, 2009). In addition, it has been shown in earlier work that this off-to-on bias is contingent on binding by the LRP protein to two adjacent sites, LRP-1 and LRP-2, within *fimS* (Kelly *et al.*, 2006) (Fig. 1.7). The two LRP binding sites are located asymmetrically within *fimS,* raising the possibility that they may contribute to a structure that distinguishes the phase-on switch from the phase-off switch. This hypothesis postulates the existence of a reference point that somehow communicates with the LRP-occupied LRP-1 and LRP-2 sites in the phase-on switch. An inspection of the DNA adjacent to *fimS* revealed that, in the on phase, the two LRP binding sites lie close to a well characterized binding site for IHF (Fig. 1.7).

The DNA sequence of this IHF binding site (IHF-1) was a strong match to the consensus for such sites (Dorman and Higgins, 1987; Hales *et al.*, 1994). IHF-1 was disrupted by site-directed mutagenesis as described (section 2.7.8) using IHF-1mut.fw/rv and pCPC2 as template. The wild-type and mutant sequences were then compared for their ability to bind purified IHF protein in an electrophoretic mobility shift assay (EMSA; see section 2.15.1). 120-bp biotinylated DNA probes containing either the wild type or mutant IHF-1 site were amplified using primers Ihf1.bs.fw.BIO and Ihf1.bs.rv with plasmids pCPC2 or pCPC2IHF-1^{-} serving as templates. The EMSA results showed that the base substitutions had abrogated the ability of IHF to bind to the mutant sequence (Fig. 3.2). The altered sequence was then substituted for the wild type version on the chromosome and

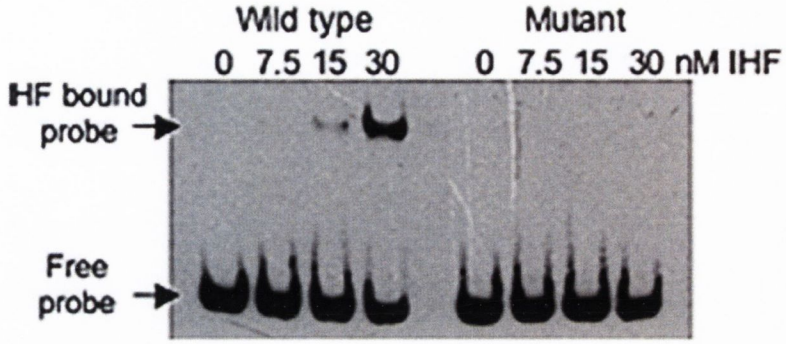
PCR amplification and DNA sequencing were used to confirm the presence of the mutated IHF binding site in the chromosomal *fim* locus.

The effect of the IHF binding site mutation on *fimS* inversion following DNA gyrase inhibition was tested using a PCR-based inversion assay (section 2.7.7) with cultures incubated in the presence of increasing concentrations of the gyrase inhibitor novobiocin. The results showed that mutation of the IHF-1 binding site caused *fimS* inversion to become strongly biased in favour of the off orientation (Fig 3.3). This outcome was identical to that seen previously when the LRP-1 and LRP-2 sites were mutated to prevent LRP binding to *fimS* (Kelly *et al.*, 2006). These data supported the hypothesis that IHF and LRP were required collectively when DNA was relaxed to maintain *fimS* in its on orientation.

It was anticipated that elimination of IHF binding to its IRL-proximal IHF-1 site, together with abrogation of LRP binding to sites LRP-1 and LRP-2, would result in a strong phase-off bias in *fimS* inversion following DNA gyrase inhibition. To test this hypothesis, mutations that prevented IHF binding to its IRL-proximal site and mutations that abolished LRP binding to sites LRP-1 and LRP-2 were all introduced into the *fim* locus on the chromosome. Application of the PCR-based *fimS* inversion assay following DNA gyrase inhibition with increasing concentrations of novobiocin showed that the *fim* switch adopted a phase-off orientation, as expected (Fig 3.3)

3.2.3 IHF binding to site IHF-2 is not required to bias *fimS* towards the on phase

IHF binds to a second site (IHF-2) that is located within *fimS* and proximal to P_{fimA} (Fig. 1.5) (Dorman and Higgins, 1987; Eisenstein *et al.*, 1987; Blomfield *et al.*, 1997). Mutations aimed at disrupting IHF binding were introduced by site-directed mutagenesis using primers ihf2mut.fw/.rv and pMMC108 as a template. The wild type and mutant sequences were then compared for their ability to bind purified IHF protein in an electrophoretic mobility shift assay (Fig. 3.4). ~150-bp biotinylated DNA probes containing either the wild type or mutant IHF-2 site were amplified using primers Ihf2.bs.fw.BIO and Ihf2.bs.rv with plasmids pMMC108 or pMMC108IHF-2$^-$ serving as templates. The mutated IHF-2 site was shown to no longer be bound by IHF at physiologically relevant concentrations and was substituted for the wild-type version on the



IHF consensus	TTTTTTTNNNNNNNNN$^{T}_{A}$ATCAANNNNTT$^{A}_{G}$
WT *fim* IHF-1 site	TTTTTTTAACTTTATTATCAATAAGTTA
Mutant *fim* IHF-1 site	TTGTTTTAACTTTATTATCGATAAGCCA

Fig. 3.2 Inactivation of the IHF-1 binding site. The DNA sequence of the IHF-1 binding site (WT) for the IHF protein is shown aligned with the consensus for IHF binding site sequences together with the base pair substitutions made in IHF-1. Matches to the consensus sequence are shown in blue. The mutated base pairs are shown in red. The effect of these mutations on IHF binding to IHF-1 site DNA was tested by electrophoretic mobility shift assay. The concentration of purified IHF used is indicated above each lane. The unshifted biotinylated DNA (free probe) and the IHF-DNA complex are shown by arrows.

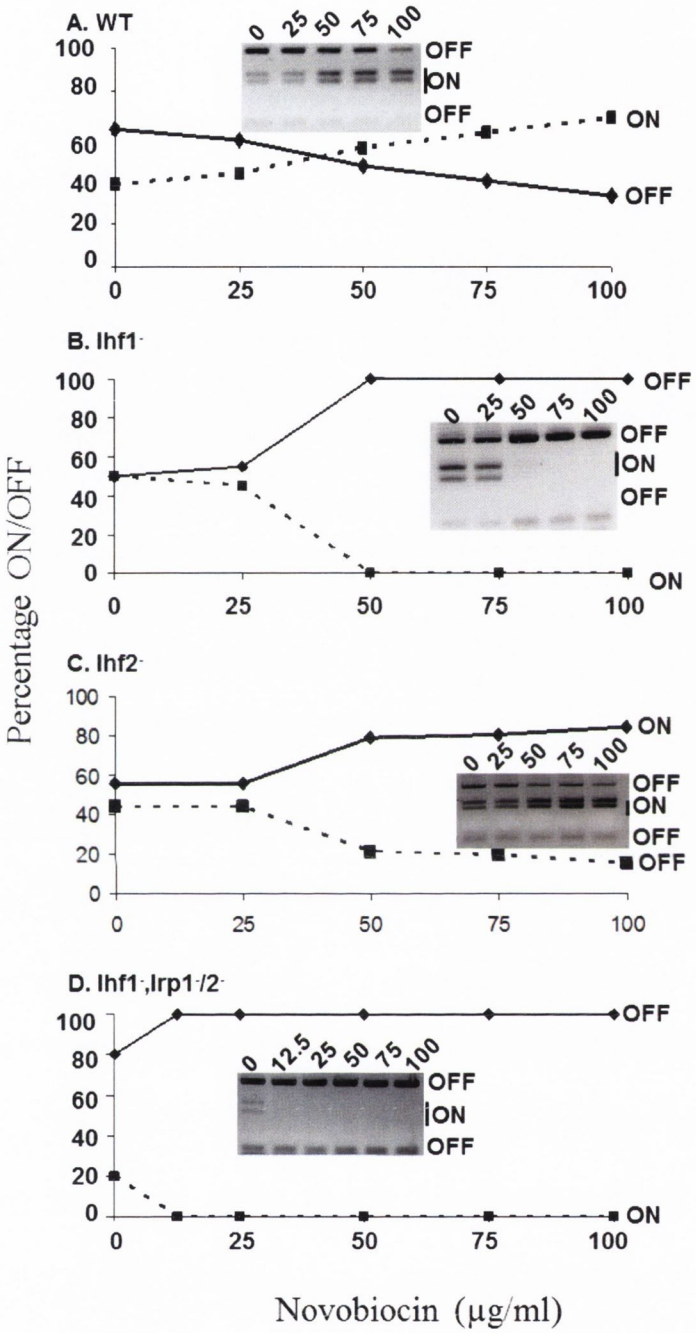
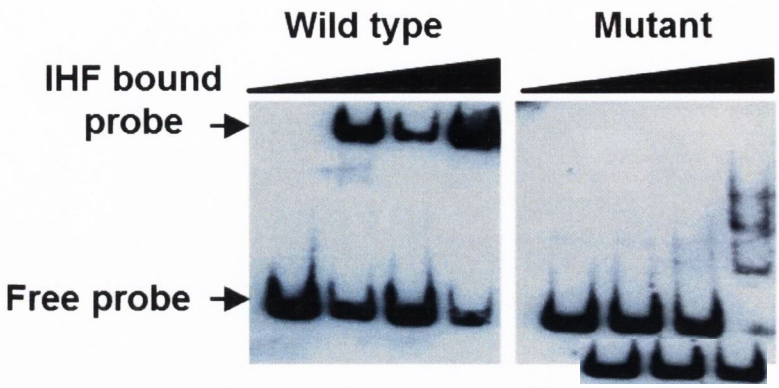


Fig. 3.3 IHF binding affects inversion preferences of *fimS*. The PCR products diagnostic of the on and off forms of the fim switch were resolved by electrophoresis on agarose gels (insets). The concentration of novobiocin (μg/ml) used to treat the culture prior to PCR is given above each lane. Phase-on-specific and phase-off-specific DNA bands (two each) are labelled. The bands were scanned with a densitometer and the data plotted to summarize the effect of novobiocin treatment on *fimS* phase biasing. Data are presented for the wild type (A), the mutant fim switch lacking the IHF-1 site (B), the IHF-2 site (C) and the IHF-1 site and the LRP-1 and LRP-2 sites (D).



IHF consensus	TTTTTTTNNNNNNNNN$^{T}_{A}$ATCAANNNNTT$^{A}_{G}$
WT *fim* IHF-2 site	TAGAAATAATTTACATATCAAACAGTTA
Mutant *fim* IHF-2 site	TAGAAATAATTTACAGGCAGCGTGTGGC

Fig. 3.4 Inactivation of the IHF-2 binding site. The DNA sequence of the IHF-2 binding site (WT) for the IHF protein is shown aligned with the consensus for IHF binding site sequences together with the base pair substitutions made in IHF-2. Matches to the consensus sequence are shown in blue. The mutated base pairs are shown in red. The effect of these mutations on IHF binding to IHF-2 site DNA was tested by electrophoretic mobility shift assay using purified IHF (left to right; 0, 30 60, 90 nM IHF). The unshifted biotinylated DNA (free probe) and the IHF-DNA complex are shown by arrows.

chromosome. IHF binding to low-affinity sites was observed when the mutated IHF-2 site was incubated with high concentrations of IHF (90nM) (Fig. 3.4). This was consistent with previous observations (Blomfield *et al.*, 1997). These low affinity sites are not occupied in the wild type system and are not believed to be physiologically relevant (Blomfield *et al.*, 1997).

The removal of IHF binding at site 2 had no effect on phase biasing of *fimS* following DNA gyrase inhibition (Fig 3.3C). This was consistent with the theory that DNA relaxation in combination with IHF bound to site 1 and LRP bound to sites 1 and 2 form a nucleoprotein complex at the IRL that is unique to the phase on orientation and is responsible for the biasing of switching towards phase on.

3.2.4 Construction of a tetracycline resistant *hns* mutant.

To test the effects of H-NS on phase biasing of *fimS*, a *hns* mutant allele that contained a compatible marker was required. The widely used *hns206*::*amp* (Dersch *et al.*, 1994) allele was unsuitable since selection for pSLD203 is through resistance to carbenicillin. The *fim* NAP site mutant strains also contain a *kan* resistance marker in *fimB*. While the NAP mutation strains were all constructed in a *recD*::Tn*10* strain the mutations were subsequently transduced into VL386 (tetracycline sensitive) making tetracycline resistance a suitable selectable marker. The *tetRA* element from Tn*10* was amplified using primers hns205.tetRA.fw/.rv producing a linear DNA fragment that contained *tetRA* flanked by regions of DNA sequence homology to the *hns* open reading frame. The *tetRA* element was integrated using lambda-*red* mediated homologous recombination (section 2.16.2) (Datsenko and Wanner, 2000) at base pair 205 of *hns* in VL386. This deleted those parts of the *hns* gene that encoded the linker and DNA binding domains of the H-NS protein and introduced an in-frame stop codon to create a 7.9 kDa H-NS truncate while also conferring tetracycline resistance. The correct integration of the *tetRA* cassette was confirmed by PCR (Fig. 3.5) and by DNA sequencing (GATC biotech). The mutant allele was P1 transduced into a fresh background (VL386) and tested for the rapid switching of *fimS* on MacConkey lactose plates. The mutant showed the expected phenotype of uniform intense red colonies (data not shown). Western blot analysis using monoclonal anti-H-NS antibodies was performed (Fig. 3.5). Whereas H-NS was detected in the wild type sample, the expected truncated proteins were not detected in *hns*205::*tetRA* or in the *hns*206::*amp* control (Fig. 3.5). This was possibly due to the specificity of the monoclonal antibody for a region of H-NS which was deleted in the mutant alleles.

3.2.5 H-NS modulates *fimS* inversion directionality.

Previous research has shown that the nucleoid-associated protein H-NS influences the inversion of the *fimS* DNA element (section 1.8.6); H-NS deficient cells show increased rates of FimB-mediated inversion (Higgins *et al.*, 1988; Kawula and Orndorff, 1991; Donato *et al.*, 1997; Schembri *et al.*, 1998). Although transcription of *fimB* increases in the absence of H-NS (Donato *et al.*, 1997) the associated increase in FimB protein levels is not the primary cause of the rapid-switching phenotype that is characteristic of *hns* null alleles (Dove and Dorman, 1996; Donato *et al.*, 1997). H-NS binding in *fimS* is believed to repress inversion directly by an unknown mechanism (Kawula and Orndorff, 1991; O' Gara and Dorman, 2000). While H-NS has been described in detail as a repressor of transcription (Dorman, 2004) (section 1.8.5) its contributions to DNA recombination have been studied much less intensively.

Inactivation of the *hns* gene led to a dramatic increase in phase-on cells in the population. H-NS deficient populations consistently contained ~80% phase-on cells and their *fimS* were unresponsive to novobiocin induced DNA relaxation (Fig. 3.6A). This ~80% phase-on bias in H-NS deficient cells was previously observed by O'Gara and Dorman (2000). These results provide genetic evidence that H-NS represses switching from phase-off to phase-on. H-NS represses transcription from P_{fimB} and thus H-NS deficient cells have increased levels of FimB (Donato *et al.*, 1997). The effect of overexpressing FimB on *fimS* inversion in the absence of H-NS was tested using pSLD203. Overexpression of FimB alleviated the 80% phase-on bias characteristic of H-NS deficient populations and restored the responsiveness of *fimS* inversion to DNA relaxation (Fig. 3.6B). These data indicated that H-NS has a direct role in repressing switching from phase-off to phase-on, which was alleviated by the over-expression of FimB.

To test the role of H-NS in creating the IHF/LRP dependent DNA relaxation phase-on bias, the *hns205*::*tetRA* allele was P1 transduced into the strains containing NAP mutations. The results showed that in the strain lacking H-NS and without a functional IHF-1 site, *fimS* displayed a strong on-to-off bias (Fig. 3.6C), a pattern that was very similar to that seen in the absence of the IHF-1 binding site alone (Fig 3.3B). These data showed that IHF was required for the maintenance of the DNA relaxation induced phase on bias regardless of the status of the *hns* gene.

In contrast, removal of H-NS had an effect on *fimS* inversion in strains containing mutations in LRP sites LRP-1 and LRP-2. Strains lacking H-NS and LRP binding at sites

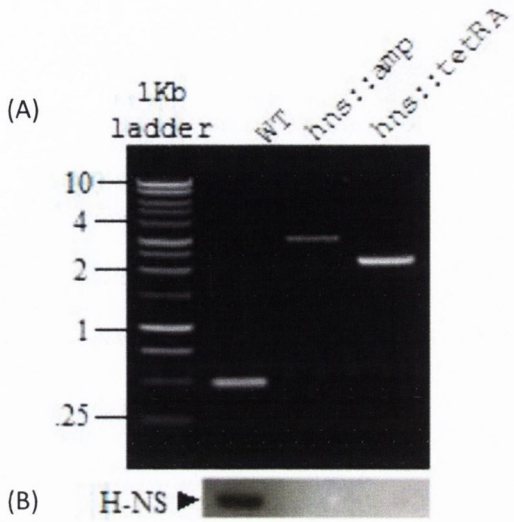


Fig. 3.5 Construction of a tetracycline resistant H-NS mutant. The presence of mutant *hns* alleles in VL386 was confirmed by (A) PCR and (B) western blot. Wild-type (WT) VL386 and VL386 containing the *hns*::*amp* allele from PD32 (Dersch *et al.*, 1993) was used as a control in both experiments.

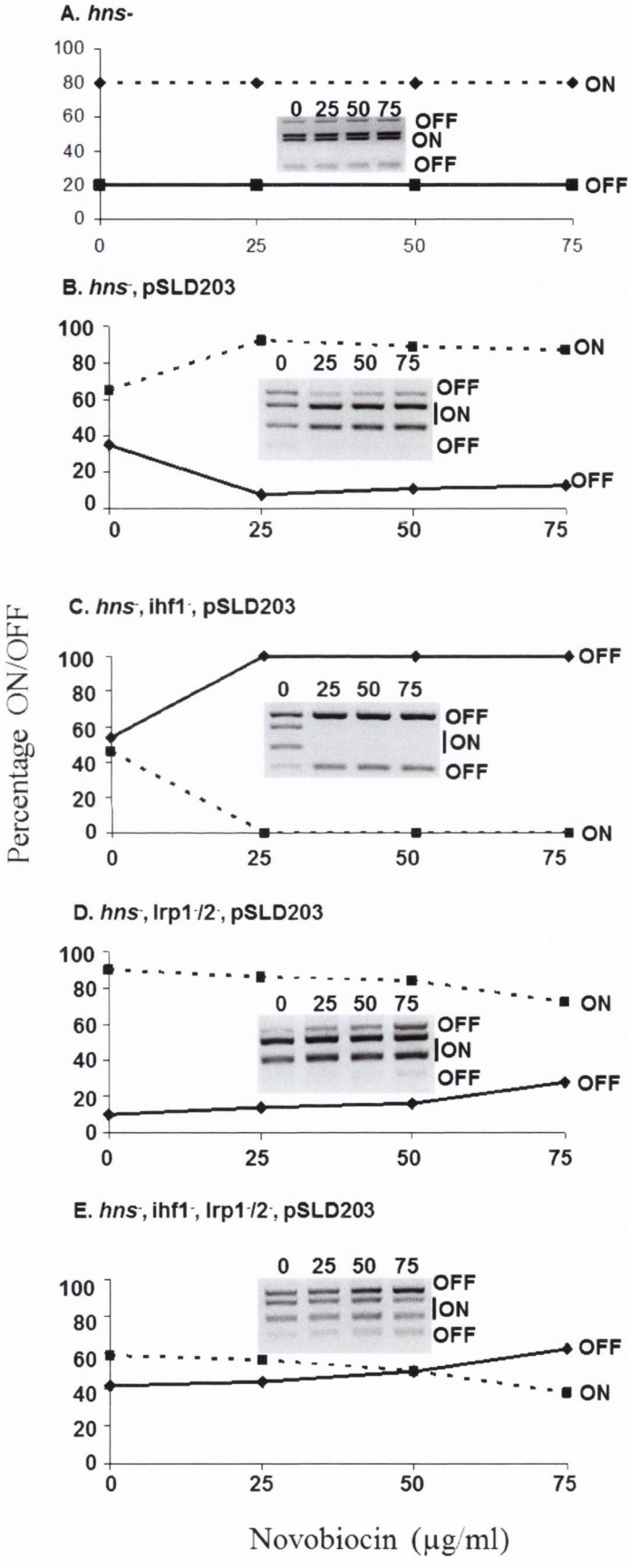

A. hns-
100
80
60
40
20
0
0 25 50 75
ON
OFF
ON
OFF
OFF
0
25
50
75
B. hns-, pSLD203
ON
OFF
ON
OFF
C. hns-, ihf1-, pSLD203
OFF
ON
D. hns-, lrp1-/2-, pSLD203
E. hns-, ihf1-, lrp1-/2-, pSLD203
Percentage ON/OFF
Novobiocin (μg/ml)

Fig. 3.6 Switch inversion preferences in the absence of H-NS. The PCR products diagnostic of the on and off forms of the *fim* switch were resolved by electrophoresis on agarose gels (insets). The concentration of novobiocin (μg/ml) used to treat the culture prior to PCR is given above each lane. Phase-on-specific and phase-off-specific DNA bands (two each) are labelled. The bands were scanned with a densitometer and the data plotted to summarize the effect of novobiocin treatment on *fimS* phase biasing. Data are presented for the *hns* mutant (A), the *hns* mutant with pSLD203 (B) the *hns* mutant with pSLD203 with *fim* switch lacking the IHF-1 site (C), the LRP-1 and LRP-2 sites (D) and the IHF-1 site and the LRP-1 and LRP-2 sites (E).

LRP-1 and LRP-2 showed reduced bias towards phase-off (evidenced by the clear increase in phase-on cells, even in the absence of novobiocin). This combination of mutations also showed a reduction in responsiveness in *fimS* inversion to novobiocin compared with results from previous studies (Kelly *et al.*, 2006). The difference in switching pattern is most clearly seen in strains lacking functional IHF-1, LRP-1 and LRP-2 sites. In the presence of H-NS *fimS* is strongly biased to phase off even in the absence of novobiocin (Fig. 3.3D). When H-NS is removed, bidirectional switching is restored (Fig. 3.6E). This is in contrast to strains containing only the IHF-1 site mutation, where the presence or absence of H-NS has little effect on the *fimS* orientational bias. These data showed that LRP and H-NS collaborated in sustaining a phase-off trap.

3.2.6 H-NS interaction with *fimS* is phase variable

To assess the interaction of H-NS with *fimS*, ~760-bp and ~670-bp regions encompassing *fimS* were PCR amplified from phase-locked substrates using primer pairs Ihf1.int.fw/fimS.rv and fimE.chip.fw/fimA.chip.rv respectively. The PCR products were then sub-divided into shorter segments by restriction enzyme digestion for use in competitive electrophoretic mobility shift assays with purified H-NS protein (section 2.15.3). Since the *fim* region with either phase-on or phase-off switches had been amplified using 2 different primer pairs, key DNA segments were present in probes of multiple sizes (Fig. 3.7A). This broad range experiment demonstrated binding of H-NS to the IRL in the phase-off orientation as both of the probes that contained this region (550-OFF, 340-OFF) were shifted with increasing H-NS concentrations (Fig. 3.7A).

H-NS binding in this region was tested in a more refined experiment where a single pair of primers was used to amplify the ~760-bp region compassing *fimS* in the phase-off orientation (Fig. 3.7B). Digestion with Hpy99I separated the PCR product into two fragments of 550-bp and 210-bp respectively, each containing one of the inverted repeats. A separate restriction enzyme digestion with MfeI cleaved the DNA within IRL to produce two fragments of 400-bp and 370-bp (Fig. 3.7B). Incubation of all of these DNA fragments with increasing amounts of H-NS clearly showed high affinity H-NS binding to the 550-bp fragment that contained the intact IRL and its flanking DNA with the P_{fimA} promoter in the phase-off orientation (as had been suggested in Fig. 3.7A). The lack of binding in the MfeI-cut DNA fragment indicated that H-NS bound to a site that overlapped the IRL in the phase-off orientation, and that *fimS* alone (i.e. the DNA extending from IRL to IRR only) was not sufficient for H-NS binding (i.e. the 400-OFF fragment did not shift).

Evidence suggesting that H-NS did not interact with IRR-proximal DNA sequences when *fimS* was in the phase-on orientation had been obtained with DNA digested with SnaBI, an enzyme that cut the DNA very close to IRR. The experiment was redesigned to leave more DNA flanking IRR in the phase-on switch. A single ~760-bp DNA fragment amplified from a phase-locked-on substrate was digested with MfeI to produce a fragment containing the entire *fimS* region (including the *fim* promoter), IRR and upstream flanking DNA (fragment 400-ON). This DNA segment was bound strongly by H-NS and resolved as the H-NS-DNA complex indicated by the open arrowhead (Fig. 3.8 A). A comparison of Fig. 3.7B and Fig. 3.8 A clearly showed the phase variable nature of H-NS binding to the *fim* switch: DNA fragment 400-ON (Fig. 3.8A) was identical to 400-OFF (Fig. 3.7B) except for *fimS* orientation, and 400-ON was bound strongly by H-NS whereas 400-OFF was poorly bound by the protein.

The invertible sequence extending from IRL to IRR is common to both the phase-on and phase-off forms of the switch. The data obtained thus far suggested that this was unlikely to support H-NS binding alone because DNA sequences in the static portion of the chromosome flanking IRL or IRR were also needed. To test this hypothesis, the invertible element was precisely excised from between IRL and IRR by restriction enzyme digestion with HaeIII and tested for H-NS binding. The results showed that H-NS was unable to bind to the invertible region when this was isolated from its flanking DNA sequences (Fig. 3.8B).

Overall, the data from the EMSA analysis showed that H-NS bound to a site that moved from one end of *fimS* to the other as the switch inverted. This site consisted of a constant component that was located within *fimS* close to the P_{fimA} promoter and a variable component that was provided by DNA sequences outside *fimS* and adjacent either to IRL or to IRR. Only the combination of the P_{fimA}-proximal portion of the binding site with the portion associated with DNA sequences flanking IRL (phase off switch) or IRR (phase on switch) produced a fully functioning H-NS binding site. This arrangement provided yet another means by which the phase-on and phase-off forms of the *fim* switch could be distinguished.

3.2.7 The effect of novobiocin on growth

A recent study investigating the effect of metabolism on *fimS* inversion also utilized novobiocin to determine the effect on DNA relaxation on *fimS* inversion (Müller *et al.*,

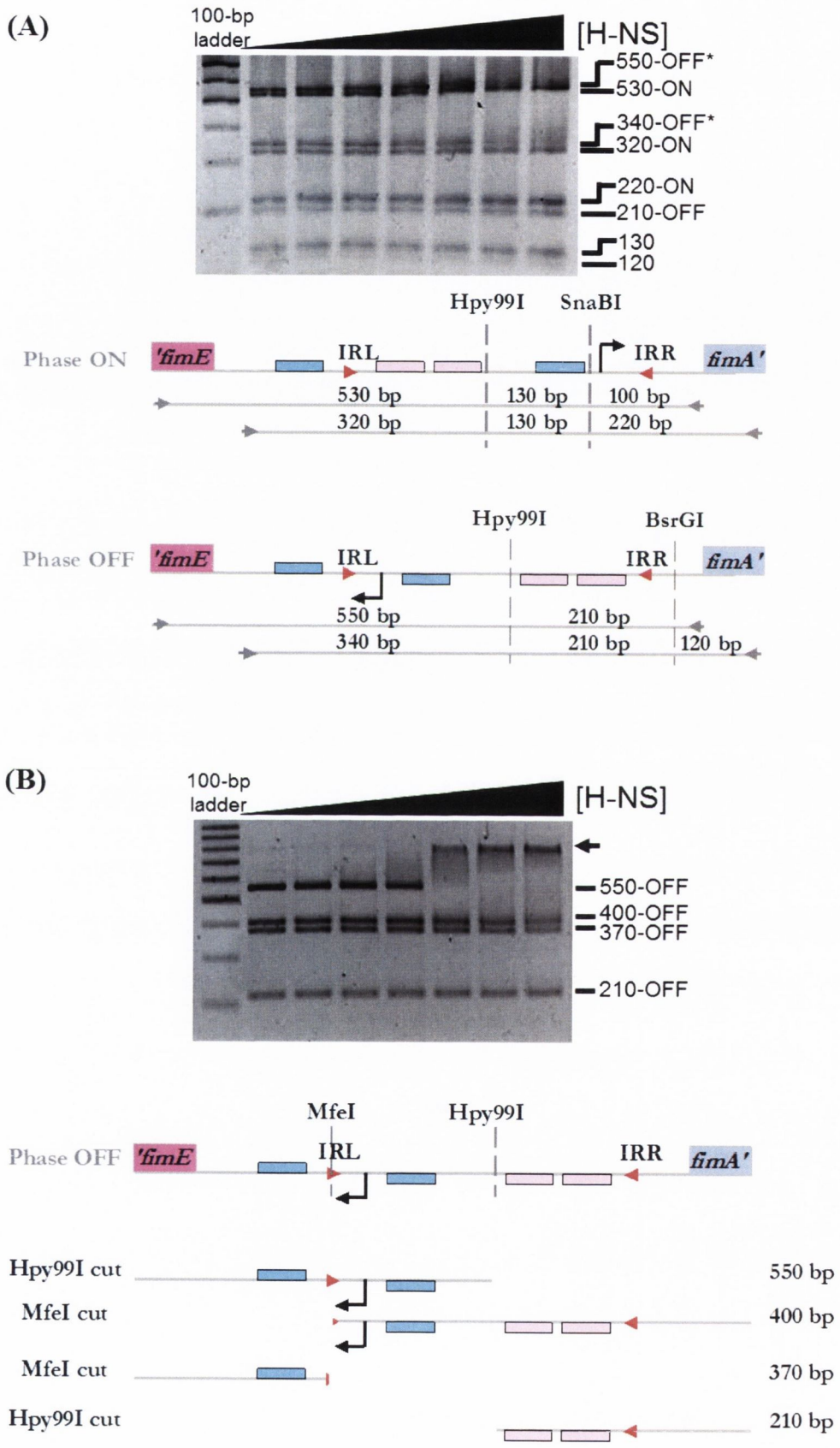

(A)
100-bp ladder
[H-NS]
550-OFF*
530-ON
340-OFF*
320-ON
220-ON
210-OFF
130
120
Hpy99I
SnaBI
Phase ON
'fimE
IRL
IRR
fimA'
530 bp
130 bp
100 bp
320 bp
130 bp
220 bp
Hpy99I
BsrGI
Phase OFF
'fimE
IRL
IRR
fimA'
550 bp
210 bp
340 bp
210 bp
120 bp
(B)
100-bp ladder
[H-NS]
550-OFF
400-OFF
370-OFF
210-OFF
MfeI
Hpy99I
Phase OFF
'fimE
IRL
IRR
fimA'
Hpy99I cut
MfeI cut
MfeI cut
Hpy99I cut
550 bp
400 bp
370 bp
210 bp

Fig. 3.7 H-NS binds overlapping the phase-off IRL. H-NS binding to the *fim* switch was assessed by a competitive *in vitro* binding assay. DNA fragments were amplified by PCR from phase-on-locked or phase-off-locked bacterial strains using two primer pairs to give products of ~770 bp and ~660 bp (panel A) or a single primer pair to give a ~770-bp product (panel B). These PCR products were then digested with restriction enzymes as shown schematically (vertical dashed line) before being mixed and incubated with increasing amounts of purified H-NS protein (final concentrations used in each gel lane, left to right: 0, 7, 35, 70, 100, 135 and 165 nM). Each DNA fragment is identified on the agarose gels by size (bp) and *fimS* orientation (e.g. 340-OFF). DNA fragments that either contained only static (non-inverting DNA) or an invertible segment not attached to static DNA are indicated by size alone. Fragments that corresponded to the phase-off IRL were bound by H-NS (A; indicated by *). Resolved nucleoprotein complexes in which H-NS is bound at the phase-off IRL (B) are indicated by a filled arrow.

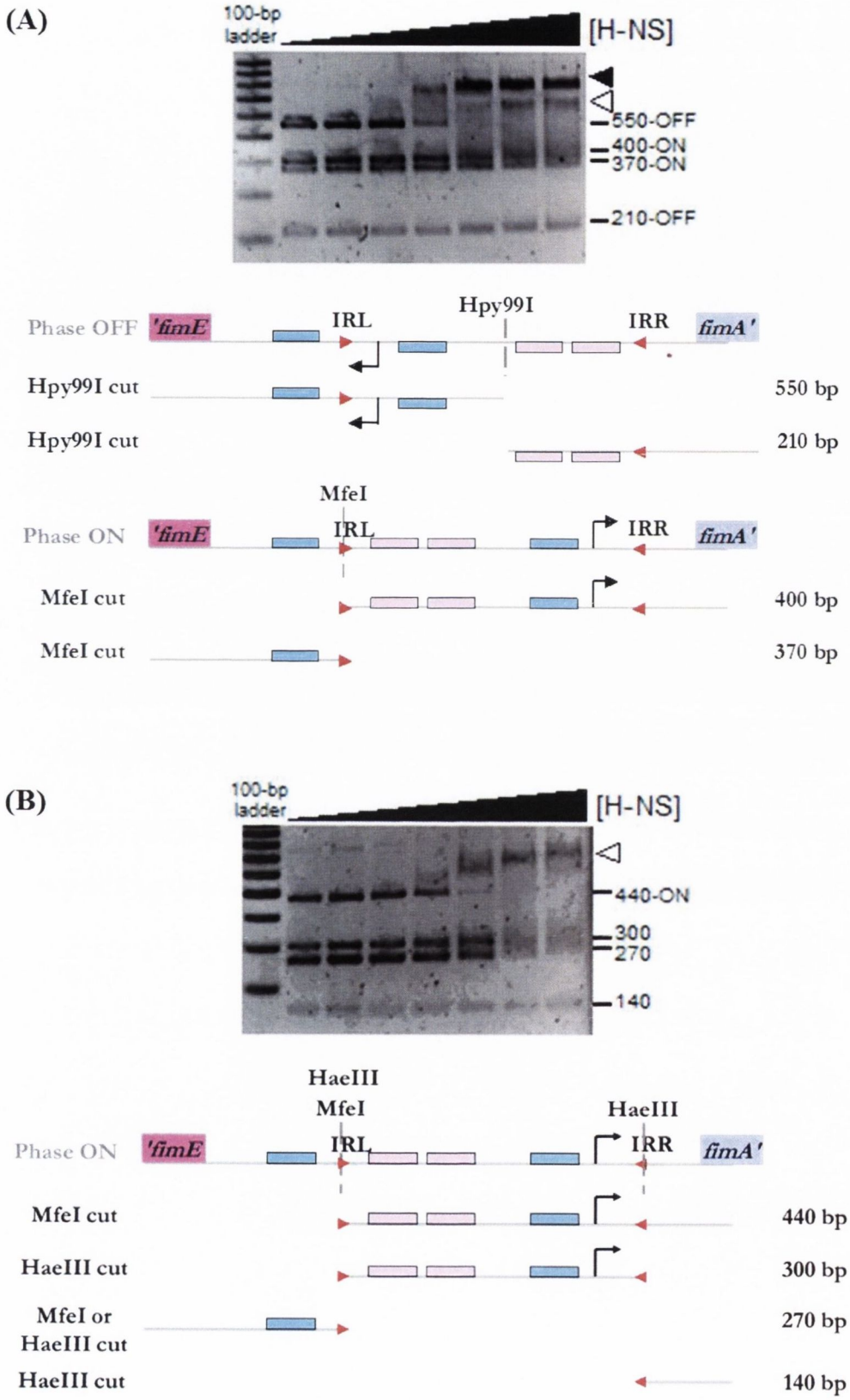

(A)
100-bp ladder
[H-NS]
550-OFF
400-ON
370-ON
210-OFF
Hpy99I
Phase OFF
'fimE
IRL
IRR
fimA'
Hpy99I cut
550 bp
Hpy99I cut
210 bp
MfeI
Phase ON
'fimE
IRL
IRR
fimA'
MfeI cut
400 bp
MfeI cut
370 bp
(B)
100-bp ladder
[H-NS]
440-ON
300
270
140
HaeIII
MfeI
HaeIII
Phase ON
'fimE
IRL
IRR
fimA'
MfeI cut
440 bp
HaeIII cut
300 bp
MfeI or HaeIII cut
270 bp
HaeIII cut
140 bp

Fig. 3.8 H-NS binds overlapping the phase-on IRR. Binding of H-NS to *fimS* is phase-variable. H-NS binding to the *fim* switch was assessed by a competitive *in vitro* binding assay. DNA fragments were amplified by PCR from phase-on-locked or phase-off-locked bacterial strains using a single primer pair to give a ~770-bp product. These PCR products were then digested with restriction enzymes as shown schematically (vertical dashed line) before being mixed and incubated with increasing amounts of purified H-NS protein (final concentrations used in each gel lane, left to right: 0, 7, 35, 70, 100, 135 and 165 nM). Each DNA fragment is identified on the agarose gels by size (bp) and *fimS* orientation (e.g. 340-OFF). DNA fragments that either contain only static (non-inverting DNA) or an invertible segment not attached to static DNA are indicated by size alone. Resolved nucleoprotein complexes in which H-NS is bound at the phase-off IRL (panel A) are indicated by a filled arrowhead while resolved nucleoprotein complexes in which H-NS is bound at the phase-on IRR (panels A and B) are indicated by open arrowheads.

2009). Their results supported the role for LRP in the DNA relaxation induced phase on bias, although their test conditions differed slightly to those used in this and previous studies. Specifically, the novobiocin levels used in the study by Müller *et al.*, (2009) (0-25 μg/ml) were lower than those used in this and previous studies (Dove and Dorman, 1994; Kelly *et al.*, 2006; Corcoran and Dorman, 2009) (0-100μg/ml). While 100μg/ml novobiocin is considered sub-inhibitory to *E. coli* K-12 cells (Gellert *et al.*, 1976b), addition of 100μg/ml novobiocin to exponentially growing cultures had a pronounced effect on growth (Fig. 3.9). Novobiocin caused a dose-dependent inhibition of growth in both wild-type and *hns*- strains (Fig. 3.9). Addition of >75 μg/ml novobiocin caused a temporary (>2hr) cessation of growth. Growth resumed after an undefined period since after 24 hr all cultures had increased dramatically in turbidity. The addition of 25μg/ml novobiocin had only minor effects on growth. The response of *fimS* inversion to low levels of novobiocin (<25 μg/ml) was tested.

3.2.8 *fimS* inversion is highly sensitive to novobiocin

The effect of low levels of novobiocin (<25μg/ml) on inversion biasing of *fimS* was tested using CSH50. A dose dependent response was observed with even the lowest concentration of novobiocin (6.25 μg/ml) resulting in an increase in phase-on cells in the population (Fig. 3.10). Since using lower levels of novobiocin had reduced effects on cell growth (Fig. 3.9) and presumably reduced pleiotropic effects (Jeong *et al.*, 2006), this low concentration regime can be considered for use in future studies.

3.3 Discussion

Those tyrosine integrase site-specific recombinases that have been characterized in most detail catalyse DNA integration and excision events such as Int-mediated bacteriophage lambda entry and exit from the chromosome of *E. coli* or DNA dimer resolution reactions such as those catalysed by XerC and XerD, or Cre, or Flp (Rajeev *et al.*, 2009). Recombination Directionality Factors (RDFs) have been described in the case of several recombination systems and appear to be an important part of the machinery that ensures that an otherwise random event proceeds in a specific direction under a given set of physiological circumstances. For example, phage often encode an excisionase (Xis)

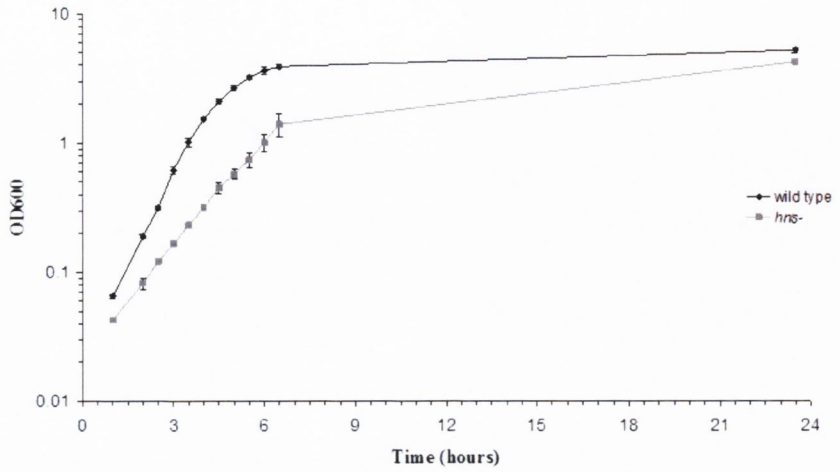
protein that binds within the phage genome altering DNA topology to promote excision of integrated phage from the host chromosome and inhibit integration of excised phage into the host chromosome (Better *et al.*, 1983; Lewis and Hatfull, 2001, 2003). In the case of the Int recombinase of phage lambda, the directionality of the reaction is also regulated by host proteins (see section 1.7) (Van Duyne, 2002). IHF binds at 3 sites (H1, H2 and H') in phage lambda and can promote Int interaction with low affinity sites flanking the core region (De Vargas *et al.*, 1989a). IHF binding to H2 and H' is required for bidirectional recombination (Bushman *et al.*, 1985; Thompson *et al.*, 1986). IHF binding to H1 promotes integration and inhibits excision but binding of IHF to H1 is of lower affinity than binding of IHF to sites H2 and H' and thus only occurs at high cellular IHF concentrations (Bushman *et al.*, 1985). Since IHF levels increase upon entry into stationary phase (Ditto *et al.*, 1994) the H1 site is more likely to be occupied during poor growth conditions. Host cell physiology therefore plays an important role in controlling the directionality of lambda phage recombination.

The direction of recombination is also controlled by the differential requirement for negative DNA supercoiling, which is required for integration but not for excision (Crisona *et al.*, 1999). DNA supercoiling levels vary with the stage of growth of the bacterial culture (Dorman *et al.*, 1988). DNA becomes more relaxed at the onset of stationary phase when the metabolic flux in the cell and the concentration of ATP are diminished (Balke and Gralla, 1987; van Workum *et al.*, 1996). Therefore the alternate requirement for DNA supercoiling in excision and integrative recombination events may act to control the directionality of recombination in response to bacterial growth.

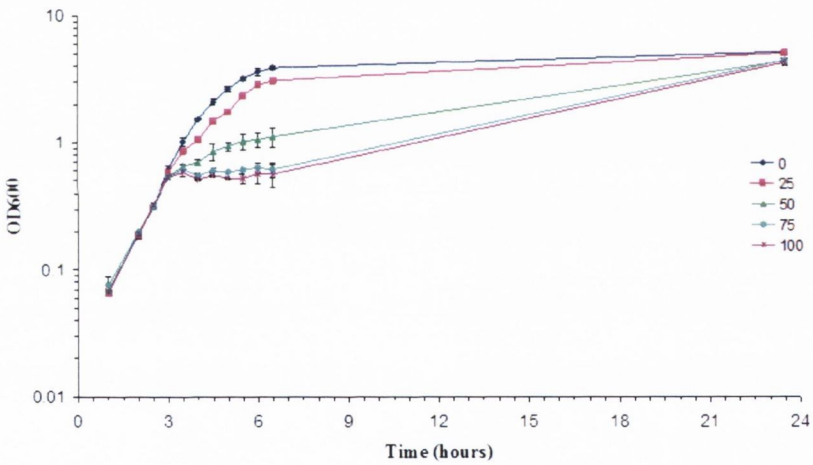
The *fim* system is unusual since it contains two recombinases, which act independently to catalyse recombination of *fimS*. The differential expression of FimB, which promotes bidirectional switching, and FimE, which promotes rapid switching from phase-on to phase-off only, provides an important method for biasing of recombination in response to changing environmental stimuli. However, mutations in FimE arise frequently and here FimB alone catalyses *fimS* inversion (Blomfield *et al.*, 1991b).

In bacteria growing in mid-exponential phase at 37°C in LB broth, the FimB protein inverts the *fimS* element from the on phase to the off phase and back again at approximately equal rates (McClain *et al.*, 1991; Gally *et al.*, 1996). However, when novobiocin is used to inhibit DNA gyrase activity, allowing DNA to become relaxed by the unopposed action of DNA topoisomerase I, the FimB-catalysed *fimS* inversion reaction adopts a pronounced

(A) No novobiocin



(B) Wild type



(C) *hns-*

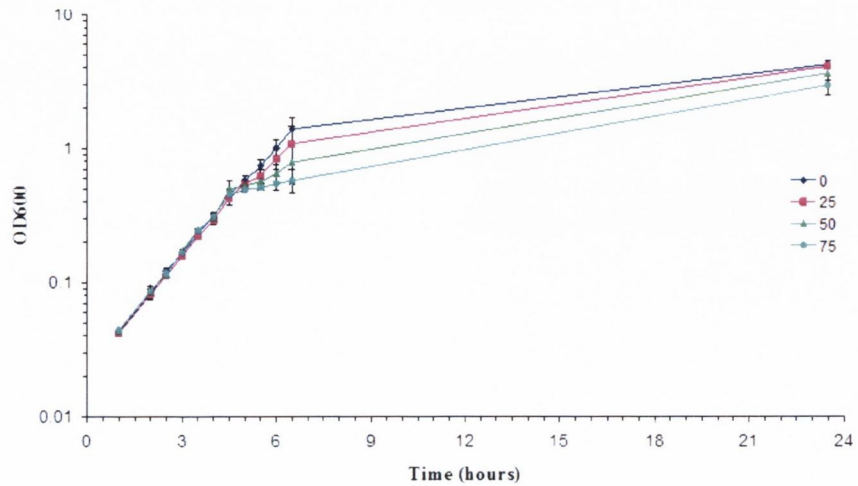


Fig. 3.9 The effect of novobiocin on growth of VL386 and VL386*hns*-. Comparison of wild-type and *hns*- growth in the absence of novobiocin (A). Novobiocin causes a dose dependent decrease in growth rate in wild type (B) and *hns*- (C) strains. Cultures were grown to exponential phase (OD600 ~0.3) before novobiocin was added to a final concentration of 0,25,50,75 or 100 µg/ml. Data shown are averages of duplicate independent experiments. Error bars are standard error of the mean (sem).

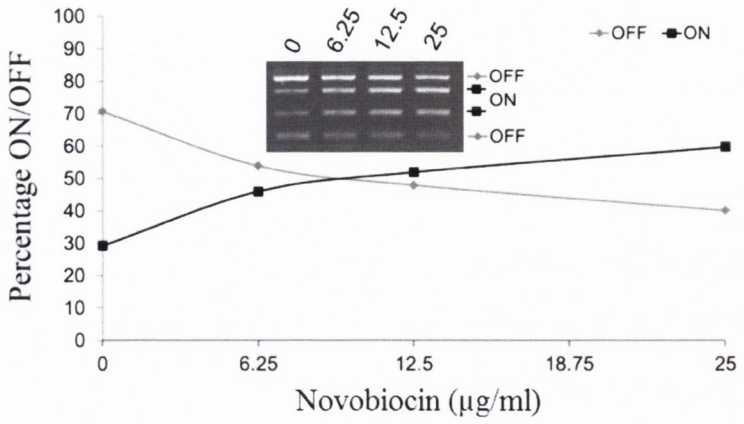


Fig. 3.10 Inversion of *fimS* is sensitive to low levels of novobiocin. Exponentially growing cells were diluted in pre-warmed broths containing increasing amounts of novobiocin (0-25µg/ml) and sampled after 15 generations (see section 2.2). The PCR products diagnostic of the on and off forms of the *fim* switch were resolved by electrophoresis on agarose gels (inset). The concentration of novobiocin (µg/ml) used to treat the culture prior to PCR is given above each lane. Phase-on-specific and phase-off-specific DNA bands (two each) are labelled. The bands were scanned with a densitometer and the data plotted to summarize the effect of novobiocin treatment on *fimS* phase biasing.

bias in favour of the on orientation and thus FimB mediated inversion of *fimS* can be considered programmable (Dove and Dorman, 1994; Kelly *et al.,* 2006; Müller *et al.,* 2009). Previously, the LRP NAP was identified as playing a role as an RDF in this reaction (Kelly *et al.*, 2006). Specifically, the LRP protein is required to maintain the *fimS* element in the on orientation and it does this by binding to the LRP-1 and LRP-2 sites within *fimS*.
In the present study (Appendix 1; Corcoran and Dorman, 2009), the IHF protein was identified as a second RDF that is required to maintain *fimS* in the on orientation when gyrase activity is inhibited. The location of the IHF binding site in the non-inverting chromosomal DNA immediately adjacent to IRL creates the potential for differential interaction with LRP bound to sites LRP-1 and LRP-2 within *fimS* because all three sites are only juxtaposed when the switch is in the on orientation (Fig. 1.5).

While changes in NAPs and DNA supercoiling clearly have an important role in controlling the directionality of *fimS* inversion it is difficult to predict the mechanism for this effect by analogy with other systems. Recombination by tyrosine recombinases may be mediated by changes in negative supercoiling through a number of different mechanisms: (i) an increase in the efficiency of formation of the synaptic complex; (ii) an increase in the catalytic recombinase-mediated reaction (Holliday junction formation); or (iii) an increase in the efficiency of resolution of the Holliday junction, possibly produced by modifications in the topology of the Holliday junction due (Mikheikin *et al.*, 2006).
DNA supercoiling can have different effects on recombination even in closely related systems. This is clearly illustrated in the alternative effects of DNA supercoiling in recombination by the Xer recombinases at the highly homologous sites *mwr* and *cer* (Trigueros *et al.*, 2009).

The Xer systems employs two tyrosine site-specific recombinases XerC and XerD that bind co-operatively to different half-sites of the core site to catalyse recombination (Blakely and Sherratt, 1996). Together they resolve multimers of chromosome and plasmid DNA ensuring proper segregation during cell division (Blakely *et al.*, 1993). Xer recombination at plasmid sites is initiated by XerC-catalyzed strand exchange of one pair of strands to form a Holliday junction. In the case of recombination at the sites *cer* (from plasmid ColE1) (Summers and Sherratt, 1988; Guhathakurta *et al.*, 1996) and *mwr* (from the multidrug resistance plasmid pJHCMW1) (Tolmasky *et al.*, 2000; Bui *et al.*, 2006), the

Xer catalyzed process stops at a Holliday junction and these intermediates are resolved by Xer-independent cellular processes (Trigueros *et al.*, 2009).

An increase in negative supercoiling of the DNA was shown to increase recombination efficiency between *mwr* sites by increasing the formation of Holliday junctions (Trigueros *et al.*, 2009). Recombination between *mwr* sites was highly sensitive to changes in DNA supercoiling with *mwr* sites becoming progressively worse substrates as DNA was relaxed (Trigueros *et al.*, 2009). In contrast, efficiency of recombination at the highly related *cer* site was not significantly modified by the superhelicity of the DNA. In this case, recombination proceeded efficiently unless on fully relaxed DNA (Trigueros *et al.*, 2009). The increased levels of Holliday junction formation implied that negative supercoiling facilitated recombination between *mwr* sites by either promoting synapsis and/or by promoting the catalytic action of XerC.

The exact mechanism by which DNA supercoiling influences phase biasing of *fimS* must be determined experimentally. Useful information could be gained by monitoring the occupancy of binding sites for IHF, LRP and FimB in response to DNA relaxation. These data, in conjunction with *in vitro* recombination assays would give further insight into the mechanistic role for each component of the phase-on 'trap'.

One can however envisage a model whereby the DNA bending activity of IHF and the DNA wrapping activity of LRP in combination with the relaxed nature of the *fim* DNA that obtains in the absence of adequate levels of gyrase activity creates a nucleoprotein complex that inhibits synapsis, possibly by preventing binding of FimB to IRL and thus preventing inversion by FimB from the on-to-off direction (Fig. 3.11).

Previous work has suggested a role for the H-NS protein in influencing the FimB-mediated inversion of *fimS* (Higgins *et al.,* 1988; Kawula and Orndorff, 1991; O'Gara and Dorman, 2000). This study revealed a role for H-NS in maintaining *fimS* in the phase-off orientation. Deletion of H-NS causes a dramatic increase in the percentage of phase-on cells in a population (Fig. 3.6A), which supports previous reports of a locked-on phenotype in H-NS deficient strains (Higgins *et al.*, 1988; Kawula and Orndorff, 1991).

Careful examination of H-NS interactions with *fimS* and its flanking regions showed that H-NS interaction with the switch was contingent on *fimS* orientation. These data showed

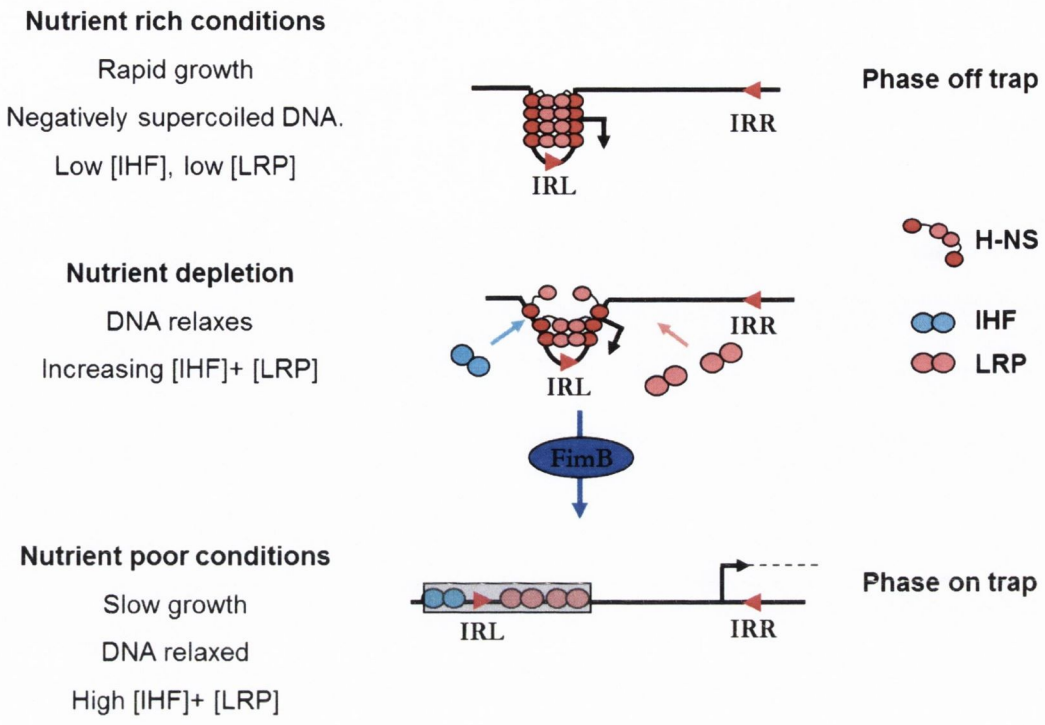


Fig. 3.11 Model for the direct roles of DNA supercoiling, H-NS, IHF and LRP in phase biasing of *fimS*. H-NS binding in *fimS* is phase-variable. H-NS binds overlapping the IRL in phase off, sequestering the IRL and preventing switching from phase off to phase on (phase off trap). The model involves the formation of a nucleoprotein complex between LRP and IHF bound at IHF-1, at the IRL in the phase on orientation where, upon DNA relaxation, the IRL is sequestered. This prevents efficient switching from phase on to phase off, which results in an increase in the amount of on cells in a population (phase on trap).

that the H-NS binding site in the *fimS* region is distributed across three incomplete sites. One of these is in the invertible *fimS* element and is adjacent to the P_{fimA} promoter. The others are located in the non-inverting parts of the chromosome immediately outside IRL and IRR. When the switch is in the off orientation, a complete site capable of binding H-NS in an electrophoretic mobility shift assay is formed at the IRL end of the switch; when *fimS* is in the on orientation, a complete H-NS interaction site is formed only at the IRR end. The physiological significance of H-NS binding at IRR when *fimS* is in the on orientation is unknown. It may play a role in modulating the activity of the P_{fimA} promoter. While this demonstrates a novel mechanism for H-NS interacting with DNA, the concept of phase variable binding in *fimS* is well established and has been studied extensively with the Fim recombinases. Each recombinase binds to a half-site located in the invertible region and a second half site external to *fimS*. Since the 4 half-sites are non-identical, certain combinations provided better targets for binding than others. FimE binds poorly to half sites flanking the IRL in the phase OFF orientation and this may explain the low rate of inversion by FimE from phase OFF to phase ON (Gally *et al.*, 1996; Kulasekara and Blomfield, 1999).

While over-expression of FimB restores the normal phase biasing of *fimS* in an *hns* mutant, incorporation of the *hns* mutant allele with the *fimS* NAP site mutations showed LRP binding within *fimS* supported H-NS in maintaining a phase off bias. In bacteria that express the H-NS protein, elimination of LRP binding to *fimS* sites LRP-1 and LRP-2 results in switching in the on-to-off direction when DNA gyrase is inhibited (Kelly *et al.*, 2006). This is not the case in the absence of H-NS: here bidirectional switching continues to be detected even though the LRP-1 and LRP-2 sites have been inactivated (Fig. 3.6D; E). Our results show that either H-NS or LRP can sustain the 'off' trap in relaxed DNA; only when both proteins fail to interact with *fimS* does the switch fail to become biased to the off orientation following gyrase inhibition. The results also show that IHF and LRP are not equivalent as RDF elements; prevention of IHF binding to the IRL-proximal site creates the 'off' trap regardless of the status of H-NS expression; abrogation of LRP binding to LRP-1 and LRP-2 only creates the 'off' trap if H-NS is present.

Multiple types of mutations arise that alter the inversion of *fimS*. For example, some uropathogenic strains contain a third site-specific recombinase termed HbiF, which counters the effect of FimE by inverting exclusively from phase off to phase on (Xie *et al.*,

2006). Mutations in FimE (Blomfield *et al.*, 1991b) and *fimS* (Leathart and Gally, 1998) that alter the inversion of *fimS* have also been isolated. While some mutations isolated have resulted in locked phenotypes, a strain with point mutations in the IHF-1 site has been isolated and was shown to have a preference for switching from phase on to phase off (Leathart and Gally, 1998). This supports our findings that the IHF-1 site is required to maintain the DNA relaxation induced phase-on bias.

Results from this study also show that the structure of *fimS* is important to maintain heterogeneity in the bacterial population. The combination of IHF-1, LRP-1 and LRP-2 site mutants effectively locked *fimS* in the off orientation and this bias was strengthened by DNA relaxation (Fig 3.3). The removal of H-NS restored heterogeneity to the system, allowing mixed phase on and phase off populations, however since mutations in H-NS result in pleiotropic changes in gene expression (Dorman, 2004), they are unlikely to be advantageous in the long term and thus unlikely to be fixed in a population (Chattopadhyay *et al.*, 2009).

Heterogeneity pre-empts the requirements for the next stage of infection and allows it to be achieved by a successful bacterial subpopulation. Feedback mechanisms play an important role in maintaining heterogeneous populations (Balaban *et al.*, 2004). For example, *fimS* contains a Rho-dependent transcription terminator that, when in the phase off orientation, causes shortening of the *fimE* transcript. The shortening of this transcript decreases the half-life of *fimE* mRNA and hence reduces the level of FimE protein in the cell and allows for FimB-mediated switching to phase on (Sohanpal *et al.*, 2001; Joyce and Dorman, 2002; Hinde *et al.*, 2005). Another method of ensuring population heterogeneity involves *fimS* in the on orientation. It has been reported that cells undergo phase variation independently of *fimS* orientation or P_{fimA} activity resulting in 10% of phase on cells remaining afimbriate (McClain *et al.*, 1993).

The phase-on nucleoprotein trap formed by IHF and LRP at *fimS* in response to DNA relaxation may also play an essential role in ensuring successful niche colonization by multiple generations of bacterial cells. Bacteria expressing a favourable combination of surface appendages will have a growth advantage. It may therefore be important for daughter cells to maintain the arrangement of surface appendages that allow for successful colonisation of the niche. In steady state growth, the levels of DNA supercoiling and nucleoid associated proteins are heritable traits and thus once the DNA is replicated, the

nucleoprotein trap at *fimS* will mimic that of the mother cell. This form of heritable memory is a key component of successful colonization (Casadesus and D'ari, 2002).

The data presented here cast new light on the issue of programmed versus stochastic phase variation of type 1 fimbriae in *E. coli*. Similar to Int-mediated recombination of phage lambda, FimB-mediated inversion of *fimS* is strongly influenced by changing growth conditions. This and previous studies have shown that the FimB site-specific recombinase acquires an orientational bias when DNA gyrase activity is impaired and DNA becomes relaxed (Dove and Dorman, 1994; Kelly *et al.*, 2006; Corcoran and Dorman, 2009; Müller *et al.*, 2009). The maintenance of negatively supercoiled DNA depends on the activity of DNA gyrase that is in turn dependent on the phosphorylation potential of the cell (van Workum *et al.*, 1996). Metabolically active bacteria typically adopt a motile or planktonic lifestyle and become sedentary when resources can no longer support this lifestyle. Attachment to and colonization of a surface represents a distinct lifestyle and the transition from motility to non-motility and attachment is known to involve complex signalling cascades and gene regulatory circuits (Müller *et al.*, 2009; Pesavento and Hengge, 2009). DNA relaxation is a feature of the stationary phase of growth that correlates with a reduction in the negative supercoiling activity of DNA gyrase (Bordes *et al.*, 2003) and is an important factor in controlling the expression of many virulence genes (Appendix 2; Dorman and Corcoran, 2009). It is possible that the *fim* switch becomes biased towards the on phase when bacteria approach a state of low metabolic flux since this will promote type 1 fimbrial expression and assist attachment and colonization. The involvement of the LRP and IHF proteins in the establishment of a phase on bias is also consistent with a need for physiological sensitivity on the part of *fim* gene expression (Fig. 3.11). IHF protein levels peak at the transition from exponential growth to stationary phase (Ditto *et al.*, 1994) making this protein an attractive growth stage reporter. LRP has been established as a regulator of genes involved in the response to nutrient stress (Brinkman *et al.*, 2003; Yokoyama *et al.*, 2006) and LRP levels increase as bacteria approach stationary phase (Landgraf *et al.*, 1996).

The variation in the supply or the activities of DNA gyrase and the NAPs described in this study offers the bacterium an excellent mechanism for fine-tuning the phase-variable expression of the *fim* structural genes to its physiological circumstances.

Chapter 4.

Development and characterization of a novel *gfp* gene.

4.1 Introduction.

Since the discovery and isolation of the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* over a decade ago there has been a rapid rise in the use of fluorescent proteins as fundamental research tools (Shaner *et al.*, 2007; Rizzo *et al.*, 2009). Transcriptional fusions to *gfp* facilitate high-throughput *in vivo* monitoring of gene expression at a single cell level and have been used in landmark studies including the identification of genes involved in virulence of *Salmonella* (Valdivia and Falkow, 1997) and metabolism in *E. coli* (Zaslaver *et al.*, 2006). *gfp* is widely used as a transcriptional reporter gene since it requires no substrate other than oxygen to convert blue light to green and promoter activity is easily monitored by changes in fluorescent signal from GFP (Heim *et al.*, 1994).

GFP variants optimized for use in flow assisted cell sorters (FACs) were developed by random mutation of a portion of *gfp* encoding the 20 amino acids that flank the chromophore (Ser-Tyr-Gly sequence at amino acids 65-67) (Cormack *et al.*, 1996). Mutants were pooled and screen by FACs for increased fluorescence when excited at 488 nm. The mutations isolated were placed in three groups (GFPmut1, 2 and 3) according to the nature of the mutations although all three groups contained mutations in S65 along with at least one other mutation. Each of the 3 classes of mutants had dramatically altered emission wavelengths that increased the intrinsic fluorescence of the GFPmut molecules compared to wild-type GFP. GFPmut1 (also known as enhanced GFP; EGFP), which was the most intrinsically fluorescent mutant, contained the mutations F64L and S65T. GFPmut2 (S65A, V68L, S72A) and GFPmut3 (S65G, S72A) exhibited approximately half the intrinsic fluorescence of GFPmut1 but were more fluorescent due to more efficient folding than GFPmut1 (Cormack *et al.*, 1996). Another approach used in the development of GFP used multiple rounds of random "DNA shuffling" followed by visual screening for increased fluorescence under UV light (Crameri *et al.*, 1996). By not screening using FACs, Crameri *et al.* avoided the selection of red-shifted variants and instead isolated mutants with increased fluorescence due to improved folding at 37°C (GFPuv; F99S, M153T, V163A).

While these fluorescent proteins (fps) are often used as transcriptional fusions on multi-copy based plasmids, studies requiring chromosomal based *gfp* fusions were limited since only high levels of promoter activity produced enough GFP for detection (Hautefort *et al.*,

2003). A *gfp* gene (*gfp*+) (Scholz *et al.*, 2000) was developed by combining the beneficial mutations from *gfpmut1* and *gfpUV*, which resulted in significantly increased fluorescence. This allowed for the monitoring of promoter activity using single copy *gfp* fusions (Hautefort *et al.*, 2003).

Measurement of promoter activity from P_{fimA} is complicated since a change in expression detected by traditional (*lacZ*, *cat*, *luxAB*) fusions to *fimA* could be due to a change in P_{fimA} activity, inversion of *fimS,* or a combination of both occurring separately in the population. To overcome this, studies have monitored P_{fimA} expression in phase locked-ON strains (Dorman and Ní Bhriain, 1992) or have monitored the orientation of *fimS* while simultaneously monitoring *lacZ* activity (Dove and Dorman, 1994). A *gfp* fusion to *fimA* would permit simultaneous measurements of promoter orientation and activity at a single cell level. Single cell analysis may also reveal new aspects of regulation within populations that could not be identified using techniques that involve assaying whole populations and reporting average values (Paulsson, 2004; Newman *et al.*, 2006).

4.2 Results.

4.2.1 Development of a novel *fimA* reporter fusion.

A highly-fluorescent variant of *gfp*, (*gfp*+) designed for use in single-copy (Scholz *et al.*, 2000; Hautefort *et al.*, 2003) was used to create the *fimA-gfp* fusion. Initially a 4.5 kb region of *fim* was amplified from CSH50*ihfB*::Tn*10* (*fimS* locked-off) using primers fimE.HindIII.fw and fimD.hindIII.rv. This region was then cloned into the HindIII site of pACYC184 to create pCPC501 (Fig. 4.1A). A 2-kb region from pZep08 was amplified (gfp+.pstI.fw and cat.pstI.rv) and cloned into the PstI site of *fimA* on pCPC501, creating pCPC502. The gfp+.pstI.fw primer introduced an in-frame stop codon (TAA) in *fimA* preventing translational fusion of *fimA* to *gfp*. The amplified *gfp* gene included a near-optimal ribosome binding site from pZep08 that ensures efficient translation. Numerous attempts were made to clone the *fimS* region in the phase on orientation but these were unsuccessful. It was previously reported that multi-copy plasmids containing *fimS* in the ON orientation are toxic and cannot be transformed (Klemm, 1986; Gunther *et al.*, 2002). Plasmid based phase on *fimS* can be obtained through FimB mediated inversion of *fimS*, however to obtain a homogenous plasmid based *fimS* phase on population, the plasmids must be maintained in a Δ*ihf* background to prevent maximal expression from P_{fimA}

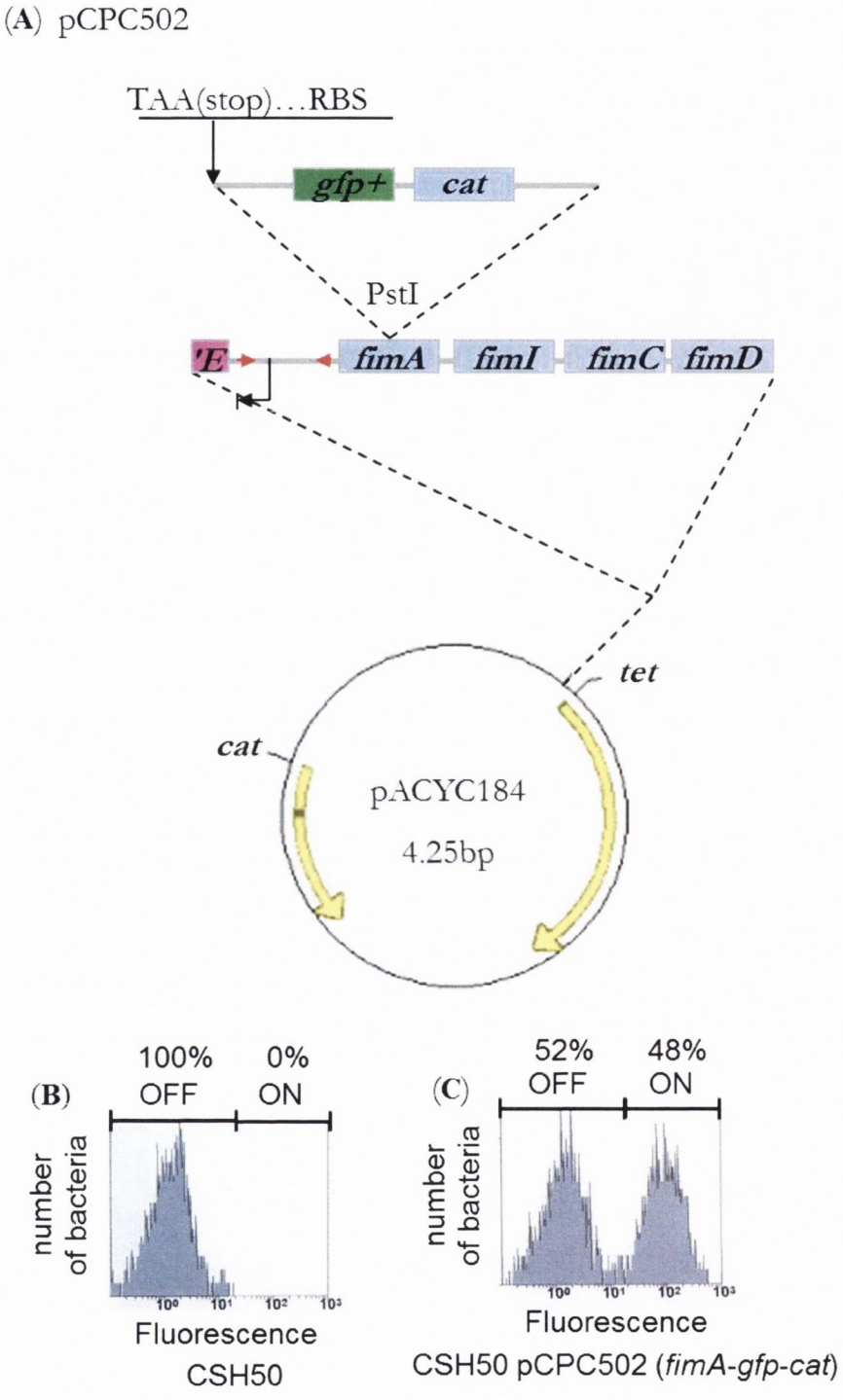


Fig. 4.1 Construction of a *fimA-gfp* fusion and determination of *fimS* orientation by flow cytometry. (A) The construction of a *fimA-gfp* fusion required multiple cloning steps (see text for further detail). (B) CSH50 lacking a *gfp* fusion is non-fluorescent. (C) Cultures of bacteria containing the *fimA-gfp* fusion contain fluorescent (ON) and non-fluorescent (OFF) populations.

(McClain *et al.*, 1991). pCPC502 was transformed into CSH50 and fluorescence was measured after overnight growth. CSH50 pCPC502 contained dual populations of non-fluorescent and fluorescent bacteria from which the percentage phase off (non-fluorescent) population and phase on (fluorescent) population could be quantified (Fig. 4.1C). The *fimA-gfp-cat* cassette was amplified from pCPC502 using (fimE.hindIII.fw and fimD.hindIII.rv) and integrated into the chromosome using lamda-red mediated homologous recombination (pKobega). Lower levels of fluorescence were detected when the *fimA-gfp* fusion was chomosomally located compared with the plasmid based fusion due a reduction in copy number upon insertion into the chromosome. The chromosomal fusion did however produce detectably fluorescent bacteria (Fig. 4.2A) that, when concentrated by centrifugation, exhibited green fluorescence visible to the naked eye (data not shown).

4.2.2 The *fimA-gfp*+ fusion has an altered response to DNA supercoiling

The chromosomal *fim-gfp* reporter strain was then tested under DNA relaxing conditions which, in VL386, creates an inversion bias of *fimS* to phase ON (Dove and Dorman, 1994). Results from flow cytometry, PCR based assay and western blot analysis using a GFP specific antibody showed the reporter fusion to respond in a manner opposite to that which has been published, switching OFF in response to DNA relaxation (Fig. 4.2). A mutation in either an IHF or LRP binding site in *fimS* could have accounted for this bias (Corcoran and Dorman, 2009). However, DNA sequencing detected no mutations in the *fimS* region. To rule out the possibility that other unlinked chromosomal mutations were affecting the switch bias, the *fimA-gfp* fusion was transduced into a fresh background, with 3 isolates tested showing the same reversed response to reduced negative DNA superhelicity (data not shown).

Previous studies describing the DNA-relaxation-induced phase ON bias were performed in VL386, a strain that contains a *lacZ* fusion to *fimA* (Dove and Dorman, 1994; Kelly *et al.*, 2006). Considering the sensitivity of the *fim* switch to local levels of DNA supercoiling, both fusions were screened using WebSIDD for strong SIDD sites which may be affecting the distribution of superhelical energy in the region. Although the inserted DNA was found to have a higher A+T content (57%) than both *fimA* (48%) and *lacZ* (44%), no strong SIDD sites were identified (data not shown).

Since it remained a formal possibility that a mutation affecting *fimS* inversion was present outside the *fimS* region but close enough to co-transduce at high frequency with the chloramphenicol resistance marker, the *gfp* fusion strain was re-made in a fresh background (CSH50). The inversion of *fimS* to DNA relaxation was tested before and after insertion of the *gfp-cat* cassette. Wild type CSH50 showed the expected phase on bias, which was then reversed with the introduction of the *gfp-cat* cassette in *fimA* (data not shown). This suggested that the *gfp-cat* cassette was itself responsible for the reversal of the DNA-relaxation-induced phase on bias in *fimS*. Although the mechanism for this was unknown, it was shown to be dependent on the H-NS protein (Fig. 4.2D). Strain CSH50Δ*hns*,*fimA-gfp-cat* (Fig. 4.2D) showed the same novobiocin independent 80% phase on bias seen in VL386 Δ*hns* (Fig. 3.6A) and in other Δ*hns* strains containing various (inactive) promoters transcribing from *fimE* into *fimS* (O' Gara and Dorman, 2000). This supported the role for H-NS in FimB-mediated phase biasing of *fimS* (O' Gara and Dorman, 2000; Corcoran and Dorman, 2009).

FimB-mediated inversion of *fimS* can be biased towards phase off by transcription from the *fimE* gene in a manner that requires the presence of H-NS (O' Gara and Dorman, 2000). For this reason the possibility that a change in DNA topology caused by transcription from the *cat* gene that had been inserted downstream of *fimS* was investigated. Firstly the orientation of the *cat* gene in *fimA* was reversed, directing transcription towards *fimS* (Fig.4.3B). The *cat* gene in both orientations was then replaced by the *kan* cassette from pKD4 (Fig. 4.3, C and D) which was excised using pCP20 (Fig 4.3, E and F) (Datsenko and Wanner, 2000). This left the *fimA-gfp* fusion intact but removed all non-native promoters from downstream of *fimS* (Fig 4.3, E and F). Although experiments to link transcriptional activity from within *fimA* with the altered regulation of *fimS* inversion proved inconclusive, the data did suggest that the presence of the *gfp* gene itself was responsible for the aberrant switching of *fimS* (data not shown).

Bioinformatic analysis of the *gfp*+ gene showed that it was highly A+T-rich (59%) and that it was likely to contain regions of strong intrinsic DNA curvature; these are the two key determinants of H-NS binding.

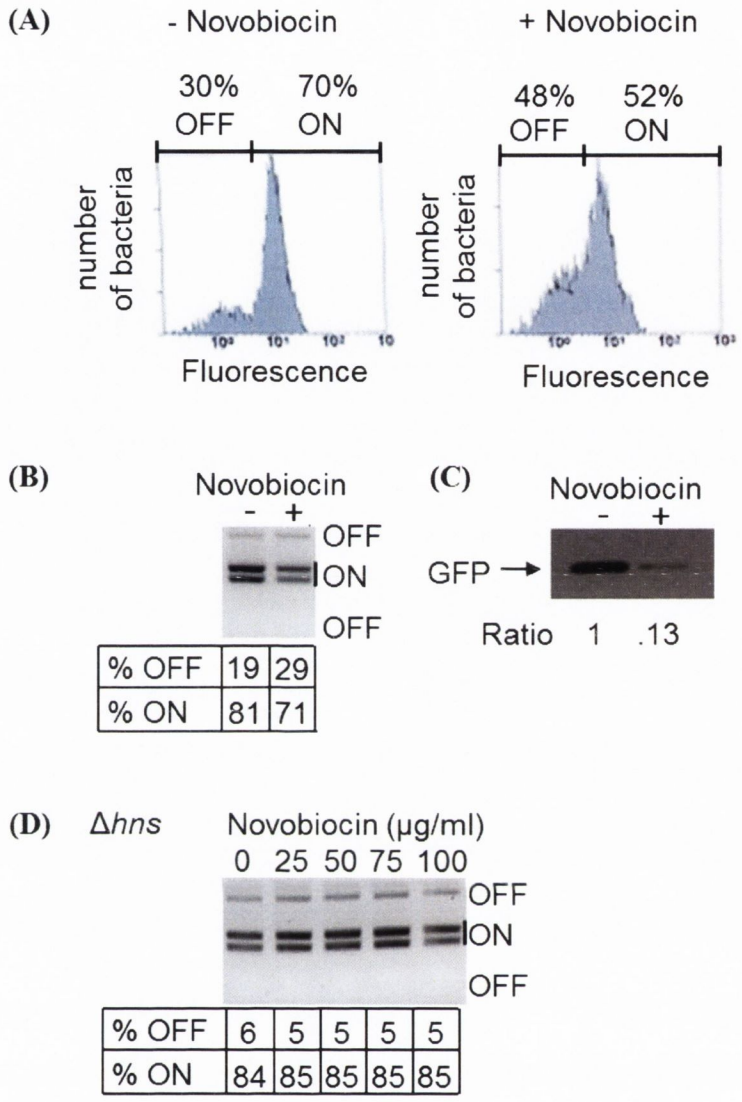


Fig. 4.2 The switching bias of *fimS* in response to novobiocin is reversed in CSH50*fimA-gfp-cat*. *fimS* orientation was determined by monitoring the levels of GFP fluorescence by flow cytometry (A), PCR based switch assay (B) and western blot using a GFP specific antibody (C). Novobiocin, when present (+), was used at 100 μg/ml. The switching bias of CSH50*fimA-gfp-cat* was also tested by PCR-based assay in the absence of H-NS (D).

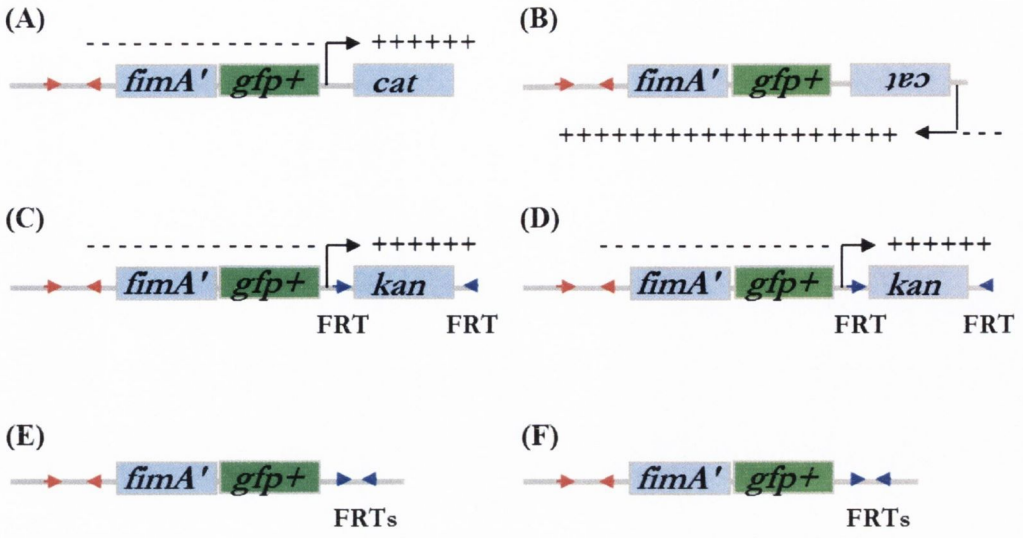


Fig. 4.3 Transcription from within *fimA* may affect inversion of *fimS*. Chromosomal *gfp* fusion strains with; the *cat* gene inserted both orientations (A and B); the *cat* gene replaced with the *kan* cassette from pKD4 (C and D); no selectable marker after excision of the *kan* cassette by site-specific recombination using the FLP recombinase (E and F) (Datsenko and Wanner, 2000). Promoters are indicated as ⌜ with the arrows showing the direction of transcription. Transcription induced positive and negative supercoils are indicated by + and – respectively. Binding site for the FLP recombinase (C-F) are indicated by blue filled arrowheads and labelled FRT (FLP recognition target). Excision leaves a ~50 bp region including the inverted repeats (E and F).

4.2.3 The *gfp+* gene is bound *in vivo* by H-NS

Mapping of H-NS binding in the genomes of *E. coli* (Oshima *et al.*, 2006) and *S.* Typhimurium (Lucchini *et al.*, 2006; Navarre *et al.*, 2006; Dillon *et al.*, 2010) has shown that H-NS binds preferentially to horizontally acquired genes of high A+T content relative to the host genome. It was therefore possible that the original *gfp* gene from *A. victoria*, which is 61% A+T, would be targeted by H-NS when used in *E. coli*, which has an average A+T content of 50%. *gfp* was also predicted to contain regions of strong intrinsic curvature, another key feature of H-NS bound DNA (Owen-Hughes *et al.*, 1992; Spurio *et al.*, 1997; Dame *et al.*, 2001). These key determinants of H-NS binding were not disrupted by the minor substitutions in the coding sequence of *gfp* needed to alter the GFP protein and create improved variants such as GFPmut2 and GFP+ (Fig. 4.4). H-NS binding in *gfp+* could dampen transcription of an associated promoter and could be responsible for the altered inversion of *fimS* (Fig. 4.2).

H-NS binding to *gfp+ in vivo* was tested by chromatin immunoprecipation using an H-NS specific monoclonal antibody and quantitative PCR (Dillon *et al.*, 2010). Chromatin immunoprecipitation involves crosslinking proteins to DNA in live cells using formaldehyde, sonication of the DNA into small fragments (100-500-bp), purifying the DNA-protein complexes and then using a specific antibody to the protein of interest (in this case H-NS) (see Materials and Methods; section 2.12). This antibody-protein-DNA complex can then be isolated, the protein removed from the DNA and the DNA quantified using quantitative PCR. This confirms that a specific piece of DNA has been bound by H-NS. The fold enrichment of the DNA is an indication of the affinity of the protein of interest for the DNA.

Five primer sets (ChIP.1-5.fw/.rv) were used to amplify regions of the *fimA-gfp+-cat* fusion before and after immunoprecipition with a monoclonal H-NS antibody (see Materials and Methods; section 2.7.9). A region of DNA located centrally in the *gfp+* gene (probe 2), was strongly enriched for H-NS binding while the other probes covering the *gfp+* and *cat* genes were not (Fig. 4.5). Probe 2, which was more than 12-fold enriched over input DNA, contained a 38-bp region that was 76% A+T (nucleotides 373 to 411 of *gfp+*) making it similar to other well characterized H-NS-bound regions (Chen *et al.*, 2001a; Chen *et al.*, 2005a; Bouffartigues *et al.*, 2007; Lang *et al.*, 2007a). That only probe 2 was enriched for H-NS binding suggested that H-NS bound in this region did not promote nucleation of other H-NS molecules along the DNA but instead, may have formed DNA-

protein-DNA bridges with a second H-NS binding site. This second site may have overlapped the phase OFF IRL in *fimS* (Corcoran and Dorman, 2009) and thus may create a new topological domain around *fimS*, altering regulation of the highly topologically sensitive system (Fig. 4.5B).

4.2.4 Reverse engineering of *gfp+*

Since the other commonly used *gfp* derivatives were predicted to be strong targets for H-NS we designed *ab initio* a gene sequence with reduced A+T content and intrinsic DNA curvature compared to *gfp+*. The genetic code uses 64 nucleotide triplets (codons) to encode 20 amino acids and stop, so that each amino acid is encoded, on average, by 3 codons (Crick *et al.*, 1961; Brenner *et al.*, 1967). The frequency with which codons are used by different organisms varies significantly, leading to variation in G+C content between genomes (Bernardi and Bernardi, 1985; Knight *et al.*, 2001).

The computer program that was used to alter the codon usage of *gfp+* (*Gene Designer*) (Villalobos *et al.*, 2006) optimized genes for expression using a codon usage table in which each codon is given a probability score based on the frequency distribution of the codons in the desired genome (in this case *E. coli*) normalized for every amino acid. For the redesign of the *gfp+* gene, we used the EColi_CII table that is derived from a collection of highly expressed *E. coli* genes (Villalobos *et al.*, 2006). This approach avoided the use of rare codons, which are strongly associated with low levels of protein expression due to ribosome stalling and abortive translation (Kurland and Gallant, 1996).

Gene Designer was used to produce iterative, equally optimized sequences with reduced A+T content (50% +/- 1%) that were then screened for reduced predicted intrinsic DNA curvature using the Bend.it algorithm (Fig. 4.6). One sequence, which differed from *gfp+* by 157-base pair (bp) substitutions across the 717-bp gene, had a 50% A+T content and showed dramatically reduced predicted intrinsic DNA curvature (Fig. 4.7). This gene, termed gfp^{TCD}, was synthesized by DNA2.0 (San Diego, California, USA). gfp^{TCD} encoded a protein that was identical in its amino acid sequence to the product of the *gfp+* gene but was predicted to be a poorer target for H-NS binding.

4.2.5 *In vitro* analysis of H-NS binding to *gfp+* and gfp^{TCD}

Purified H-NS, was used for comparative analysis of H-NS binding affinity to *gfp+* and gfp^{TCD}. To determine protein binding affinity accurately, the protein must be present in excess of the DNA. Therefore, biotinylated probes were used to allow visualization of

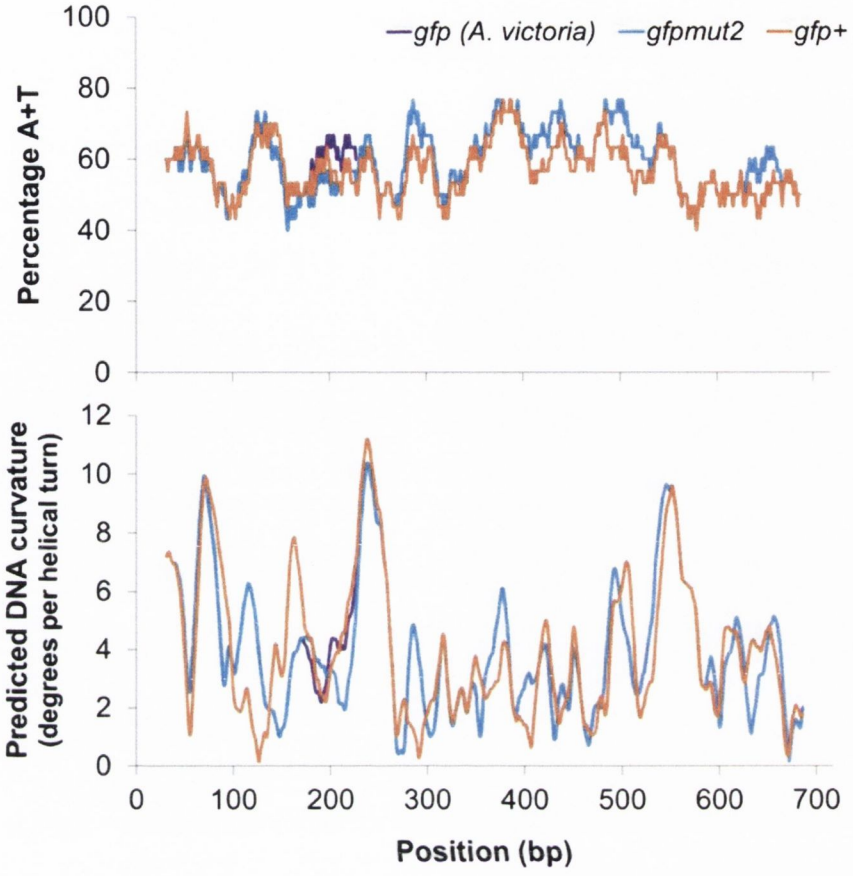


Fig. 4.4 DNA sequence features of the original *gfp* gene and two widely used derivatives (*gfpmut2* and *gfp+*) are conducive to H-NS binding. Data shown are A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot). Each GFP protein has dramatically altered fluorescence (illustrated by different shades of green) but have virtually identical DNA curvature and A+T content profiles.

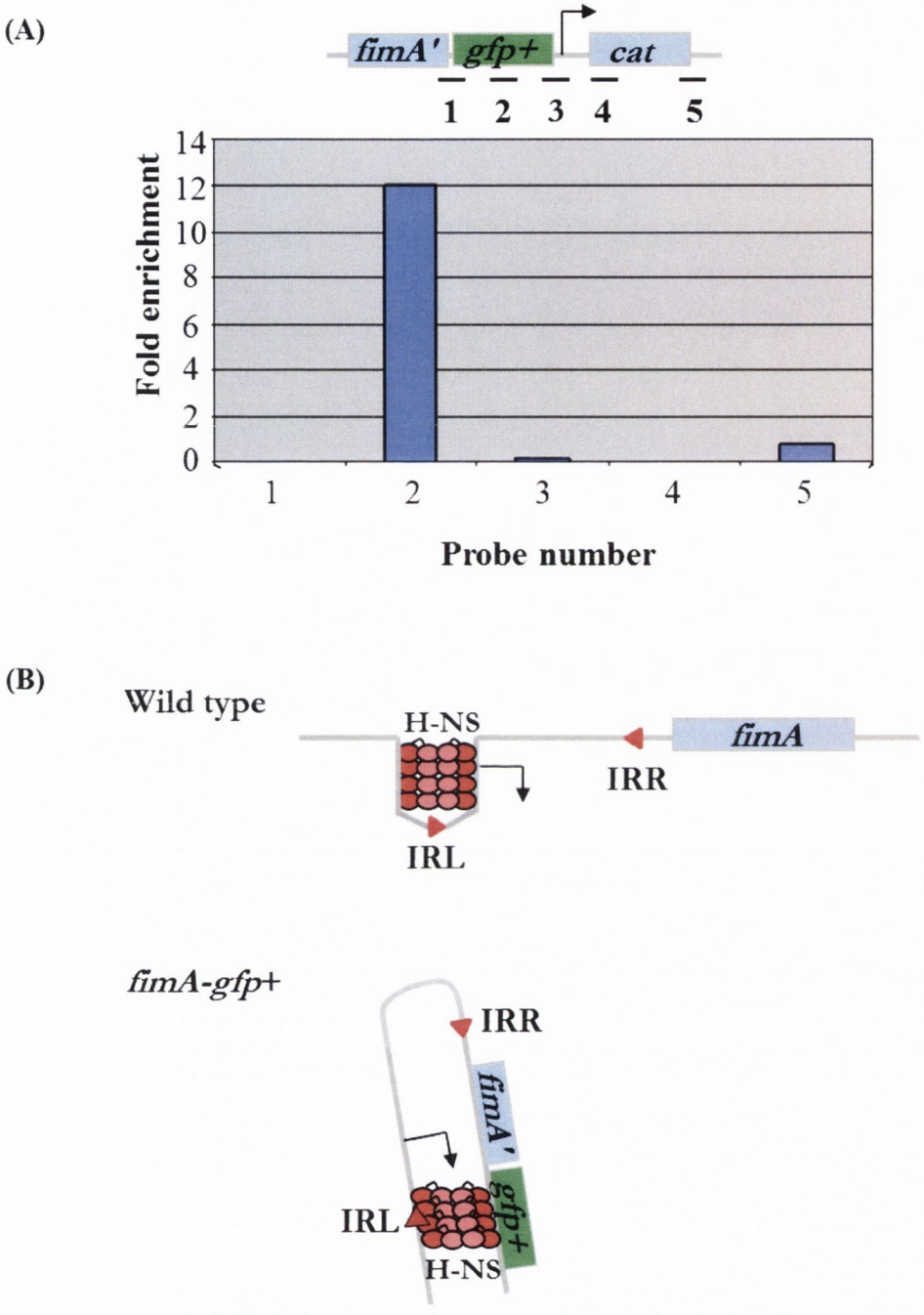


Fig. 4.5 H-NS binds to *gfp+ in vivo*. Chromatin immunoprecipitation (ChIP) using an H-NS specific monoclonal antibody was followed by quantitative PCR (qPCR) (A). The Y-axis indicates fold enrichment relative to input DNA (DNA before addition of the H-NS specific antibody). These data are derived from a single ChIP experiment. Regions targeted for PCR amplification are referred to as probes. The approximate locations of these probes (1-5) are indicated in the illustration above the histogram. (B) H-NS bound in *gfp+* may interact with H-NS bound in *fimS* to form a new topological domain.

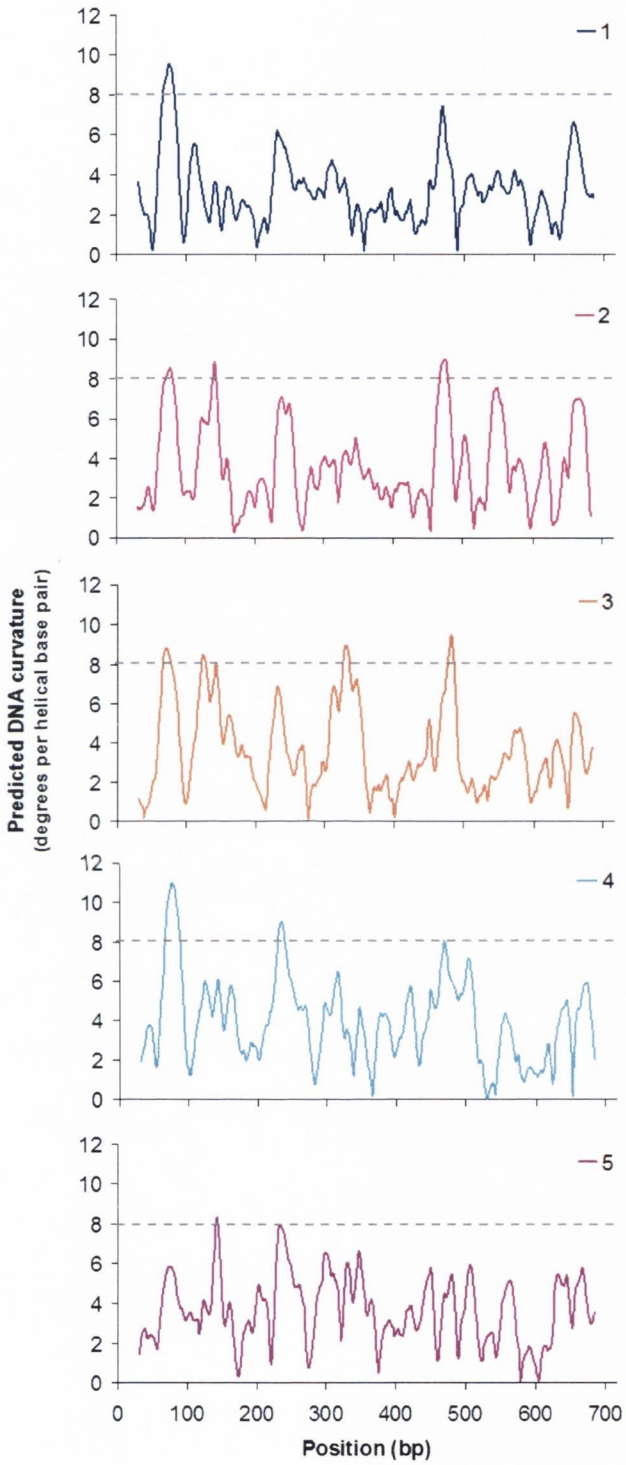


Fig. 4.6 Predicted DNA curvature varies among genes of similar A+T content. Multiple codon optimizations of *gfp+* using *GeneDesigner* produced genes of varying intrinsic curvature. All 5 genes encode an identical GFP+ protein and have 50 % (+/- 1%) A+T content. A dashed grey line highlights all peaks above 8° per helical turn.

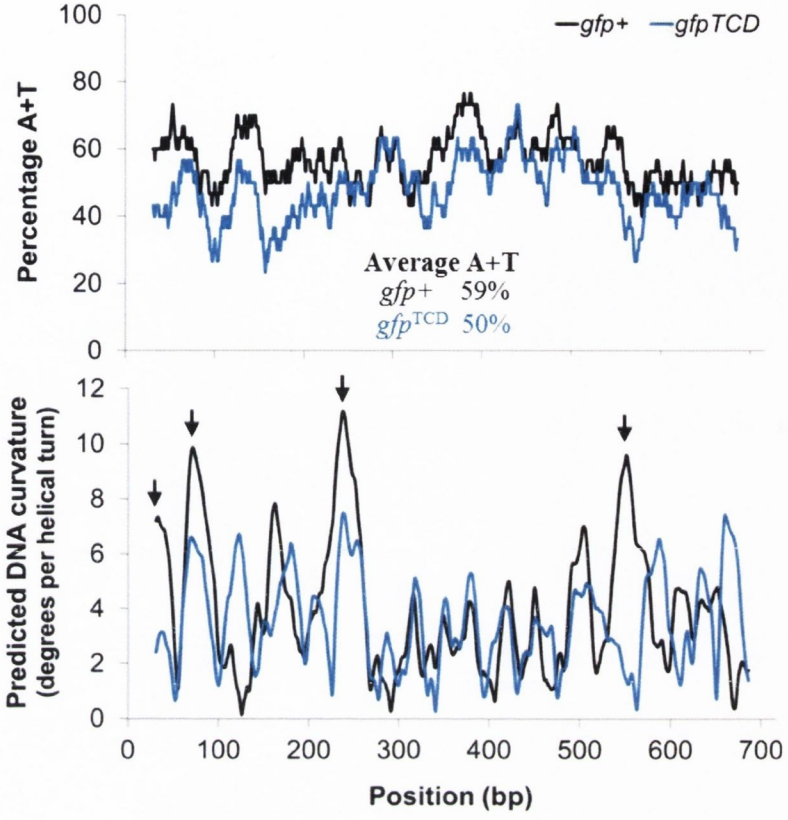


Fig. 4.7 The DNA sequence features of the *gfp*+ gene conducive to H-NS binding are not present in gfp^{TCD}. A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot) of *gfp*+ and gfp^{TCD}. Arrows highlight regions of strong predicted curvature in *gfp*+ that are reduced in gfp^{TCD}.

picomolar concentrations of DNA while H-NS was used at nanomolar concentrations. This allowed estimation of the apparent dissociation constant (K_{app}), a measurement which requires the DNA concentration to be negligible compared to that of the protein at the midpoint (i.e. at least 10-, and preferably 100-fold, lower) (Carey, 1991).

4.2.6 Temperature affects H-NS binding *in vitro*

Since H-NS binding to DNA has been shown previously to be affected by temperature (Badaut *et al.*, 2002; Amit *et al.*, 2003; Bouffartigues *et al.*, 2007), the effect of temperature on H-NS binding to both *gfp*+ and *gfp*^{TCD} was tested by incubation and electrophoresis of identical reactions at 4°C or 20°C (Fig. 4.8). Temperature was found to have a major effect on H-NS binding to both probes consistent with previous observations that H-NS binding co-operativity is increased at lower temperatures (Amit *et al.*, 2003; Bouffartigues *et al.*, 2007). The association of H-NS with the DNA seemed to be unaffected by temperature since both probes at both temperatures were completely bound (no unbound probe visible) at 12.5 nM H-NS (Fig. 4.8). There was however a dramatic temperature dependent difference in resolution of the bound complexes. While shifting to a single complex represented high-affinity interactions, smearing of the DNA suggested shorter, lower affinity interactions where H-NS disassociated from the DNA faster than it re-associated. The dissociated DNA migrated faster than the strongly bound DNA but did not catch up with the DNA that had never bound the protein, which led to a smearing effect. At 4°C, resolution of a single complex occurred at considerably lower H-NS concentrations (*gfp*+ 12.5-25nM; *gfp*^{TCD} 25-50nM) than at 20 °C, which only fully resolved as a single complex at 400 nM H-NS (Fig. 4.8). A difference in affinity at 4°C between *gfp*+ and *gfp*^{TCD} was also evident since the *gfp*+ probe was almost fully resolved as a single H-NS bound complex at 12.5 nM while the H-NS-*gfp*^{TCD} complexes were smeared, indicating lower affinity binding. Since performing the reactions at lower temperature allowed more accurate analysis of H-NS affinity for DNA, all subsequent bandshifts were carried out at 4°C.

4.2.7 H-NS preferentially binds to curved DNA

The regulatory regions flanking the *proU* promoter are A+T-rich, highly intrinsically curved (Owen-Hughes *et al.*, 1992) and contain multiple high affinity H-NS binding sites

(Lucht *et al.*, 1994; Bouffartigues *et al.*, 2007). This DNA was therefore used as a positive control for H-NS binding (Fig. 4.9). The *lacZ* reporter gene is a poor target for H-NS as it is relatively G+C rich and not intrinsically curved (Fig4.9A; Owen-Hughes *et al.*, 1992) and was used as a negative control for H-NS binding (Fig. 4.9). While the standard method for determining the intrinsic curvature of DNA used in this study (Bend.It) was computationally based , intrinsic curvature can also be observed by altered migration of DNA during gel electrophoresis (Wu and Crothers, 1984). As this aberrant mobility is removed by ethidium bromide intercalation (Calladine *et al.*, 1991), DNA fragments containing curved DNA can be recognised by reduced mobility in the absence of ethidium bromide compared to their mobility in its presence. Fig. 4.9B shows the identically sized (717-bp) *proV* and *lacZ* probes after electrophoresis through a 1% TAE/agarose gel containing ethidium bromide, conditions which cause DNA fragments to migrate according to size and not curvature. Fig. 4.9C shows the same probes after migration through a 5% polyacrylamide gel not containing ethidium bromide, conditions which cause separation of identically sized DNA fragments by intrinsic curvature. The *proV* probe (0 nM H-NS), which migrated identically to *lacZ* in the presence of ethidium bromide (Fig. 4.9B), had clearly reduced mobility compared to *lacZ* (0 nM H-NS) when ethidium bromide was absent (Fig. 4.9C). The differences in A+T content and intrinsic curvature between *proV* and *lacZ* resulted in a dramatic difference in H-NS binding affinity for the two probes (Fig. 4.9C).

Electrophoretic mobility shift assays (EMSA) were performed to determine the affinity (K_{app}) of H-NS for *gfp*+ and gfp^{TCD} *in vitro*. Since H-NS binds with low specificity and affinity and H-NS binding is highly co-operative, in order to assess H-NS binding affinity accurately for the two *gfp* genes a narrow range of protein concentrations was chosen (0.75 dilutions from 25 nM to 1.4 nM). H-NS was found to bind *gfp*+ strongly with a K_{app} of 4.9 ± 0.1 nM (Fig. 4.10A). A further indication of the high affinity of H-NS for *gfp*+ is the narrow range of protein (4.5-10.55 nM) required for the transition from initial binding to fully bound probe, resulting in a single high molecular mass complex . This also illustrates the highly co-operative nature of H-NS binding. H-NS had a lower affinity for gfp^{TCD} Fig. 4.10A; K_{app}, 7.5 ± 0.5 nM**)**. The lower affinity of H-NS for gfp^{TCD} resulted in smearing of the DNA over a wide range of protein concentrations (7.9-18.75nM) with the gfp^{TCD} probe only resolving as a single bound complex at 25nM H-NS.

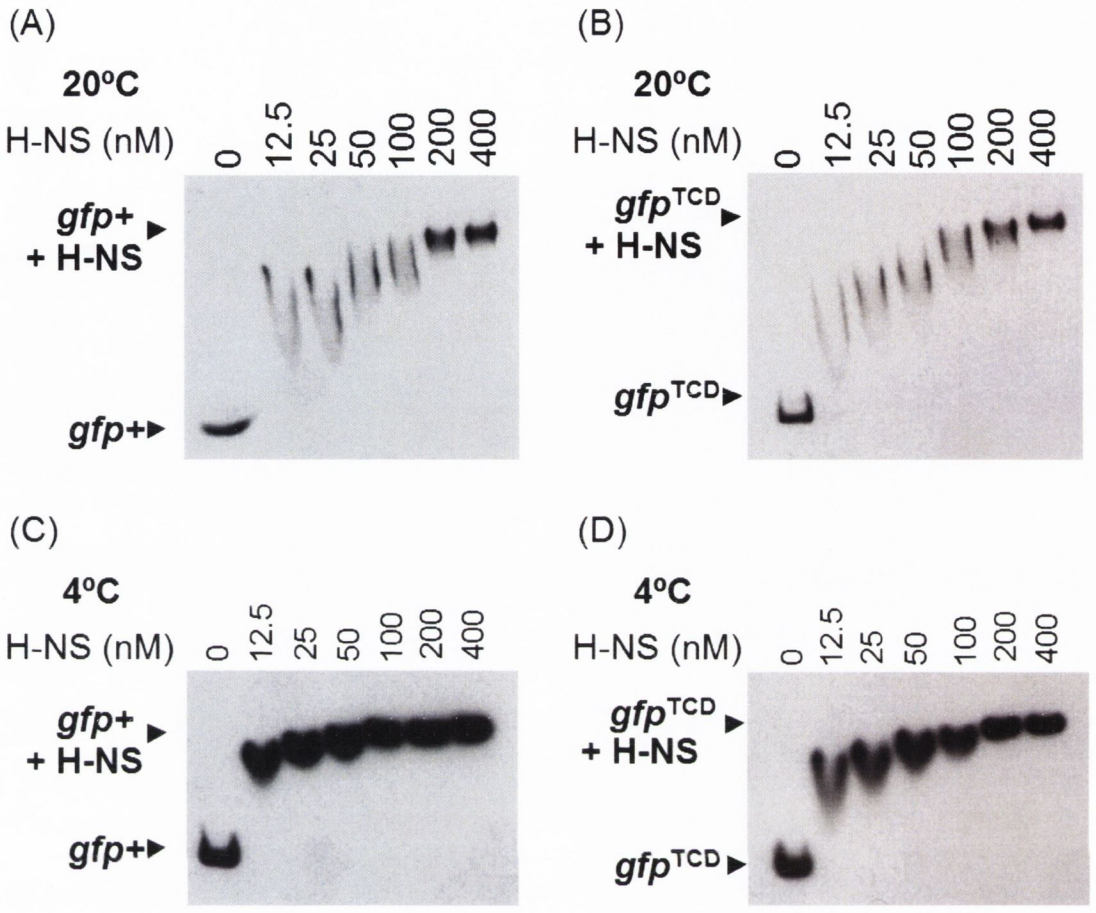


Fig. 4.8 Temperature affects H-NS binding to DNA *in vitro*. Electrophoretic mobility shift assay (EMSA) analysis of H-NS binding to *gfp*+, and *gfp*^{TCD} at 20°C (upper gels) and at 4°C (lower gels). The concentration of purified H-NS used is indicated above each lane.

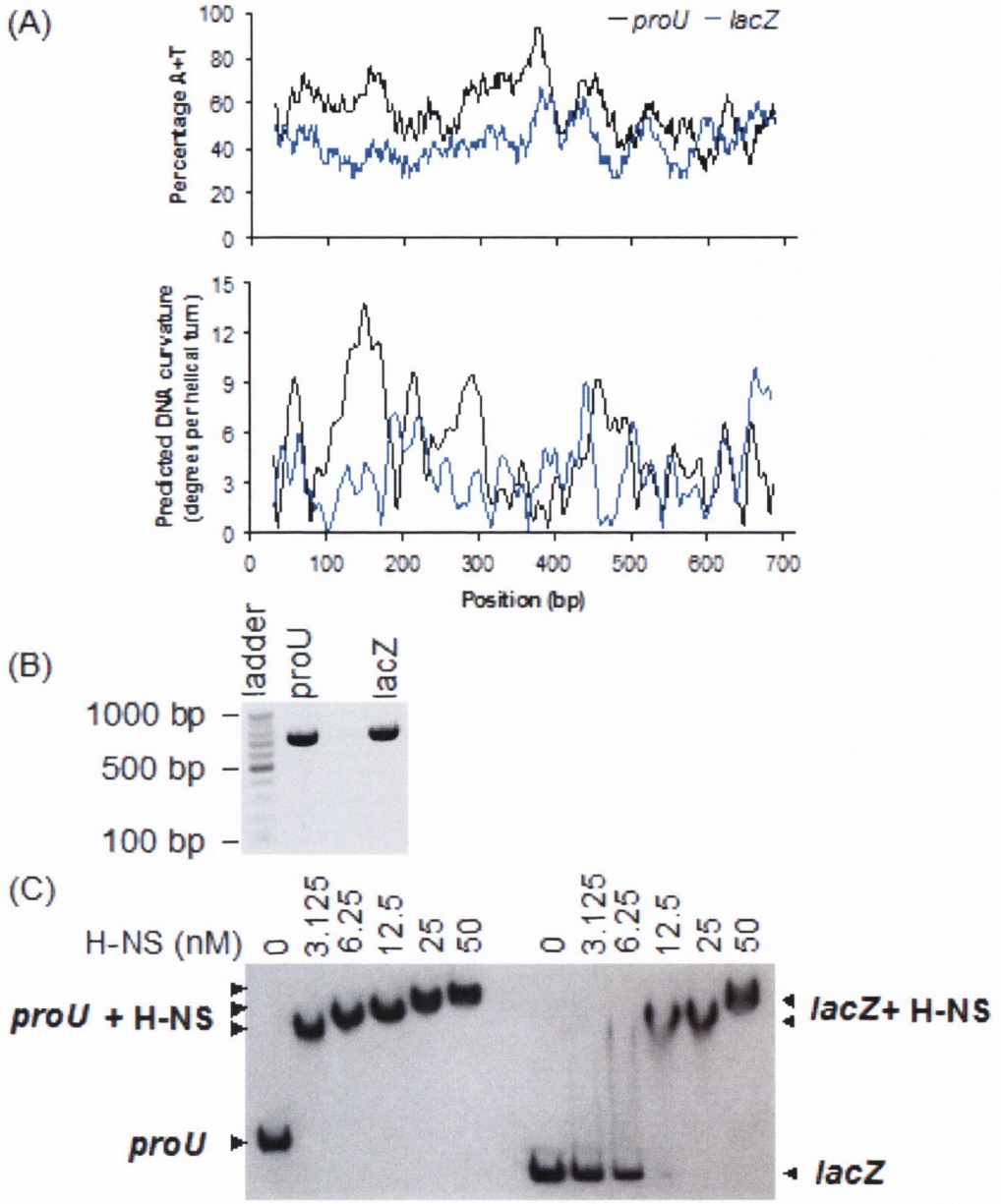


Fig. 4.9 H-NS binds preferentially to A+T-rich intrinsically curved DNA. (A) DNA sequence features, A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot), of the *proU* and *lacZ* regions used for electrophoretic mobility shift assay (EMSA) analysis. (B) *proU* and *lacZ* DNA electrophoresed in a 1% agarose gel containing ethidium bromide. (C) EMSA analysis of *proU* and *lacZ* probes electrophoresed through a 5% polyacrylamide gel in the absence of ethidium bromide. The concentration of purified H-NS used is indicated above each lane.

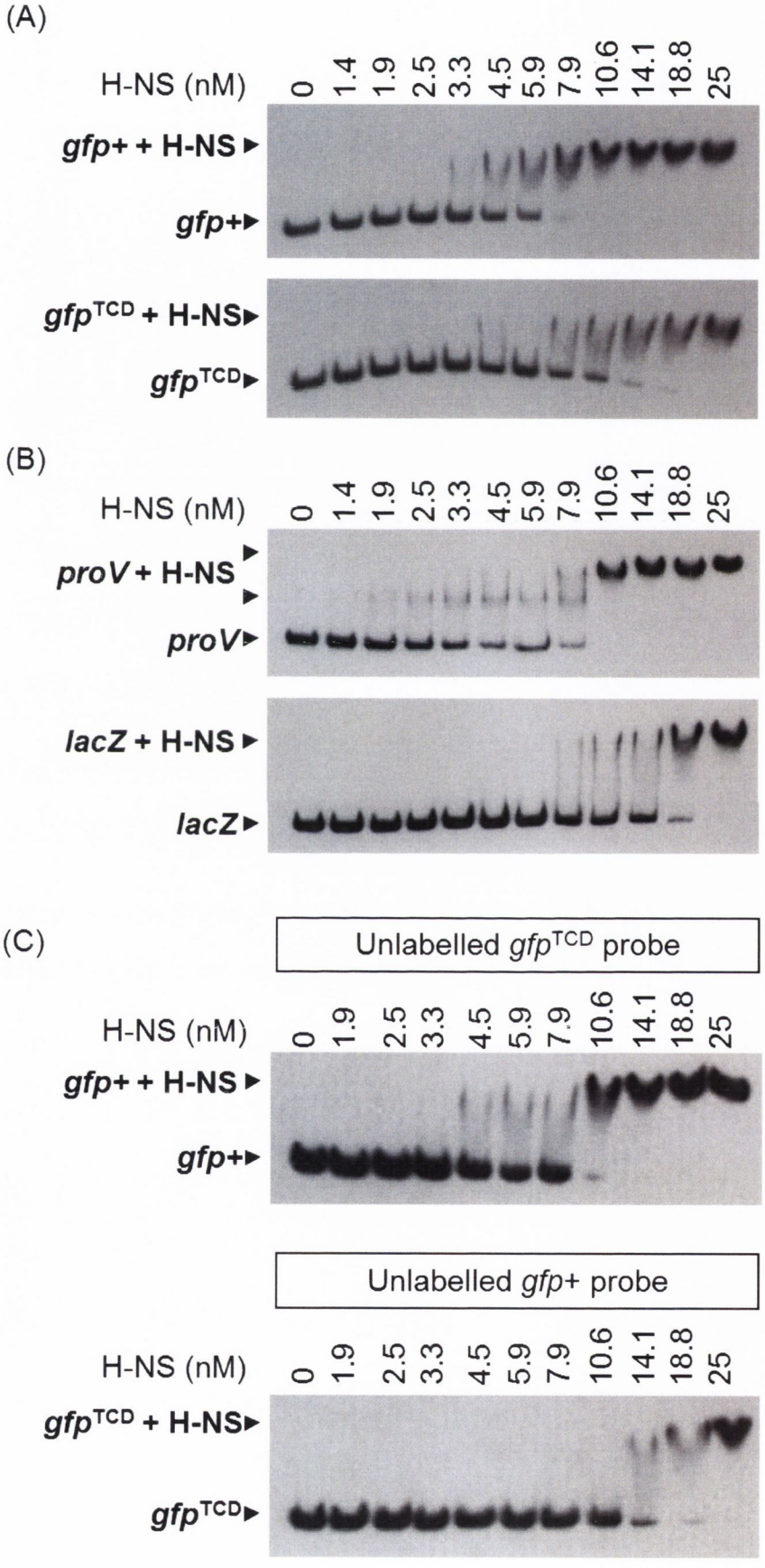

(A)
H-NS (nM)
0
1.4
1.9
2.5
3.3
4.5
5.9
7.9
10.6
14.1
18.8
25
gfp+ + H-NS
gfp+
gfpTCD + H-NS
gfpTCD
(B)
H-NS (nM)
0
1.4
1.9
2.5
3.3
4.5
5.9
7.9
10.6
14.1
18.8
25
proV + H-NS
proV
lacZ + H-NS
lacZ
(C)
Unlabelled gfpTCD probe
H-NS (nM)
0
1.9
2.5
3.3
4.5
5.9
7.9
10.6
14.1
18.8
25
gfp+ + H-NS
gfp+
Unlabelled gfp+ probe
H-NS (nM)
0
1.9
2.5
3.3
4.5
5.9
7.9
10.6
14.1
18.8
25
gfpTCD + H-NS
gfpTCD

Fig. 4.10 H-NS binds to *gfp*+ with higher affinity than gfp^{TCD} *in vitro*. Electrophoretic mobility shift assay (EMSA) analysis using purified H-NS and biotinylated *gfp*+ and gfp^{TCD} probes (A). Biotinylated *proU* and *lacZ* probes are used as positive and negative controls, respectively (B). H-NS binding to *gfp*+ (C) and gfp^{TCD} (D) in the presence of equal amounts of non biotinylated (unlabelled) probes. The concentration of purified H-NS used is indicated above each lane.

The *proU* regulatory region was used as a positive control for H-NS binding (Fig. 4.10B). As expected, the *proU* probe was strongly bound by H-NS (K_{app} 6.2 ± 0.5 nM). The *proU* region contains a number of well characterized H-NS binding sites (Bouffartigues *et al.*, 2007) and resolved as two separate high affinity complexes (arrowed). The *lacZ* gene, which is not bound by H-NS *in vivo* (Oshima *et al.*, 2006) and poorly bound *in vitro* (Owen-Hughes *et al.*, 1992), was used as a negative control for H-NS binding. *lacZ* was poorly bound by H-NS (K_{app} 16.3 ±1.8 nM) and the resolution of a single protein-DNA complex at high H-NS concentrations (25 nM) simply highlights the low specificity of H-NS, which at saturating concentrations binds independently of sequence to DNA (Tupper *et al.*, 1994).

Although the change in K_{app} between *gfp*+ and gfp^{TCD} measured *in vitro* was relatively small, coupled with the altered migration of the bound DNA we predicted that this difference was highly significant *in vivo*. H-NS affinity for *gfp*+ and gfp^{TCD} was also compared in the same reaction using biotinylated and unlabelled DNA in equal amounts (50 pM). (Fig. 4.10C). These data showed that when both genes were present, H-NS bound specifically to *gfp*+ and only bound gfp^{TCD} when all the *gfp*+ DNA had been bound (14.1 nM).

4.2.8 Altering H-NS binding using divalent cations

H-NS binding to DNA can either nucleate along the length of the DNA molecule, thus stiffening the DNA (Amit *et al.*, 2003) or promote inter- or intra-molecular bridging (Dame *et al.*, 2006). The switching between these modes of binding was shown recently to involve a role for divalent cations such as Mg^{2+} (Liu *et al.*, 2010). This may be due to the compaction of DNA by divalent cations (Zinchenko *et al.*, 2004), which could aid H-NS bridging of opposing segments of DNA. Since all the EMSA assays described so far have been performed in the absence of divalent cations, the effect of their addition in the form of $MgCl_2$ was tested. The effect of 10 mM $MgCl_2$ on binding of H-NS to *proV* suggested that the addition of $MgCl_2$ lowered the affinity of H-NS for DNA (Fig. 4.11A). On a technical note, the addition of the $MgCl_2$ considerably slowed migration of the DNA in the gel. All 4 probes were then tested in the presence of 10 mM $MgCl_2$ and electrophoresed for 4 hr instead of 1 hr to ensure migration of all H-NS-DNA complexes into the gel. The addition of $MgCl_2$ uniformly reduced H-NS affinity for DNA and thus under these conditions, *gfp*+ is still bound by H-NS with higher affinity than gfp^{TCD} (Fig. 4.11B). The reduction in

affinity for H-NS may be due to the presence of negatively charged chloride ions in solution masking the basic charges of H-NS and reducing the attractive force between the H-NS and the negatively charged DNA. More interestingly, the addition of $MgCl_2$ promoted the formation of new high affinity complexes (indicated by the white arrows in *gfp*+, gfp^{TCD} and *lacZ* images; Fig. 4.11B). The newly resolved complexes that have increased mobility compared to H-NS-DNA complexes resolved in the absence of $MgCl_2$, may represent bridged DNA molecules. DNA is present in pM quantities making it likely that both DNA-binding domains of an H-NS dimer will bind intra-molecularly as opposed to binding multiple DNA molecules and forming inter-molecular bridges (Wiggins *et al.*, 2009) . This internally bridged DNA molecule may represent the lower H-NS bound complexes Fig. 4.11B). The upper resolved complex which is formed at very high H-NS concentrations and migrates at a similar rate to H-NS-DNA complexes resolved in the absence of $MgCl_2$, probably represents saturation of the DNA molecules with H-NS. Saturation of DNA with H-NS disrupts the formation of DNA-H-NS-DNA bridges (bridged complex) and instead results in nucleation along the full length of the DNA (stiffened complex) (Tupper *et al.*, 1994). This would also explain the subtle shifting of fully resolved H-NS-DNA complexes in the absence of $MgCl_2$ whereby a dramatic shift occurs in the initial binding of H-NS to the DNA but increasing the amount of H-NS, and thus coating the complete DNA molecule, has only a minor effect on the migration of the DNA. The high affinity H-NS-DNA complex that was initially resolved is a stiffened DNA molecule and did not contain any looping due to DNA bridging. The transition from looped DNA to stiffened DNA results in a larger change in DNA shape which is reflected in the dramatically altered migration of the saturated H-NS-DNA complex (Fig. 4.12).

That a large number of the H-NS bound DNA molecules are not resolved at all indicates that DNA structures incapable of entering the polyacrylamide matrix are forming. This is most clearly visible for *proV*. As H-NS concentration increases, two distinct complexes are formed (Fig. 4.11B). The detected signal from these complexes is +17% relative to signal from the 0 nM DNA probe. This indicates that most of the DNA in the reaction has entered the gel (a positive value is reasonable considering the 0 nM DNA band is saturated). As the amount of H-NS in the reaction increases, first the lower 'bridged' complex and lastly the higher 'stiffened' complex disappear. At 100 nM H-NS <40% of the DNA resolved at 25 nM H-NS is detected (analysed by densitometry). This formation of higher molecular weight complexes incapable of entering the gel matrix is dependent on

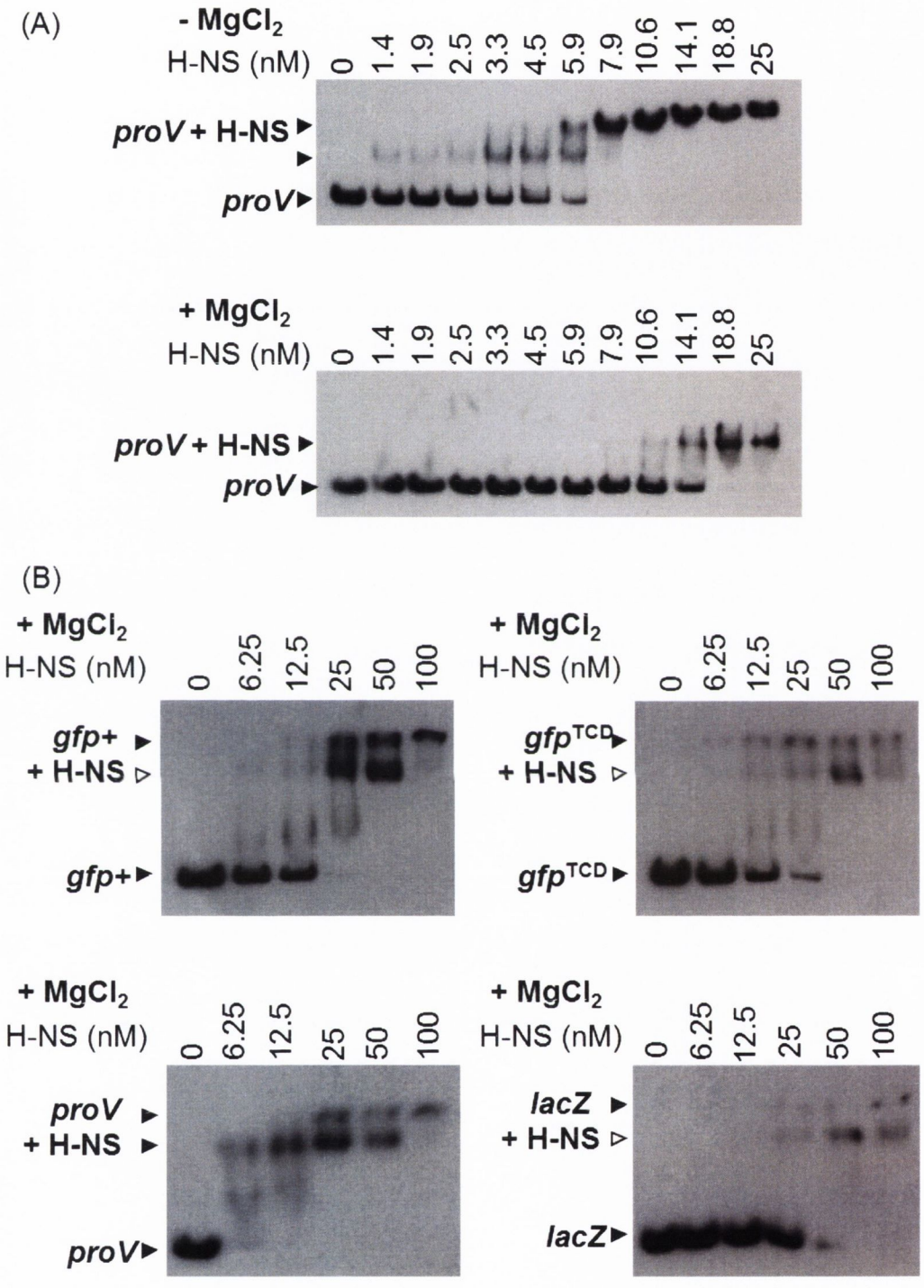


Fig. 4.11 H-NS binding is altered by the addition of $MgCl_2$. (A) Electrophoretic mobility shift assay (EMSA) analysis using purified H-NS to *proV* in the presence and absence of 10 mM $MgCl_2$. (B) *gfp*+, gfp^{TCD}, *proU* and *lacZ* probes incubated in the presence of 10 mM $MgCl_2$. Complexes that form only in the presence of $MgCl_2$ are indicated by unfilled arrows. The concentration of purified H-NS used is indicated above each lane.

$MgCl_2$ since in Fig. 4.12B *gfp+* and *gfp*^{TCD} probes incubated with 400 nM were clearly resolved.

The resolution of multiple *proV*-H-NS complexes in the absence of $MgCl_2$ could be due to the sequential filling of multiple binding sites. It could however also indicate that H-NS can form DNA-bridges within *proV* in the absence of $MgCl_2$, although $MgCl_2$ does still promote formation of the bridged complex.

4.2.9 *gfp+*, but not *gfp*^{TCD}, represses transcription *in vivo*.

Since H-NS binding is often associated with transcriptional repression, the effect of H-NS binding in *gfp+* on transcription from *proU* was tested. In *E. coli*, the products of the *proU* operon mediate the cytoplasmic accumulation of compatible solutes following osmotic upshock, an environmental stress that activates transcription from the *proU* promoter (Gowrishankar, 1985). Transcriptional repression of the *proU* operon in low osmolarity conditions involves binding by H-NS to the A+T-rich, highly curved DNA located in the upstream and downstream regulatory elements (URE and DRE, respectively) that flank the *proU* promoter (Owen-Hughes *et al.*, 1992; Lucht *et al.*, 1994; Bouffartigues *et al.*, 2007). H-NS bound at the DRE represses transcription by preventing open-complex formation (Ueguchi and Mizuno, 1993). This repressive complex is disrupted by increasing the osmolarity of the medium. While H-NS binding in the DRE is essential for osmoregulation of *proU*, it has previously been shown that the native DRE can be functionally replaced by an unrelated piece of DNA that is bound by H-NS (Owen-Hughes *et al.*, 1992). It was also shown that the *lacZ* gene cannot functionally replace the DRE and thus its presence in place of the DRE leads to derepression of *proU* in low osmolarity conditions (100 mM NaCl) (Owen-Hughes *et al.*, 1992). Therefore we tested the ability of *gfp+*, *gfp*^{TCD}, or *lacZ* to functionally replace the DRE by insertion of each gene within the DRE at +98-bp downstream from the transcriptional start site (Fig. 4.13). Expression of the *proU-gfp+*(+98) fusion was repressed relative to expression of *proU-gfp*^{TCD}(+98), indicating that *gfp+*, but not *gfp*^{TCD}, functionally replaced the DRE as a binding site for H-NS and thus repressed *proU* expression. As previously described, the *proU-lacZ*(+98) fusion was not repressed at low NaCl concentrations (Owen-Hughes *et al.*, 1992). Reporter fusions that did not disrupt the DRE were generated by insertion of each of the three reporter genes (*gfp+*, *gfp*^{TCD} or *lacZ*) at +936-bp downstream from the transcriptional start site (Fig.

4.13A). As predicted, expression of all three *proU*(+936) fusions was repressed by H-NS binding to the DRE at NaCl concentrations below 200 mM however, the *proU-gfp+* construct required more NaCl to antagonize the H-NS mediated repression of *proU* (compare *proU-gfp+* with *proU-gfp*^{TCD} at 100-200mM NaCl in Fig. 4.13A). This suggests that the presence of a second A+T-rich curved piece of DNA (*gfp+*) downstream of the DRE acts to reinforce the repression of *proU*.

To confirm that H-NS binding to *gfp+* accounted for repression of *proU-gfp+*(+98) *in vivo*, the expression of all *proU* fusions was tested in a Δ*hns* background. Cells were cultured in the repressive conditions of 100 mM NaCl, and the data expressed as a percentage of maximal derepressed expression to facilitate comparisons between GFP fluorescence and β-galactosidase activity (Fig. 4.13B). For all three fusions, *proU* expression was elevated in the absence of H-NS. This revealed that even in the absence of the DRE, H-NS continued to bind the URE and repress the *proU-gfp*^{TCD}(+98) and *proU-lacZ*(+98) fusions. These data allowed an assessment of the relative effects of replacing only the DRE compared to elimination of H-NS protein from the cell (Fig. 4.13C). As shown in Fig. 4.13A, replacing the DRE with *gfp*^{TCD} resulted in an 8-fold increase in expression relative to expression from the *proU-gfp*^{TCD}(+936) fusion. A similar comparison between *proU-gfp+* at position +98 and +936 showed that *gfp+* inserted in the DRE maintained full repression. Consequently, eliminating H-NS genetically resulted in a 7.3 to 9.4-fold increase in expression from the *gfp+* fusions while the expression of *proU-gfp*^{TCD}(+98) improved only 2.2-fold upon removal of H-NS (Fig. 4.13C). This also confirmed that H-NS-mediated repression of *proU* occurs primarily through binding to the DRE.

Even in the absence of H-NS, *proU-gfp*^{TCD}(+98) produced 3-fold higher levels of GFP than *proU-gfp+*(+98). *gfp*^{TCD} is codon optimized for translation in *E. coli*, and these data suggest that *gfp*^{TCD} is translated 3-fold more efficiently than *gfp+* (Corcoran *et al.*, 2010). *gfp*^{TCD} mitigates the H-NS associated repression by *gfp+* and is optimally translated in *E. coli*, providing an ideal template for the directed evolution of novel *gfp* variants.

The increased fluorescence of *proU-gfp*^{TCD}(+98) compared to *proU-gfp+*(+98) was analysed by fluorescence microscopy, confirming that GFP^{TCD} is present at higher levels than GFP+ under high-osmolarity conditions (Fig. 4.14).

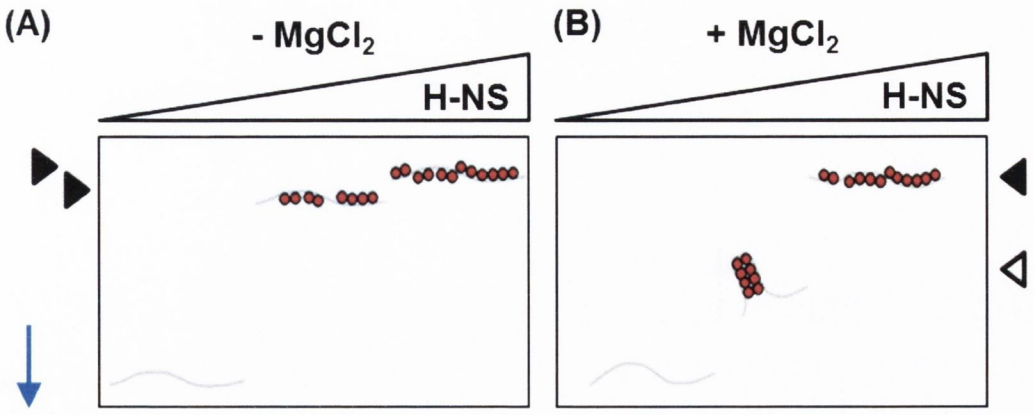
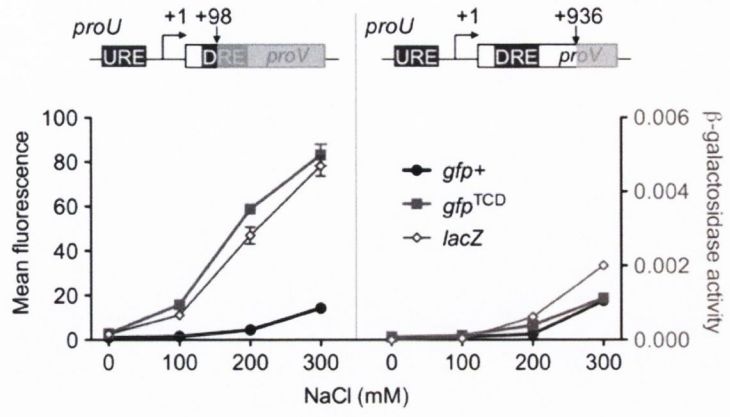
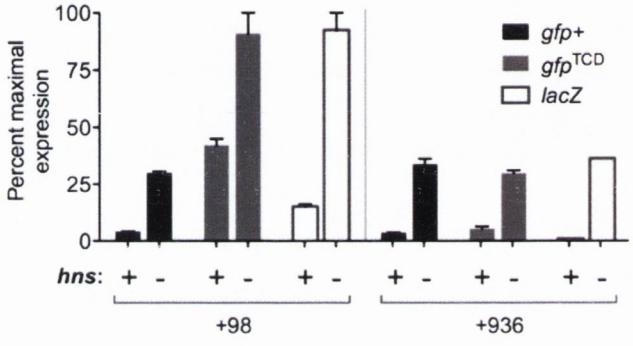


Fig. 4.12 Model for the effect of $MgCl_2$ on H-NS binding *in vitro*. (A) In the absence of $MgCl_2$, H-NS (red balls) binds to DNA (grey line) according to the stiffening model (filled arrowheads). An increase in H-NS binding to H-NS-stiffened DNA results in minor changes to DNA topology; reflected in the minor change in electrophoretic migration of the DNA. (B) $MgCl_2$ promotes H-NS-mediated DNA bridging (open arrow head). An increase in H-NS binding causes disruption of the DNA-H-NS-DNA bridge and formation of a H-NS-stiffened DNA molecule (filled arrowhead). This results in a large change in DNA topology which is reflected in a large change in electrophoretic mobility. The direction of electrophoresis is indicated by a blue arrow.

(A)



(B)



(C)

Fold increase in expression			
		gfp+	*gfp*^{TCD}
When DRE is replaced (+98 / +936)	wildtype	1.1	8.0
	Δ*hns*	0.9	3.1
When H-NS is removed (Δ*hns* / wildtype)	+98	7.3	2.2
	+936	9.4	5.6

Fig. 4.13 *gfp*+, but not *gfp*^{TCD}, has repressive effects on transcription.(A) The effect of increasing NaCl concentrations on the expression of reporter gene fusions inserted at +98 and +936 base pairs from the *proU* transcriptional start site. Mean values and ranges are plotted. Diagrams not to scale. (B) The effect of removing H-NS on fusion gene expression in 100 mM NaCl. Mean values and ranges are plotted. (C) Fold-increases in fusion gene expression caused by changes in H-NS occupancy.

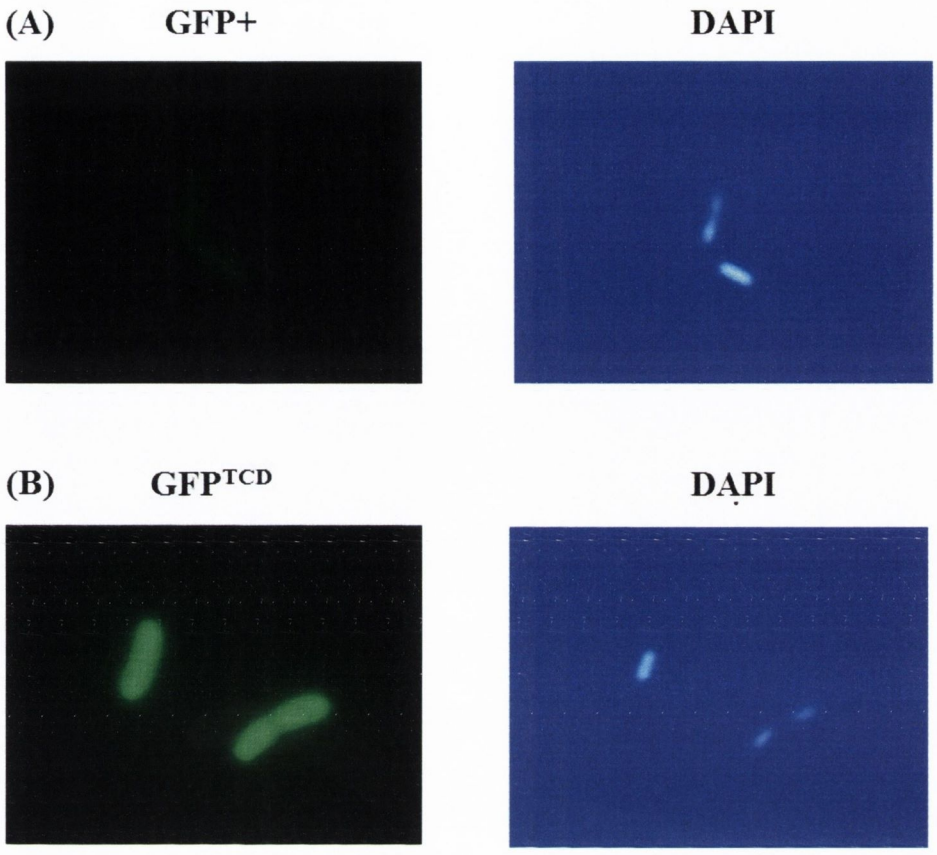


Fig. 4.14 Analysis of *gfp*+ and *gfp*^{TCD} containing strains using fluorescent microscopy. Strains containing either *gfp*+ or *gfp*^{TCD} inserted at +98 base pairs from the transcriptional start site of *proU* were grown overnight in the presence of 300mM NaCl. Samples were fixed and imaged under identical conditions (see Materials and Methods; section 2.11). Since the GFP+ and GFP^{TCD} proteins are identical, the variation in fluorescence intensity between GFP+ and GFP^{TCD} must be due to increased levels of GFP^{TCD}. Nucleoid staining using DAPI was performed as an control.

4.2.10 Mutation of the *gfp* chromophore

To highlight the utility of pCoC1 (*gfp*^{TCD}-*kan*, cloned in pUC derivative pJ204) in creating new fluorescent proteins a Y66H mutation was incorporated into the *gfp*^{TCD} chromophore by site directed mutagenesis using primers gfpTCD.Y66H.fw/.rv. Two point mutations were introduced to convert a tyrosine residue to a histidine at amino acid 66, which is within the GFP chromophore. This mutation lowered the emission maxima from 508 nm to 448 nm thus producing a blue light (Fig. 4.15A) (Heim *et al.*, 1994). As expected, the minor mutations required to dramatically alter the characteristics of the protein had no effect on the predicted intrinsic curvature or A+T content of the DNA (Fig. 4.15B).

4.2.11 Fluorescent proteins from other species are A+T-rich.

The red fluorescent protein FP583 (commercially known as DsRed) is a recently cloned 28-kDa fluorescent protein isolated from the coral of the *Discoma* genus (Matz *et al.*, 1999). DsRed has an emission maximum of 583 nm, which can be further extended to 602 nm by mutation of Lys-83 to Met (Matz *et al.*, 1999; Baird *et al.*, 2000). This emission spectrum makes it ideal as a fluorescent partner for GFP, as fluorescence of both proteins can be individually measured in a single cell (Ayoob *et al.*, 2001; Maksimow *et al.*, 2002; Hakkila *et al.*, 2003). Similar to the early studies of GFP, a number of the limitations of this protein were addressed through random mutagenesis of the coding sequence and screening for improved variants (Baird *et al.*, 2000; Campbell *et al.*, 2002; Verkhusha and Sorkin, 2005). Initially the major issues with DsRed included the long maturation time before a red signal is detected and that it forms tetramers, both undesirable characteristics for a transcriptional or translational fusion.

While providing a potential alternative to *gfp*+, *dsred* contains the same intrinsic features make *gfp*+ a target for H-NS binding; the *dsred* gene is 55% A+T and is predicted to contain strong intrinsic curvature (Fig. 4.16). The same method of optimization that was applied to *gfp*+ was used to alter the *dsred* gene *in silico*. The selected optimized gene, *dsred*^{TCD}, differs from *dsred* by 155 nucleotide substitutions across the 678 bp gene has reduced A+T content (49%) and reduced predicted DNA curvature (Fig. 4.16). While *dsred*^{TCD} was not synthesized or tested *in vivo*, it contains similar intrinsic characteristics to *gfp*^{TCD} and is therefore a poor target for H-NS (Fig. 4.17).

4.3 Discussion

In many bacteria, including *E. coli*, H-NS selectively targets horizontally acquired genes of high A+T content preventing transcription of the potentially deleterious foreign DNA (Dorman, 2007). It was shown in this study that the gene encoding the green fluorescent protein (*gfp*), and many of its derivatives, contains the key motifs required for H-NS silencing of horizontally acquired DNA. It was shown that H-NS binds strongly to a modern variant of the green fluorescent protein gene (*gfp+*) both *in vivo* and *in vitro*. The introduction of A+T-rich plasmid DNA can titrate H-NS away from native chromosomal locations leading to a mild Δ*hns* phenotype with pleiotropic effects including reduced virulence and motility (Doyle *et al.*, 2007; Dillon *et al.*, 2010). Titration of H-NS may therefore explain a recent report of reduced invasiveness of *Salmonella* due solely to the presence of a promoter-less *gfp+* gene on a multi-copy plasmid (Clark *et al.*, 2009).

We have shown that H-NS binding in *gfp+* can repress transcription from a proximal promoter and may also be responsible for altered regulation of *fimS* inversion in response to DNA relaxation in a *fimA-gfp+* fusion strain. This could have wide-spread implications for researchers utilizing *gfp* as the sole reporter of transcriptional activity. Over the years, many bacterial studies have relied solely on plasmid-based *gfp* fusions, and some of these studies may have experienced unrecognized local and global effects due to H-NS repression and H-NS titration, respectively.

The *gfp* gene has been used in conjunction with plasmid based high-throughput genetic screens in many landmark studies including the identification of promoters involved in the intracellular survival of *Salmonella enterica* serovar Typhimurium (Valdivia and Falkow, 1997) and the adaptation of the metabolism of *E. coli* to different carbon sources (Zaslaver *et al.*, 2006). Bacterial systems, particularly those based on *E. coli*, are commonly used to study the fundamental processes common to all forms of life. A recent example involving *E. coli* used a plasmid based *gfp* as a reporter to study the coding-sequence determinants of gene expression (Kudla *et al.*, 2009). Kudla *et al.*, created 154 variants of *gfp* that encoded an identical GFP protein but differed dramatically in coding sequence. They deduced that 5' mRNA structure was the key codon sequence determinant of gene expression, accounting for 59% of variation in fluorescence when under the control of a bacterial promoter (Kudla *et al.*, 2009). They also showed that the addition of a 28-codon leader sequence that had low mRNA secondary structure increased the expression of all constructs to a uniformly high level of expression. The variation in coding sequence

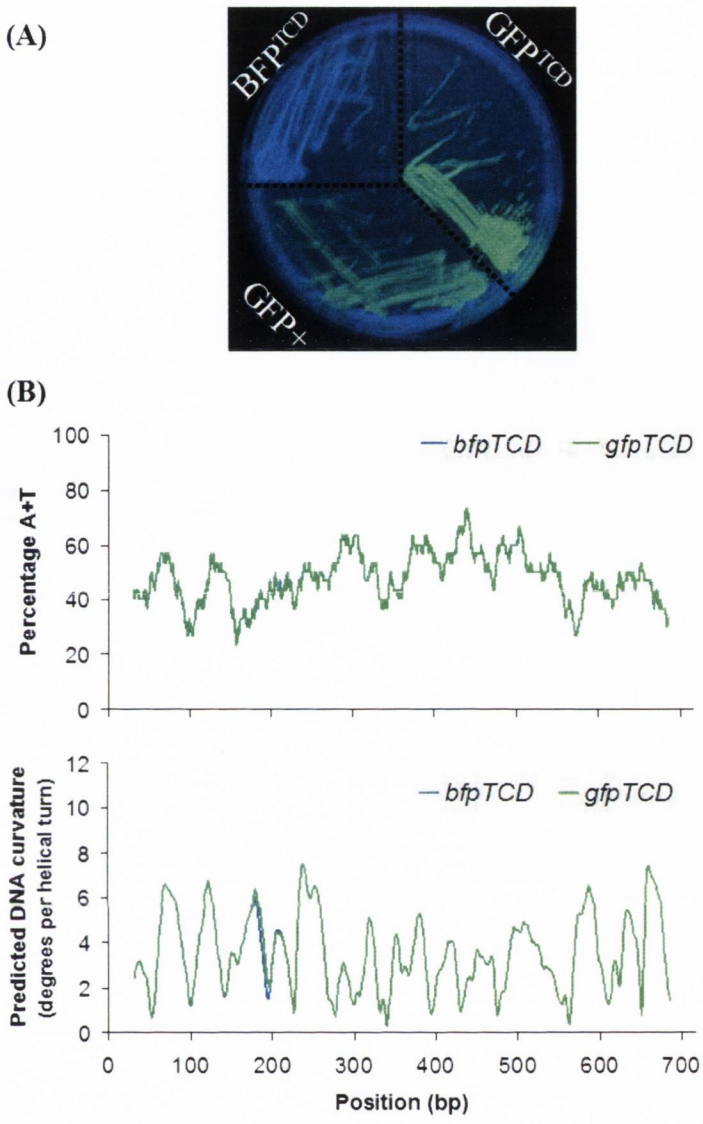


Fig. 4.15 Mutation of *gfp*^{TCD} to create a blue shifted variant (*bfp*^{TCD}). (A) Visual comparison of fluorescent proteins GFP+, GFP^{TCD} and BFP^{TCD} (GFP^{TCD} containing a Y66H mutation). All three genes are under the control of the *prp* promoter (Lee and Keasling, 2005). Strains were grown overnight in the presence of 25 µg/ml proprionate. Fluorescence was visualized using a handheld UV lamp (UVGL-58, UVP, Cambridge, UK) at 365nm excitation. Images were captured using a digital camera (Canon, SD1100IS). (B) Comparison of *gfp*^{TCD} and *bfp*^{TCD} A+T content and predicted intrinsic DNA curvature.

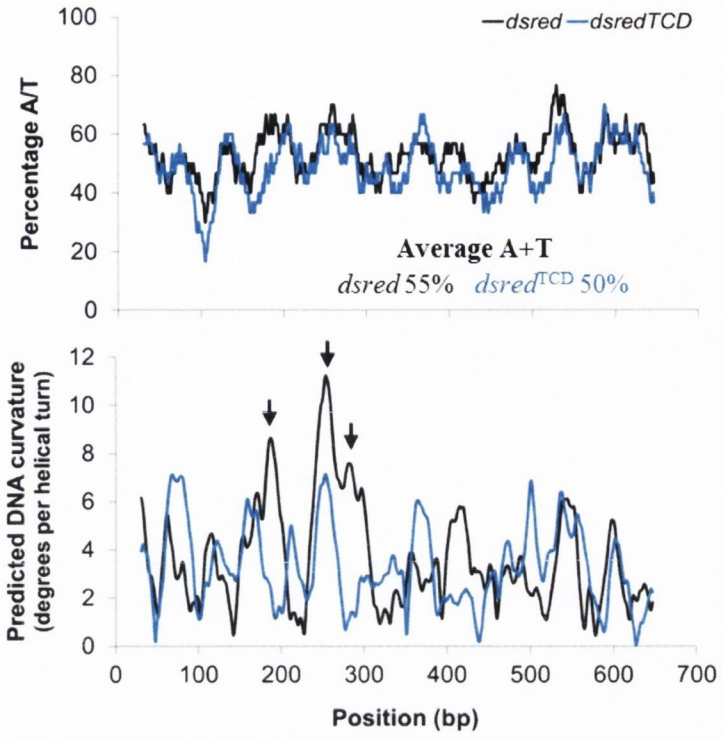


Fig. 4.16 DNA sequence features of the *dsred* gene are conducive to H-NS binding. A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot) of *dsred* and *dsred*^{TCD}. Arrows highlight regions of strong predicted curvature in *dsred* that are reduced in *dsred*^{TCD}.

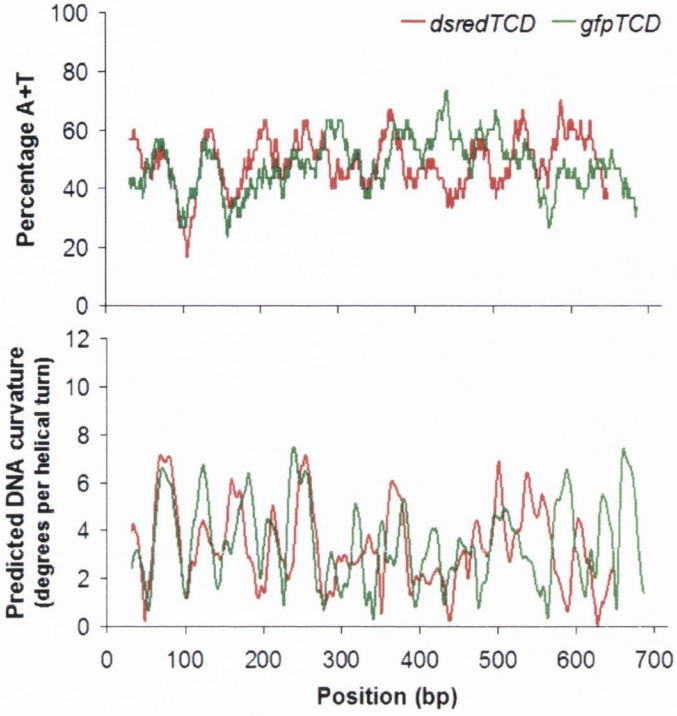


Fig. 4.17 *gfp*^{TCD} and *dsred*^{TCD} contain similar intrinsic DNA features that are not conducive to H-NS binding. Data shown are percentage A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot).

resulted in dramatic changes in both A+T content and intrinsic curvature. While 5' mRNA structure accounts for variation in expression from a strong, actively transcribing promoter, H-NS binding could affect gene expression when lower amounts of inducer are present or the leakiness (expression in the absence of inducer) of the promoter. In this study, 80 of the Kudla *et al. gfp* constructs were screened for A+T content and predicted intrinsic curvature in an attempt find a correlation between expression of the *gfp* variants that were strong targets for H-NS and variants that were weak targets for H-NS (Fig. 4.18). A+T content and predicted intrinsic curvature varied widely among the *gfp* constructs including some with an average A+T content >60% containing multiple highly curved regions similar to the wild type *gfp* (e.g. construct 3, Fig. 4.18). Consistent with an effect of H-NS repression construct 3 construct was less fluorescent than genes of lower A+T content and intrinsic curvature (constructs 2 and 12, Fig. 4.18). Conversely, multiple examples of low A+T constructs were shown to be poorly fluorescent (e.g. construct 14, Fig. 4.18) compared with construct 3 making deciphering a role for H-NS in this system by *in silico* methods alone difficult. While no consistent correlation was observed between fluorescence of ideal H-NS targets and fluorescence of poor H-NS targets, H-NS binding within coding sequences can repress transcription (Nagarajavel *et al.*, 2007) and thus should be considered when studying the coding-sequence determinants of gene expression.

We have designed a new *gfp* that encodes an identical protein to GFP+ but is designed to resemble an *E. coli* gene and thus circumvent H-NS silencing activity of xenogenic DNA (Appendices 3 and 4; Corcoran *et al.*, 2010). This new *gfp*, gfp^{TCD}, is bound with lower affinity by H-NS (Fig. 4.10) and does not repress local transcription (Fig. 4.13).

High levels of GFP, usually associated with expression from a multi-copy plasmid, have previously been shown to be toxic to the host cell, leading to dramatic plasmid loss or mutation in order to reduce the amount of GFP being produced (Hautefort *et al.*, 2003). It is possible that a part of this toxicity is due to the sub-optimal codon usage of *gfp*+ leading to reduced translation efficiency. This would cause an increased burden on the translational machinery (including the ribosomal components), preventing efficient translation of essential host genes and thus, create toxic effects on the host (Andersson and Kurland, 1990; Kudla *et al.*, 2009). gfp^{TCD} is optimized for high expression in *E. coli* and thus may have reduced toxicity when highly expressed.

Since nucleoid associated proteins with a preference for A+T-rich DNA are found in most cell types, including archaea and eukaryotes (Dillon and Dorman, 2010), our discovery of

H-NS binding and repression of *gfp* may have far-reaching implications for the use of *gfp* in all domains of life. For example, in eukaryotes the introduction of an A+T-rich curved DNA region proximal to a promoter facilitated histone binding and nucleosome formation and lead to transcription repression (Wolffe and Drew, 1989). Also, the murine *btcD* (binds to curved DNA) gene can complement an *hns* deletion mutation in *E. coli* (Timchenko *et al.*, 1996) highlighting how these basic proteins are conserved in all domains of life.

This raises the possibility that many studies utilizing *gfp* have exhibited unrecognised silencing of *gfp*.

The high A+T content of *gfp*+ may also target it for post-transcriptional regulation by the Sm-like global regulator Hfq (Brennan and Link, 2007). Hfq binds to A+U-rich RNA that is flanked by a stem-loop structure altering translation of mRNA both directly and through facilitating the action of small non-coding RNAs (Brennan and Link, 2007). Since gfp^{TCD} has reduced A+T content compared to other *gfp* variants, it is expected to be less susceptible to post-transcriptional regulation by Hfq.

The initial aim of this body of work was to create a *fimA-gfp* fusion that could be used to rapidly determine *fimA* promoter activity and *fimS* orientation. A *fimA-*gfp^{TCD} fusion was therefore created that will allow direct monitoring of *fimS* inversion without the added layer of regulation associated with previous versions of *gfp*.

This study shows that with its high A+T content and intrinsic curvature, the DNA in the *gfp* gene makes it a favourable target for the bacterial transcriptional repressor H-NS. Previously the *luxAB* was shown to be bound strongly by H-NS leading to transcriptional repression (Owen-Hughes *et al.*, 1992; Forsberg *et al.*, 1994). This subsequently led to its reduced use in bacterial systems. Here we provide a novel solution to the H-NS-mediated silencing of a reporter gene. This method can be applied to any DNA from any genetic background to prevent binding of primordial gene silencing proteins, such as H-NS, as thus reduce unrecognised regulatory inputs. To illustrate this a red fluorescent protein from a coral of the *Discosoma* genus, which was shown in this study to be a favourable target for H-NS silencing, was successfully reverse-engineered to reduce H-NS affinity (Fig. 4.16). Using this method 'H-NS-proofing' of reporter genes is relatively easy and cost-effective and thus should be considered for all reporter genes. The development of gfp^{TCD} that is both a poor target for H-NS and a more accurate reporter of transcription provides a new tool for future studies free from artefacts caused by H-NS interference.

(A)

No.	Fluorescence (A/U)	Average A+T (%)
2	7673	53
3	5388	62
12	6350	53
14	978	46

(B)

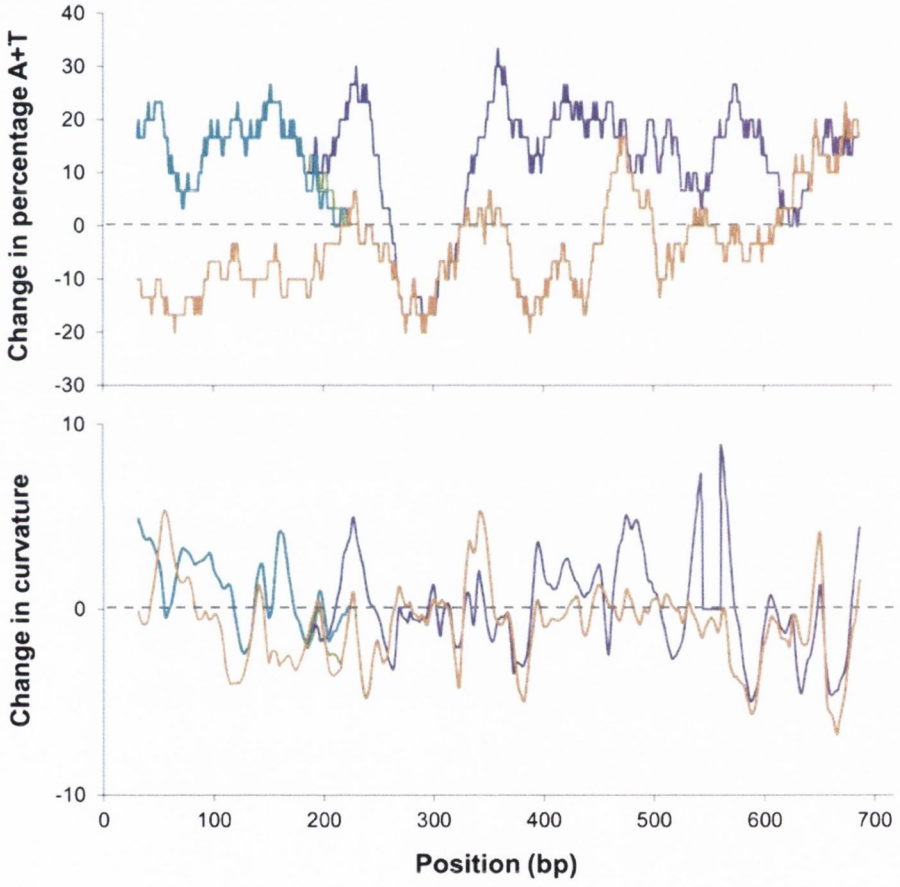


Fig. 4.18. Predicted H-NS binding does not directly correlate with fluorescence levels. The fluorescence values and average A+T content of 4 representative samples of the 154 *gfp* variants generated by Kudla *et al.*, (2009) are shown in (A). The change in percentage A+T content and change in predicted intrinsic curvature of the constructs detailed in (A) relative to gfp^{TCD} are shown in (B).

Chapter 5.

General Discussion.

Type 1 fimbriae are typically expressed in nutrient poor environments and facilitate colonization through attachment when the bacterium can no longer support a motile lifestyle that demands a high metabolic flux (Müller *et al.,* 2009). They play a key role in the colonization of various host tissues by *E. coli* (Yamamoto *et al.*, 1990; Bloch *et al.*, 1992) and in biofilm formation on abiotic surfaces such as catheters (Pratt and Kolter, 1998; Schembri and Klemm, 2001a).

At a molecular level, the connection between fimbrial expression and low metabolic flux is known to involve the general stress response regulator ppGpp, which is induced in poor growth conditions and acts positively at the P_{fimB} promoter to increase fimbriation (Aberg *et al.*, 2006, 2008). A second feature of bacteria growing in poor growth conditions is the global reduction in DNA supercoiling levels, which correlates with an increase in fimbriation (Dove and Dorman, 1994; Kelly *et al.*, 2006; Corcoran and Dorman, 2009; Müller *et al.*, 2009).

In bacteria growing in mid-exponential phase at 37°C in LB broth, the FimB protein inverts the *fimS* element from the on phase to the off phase and back again at approximately equal rates (McClain *et al.*, 1991). However, when novobiocin was used to inhibit DNA gyrase activity, causing DNA to become more relaxed, the FimB-catalyzed *fimS* inversion reaction adopts a pronounced bias in favour of the ON orientation (Dove and Dorman, 1994; Kelly *et al.*, 2006; Corcoran and Dorman, 2009; Müller *et al.*, 2009). Previously, LRP was identified as playing a role in determining the directionality in this reaction (Kelly *et al.*, 2006). Specifically, the LRP protein is required to maintain the *fimS* element in the ON orientation and it does this by binding to the LRP-1 and LRP-2 sites within *fimS* (Fig. 1.7). In the present study (Corcoran and Dorman, 2009), the IHF protein bound to the IHF-1 site was identified as a second recombination directionality factor (RDF) that is required to maintain *fimS* in the on orientation when gyrase activity is inhibited (Fig 3.3B). The location of the IHF binding site in the non-inverting chromosomal DNA immediately adjacent to IRL explains the differential interaction of IHF with LRP bound to sites LRP-1 and LRP-2 within *fimS* because all three sites are only juxtaposed when the switch is in the on orientation (Fig. 1.5). This work supports a model whereby the DNA bending activity of IHF and the DNA wrapping activity of LRP in combination with the relaxed nature of the *fim* DNA that obtains in the absence of adequate levels of DNA gyrase activity creates a

nucleoprotein complex that prevents DNA inversion by FimB in the off-to-off direction (Fig. 3.11). This constitutes a phase-on 'trap'.

DNA relaxation is a feature of the stationary phase of growth that correlates with a reduction in the negative supercoiling activity of DNA gyrase (Bordes *et al.,* 2003) and is an important factor in controlling the expression of many virulence genes (Dorman and Corcoran, 2009). It is possible that the *fim* switch becomes biased towards the on phase when bacteria approach a state of low metabolic flux since this will promote type 1 fimbrial expression and assist attachment and colonization. The involvement of the LRP and IHF proteins in the establishment of a phase-on bias is also consistent with a need for physiological sensitivity on the part of *fim* gene expression. IHF protein levels peak at the transition from exponential growth to stationary phase (Ditto *et al.,* 1994) making this protein an attractive growth stage reporter. LRP has been established as a regulator of genes involved in the response to nutrient stress (Brinkman *et al.,* 2003; Yokoyama *et al.,* 2006) and LRP levels increase as bacteria approach stationary phase (Landgraf *et al.*, 1996).

Nucleoid associated proteins and DNA topology combine to bias the outcome of other site-specific recombination events. For example, IHF binding to a site within phage λ plays a key role in promoting integration into the host chromosome while simultaneously inhibiting excision from the host chromosome (Bushman *et al.*, 1985; Thompson *et al.*, 1986). Changes in DNA supercoiling levels also impart directionality to the recombination reaction since integration, but not excision, is dependent on negative DNA supercoiling (Crisona *et al.*, 1999).

IHF also directly controls the formation of the Tn*10* transposome in response to DNA supercoiling (Chalmers *et al.*, 1998); a role that is directly antagonized by H-NS (Singh *et al.*, 2008), and IHF is important for modulating the responsiveness of Mu transposition to DNA supercoiling (Surette and Chaconas, 1989; Surette *et al.*, 1989). This study therefore provides another detailed example of IHF and DNA supercoiling influencing a site-specific recombination event that involves DNA inversion rather than integration and excision.

Previous work has suggested a role for the H-NS protein in influencing the FimB-mediated inversion of *fimS* (Higgins *et al.*, 1988; Kawula and Orndorff, 1991; Dorman and Ní Bhriain, 1992; O'Gara and Dorman, 2000). Careful examination of H-NS interactions with

fimS and its flanking regions showed that H-NS interaction with the switch was contingent on *fimS* orientation (section 4.2.7). This study showed that the H-NS binding site in the *fimS* region is distributed across three incomplete sites. One of these is in the invertible *fimS* element and is adjacent to the P_{fimA} promoter. The others are located in the non-inverting parts of the chromosome immediately outside IRL and IRR. When the switch is in the off orientation, a complete site capable of binding H-NS in an electrophoretic mobility shift assay is formed at the IRL end of the switch (Fig. 3.7); when *fimS* is in the on orientation, a complete H-NS interaction site is formed only at the IRR end (Fig. 3.8). The data obtained in this study show that the H-NS-DNA complex involving IRL maintains *fimS* in the off orientation and constitutes an 'off trap'. The absence of H-NS causes a dramatic increase in the percentage of phase-on cells in a population (Fig. 3.6A), which supports previous reports of a locked-on phenotype in H-NS deficient strains (Higgins *et al.*, 1988; Kawula and Orndorff, 1991).The physiological significance of H-NS binding at IRR when *fimS* is in the on orientation is unknown, although it may play a role in modulating the activity of the P_{fimA} promoter. The modulation of H-NS binding by DNA inversion demonstrates a novel mechanism for H-NS interacting with DNA.

By furthering our understanding of *fimS* regulation, this study may aid the development of new biological tools. The *fim* system has recently been used in conjunction with the *hin* invertible system to create a heritable system that 'remembers' its previous orientation (Ham *et al.*, 2008). The *hin* system was the basis for a 'bacterial computer' which has been used to solve complex mathematical problems (Baumgardner *et al.*, 2009). The *fim* system has also been utilized to create a tightly-controlled expression vector for the expression of toxic genes (Ham *et al.*, 2006). The effective use of *fimS* in biotechnology applications requires a thorough knowledge of the underlying genetic regulation of the system. This study has made important contributions to furthering this understanding.

A major finding of this study was that the ubiquitous reporter gene *gfp* was bound by the global repressor protein H-NS. H-NS binding in *gfp* caused a repression of transcription *in vivo* and thus previous studies utilizing *gfp* may be complicated by added H-NS regulation. Transcriptional fusions to *gfp* facilitates the high-throughput *in vivo* monitoring of gene expression at a single-cell level and have been used in landmark studies including the identification of genes involved in virulence of *Salmonella* (Valdivia and Falkow, 1997)

and metabolism in *E. coli* (Zaslaver *et al.*, 2006). Transcriptional fusions to *gfp* are also used as biological sensors for signals such as water availability (Axtell and Beattie, 2002) and environmental pollutants (Belkin, 2003). In this study, a highly fluorescent, fast-folding variant of *gfp* (*gfp*+) was reverse-engineered to reduce H-NS binding affinity without altering the amino acid sequence of the protein (Chapter 4). The reverse engineering method for reducing H-NS affinity for a region of DNA was developed as part of this study and was the subject of a patent application (Appendix 3). This novel method can be applied to any DNA, including other fluorescent proteins (Fig. 4.16). The 'H-NS-proofing' of reporter genes alleviates unforeseen complications associated with H-NS binding, such as the titration of H-NS from chromosomal genes to *gfp* on a multi-copy plasmid (Doyle *et al.*, 2007; Dillon *et al.*, 2010).

One aim of this project was to create a *fimA-gfp* fusion that would allow monitoring of *fimS* orientation and P_{fimA} activity in single cells. Creation of a *fimA* transcriptional fusion to the *gfp*+ gene caused an alteration in the DNA relaxation-induced phase-on bias of *fimS*, possibly through the interaction of H-NS bound in the *gfp*+ gene with H-NS bound in *fimS* and the subsequent creation of a new topological domain surrounding *fimS* (Fig. 4.5B). However, this aspect of this study requires further investigation since attempts to more accurately define the response of the *fimA-gfp*+ fusion (and the multiple variants described in Fig. 4.3) using the low-level novobiocin treatment failed to mimic the altered response to DNA relaxation that was seen in the intial experiments. However, a *fimA-gfp*^{TCD} fusion was created which conclusively mimicked the wild-type response to DNA relaxation (Appendix 4 and data not shown). While this strain will be of use in future studies, the inherent stability of GFP may be problematic when using high-throughput methods such as flow cytometry to determine the strength and orientation of the P_{fimA} or when monitoring switching in a single cell using fluorescent microscopy (Adiciptaningrum *et al.*, 2009). For example, a cell which switches from phase on to phase off will be visualized as phase on until GFP is diluted out of the cell by growth. A GFP variant with a shorter half-life (Andersen *et al.*, 1998) would more accurately reflect changes in the strength and orientation of the P_{fimA}. A second fluorescent protein gene, such as *bfp*^{TCD} or *dsred*^{TCD}, could be incorporated into *fimE* so that it is transcribed by P_{fimA} in the off orientation. This would provide a robust system for monitoring *fimS* inversion in a single cell and permit the rapid simultaneous measurement of promoter orientation and activity at a single-cell level. Single-cell analysis may also reveal new aspects of regulation within populations that could

not be identified using techniques that involve assaying whole populations and reporting average values (Paulsson, 2004; Newman *et al.*, 2006). That 10% of phase-on cells do not produce fimbriae suggests that production of fimbriae is bistable (McClain *et al.*, 1993). Random variation is inherent in biological systems. Many systems have evolved to reduce this variation or noise while in other systems the noise is amplified and results in the generation of phenotypic diversity termed bistability (Paulsson, 2004). Single-cell studies may provide insights into this poorly characterized aspect of fimbrial regulation.

This study has made important fundamental and applied contributions to prokaryotic research both through the novel insights gained in the complex regulation of the *fim* invertible element and through the development of a novel *gfp* gene that is a more accurate reporter of transcriptional activity than previous *gfp* variants. By furthering our understanding of the regulation of an important determinant of bacterial virulence and increasing the accuracy of a ubiquitous reporter gene, this study will hopefully contribute to the ultimate aim of developing improved therapeutic or preventive interventions in infection.

Bibliography

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

Aberg, A., Shingler, V. and Balsalobre, C. (2008) Regulation of the *fimB* promoter: a case of differential regulation by ppGpp and DksA *in vivo*. *Molecular Microbiology* **67**: 1223-1241.

Abraham, J. M., Freitag, C. S., Clements, J. R. and Eisenstein, B. I. (1985) An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proceedings of the National Academy of Science U S A* **82**: 5724-5727.

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

Adiciptaningrum, A. M., Blomfield, I. C. and Tans, S. J. (2009) Direct observation of type 1 fimbrial switching. *EMBO Reports* **10**: 527-532.

Aihara, H., Kwon, H. J., Nunes-Duby, S. E., Landy, A. and Ellenberger, T. (2003) A conformational switch controls the DNA cleavage activity of lambda integrase. *Moecular Cell* **12**: 187-198.

Ali, B. M., Amit, R., Braslavsky, I., Oppenheim, A. B., Gileadi, O. and Stavans, J. (2001) Compaction of single DNA molecules induced by binding of integration host factor (IHF). *Proceedings of the National Academy of Science U S A.* **19**: 10658-10663.

Amit, R., Oppenheim, A. B. and Stavans, J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *The Biophysical Journal* **84**: 2467-2473.

Amundsen, S. K., Taylor, A. F., Chaudhury, A. M. and Smith, G. R. (1986) recD: the gene for an essential third subunit of exonuclease V. *Proceedings of the National Academy of Science U S A* **83**: 5558-5562.

Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M. and Molin, S. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and Environmental Microbiology* **64**: 2240-2246.

Andersson, S. G. E. and Kurland, C. G. (1990) Codon preferences in free-living microorganisms. *Microbiological Reviews* **54**: 198-210.

André, C. (2009) New developments in PCR. *BioTechniques* **46**: 375-376.

Artsimovitch, I., Patlan, V., Sekine, S., Vassylyeva, M. N., Hosaka, T., Ochi, K., Yokoyama, S. and Vassylyev, D. G. (2004) Structural basis for transcription regulation by alarmone ppGpp. *Cell* **117**: 299-310.

Aviv, M., Giladi, H., Schreiber, G., Oppenheim, A. B. and Glaser, G. (1994) Expression of the genes coding for the *Escherichia coli* integration host factor are controlled by growth phase, *rpoS*, ppGpp and by autoregulation. *Molecular Microbiology* **14**: 1021-1031.

Axtell, C. A. and Beattie, G. A. (2002) Construction and characterization of a *proU-gfp* transcriptional fusion that measures water availability in a microbial habitat. *Applied and Environmental Microbiology* **68**: 4604-4612.

Ayoob, J., Shaner, N., Sanger, J. and Sanger, J. (2001) Expression of green or red fluorescent protein (GFP or DsRed) linked proteins in non-muscle and muscle cells. *Molecular Biotechnology* **17**: 65-71.

Azam, T. A. and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *Journal of Biological Chemistry* **274**: 33105-33113.

Badaut, C., Williams, R., Arluison, V., Bouffartigues, E., Robert, B., Buc, H. and Rimsky, S. (2002) The degree of oligomerization of the H-NS nucleoid structuring protein is related to specific binding to DNA. *Journal of Biological Chemistry* **277**: 41657-41666.

Baird, G., Zacharias, D. and Tsien, R. (2000) Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral. *Proceedings of the National Academy of Sciences U. S. A.* **97**: 11984-11989.

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. and Leibler, S. (2004) Bacterial persistence as a phenotypic switch. *Science* **305**: 1622-1625.

Balke, V. L. and Gralla, J. D. (1987) Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *Journal of Bacteriology* **169**: 4499-4506.

Ball, C. A. and Johnson, R. C. (1991a) Efficient excision of phage lambda from the *Escherichia coli* chromosome requires the Fis protein. *Journal of Bacteriology* **173**: 4027-4031.

Ball, C. A. and Johnson, R. C. (1991b) Multiple effects of Fis on integration and the control of lysogeny in phage lambda. *Journal of Bacteriology* **173**: 4032-4038.

Ball, C. A., Osuna, R., Ferguson, K. C. and Johnson, R. C. (1992) Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. *Journal of Bacteriology* **174**: 8043-8056.

Baumgardner, J., Acker, K., Adefuye, O., Crowley, S. T., Deloache, W., Dickson, J. O., Heard, L., Martens, A. T., Morton, N., Ritter, M., Shoecraft, A., Treece, J., Unzicker, M., Valencia, A., Waters, M., Campbell, A. M., Heyer, L. J., Poet, J. L. and Eckdahl, T. T. (2009) Solving a Hamiltonian Path Problem with a bacterial computer. *Journal of Biological Engineering* **3**: 11.

Belkin, S. (2003) Microbial whole-cell sensing systems of environmental pollutants. *Current Opinion in Microbiology* **6**: 206-212.

Benham, C. J. (1979) Torsional stress and local denaturation in supercoiled DNA. *Proceedings of the National Academy of Science U S A* **76**: 3870-3874.

Benjamin, H. W. and Cozzarelli, N. R. (1988) Isolation and characterization of the Tn3 resolvase synaptic intermediate. *EMBO Journal* **7**: 1897-1905.

Bernardi, G. and Bernardi, G. (1985) Codon usage and genome composition. *Journal of Molecular Evolution* **22**: 363-365.

Better, M., Wickner, S., Auerbach, J. and Echols, H. (1983) Role of the Xis protein of bacteriophage lambda in a specific reactive complex at the *attR* prophage attachment site. *Cell* **32**: 161-168.

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

Biek, D. P. and Cohen, S. N. (1986) Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. *Journal of Bacteriology* **167**: 594-603.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A. and Ellenberger, T. (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature* **435**: 1059-1066.

Blakely, G., May, G., McCulloch, R., Arciszewska, L. K., Burke, M., Lovett, S. T. and Sherratt, D. J. (1993) Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75**: 351-361.

Blakely, G. and Sherratt, D. (1996) Determinants of selectivity in Xer site-specific recombination. *Genes and Development* **10**: 762-773.

Blattner, F. R., Plunkett, G., 3rd, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277**: 1453-1462.

Bloch, C. A., Stocker, B. A. and Orndorff, P. E. (1992) A key role for type 1 pili in enterobacterial communicability. *Molecular Microbiology* **6**: 697-701.

Bloch, V., Yang, Y., Margeat, E., Chavanieu, A., Auge, M. T., Robert, B., Arold, S., Rimsky, S. and Kochoyan, M. (2003) The H-NS dimerization domain defines a new fold contributing to DNA recognition. *Nature Structural Biology* **10**: 212-218.

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

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

Blomfield, I. C., McClain, M. S. and Eisenstein, B. I. (1991a) Type 1 fimbriae mutants of *Escherichia coli* K12: characterization of recognized afimbriate strains and construction of new *fim* deletion mutants. *Molecular Microbiology* **5**: 1439-1445.

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

Bouffartigues, E., Buckle, M., Badaut, C., Travers, A. and Rimsky, S. (2007) H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing. *Nataure Structural and Molecular Biology* **14**: 441-448.

Braaten, B. A., Platko, J. V., van der Woude, M. W., Simons, B. H., de Graaf, F. K., Calvo, J. M. and Low, D. A. (1992) Leucine-responsive regulatory protein controls the

expression of both the *pap* and *fan* pili operons in *Escherichia coli*. *Proceedings of the National Academy of Science U S A* **89**: 4250-4254.

Brennan, R. G. and Link, T. M. (2007) Hfq structure, function and ligand binding. *Current Opinion in Microbiology* **10**: 125-133.

Brenner, S., Barnett, L., Katz, E. R. and Crick, F. H. C. (1967) Uga - a third nonsense triplet in genetic code. *Nature* **213**: 449-&.

Brinton, C. C., Jr. (1965) The structure, function, synthesis and genetic control of bacterial pili and a molecular model for DNA and RNA transport in gram negative bacteria. *Transactions of the New York Academy of Sciences* **27**: 1003-1054.

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

Brinton, C. C. J., Gemski, P., Falkow, S. and Baron, L. S. (1961) Location of the piliation factor on the chromosome of *Escherichia coli*. *Biochemical and Biophysical Research Communications* **5**: 293-.

Brown, P. O., Peebles, C. L. and Cozzarelli, N. R. (1979) A topoisomerase from *Escherichia coli* related to DNA gyrase. *Proceedings of the National Academy of Science USA* **76**: 6110-6114.

Bui, D., Ramiscal, J., Trigueros, S., Newmark, J. S., Do, A., Sherratt, D. J. and Tolmasky, M. E. (2006) Differences in resolution of mwr-containing plasmid dimers mediated by the *Klebsiella pneumoniae* and *Escherichia coli* XerC recombinases: potential implications in dissemination of antibiotic resistance genes. *Journal of Bacteriology* **188**: 2812-2820.

Burland, V., Plunkett, G., 3rd, Sofia, H. J., Daniels, D. L. and Blattner, F. R. (1995) Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. *Nucleic Acids Research* **23**: 2105-2119.

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

Bushman, W., Thompson, J. F., Vargas, L. and Landy, A. (1985) Control of directionality in lambda site specific recombination. *Science* **230**: 906-911.

Cabrera, J. E. and Jin, D. J. (2003) The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. *Molecular Microbiology* **50**: 1493-1505.

Cairns, J. (1963) The bacterial chromosome and its manner of replication as seen by autoradiography. *Journal of Molecular Biology* **6**: 208-213.

Calladine, C. R., Collis, C. M., Drew, H. R. and Mott, M. R. (1991) A study of electrophoretic mobility of DNA in agarose and polyacrylamide gels. *Journal of Molecular Biology* **221**: 981-1005.

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

Campbell, R., Tour, O., Palmer, A., Steinbach, P., Baird, G., Zacharias, D. and Tsien, R. (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences U. S. A.* **99**: 7877-7882.

Carey, J. (1991) Gel retardation. *Methods in Enzymology* **208**: 103-117.

Casadesus, J. and D'ari, R. (2002) Memory in bacteria and phage. *Bioessays* **24**: 512-518.

Castell, S. E. and Halford, S. E. (1989) DNA supercoiling determines the activation energy barrier for site specific recombination by Tn21 resolvase. *Nucleic Acids Research* **17**: 7045-7058.

Chalmers, R., Guhathakurta, A., Benjamin, H. and Kleckner, N. (1998) IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93**: 897-908.

Champoux, J. J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annual Reviews in Biochemistry* **70**: 369-413.

Chatterji, D. and Ojha, A. K. (2001) Revisiting the stringent response, ppGpp and starvation signaling. *Current Opinion in Microbiology* **4**: 160-165.

Chattopadhyay, S., Weissman, S. J., Minin, V. N., Russo, T. A., Dykhuizen, D. E. and Sokurenko, E. V. (2009) High frequency of hotspot mutations in core genes of *Escherichia coli* due to short-term positive selection. *Proceedings of the National Acadamy of Science U S A* **106**: 12412-12417.

Chaudhury, A. M. and Smith, G. R. (1984) A new class of *Escherichia coli recBC* mutants: implications for the role of RecBC enzyme in homologous recombination. *Proceedings of the National Acadamy of Science U S A* **81**: 7850-7854.

Chen, C., Chou, M., Huang, C., Majumder, A. and Wu, H. (2005a) A cis-spreading nucleoprotein filament is responsible for the gene silencing activity found in the promoter relay mechanism. *Journal of Biological Chemistry* **280**: 5101-5112.

Chen, C., Fang, M., Majumder, A. and Wu, H. (2001a) A 72-base pair AT-rich DNA sequence element functions as a bacterial gene silencer. *Journal of Biological Chemistry* **276**: 9478-9485.

Chen, S. and Calvo, J. M. (2002) Leucine-induced dissociation of Escherichia coli Lrp hexadecamers to octamers. *Journal of Molecular Biology* **318**: 1031-1042.

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

Chen, S., Iannolo, M. and Calvo, J. M. (2005b) Cooperative binding of the leucine-responsive regulatory protein (Lrp) to DNA. *Journal of Molecular Biology* **345**: 251-264.

Chen, S., Rosner, M. H. and Calvo, J. M. (2001c) Leucine-regulated self-association of leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. *Journal of Molecular Biology* **312**: 625-635.

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

Cho, B. K., Barrett, C. L., Knight, E. M., Park, Y. S. and Palsson, B. O. (2008) Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. *Proceedings of the National Academy of Science U S A* **105**: 19462-19467.

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

Chu, D. and Blomfield, I. C. (2007) Orientational control is an efficient control mechanism for phase switching in the *E. coli fim* system. *Journal of Theoretical Biology* **244**: 541-551.

Clark, L., Martinez-Argudo, I., Humphrey, T. J. and Jepson, M. A. (2009) GFP plasmid-induced defects in *Salmonella* invasion depend on plasmid architecture, not protein expression. *Microbiology* **155**: 461-467.

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

Connell, H., Agace, W., Klemm, P., Schembri, M., Marild, S. and Svanborg, C. (1996) Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. *Proceedings of the National Academy of Sciences USA* **93**: 9827-9832.

Corcoran, C. P., Cameron, A. D. and Dorman, C. J. (2010) H-NS silences *gfp*, the Green Fluorescent Protein gene: gfp^{TCD} is a genetically remastered *gfp* gene with reduced susceptibility to H-NS-mediated transcription silencing and with enhanced translation. *Journal of Bacteriology*.

Corcoran, C. P. and Dorman, C. J. (2009) DNA relaxation-dependent phase biasing of the *fim* genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP. *Molecular Microbiology* **74**: 1071-1082.

Costerton, J. W., Stewart, P. S. and Greenberg, E. P. (1999) Bacterial biofilms: a common cause of persistent infections. *Science* **284**: 1318-1322.

Cozzarelli, N. R. (1980) DNA gyrase and the supercoiling of DNA. *Science* **207**: 953-960.

Crellin, P., Sewitz, S. and Chalmers, R. (2004) DNA looping and catalysis; the IHF-folded arm of Tn10 promotes conformational changes and hairpin resolution. *Molecular Cell* **13**: 537-547.

Crick, F. H., Brenner, S., Watstobi.Rj and Barnett, L. (1961) General nature of genetic code for proteins. *Nature* **192**: 1227.

Crisona, N. J., Weinberg, R. L., Peter, B. J., Sumners, D. W. and Cozzarelli, N. R. (1999) The topological mechanism of phage lambda integrase. *Journal of Molecular Biology* **289**: 747-775.

Cui, Y., Midkiff, M. A., Wang, Q. and Calvo, J. M. (1996) The leucine-responsive regulatory protein (Lrp) from *Escherichia coli*. Stoichiometry and minimal requirements for binding to DNA. *Journal of Biological Chemistry* **271**: 6611-6617.

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

D'Ari, R., Lin, R. T. and Newman, E. B. (1993) The leucine-responsive regulatory protein: more than a regulator? *Trends in Biochemical Sciences* **18**: 260-263.

Dame, R. T., Noom, M. C. and Wuite, G. J. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature* **444**: 387-390.

Dame, R. T., Wyman, C. and Goosen, N. (2001) Structural basis for preferential binding of H-NS to curved DNA. *Biochimie* **83**: 231-234.

Danese, P. N., Pratt, L. A., Dove, S. L. and Kolter, R. (2000) The outer membrane protein, antigen 43, mediates cell-to-cell interactions within *Escherichia coli* biofilms. *Molecular Microbiology* **37**: 424-432.

Datsenko, K. A. and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences U. S. A.* **97**: 6640-6645.

de los Rios, S. and Perona, J. J. (2007) Structure of the *Escherichia coli* leucine-responsive regulatory protein Lrp reveals a novel octameric assembly. *Journal of Molecular Biology* **366**: 1589-1602.

De Vargas, L. M., Kim, S. and Landy, A. (1989a) DNA Looping Generated by DNA Bending Protein Ihf and the 2 Domains of Lambda Integrase. *Science* **244**: 1457-1461.

de Vargas, L. M., Kim, S. and Landy, A. (1989b) DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* **244**: 1457-1461.

de Vargas, L. M. and Landy, A. (1991) A switch in the formation of alternative DNA loops modulates lambda site-specific recombination. *Proceedings of the National Acadamy of Science U S A* **88**: 588-592.

Dersch, P., Kneip, S. and Bremer, E. (1994) The nucleoid-associated DNA-binding protein H-NS is required for the efficient adaptation of *Escherichia coli* K-12 to a cold environment. *Molecular Genomics and Genetics* **245**: 255-259.

Dersch, P., Schmidt, K. and Bremer, E. (1993) Synthesis of the *Escherichia coli* K-12 nucleoid-associated DNA-binding protein H-NS is subjected to growth-phase control and autoregulation. *Molecular Microbiology* **8**: 875-889.

Dhar, G., Heiss, J. K. and Johnson, R. C. (2009) Mechanical constraints on Hin subunit rotation imposed by the Fis/enhancer system and DNA supercoiling during site-specific recombination. *Molecular Cell* **34**: 746-759.

Dillon, S. C., Cameron, A. D. S., Hokamp, K., Lucchini, S., Hinton, J. C. and Dorman, C. J. (2010) Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Molecular Microbiology* **76**: 1250-1265.

Dillon, S. C. and Dorman, C. J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology* **8**: 185-195.

Ditto, M. D., Roberts, D. and Weisberg, R. A. (1994) Growth phase variation of integration host factor level in *Escherichia coli*. *Journal of Bacteriology* **176**: 3738-3748.

Dodson, K. W., Jacob-Dubuisson, F., Striker, R. T. and Hultgren, S. J. (1993) Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proceedings of the National Academy of Science U S A* **90**: 3670-3674.

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

Dorman, C., Bhriain, N. and Higgins, C. (1990) DNA supercoiling and environmental regulation of virulence gene expression in *Shigella flexneri*. *Nature* **344**: 789-792.

Dorman, C. J. (2004) H-NS: a universal regulator for a dynamic genome. *Nature Reviews Microbiology* **2**: 391-400.

Dorman, C. J. (2006) DNA supercoiling and bacterial gene expression. *Science progress* **89**: 151–166.

Dorman, C. J. (2007) H-NS, the genome sentinel. *Nature Reviews Microbiology* **5**: 157-161.

Dorman, C. J., Barr, G. C., Bhriain, N. N. and Higgins, C. F. (1988) DNA supercoiling and the anaerobic and growth phase regulation of *tonB* gene expression. *Journal of Bacteriology* **170**: 2816-2826.

Dorman, C. J. and Corcoran, C. P. (2009) Bacterial DNA topology and infectious disease. *Nucleic Acids Research* **37**: 672-678.

Dorman, C. J. and Higgins, C. F. (1987) Fimbrial phase variation in *Escherichia coli* - dependence on Integration Host Factor and homologies with other site-specific recombinases. *Journal of Bacteriology* **169**: 3840-3843.

Dorman, C. J. and Ní Bhriain, N. N. (1992) Thermal regulation of *fimA*, the *Escherichia coli* gene coding for the type 1 fimbrial subunit protein. *FEMS Microbiological Letters* **78**: 125-130.

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

Dove, S. L. and Dorman, C. J. (1996) Multicopy *fimB* gene expression in *Escherichia coli*: binding to inverted repeats *in vivo*, effect on *fimA* gene transcription and DNA inversion. *Molecular Microbiology* **21**: 1161–1173.

Doyle, M., Fookes, M., Ivens, A., Mangan, M. W., Wain, J. and Dorman, C. J. (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* **315**: 251-252.

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

Eisenstein, B. I., Clements, J. R. and Dodd, D. C. (1983) Isolation and characterization of a monoclonal antibody directed against type 1 fimbriae organelles from *Escherichia coli*. *Infection and Immunity* **42**: 333-340.

Eisenstein, B. I., Sweet, D. S., Vaughn, V. and Friedman, D. I. (1987) Integration host factor is required for the DNA inversion that controls phase variation in *Escherichia coli*. *Proceedings of the National Acadamy of Science U S A* **84**: 6506-6510.

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

Eshdat, Y., Speth, V. and Jann, K. (1981) Participation of pili and cell wall adhesion in the yeast agglutination activity of *Escherichia coli*. *Infection and Immunity* **34**: 980-986.

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

Fang, M. and Wu, H. Y. (1998a) A promoter relay mechanism for sequential gene activation. *Journal of Bacteriology* **180**: 626-633.

Fang, M. and Wu, H. Y. (1998b) Suppression of *leu-500* mutation in *topA+ Salmonella* Typhimurium strains. The promoter relay at work. *Journal of Biological Chemistry* **273**: 29929-29934.

Forsberg, A., Pavitt, G. and Higgins, C. (1994) Use of transcriptional fusions to monitor gene expression: a cautionary tale. *Journal of Bacteriology* **176**: 2128-2132.

Forties, R. A., Bundschuh, R. and Poirier, M. G. (2009) The flexibility of locally melted DNA. *Nucleic Acids Research* **37**: 4580-4586.

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

Fuller, F. B. (1971) The writhing number of a space curve. *Proceedings of the National Academy of Science U S A* **68**: 815-819.

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

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

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

Gellert, M., Mizuuchi, K., O'Dea, M. H. and Nash, H. A. (1976a) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Acadamy of Science U S A* **73**: 3872-3876.

Gellert, M., O'Dea, M. H., Itoh, T. and Tomizawa, J. (1976b) Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proceedings of the National Acadamy of Science U S A* **73**: 4474-4478.

Giladi, H., Gottesman, M. and Oppenheim, A. B. (1990) Integration host factor stimulates the phage lambda pL promoter. *Journal of Molecular Biology* **213**: 109-121.

Goldstein, E. and Drlica, K. (1984) Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. *Proceedings of the National Academy of Science U S A* **81**: 4046-4050.

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

Goodman, S. D., Nicholson, S. C. and Nash, H. A. (1992) Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. *Proceedings of the National Academy of Science U S A* **89**: 11910-11914.

Goodrich, J. A., Schwartz, M. L. and McClure, W. R. (1990) Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the

binding sites for *Escherichia coli* integration host factor (IHF). *Nucleic Acids Research* **18**: 4993-5000.

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

Gowrishankar, J. (1985) Identification of osmoresponsive genes in *Escherichia coli*: evidence for participation of potassium and proline transport systems in osmoregulation. *Journal of Bacteriology* **164**: 434-445.

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

Grindley, N. D., Whiteson, K. L. and Rice, P. A. (2006) Mechanisms of site-specific recombination. *Annual Reviews in Biochemistry* **75**: 567-605.

Guhathakurta, A., Viney, I. and Summers, D. (1996) Accessory proteins impose site selectivity during ColE1 dimer resolution. *Molecular Microbiology* **20**: 613-620.

Gunther, N. W., Lockatell, V., Johnson, D. E. and Mobley, H. L. T. (2001) *In vivo* dynamics of type 1 fimbria regulation in uropathogenic *Escherichia coli* during experimental urinary tract infection. *Infection and Immunity* **69**: 2838-2846.

Gunther, N. W., Snyder, J. A., Lockatell, V., Blomfield, I., Johnson, D. E. and Mobley, H. L. T. (2002) Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off. *Infection and Immunity* **70**: 3344-3354.

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

Guo, F., Gopaul, D. N. and Van Duyne, G. D. (1999) Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proceedings of the National Acadamy of Science U S A* **96**: 7143-7148.

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

Hakkila, K., Maksimow, M., Rosengren, A., Karp, M. and Virta, M. (2003) Monitoring promoter activity in a single bacterial cell by using green and red fluorescent proteins. *Journal of Microbiological Methods* **54**: 75-79.

Hales, L. M., Gumport, R. I. and Gardner, J. F. (1994) Determining the DNA sequence elements required for binding integration host factor to two different target sites. *Journal of Bacteriology* **176**: 2999-3006.

Hales, L. M., Gumport, R. I. and Gardner, J. F. (1996) Examining the contribution of a dA+dT element to the conformation of *Escherichia coli* integration host factor-DNA complexes. *Nucleic Acids Research* **24**: 1780-1786.

Hallet, B. and Sherratt, D. J. (1997) Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiological Reviews* **21**: 157-178.

Ham, T. S., Lee, S. K., Keasling, J. D. and Arkin, A. P. (2006) A tightly regulated inducible expression system utilizing the *fim* inversion recombination switch. *Biotechnology and Bioengineering* **94**: 1-4.

Ham, T. S., Lee, S. K., Keasling, J. D. and Arkin, A. P. (2008) Design and construction of a double inversion recombination switch for heritable sequential genetic memory. *PLoS One* **3**: e2815.

Hanson, M. S. and Brinton, C. C., Jr. (1988) Identification and characterization of *E. coli* type-1 pilus tip adhesion protein. *Nature* **332**: 265-268.

Hardy, C. D. and Cozzarelli, N. R. (2005) A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Molecular Microbiology* **57**: 1636-1652.

Hasman, H., Chakraborty, T. and Klemm, P. (1999) Antigen-43-mediated autoaggregation of *Escherichia coli* is blocked by fimbriation. *Journal of Bacteriology* **181**: 4834-4841.

Hasman, H., Schembri, M. A. and Klemm, P. (2000) Antigen 43 and type 1 fimbriae determine colony morphology of *Escherichia coli* K-12. *Journal of Bacteriology* **182**: 1089-1095.

Hautefort, I., Proenca, M. J. and Hinton, J. C. D. (2003) Single-copy green fluorescent protein gene fusions allow accurate measurement of *Salmonella* gene expression *in vitro* and during infection of mammalian cells. *Applied and Environmental Microbiology* **69**: 7480-7491.

Heddle, J. G., Barnard, F. M., Wentzell, L. M. and Maxwell, A. (2000) The interaction of drugs with DNA gyrase: a model for the molecular basis of quinolone action. *Nucleosides Nucleotides Nucleic Acids* **19**: 1249-1264.

Heim, R., Prasher, D. C. and Tsien, R. Y. (1994) Wavelength mutations and post-translational autoxidation of green fluorescent protein. *Proceedings of the National Academy of Sciences U. S. A.* **91**: 12501-12504.

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

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

Hiszczynska-Sawicka, E. and Kur, J. (1997) Effect of *Escherichia coli* IHF mutations on plasmid p15A copy number. *Plasmid* **38**: 174-179.

Hoekstra, W. P., Bergmans, J. E. and Zuidweg, E. M. (1980) Role of RecBC nuclease in *Escherichia coli* transformation. *Journal of Bacteriology* **143**: 1031-1032.

Holden, N., Blomfield, I. C., Uhlin, B. E., Totsika, M., Kulasekara, D. H. and Gally, D. L. (2007) Comparative analysis of FimB and FimE recombinase activity. *Microbiology* **153**: 4138-4149.

Holden, N. J. and Gally, D. L. (2004) Switches, cross-talk and memory in *Escherichia coli* adherence. *Journal of Medical Microbiology* **53**: 585-593.

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

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

Hsieh, L. S., Burger, R. M. and Drlica, K. (1991a) Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *Journal of Molecular Biology* **219**: 443-450.

Hsieh, L. S., Rouviere-Yaniv, J. and Drlica, K. (1991b) Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. *Journal of Bacteriology* **173**: 3914-3917.

Hultgren, S., Schwan, W., Schaeffer, A. and Duncan, J. (1986) Regulation of production of type 1 pili among urinary tract isolates of *Escherichia coli*. *Infection and Immunity* **54**: 613-620.

Hung, S. P., Baldi, P. and Hatfield, G. W. (2002) Global gene expression profiling in *Escherichia coli* K12. The effects of leucine-responsive regulatory protein. *Journal of Biological Chemistry* **277**: 40309-40323.

Iwahi, T., Abe, Y., Nakao, M., Imada, A. and Tsuchiya, K. (1983) Role of type 1 fimbriae in the pathogenesis of ascending urinary tract infection induced by *Escherichia coli* in mice. *Infection and Immunity* **39**: 1307-1315.

Jeong, K. S., Xie, Y., Hiasa, H. and Khodursky, A. B. (2006) Analysis of pleiotropic transcriptional profiles: a case study of DNA gyrase inhibition. *PLoS Genetics* **2**: e152.

Jones, C. H., Pinkner, J. S., Roth, R., Heuser, J., Nicholes, A. V., Abraham, S. N. and Hultgren, S. J. (1995) FimH adhesin of type 1 pili is assembled into a fibrillar tip structure in the *Enterobacteriaceae*. *Proceedings of the National Academy of Science U S A* **92**: 2081-2085.

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

Kaper, J. B., Nataro, J. P. and Mobley, H. L. (2004) Pathogenic *Escherichia coli*. *Nature Reviews in Microbiology* **2**: 123-140.

Kawula, T. H. and Orndorff, P. E. (1991) Rapid site-specific DNA inversion in *Escherichia coli* mutants lacking the histonelike protein H-NS. *Journal of Bacteriology* **173**: 4116-4123.

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

Kim, S. and Landy, A. (1992) Lambda-Int protein bridges between higher-order complexes at 2 distant chromosomal loci Attl and Attr. *Science* **256**: 198-203.

Kjaergaard, K., Schembri, M. A., Ramos, C., Molin, S. and Klemm, P. (2000) Antigen 43 facilitates formation of multispecies biofilms. *Environmental microbiology* **2**: 695-702.

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

Klemm, P. and Christiansen, G. (1990) The *fimD* gene required for cell surface localization of *Escherichia coli* type 1 fimbriae. *Molecular Genomics and Genetics* **220**: 334-338.

Knight, R. D., Freeland, S. J. and Landweber, L. F. (2001) A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biology* **2**: RESEARCH0010.

Korhonen, T. K., Virkola, R. and Holthofer, H. (1986) Localization of binding sites for purified *Escherichia coli* P fimbriae in the human kidney. *Infection and Immunity* **54**: 328-332.

Kowalczykowski, S. C., Dixon, D. A., Eggleston, A. K., Lauder, S. D. and Rehrauer, W. M. (1994) Biochemistry of homologous recombination in *Escherichia coli*. *Microbiological Reviews* **58**: 401-465.

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

Krogfelt, K. A., Bergmans, H. and Klemm, P. (1990) Direct evidence that the FimH protein is the mannose-specific adhesin of *Escherichia coli* type 1 fimbriae. *Infection and Immunity* **58**: 1995-1998.

Kudla, G., Murray, A. W., Tollervey, D. and Plotkin, J. B. (2009) Coding-sequence determinants of gene expression in *Escherichia coli*. *Science* **324**: 255-258.

Kulasekara, H. D. and Blomfield, I. C. (1999) The molecular basis for the specificity of FimE in the phase variation of type 1 fimbriae of *Escherichia coli* K-12. *Molecular Microbiology* **31**: 1171-1181.

Kurland, C. and Gallant, J. (1996) Errors of heterologous protein expression. *Current Opinion in Biotechnology* **7**: 489-493.

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

Landgraf, J. R., Wu, J. and Calvo, J. M. (1996) Effects of nutrition and growth rate on Lrp levels in *Escherichia coli*. *Journal of Bacteriology* **178**: 6930-6936.

Landy, A. (1989) Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annual Reviews in Biochemistry* **58**: 913-949.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., Mavathur, R., Muskhelishvili, G., Pon, C. L., Rimsky, S., Stella, S., Babu, M. M. and Travers, A. (2007a) High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Research* **35**: 6330-6337.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., Mavathur, R., Muskhelishvili, G., Pon, C. L., Rimsky, S., Stella, S., Babu, M. M. and Travers, A. (2007b) High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Research* **35**: 6330-6337.

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

Lee, S. K. and Keasling, J. D. (2005) A propionate-inducible expression system for enteric bacteria. *Applied and Environmental Microbiology* **71**: 6856-6862.

Lewis, J. A. and Hatfull, G. F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Research* **29**: 2205-2216.

Lewis, J. A. and Hatfull, G. F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *Journal of Molecular Biology* **326**: 805-821.

Liu, L. F. and Wang, J. C. (1987) Supercoiling of the DNA template during transcription. . *Proceedings of the National Academy of Sciences U. S. A.* **83**: 7024-7027.

Liu, Y., Chen, H., Kenney, L. J. and Yan, J. (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes and Development* **24**: 339-344.

Lloyd, R. G. and Buckman, C. (1985) Identification and genetic analysis of *sbcC* mutations in commonly used *recBC sbcB* strains of *Escherichia coli* K-12. *Journal of Bacteriology* **164**: 836-844.

Low, A. S., Holden, N., Rosser, T., Roe, A. J., Constantinidou, C., Hobman, J. L., Smith, D. G., Low, J. C. and Gally, D. L. (2006) Analysis of fimbrial gene clusters and their expression in enterohaemorrhagic *Escherichia coli* O157:H7. *Environmental Microbiology* **8**: 1033-1047.

Lucchini, S., Rowley, G., Goldberg, M. D., Hurd, D., Harrison, M. and Hinton, J. C. (2006) H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathogens* **2**: e81.

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

Magnusson, L. U., Farewell, A. and Nystrom, T. (2005) ppGpp: a global regulator in *Escherichia coli*. *Trends in Microbiology* **13**: 236-242.

Magnusson, L. U., Gummesson, B., Joksimovic, P., Farewell, A. and Nystrom, T. (2007) Identical, independent, and opposing roles of ppGpp and DksA in *Escherichia coli*. *Journal of Bacteriology* **189**: 5193-5202.

Maksimow, M., Hakkila, K., Karp, M. and Virta, M. (2002) Simultaneous detection of bacteria expressing GFP and DsRed genes with a flow cytometer. *Cytometry* **47**: 243-247.

Mangan, M., Lucchini, S., Danino, V., Cróinín, T., Hinton, J. and Dorman, C. (2006) The integration host factor (IHF) integrates stationary-phase and virulence gene expression in *Salmonella enterica* serovar Typhimurium. *Molecular Microbiology* **59**: 1831-1847.

Matz, M., Fradkov, A., Labas, Y., Savitsky, A., Zaraisky, A., Markelov, M. and Lukyanov, S. (1999) Fluorescent proteins from nonbioluminescent Anthozoa species. *Nature Biotechnology* **17**: 969-973.

Maurer, S., Fritz, J. and Muskhelishvili, G. (2009) A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. *Journal of Molecular Biology* **387**: 1261-1276.

May, G., Dersch, P., Haardt, M., Middendorf, A. and Bremer, E. (1990) The *osmZ* (*bglY*) gene encodes the DNA-binding protein H-NS (H1a), a component of the *Escherichia coli* K12 nucleoid. *Molecular Genomics and Genetics* **224**: 81-90.

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

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

McCusker, M. P., Turner, E. C. and Dorman, C. J. (2008) DNA sequence heterogeneity in Fim tyrosine-integrase recombinase-binding elements and functional motif asymmetries determine the directionality of the *fim* genetic switch in *Escherichia coli* K-12. *Molecular Microbiology* **67**: 171-187.

Menzel, R. and Gellert, M. (1983) Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell* **34**: 105-113.

Mikheikin, A. L., Lushnikov, A. Y. and Lyubchenko, Y. L. (2006) Effect of DNA supercoiling on the geometry of Holliday junctions. *Biochemistry* **45**: 12998-13006.

Miller, H. I. and Nash, H. A. (1981) Direct role of the *himA* gene product in phage lambda integration. *Nature* **290**: 523-526.

Miller, J. H. (1992) *A short course in bacterial genetics*. Cold Spring Harbor Laboratory Press, Plainview, N.Y.

Moskowitz, I. P., Heichman, K. A. and Johnson, R. C. (1991) Alignment of recombination sites in Hin-mediated site-specific DNA recombination. *Genes Dev* **5**: 1635-1645.

Müller, C. M., Aberg, A., Straseviciene, J., Emody, L., Uhlin, B. E. and Balsalobre, C. (2009) Type 1 fimbriae, a colonization factor of uropathogenic *Escherichia coli*, are controlled by the metabolic sensor CRP-cAMP. *PLoS Pathogens* **5**: e1000303.

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

Murphy, K. C. (1998) Use of bacteriophage lambda recombination functions to promote gene replacement in *Escherichia coli*. *Journal of Bacteriology* **180**: 2063-2071.

Murtin, C., Engelhorn, M., Geiselmann, J. and Boccard, F. (1998) A quantitative UV laser footprinting analysis of the interaction of IHF with specific binding sites: re-evaluation of the effective concentration of IHF in the cell. *Journal of Molecular Biology* **284**: 949-961.

Nagarajavel, V., Madhusudan, S., Dole, S., Rahmouni, A. R. and Schnetz, K. (2007) Repression by binding of H-NS within the transcription unit. *Journal of Biological Chemistry* **282**: 23622-23630.

Nash, H. A. (1981) Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. *Annual Reviews in Genetics* **15**: 143-167.

Nash, H. A., Robertson, C. A., Flamm, E., Weisberg, R. A. and Miller, H. I. (1987) Overproduction of *Escherichia coli* integration host factor, a protein with nonidentical subunits. *Journal of Bacteriology* **169**: 4124-4127.

Navarre, W. W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S. J. and Fang, F. C. (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. *Science* **313**: 236-238.

Newman, J., Ghaemmaghami, S., Ihmels, J., Breslow, D., Noble, M., DeRisi, J. and Weissman, J. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441**: 840-846.

Noom, M. C., Navarre, W. W., Oshima, T., Wuite, G. J. and Dame, R. T. (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Current Biology* **17**: R913-914.

Nunes-Duby, S. E., Azaro, M. A. and Landy, A. (1995) Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Current Biology* **5**: 139-148.

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

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

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

Oppenheim, A. B., Rudd, K. E., Mendelson, I. and Teff, D. (1993) Integration host factor binds to a unique class of complex repetitive extragenic DNA sequences in *Escherichia coli*. *Molecular Microbiology* **10**: 113-122.

Oshima, T., Ishikawa, S., Kurokawa, K., Aiba, H. and Ogasawara, N. (2006) *Escherichia coli* histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Research* **13**: 141-153.

Owen-Hughes, T., Pavitt, G., Santos, D., Sidebotham, J., Hulton, C., Hinton, J. and Higgins, C. (1992) The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. *Cell* **71**: 255-265.

Pallesen, L., Madsen, O. and Klemm, P. (1989) Regulation of the phase switch controlling expression of type 1 fimbriae in *Escherichia coli*. *Molecular Microbiology* **3**: 925-931.

Pan, C. Q., Feng, J. A., Finkel, S. E., Landgraf, R., Sigman, D. and Johnson, R. C. (1994) Structure of the Escherichia coli Fis-DNA complex probed by protein conjugated with 1,10-phenanthroline copper(I) complex. *Proceedings of the National Acadamy of Science U S A* **91**: 1721-1725.

Pan, C. Q., Finkel, S. E., Cramton, S. E., Feng, J. A., Sigman, D. S. and Johnson, R. C. (1996) Variable structures of Fis-DNA complexes determined by flanking DNA-protein contacts. *Journal of Molecular Biology* **264**: 675-695.

Papagiannis, C. V., Sam, M. D., Abbani, M. A., Yoo, D., Cascio, D., Clubb, R. T. and Johnson, R. C. (2007) Fis targets assembly of the Xis nucleoprotein filament to promote excisive recombination by phage lambda. *Journal of Molecular Biology* **367**: 328-343.

Parekh, B. S. and Hatfield, G. W. (1996) Transcriptional activation by protein-induced DNA bending: evidence for a DNA structural transmission model. *Proceedings of the National Acadamy of Science U S A* **93**: 1173-1177.

Parekh, B. S., Sheridan, S. D. and Hatfield, G. W. (1996) Effects of integration host factor and DNA supercoiling on transcription from the *ilvPG* promoter of *Escherichia coli*. *Journal of Biological Chemistry* **271**: 20258-20264.

Paul, B. J., Barker, M. M., Ross, W., Schneider, D. A., Webb, C., Foster, J. W. and Gourse, R. L. (2004) DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* **118**: 311-322.

Paulsson, J. (2004) Summing up the noise in gene networks. *Nature* **427**: 415-418.

Perederina, A., Svetlov, V., Vassylyeva, M. N., Tahirov, T. H., Yokoyama, S., Artsimovitch, I. and Vassylyev, D. G. (2004) Regulation through the secondary

channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* **118**: 297-309.

Peterson, S. N., Dahlquist, F. W. and Reich, N. O. (2007) The role of high affinity non-specific DNA binding by Lrp in transcriptional regulation and DNA organization. *Journal of Molecular Biology* **369**: 1307-1317.

Plasterk, R. H., Kanaar, R. and van de Putte, P. (1984) A genetic switch *in vitro*: DNA inversion by Gin protein of phage Mu. *Proceedings of the National Acadamy of Science U S A* **81**: 2689-2692.

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

Ponniah, S., Endres, R. O., Hasty, D. L. and Abraham, S. N. (1991) Fragmentation of *Escherichia coli* type 1 fimbriae exposes cryptic D-mannose-binding sites. *Journal of Bacteriology* **173**: 4195-4202.

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

Prosseda, G., Falconi, M., Giangrossi, M., Gualerzi, C. O., Micheli, G. and Colonna, B. (2004) The *virF* promoter in *Shigella*: more than just a curved DNA stretch. *Molecular Microbiology* **51**: 523-537.

Rajeev, L., Malanowska, K. and Gardner, J. F. (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiology and Molecular Biology Reviews* **73**: 300-309.

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

Richet, E., Abcarian, P. and Nash, H. A. (1986) The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* **46**: 1011-1021.

Richet, E., Abcarian, P. and Nash, H. A. (1988) Synapsis of attachment sites during lambda integrative recombination involves capture of a naked DNA by a protein-DNA complex. *Cell* **52**: 9-17.

Rimsky, S. (2004) Structure of the histone-like protein H-NS and its role in regulation and genome superstructure. *Current Opinion in Microbiology* **7**: 109-114.

Rizzo, M. A., Davidson, M. W. and Piston, D. W. (2009) Fluorescent protein tracking and detection: fluorescent protein structure and color variants. *Cold Spring Harbor Protocols*.

Robertson, C. A. and Nash, H. A. (1988) Bending of the bacteriophage lambda attachment site by *Escherichia coli* integration host factor. *Journal of Biological Chemistry* **263**: 3554-3557.

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

Russell, C. B., Thaler, D. S. and Dahlquist, F. W. (1989) Chromosomal transformation of *Escherichia coli recD* strains with linearized plasmids. *Journal of Bacteriology* **171**: 2609-2613.

Russell, P. W. and Orndorff, P. E. (1992) Lesions in two *Escherichia coli* type 1 pilus genes alter pilus number and length without affecting receptor binding. *Journal of Bacteriology* **174**: 5923-5935.

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

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-491.

Salit, I. E. and Gotschlich, E. C. (1977) Hemagglutination by purified type I *Escherichia coli* pili. *Journal of Experimental Medicine* **146**: 1169-1181.

Sambrook, J. and Russell, D. W. (2001) *Molecular cloning : a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Samul, R. and Leng, F. (2007) Transcription-coupled hypernegative supercoiling of plasmid DNA by T7 RNA polymerase in *Escherichia coli* topoisomerase I-deficient strains. *Journal of Molecular Biology* **374**: 925-935.

Sanders, E. R. and Johnson, R. C. (2004) Stepwise dissection of the Hin-catalyzed recombination reaction from synapsis to resolution. *Journal of Molecular Biology* **340**: 753-766.

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

Schembri, M. A., Christiansen, G. and Klemm, P. (2001) FimH-mediated autoaggregation of *Escherichia coli*. *Molecular Microbiology* **41**: 1419-1430.

Schembri, M. A. and Klemm, P. (2001a) Biofilm formation in a hydrodynamic environment by novel *fimH* variants and ramifications for virulence. *Infection and Immunity* **69**: 1322-1328.

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

Schembri, M. A., Olsen, P. B. and Klemm, P. (1998) Orientation-dependent enhancement by H-NS of the activity of the type 1 fimbrial phase switch promoter in *Escherichia coli*. *Molecular Genomics and Genetics* **259**: 336-344.

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

Scholz, O., Thiel, A., Hillen, W. and Niederweis, M. (2000) Quantitative analysis of gene expression with an improved green fluorescent protein. p6. *European Journal of Biochemistry* **267**: 1565-1570.

Schwan, W., Seifert, H. and Duncan, J. (1992) Growth conditions mediate differential transcription of *fim* genes involved in phase variation of type 1 pili. *Journal of Bacteriology* **174**: 2367-2375.

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

Segall, A. M. and Nash, H. A. (1993) Synaptic intermediates in bacteriophage lambda site-specific recombination: integrase can align pairs of attachment sites. *EMBO Journal* **12**: 4567-4576.

Shaner, N., Patterson, G. and Davidson, M. (2007) Advances in fluorescent protein technology. *Journal of Cell Sciences* **120**: 4247-4260.

Sheridan, S. D., Benham, C. J. and Hatfield, G. W. (1998) Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced

DNA duplex destabilization in an upstream activating sequence. *Journal of Biological Chemistry* **273**: 21298-21308.

Shindo, H., Iwaki, T., Ieda, R., Kurumizaka, H., Ueguchi, C., Mizuno, T., Morikawa, S., Nakamura, H. and Kuboniwa, H. (1995) Solution structure of the DNA binding domain of a nucleoid-associated protein, H-NS, from *Escherichia coli*. *FEBS Letters* **360**: 125-131.

Simon, M., Zieg, J., Silverman, M., Mandel, G. and Doolittle, R. (1980) Phase variation: evolution of a controlling element. *Science* **209**: 1370-1374.

Singh, R. K., Liburd, J., Wardle, S. J. and Haniford, D. B. (2008) The nucleoid binding protein H-NS acts as an anti-channeling factor to favor intermolecular Tn10 transposition and dissemination. *Journal of Molecular Biology* **376**: 950-962.

Smith, G. R. (1988) Homologous recombination in procaryotes. *Microbiological Reviews* **52**: 1-28.

Smith, H. O., Danner, D. B. and Deich, R. A. (1981) Genetic transformation. *Annual Reviews in Biochemistry* **50**: 41-68.

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

Sohanpal, B. K., El-Labany, S., Lahooti, M., Plumbridge, J. A. and Blomfield, I. C. (2004) Integrated regulatory responses of *fimB* to N-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12. *Proceedings of the National Acadamy of Science U S A* **101**: 16322-16327.

Sohanpal, B. K., Friar, S., Roobol, J., Plumbridge, J. A. and Blomfield, I. C. (2007) Multiple co-regulatory elements and IHF are necessary for the control of *fimB* expression in response to sialic acid and N-acetylglucosamine in *Escherichia coli* K-12. *Molecular Microbiology* **63**: 1223-1236.

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

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, C., Schembri, M. A. and Hasty, D. L. (1998) Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proceedings of the National Acadamy of Science U S A* **95**: 8922-8926.

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

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

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

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

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

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

Stark, W. M., Boocock, M. R. and Sherratt, D. J. (1992) Catalysis by site-specific recombinases. *Trends in Genetics* **8**: 432-439.

Stella, S., Spurio, R., Falconi, M., Pon, C. L. and Gualerzi, C. O. (2005) Nature and mechanism of the *in vivo* oligomerization of nucleoid protein H-NS. *EMBO Journal* **24**: 2896-2905.

Sternberg, N. and Hoess, R. (1983) The molecular genetics of bacteriophage P1. *Annual Reviews in Genetics* **17**: 123-154.

Stoebel, D. M., Free, A. and Dorman, C. J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154**: 2533-2545.

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

Summers, D. K. and Sherratt, D. J. (1988) Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *EMBO Journal* **7**: 851-858.

Surette, M. G. and Chaconas, G. (1989) A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is *Escherichia coli* integration host factor. *Journal of Biological Chemistry* **264**: 3028-3034.

Surette, M. G., Lavoie, B. D. and Chaconas, G. (1989) Action at a distance in Mu DNA transposition: an enhancer-like element is the site of action of supercoiling relief activity by integration host factor (IHF). *EMBO Journal* **8**: 3483-3489.

Swaney, L. M., Liu, Y. P., Ippen-Ihler, K. and Brinton, C. C., Jr. (1977) Genetic complementation analysis of *Escherichia coli* type 1 somatic pilus mutants. *Journal of Bacteriology* **130**: 506-511.

Tani, T. H., Khodursky, A., Blumenthal, R. M., Brown, P. O. and Matthews, R. G. (2002) Adaptation to famine: a family of stationary-phase genes revealed by microarray analysis. *Proceedings of the National Acadamy of Science U S A* **99**: 13471-13476.

Tchesnokova, V., Aprikian, P., Yakovenko, O., Larock, C., Kidd, B., Vogel, V., Thomas, W. and Sokurenko, E. (2008) Integrin-like allosteric properties of the catch bond-forming FimH adhesin of *Escherichia coli*. *Journal of Biological Chemistry* **283**: 7823-7833.

Thanassi, D. G., Saulino, E. T., Lombardo, M. J., Roth, R., Heuser, J. and Hultgren, S. J. (1998) The PapC usher forms an oligomeric channel: implications for pilus biogenesis across the outer membrane. *Proceedings of the National Academy of Science U S A* **95**: 3146-3151.

Thomas, W. E., Trintchina, E., Forero, M., Vogel, V. and Sokurenko, E. V. (2002) Bacterial adhesion to target cells enhanced by shear force. *Cell* **109**: 913-923.

Thomason, L. C., Costantino, N. and Court, D. L. (2007) *E. coli* Genome Manipulation by P1 Transduction. In: Current Protocols in Molecular Biology. pp.

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

Thompson, J. F., Moitoso de Vargas, L., Koch, C., Kahmann, R. and Landy, A. (1987) Cellular factors couple recombination with growth phase: characterization of a new component in the lambda site-specific recombination pathway. *Cell* **50**: 901-908.

Thompson, J. F., Waechter-Brulla, D., Gumport, R. I., Gardner, J. F., Moitoso de Vargas, L. and Landy, A. (1986) Mutations in an integration host factor-binding site: effect on lambda site-specific recombination and regulatory implications. *Journal of Bacteriology* **168**: 1343-1351.

Tolmasky, M. E., Colloms, S., Blakely, G. and Sherratt, D. J. (2000) Stability by multimer resolution of pJHCMW1 is due to the Tn1331 resolvase and not to the *Escherichia coli* Xer system. *Microbiology* **146 (Pt 3)**: 581-589.

Traxler, M. F., Summers, S. M., Nguyen, H. T., Zacharia, V. M., Hightower, G. A., Smith, J. T. and Conway, T. (2008) The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Molecular Microbiology* **68**: 1128-1148.

Trigueros, S., Tran, T., Sorto, N., Newmark, J., Colloms, S. D., Sherratt, D. J. and Tolmasky, M. E. (2009) mwr Xer site-specific recombination is hypersensitive to DNA supercoiling. *Nucleic Acids Research* **37**: 3580-3587.

Tse-Dinh, Y. C. (1985) Regulation of the *Escherichia coli* DNA topoisomerase I gene by DNA supercoiling. *Nucleic Acids Research* **13**: 4751-4763.

Tupper, A., Owen-Hughes, T., Ussery, D., Santos, D., Ferguson, D., Sidebotham, J., Hinton, J. and Higgins, C. (1994) The chromatin-associated protein H-NS alters DNA topology *in vitro*. *EMBO Journal* **13**: 258-268.

Ueguchi, C. and Mizuno, T. (1993) The *Escherichia coli* nucleoid protein H-NS functions directly as a transcriptional repressor. *EMBO Journal* **12**: 1039-1046.

Ulett, G. C., Valle, J., Beloin, C., Sherlock, O., Ghigo, J. M. and Schembri, M. A. (2007) Functional analysis of antigen 43 in uropathogenic *Escherichia coli* reveals a role in long-term persistence in the urinary tract. *Infection and Immunity* **75**: 3233-3244.

Valdivia, R. H. and Falkow, S. (1997) Fluorescence-based isolation of bacterial genes expressed within host cells. *Science* **277**: 2007-2011.

Van Duyne, G. D. (2001) A structural view of *cre-loxP* site-specific recombination. *Annual Reviews in Biophysics and Biomolecular Structure* **30**: 87-104.

Van Duyne, G. D. (2002) A structural view of tyrosine recombinase site-specific recombination. In: Mobile DNA II. N. L. Craig, R. Craigie, M. Gellert & A. M. Lambowitz (eds). Washington, D.C.: ASM Press, pp. 93-117.

Van Duyne, G. D. (2005) Lambda integrase: armed for recombination. *Current Biology* **15**: R658-660.

van Workum, M., van Dooren, S. J., Oldenburg, N., Molenaar, D., Jensen, P. R., Snoep, J. L. and Westerhoff, H. V. (1996) DNA supercoiling depends on the phosphorylation potential in *Escherichia coli*. *Molecular Microbiology* **20**: 351-360.

Verkhusha, V. V. and Sorkin, A. (2005) Conversion of the monomeric red fluorescent protein into a photoactivatable probe. *Chemical Biology* **12**: 279-285.

Villalobos, A., Ness, J., Gustafsson, C., Minshull, J. and Govindarajan, S. (2006) Gene Designer: a synthetic biology tool for constructing artificial DNA segments. *BMC Bioinformatics* **7**: 285.

Vologodskii, A. and Cozzarelli, N. R. (1996) Effect of supercoiling on the juxtaposition and relative orientation of DNA sites. *Biophysical Journal* **70**: 2548-2556.

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

Wallecha, A., Munster, V., Correnti, J., Chan, T. and van der Woude, M. (2002) Dam- and OxyR-dependent phase variation of agn43: essential elements and evidence for a new role of DNA methylation. *Journal of Bacteriology* **184**: 3338-3347.

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

Wang, Q., Sacco, M., Ricca, E., Lago, C. T., De Felice, M. and Calvo, J. M. (1993) Organization of Lrp-binding sites upstream of *ilvIH* in *Salmonella* typhimurium. *Molecular Microbiology* **7**: 883-891.

Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M. and Vander Horn, P. B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Research* **32**: 1197-1207.

Weglenska, A., Jacob, B. and Sirko, A. (1996) Transcriptional pattern of *Escherichia coli ihfB* (*himD*) gene expression. *Gene* **181**: 85-88.

Weisberg, R. A., Freundlich, M., Friedman, D. I., Gardner, J. F., Goosen, N., Nash, H. A., Oppenheim, A. and Rouvière-Yaniv, J. (1996) Nomenclature of the genes encoding IHF. *Molecular Microbiology* **19**: 642.

Weissman, S. J., Beskhlebnaya, V., Chesnokova, V., Chattopadhyay, S., Stamm, W. E., Hooton, T. M. and Sokurenko, E. V. (2007) Differential stability and trade-off

effects of pathoadaptive mutations in the *Escherichia coli* FimH adhesin. *Infection and Immunity* **75**: 3548-3555.

Wendrich, T. M., Blaha, G., Wilson, D. N., Marahiel, M. A. and Nierhaus, K. H. (2002) Dissection of the mechanism for the stringent factor RelA. *Molecular Cell* **10**: 779-788.

Werner, M. H., Clore, G. M., Gronenborn, A. M. and Nash, H. A. (1994) Symmetry and asymmetry in the function of *Escherichia coli* integration host factor: implications for target identification by DNA-binding proteins. *Current Biology* **4**: 477-487.

Wiggins, P. A., Dame, R. T., Noom, M. C. and Wuite, G. J. L. (2009) Protein-mediated molecular bridging: A key mechanism in biopolymer organization. *Biophysical Journal* **97**: 1997-2003.

Willetts, N. S. and Mount, D. W. (1969) Genetic analysis of recombination-deficient mutants of *Escherichia coli* K-12 carrying *rec* mutations cotransducible with *thyA*. *Journal of Bacteriology* **100**: 923-934.

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

Wolf, D. M. and Arkin, A. P. (2002) Fifteen minutes of *fim*: control of type 1 pili expression in *E. coli*. *OMICS* **6**: 91-114.

Wolffe, A. P. and Drew, H. R. (1989) Initiation of transcription on nucleosomal templates. *Proceedings of the National Academy of Sciences U. S. A.* **86**: 9817-9821.

Wright, K. J., Seed, P. C. and Hultgren, S. J. (2007) Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol* **9**: 2230-2241.

Wu, H. and Crothers, D. (1984) The locus of sequence-directed and protein-induced DNA bending. *Nature* **308**: 509-513.

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

Xie, Y., Yao, Y. F., Kolisnychenko, V., Teng, C. H. and Kim, K. S. (2006) HbiF regulates type 1 fimbriation independently of FimB and FimE. *Infection and Immunity* **74**: 4039-4047.

Yakovenko, O., Sharma, S., Forero, M., Tchesnokova, V., Aprikian, P., Kidd, B., Mach, A., Vogel, V., Sokurenko, E. and Thomas, W. E. (2008) FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *Journal of Biological Chemistry* **283**: 11596-11605.

Yamada, H., Muramatsu, S. and Mizuno, T. (1990) An *Escherichia coli* protein that preferentially binds to sharply curved DNA. *Journal of Biochemistry* **108**: 420-425.

Yamamoto, N. and Droffner, M. L. (1985) Mechanisms determining aerobic or anaerobic growth in the facultative anaerobe *Salmonella* typhimurium. *Proceedings of the National Academy of Science U S A* **82**: 2077-2081.

Yamamoto, T., Fujita, K. and Yokota, T. (1990) Adherence characteristics to human small intestinal mucosa of *Escherichia coli* isolated from patients with diarrhea or urinary tract infections. *Journal of Infectious Diseases* **162**: 896-908.

Yang, C. C. and Nash, H. A. (1989) The interaction of *E. coli* IHF protein with its specific binding sites. *Cell* **57**: 869-880.

Yang, S. W. and Nash, H. A. (1995a) Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. *EMBO Journal* **14**: 6292-6300.

Yang, S. W. and Nash, H. A. (1995b) Comparison of protein binding to DNA in vivo and in vitro: defining an effective intracellular target. . *EMBO J.* **14**: 6292-6300.

Zablewska, B. and Kur, J. (1995) Mutations in HU and IHF affect bacteriophage T4 growth: HimD subunits of IHF appear to function as homodimers. *Gene* **160**: 131-132.

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G. and Alon, U. (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods* **3**: 623-628.

Zinchenko, A. A., Sergeyev, V. G., Yamabe, K., Murata, S. and Yoshikawa, K. (2004) DNA compaction by divalent cations: Structural specificity revealed by the potentiality of designed quaternary diammonium salts. *ChemBioChem* **5**: 360-368.

Zinser, E. R. and Kolter, R. (2000) Prolonged stationary-phase incubation selects for *lrp* mutations in *Escherichia coli* K-12. *Journal of Bacteriology* **182**: 4361-4365.

Zulianello, L., de la Gorgue de Rosny, E., van Ulsen, P., van de Putte, P. and Goosen, N. (1994) The HimA and HimD subunits of integration host factor can specifically bind to DNA as homodimers. *Embo Journal* **13**: 1534-1540.

Appendix 1: Corcoran C.P. and Dorman C.J. (2009). DNA relaxation dependent phase biasing of the *fim* genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP. *Molecular Microbiology* 74: 1071-1082.

Molecular Microbiology (2009) 74(5), 1071–1082 ■

doi:10.1111/j.1365-2958.2009.06919.x
First published online 13 November 2009

DNA relaxation-dependent phase biasing of the *fim* genetic switch in *Escherichia coli* depends on the interplay of H-NS, IHF and LRP

Colin P. Corcoran and Charles J. Dorman*
Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin 2, Ireland.

Summary

Reversible inversion of the DNA element *fimS* is responsible for the phase variable expression of type 1 fimbriae in *Escherichia coli*. The FimB tyrosine integrase site-specific recombinase inverts *fimS* in the on-to-off and off-to-on directions with approximately equal efficiencies. However, when DNA supercoiling is relaxed, *fimS* adopts predominantly the on orientation. This orientational bias is known to require binding of the nucleoid-associated protein LRP within *fimS*. Here we show that binding of the IHF protein to a site immediately adjacent to *fimS* is also required for phase-on orientational bias. In the absence of both LRP and IHF binding, *fimS* adopts the off orientation and the H-NS protein is required to maintain this alternative orientational bias. Thus, *fimS* has three Recombination Directionality Factors, H-NS, IHF and LRP. The relevant H-NS binding site straddles the left inverted repeat in phase-off *fimS* and this site is disrupted when *fimS* inverts to the on orientation. The inversion of *fimS* with the associated creation and removal of an H-NS binding site required for DNA inversion biasing represents a novel mechanism for modulating the interaction of H-NS with a DNA target and for influencing a site-specific recombination reaction.

Introduction

Type 1 fimbriae contribute to virulence in pathogenic Gram-negative bacteria and are associated with the establishment of an attached, as opposed to a planktonic, lifestyle (Wright *et al.*, 2007). Fimbriae also contribute to biofilm formation. Taken together, fimbriae can be seen as aiding niche colonization through attachment, most likely under conditions when the bacterium can no longer support a motile lifestyle that demands a high metabolic flux (Müller *et al.*, 2009). Expression of type 1 fimbrial genes (*fim*) is phase-variable in many bacteria although the mechanism by which this is achieved can vary between species (Abraham *et al.*, 1985; McFarland *et al.*, 2008). This has been rationalized as a mechanism for immune avoidance during infection. An important question concerns the degree to which phase variation is programmed and the degree to which it is stochastic. Knowledge of the underlying molecular mechanisms should help to answer this question.

In *Escherichia coli* the promoter for transcription of the *fim* structural genes is carried within a segment of invertible chromosomal DNA that alternates its orientation between on and off *fim* expression states (Abraham *et al.*, 1985; Eisenstein, 1981). Inversion of the promoter segment, known as the *fim* switch or *fimS*, is catalysed by the tyrosine integrase site-specific recombinases FimB and FimE (Klemm, 1986; Dorman and Higgins, 1987). These are among the smallest known members of the integrase family and they act independently of one another at *fimS* (McCusker *et al.*, 2008). FimB and FimE bind to DNA sites that flank and partially overlap the 9 bp inverted repeats that flank the 314 bp *fim* switch (Fig. 1) and site-specific recombination involves DNA cleavage, strand exchange and religation within the 9 bp repeats (McCusker *et al.*, 2008).

The FimB recombinase inverts the *fim* switch in the on-to-off and the off-to-on directions with approximately equal efficiency while FimE inverts it rapidly in the on-to-off direction (Holden *et al.*, 2007; McCusker *et al.*, 2008). Mutants deficient in FimE arise frequently and here FimB alone catalyses *fimS* inversion (Blomfield *et al.*, 1991). Although the FimB recombinase has the ability to invert the switch in either direction with approximately equal efficiencies, there are circumstances when it displays a marked orientational bias as, for example, when the balance of the antagonistic activities of DNA gyrase and DNA topoisomerase I is disturbed (Dove and Dorman, 1994; 1996; Kelly *et al.*, 2006; Müller *et al.*, 2009).

Accepted 8 October, 2009. *For correspondence. E-mail cjdorman@tcd.ie; Tel. (+353) 1 896 2013; Fax (+353) 1 679 9294.

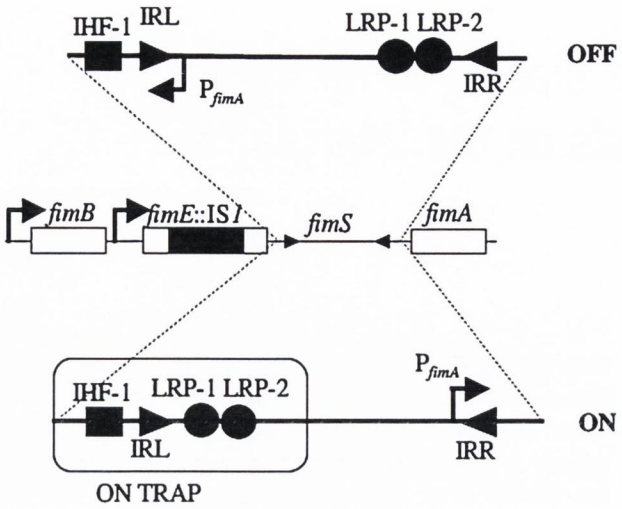


Fig. 1. Organization of the *E. coli fim* switch. The *fimB* gene encodes the tyrosine integrase that catalyses on-to-off and off-to-on inversion of the *fimS* element. This 314 bp segment of DNA is bordered by the 9 bp inverted repeats IRL and IRR and harbours the P_{fimA} promoter, represented by the angled arrow that is directed towards the *fimA* structural gene in the phase-on form of the switch and away from *fimA* in the off phase. Binding sites for IHF and LRP are also shown. The *fimE* gene is inert due to an IS*1* insertion mutation. The expanded sections show the *fimS* element is shown in its phase-on and phase-off orientations. In phase-on, the binding sites for IHF and LRP are juxtaposed at either side of IRL in a region of the switch that has been proposed previously to form a trap (the 'on' trap) that prevents its inversion back to the off phase when DNA becomes relaxed. Not to scale.

DNA gyrase introduces negative supercoils into relaxed DNA by a double-strand breakage, passage and religation mechanism that uses ATP as a source of energy; it removes positive supercoils by the same mechanism (Bates and Maxwell, 2007). DNA topoisomerase I removes negative supercoils by a DNA single-strand cleavage, swivelase and religation mechanism that relies on the energy stored in the negatively supercoiled DNA (Forterre *et al.*, 2007). DNA supercoiling levels vary with the stage of growth of the bacterial culture, with DNA becoming more relaxed at the onset of stationary phase when the metabolic flux in the cell and the concentration of ATP are diminished (Dorman *et al.*, 1988; van Workum *et al.*, 1996). These physiological conditions correlate with a reduction in the negative supercoiling activity of DNA gyrase. ATP hydrolysis by gyrase can be inhibited by treatment with the drug novobiocin. As the concentration of novobiocin increases, the level of gyrase activity in the cell is reduced, allowing DNA topoisomerase I to relax the DNA (Wigley, 1995; Heddle *et al.*, 2000). FimB-mediated inversion of the *fim* switch is highly sensitive to gyrase inhibition. As the DNA relaxes, the *fimS* element adopts a pronounced phase-on bias (Dove and Dorman, 1994; 1996; Müller *et al.*, 2009). The nucleoid-associated protein (NAP) LRP (the leucine-responsive regulatory protein) is essential for this orientational bias (Kelly *et al.*, 2006).

LRP binds to two closely adjacent sites within *fimS* (Fig. 1) and the elimination of these sites results in a dramatic reversal of FimB-mediated switching directionality in novobiocin-treated bacteria with the *fimS* element inverting in the on-to-off instead of the off-to-on direction. Our previous data led us to hypothesize the existence of a reference point adjacent to the left 9 bp inverted repeat (IRL) that allowed the inversion mechanism to distinguish phase-off from phase-on switches (Fig. 1). According to this hypothesis, the juxtaposing of the LRP-occupied sites within *fimS* and this reference point would be a characteristic of the phase-on switch and the resulting nucleoprotein complex, in combination with DNA relaxation, would constitute a 'trap' that prevented inversion of *fimS* to the off orientation. As the *fim* switch in more and more cells enters the 'trap', the subpopulation of phase-on bacteria expands. Elimination of LRP binding to *fimS*, either by introducing base substitution mutations at both LRP binding sites or by eliminating LRP from the cell by mutating the *lrp* gene, removes the trap and allows *fimS* to be switched off by FimB (Kelly *et al.*, 2006) (Fig. 1).

Here we report that the establishment of the phase-on 'trap' also involves the integration host factor, IHF. Like LRP, the IHF protein is required to hold the *fim* switch in the on phase when DNA is relaxed. The IHF binding site is located just outside the IRL sequence at one end of *fimS*, allowing IHF to act there as the reference point that distinguishes the on and off orientations of the switch. We also report that the NAP H-NS is required to establish an alternative phase-off 'trap' in bacteria that do not express LRP. The sites to which H-NS binds are created and removed by the inversion of *fimS*, the very switch element that H-NS regulates. This is a feature that is novel in H-NS biology and that has not been described hitherto in site-specific recombination systems.

Results

IHF binding is required to bias fimS *inversion to the on phase*

Previous work has shown that when DNA gyrase is progressively inhibited by novobiocin treatment, the *fimS* element shows a novobiocin-dose-dependent bias towards the on phase (Dove and Dorman, 1994; 1996; Kelly *et al.*, 2006; Müller *et al.*, 2009). In addition, it has been shown in earlier work that this off-to-on bias is contingent on binding by the LRP protein to two adjacent sites, LRP-1 and LRP-2, within *fimS* (Kelly *et al.*, 2006) (Fig. 1). The two LRP binding sites are located asymmetrically within *fimS*, raising the possibility that they may contribute to a structure that distinguishes the on phase

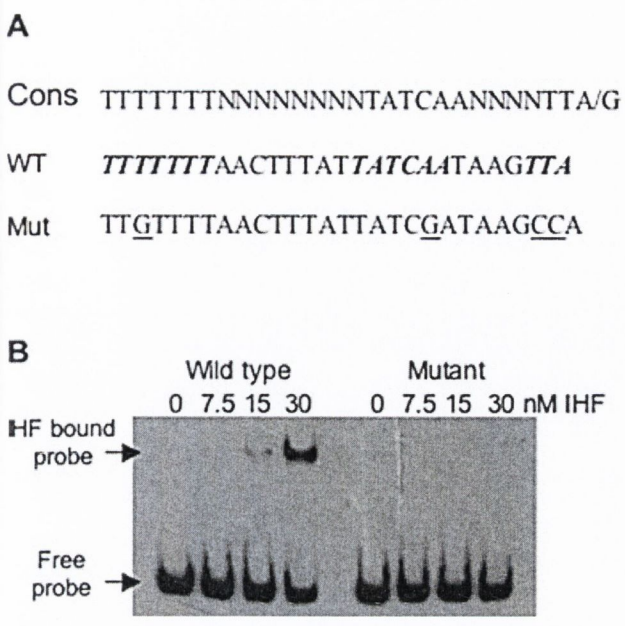


Fig. 2. Inactivation of the IHF-1 binding site. The DNA sequence of the IHF-1 binding site (WT) for the IHF protein is shown aligned with the consensus for IHF binding site sequences (Cons) together with the base-pair substitutions made in IHF-1 (Mut) (A). The effect of these mutations on IHF binding to IHF-1 site DNA was tested by electrophoretic mobility shift assay (B). The concentration of purified IHF used is indicated above each lane. The unshifted biotinylated DNA (free probe) and the IHF–DNA complex are shown by arrows.

switch from the off phase. This hypothesis postulates the existence of a reference point that somehow communicates with the LRP-occupied LRP-1 and LRP-2 sites in the phase-on switch. An inspection of the DNA adjacent to *fimS* revealed that, in the on phase, the two LRP binding sites lie close to a well-characterized binding site for IHF (Fig. 1).

The DNA sequence of this IHF binding site (IHF-1) was a strong match to the consensus for such sites (Dorman and Higgins, 1987; Hales *et al.*, 1994) and it was disrupted by site-directed mutagenesis (Fig. 2). The wild-type and mutant sequences were then compared for their ability to bind purified IHF protein in an electrophoretic mobility shift assay (EMSA). The results showed that the base substitutions had abrogated the ability of the mutant sequence to bind IHF (Fig. 2). The altered sequence was then substituted for the wild-type version on the chromosome and polymerase chain reaction (PCR) amplification and DNA sequencing were used to confirm the presence of the mutated IHF binding site in the chromosomal *fim* locus.

The effect of the IHF binding site mutation on *fimS* inversion following DNA gyrase inhibition was tested using a PCR-based inversion assay with cultures incubated in the presence of increasing concentrations of the gyrase inhibitor novobiocin. In this assay the FimB recombinase is expressed *in trans* from a multicopy plasmid and the orientation of the *fim* switch is monitored by amplifying it from the chromosome by PCR and digesting the DNA with a restriction enzyme (BstUI) that creates two DNA bands that are diagnostic of the switch in the on phase and two distinct bands that are diagnostic of the switch in the off phase (Smith and Dorman, 1999; Kelly *et al.*, 2006). The results showed that mutation of the IHF binding site immediately adjacent to the IRL sequence, caused *fimS* inversion to become strongly biased in favour of the off orientation (Fig. 3A and B). This outcome was identical to that seen previously when the LRP-1 and LRP-2 sites were mutated to prevent LRP binding to *fimS*

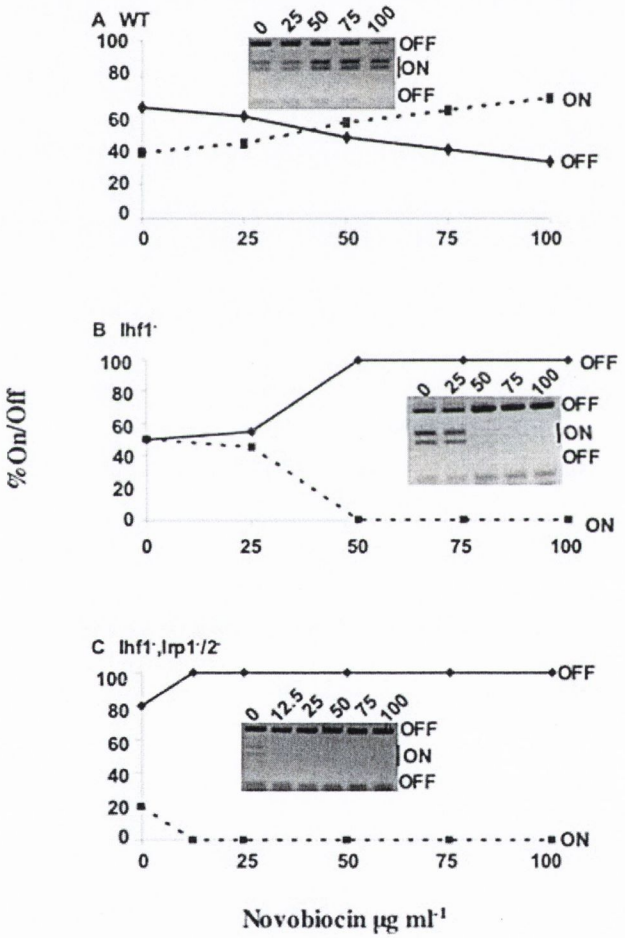


Fig. 3. Switch inversion preferences with and without the IHF-1 site. The PCR products diagnostic of the on and off forms of the *fim* switch were resolved by electrophoresis on agarose gels (insets). The concentration of novobiocin (μg ml^{-1}) used to treat the culture prior to PCR is given above each lane. Phase-on-specific and phase-off-specific DNA bands (two each) are labelled. The bands were scanned with a densitometer and the data plotted to summarize the effect of novobiocin treatment on *fimS* phase biasing. Data are presented for the wild type (A), the mutant *fim* switch lacking the IHF-1 site (B) and the mutant *fim* switch lacking the IHF-1 site and the LRP-1 and LRP-2 sites (C). The experiment was repeated three times and typical data are shown.

(Kelly *et al.*, 2006). These data supported the hypothesis that IHF and LRP were required collectively when DNA is relaxed to maintain *fimS* in its on orientation.

It was anticipated that elimination of IHF binding to its IRL-proximal IHF-1 site, together with abrogation of LRP binding to sites LRP-1 and LRP-2, would result in a strong phase-off bias in *fimS* inversion following DNA gyrase inhibition. To test this hypothesis, mutations that prevented IHF binding to its IRL-proximal site and mutations that abolished LRP binding to sites LRP-1 and LRP-2 were all introduced into the *fim* locus on the chromosome. Application of the PCR-based *fimS* inversion assay following DNA gyrase inhibition with increasing concentrations of novobiocin showed that the *fim* switch adopted a phase-off orientation, as expected (Fig. 3C).

H-NS modulates fimS *inversion directionality*

Previous research has shown that the nucleoid-associated protein H-NS influences the inversion of the *fimS* DNA element (Higgins *et al.*, 1988; Kawula and Orndorff, 1991; O'Gara and Dorman, 2000). While this protein has been described in detail as a repressor of transcription (Dorman, 2004) its contributions to DNA recombination have been studied much less intensively.

Inactivation of the *hns* gene resulted in no change in the sensitivity of *fimS* inversion directionality to gyrase inhibition in the standard assay used in this study (compare Fig. 3A with Fig. 4A). We also wished to assess the contribution of H-NS to *fimS* inversion when DNA was relaxed in the absence of the NAPs IHF and LRP. Therefore, the effect of eliminating IHF binding to its IRL-proximal site was studied in an *hns* null mutant by PCR assay. The results showed that in the absence of both IHF and H-NS binding, the *fim* switch displayed a strong on-to-off bias (Fig. 4B), a pattern that was very similar to that seen in the absence of IHF binding alone (Fig. 3B). These data showed that IHF was required for the maintenance of the phase-on *fim* switch regardless of the status of the *hns* gene.

We next assessed the effect on *fimS* inversion of eliminating LRP binding to the LRP-1 and LRP-2 sites in a mutant that was also deficient in H-NS. Results from the PCR switch assay showed that DNA gyrase inhibition resulted in no bias in favour of the off orientation (Fig. 4C). Theses data contrasted sharply with those obtained previously with the LRP-1 and LRP-2 binding site mutant in the presence of H-NS, where *fimS* became strongly biased towards the off orientation following gyrase inhibition (Kelly *et al.*, 2006). The results established a role for H-NS in maintaining the *fim* switch in its off orientation in the LRP binding sites mutant.

The *fim* switch lacking the IRL-proximal IHF binding site and the LRP-1 and LRP-2 binding sites was studied in an

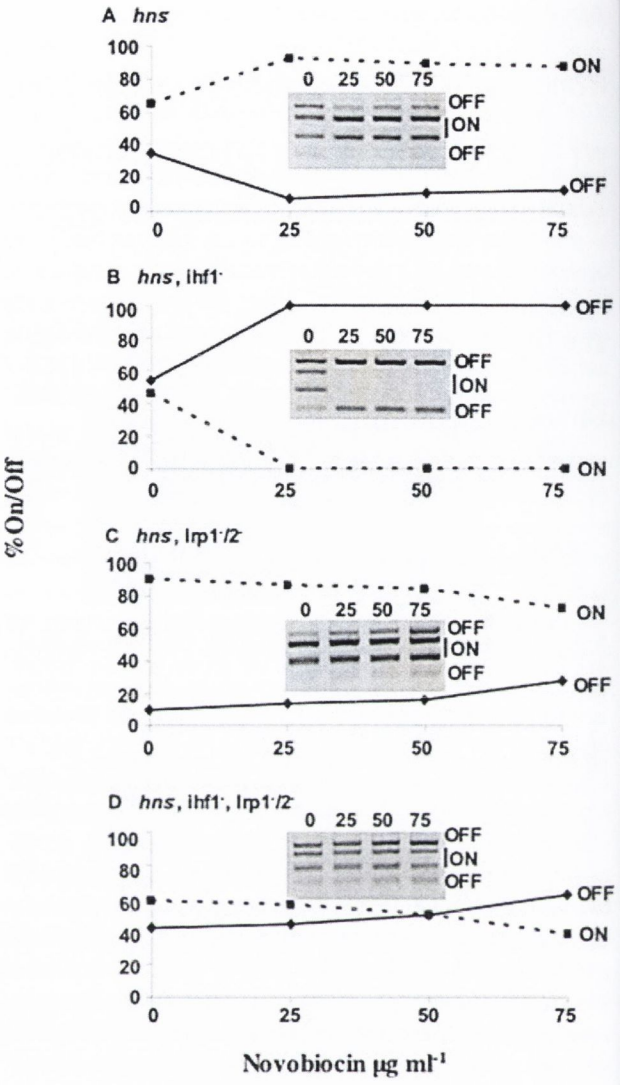


Fig. 4. Inversion of *fimS* in an *hns* mutant. The PCR products diagnostic of the on and off forms of the *fim* switch were resolved by electrophoresis on agarose gels (insets). The concentration of novobiocin (μg ml^{-1}) used to treat the culture prior to PCR is given above each lane. Phase-on-specific and phase-off-specific DNA bands (two each) are labelled. The bands were scanned with a densitometer and the data plotted to summarize the effect of novobiocin treatment on *fimS* phase biasing. Data are presented for the *hns* mutant (A), the *hns* mutant lacking the IHF-1 site in the *fim* switch (B), the *hns* mutant lacking the LRP-1 and LRP-2 binding sites in the *fim* switch (C) and the *hns* mutant lacking the IHF-1 site and the LRP-1 and LRP-2 sites in the *fim* switch (D). The experiment was repeated three times and typical data are shown.

hns mutant. Progressive inhibition of DNA gyrase activity with novobiocin did not result in a strong bias of *fimS* inversion in favour of the off orientation (Fig. 4D). This showed that LRP and H-NS collaborated in sustaining the 'off' trap. We next examined the interaction of H-NS with *fimS* in order to obtain further insight into the mechanism underlying orientational bias.

Differential binding of H-NS to phase-on and phase-off fimS

To assess the interaction of H-NS with *fimS*, a ~760 bp region encompassing *fimS* was PCR amplified from phase-locked substrates (Fig. 5). These were then subdivided into shorter segments by restriction enzyme digestion for use in competitive EMSAs with purified H-NS protein.

H-NS binding was tested in the ~760 bp region encompassing *fimS* in the phase-off orientation (Fig. 5A). Digestion with Hpy99I separated the PCR product into two fragments of 550 and 210 bp, respectively, each containing one of the inverted repeats. A separate restriction enzyme digestion with MfeI cleaved the DNA within IRL to produce two fragments of 400 and 370 bp (Fig. 5A). Incubation of all of these DNA fragments with increasing amounts of H-NS clearly showed high-affinity H-NS binding to the 550 bp fragment that contained the intact IRL and its flanking DNA with the P_{fimA} promoter in the phase-off orientation. The lack of binding in the MfeI-cut DNA fragment indicated that H-NS bound to a site that overlapped the IRL in the phase-off orientation, and that *fimS* alone (i.e. the DNA extending from IRL to IRR only) was not sufficient for H-NS binding (i.e. the 400-OFF fragment did not shift).

Evidence suggesting that H-NS did not interact with IRR-proximal DNA sequences when *fimS* was in the phase-on orientation had been obtained with DNA digested with SnaBI, an enzyme that cut the DNA very close to IRR (data not shown). The experiment was redesigned to leave more DNA flanking IRR in the phase-on switch (Fig. 5B). A ~760 bp DNA fragment amplified from a phase-locked-on substrate was digested with MfeI to produce a fragment containing the entire *fimS* region (including the *fim* promoter), IRR and upstream flanking DNA (fragment 400-ON). This DNA segment was bound strongly by H-NS and resolved as the H-NS–DNA complex indicated by the open arrowhead (Fig. 5B). A comparison of Fig. 5A and B clearly showed the phase variable nature of H-NS binding to the *fim* switch: DNA fragment 400-ON (Fig. 5B) was identical to 400-OFF (Fig. 5A) except for *fimS* orientation, and 400-ON was bound strongly by H-NS whereas 400-OFF bound the protein poorly.

The invertible sequence extending from IRL to IRR is common to both the phase-on and phase-off forms of the switch. The data obtained thus far suggested that this was unlikely to support H-NS binding alone because DNA sequences in the static portion of the chromosome flanking IRL or IRR were also needed. To test this hypothesis, the invertible element was precisely excised from between IRL and IRR by restriction enzyme digestion with HaeIII and tested for H-NS binding. The results showed that H-NS was unable to bind to the invertible region when this was isolated from its flanking DNA sequences (Fig. 5C).

Overall, the data from the EMSA analysis showed that H-NS bound to a site that moved from one end of *fimS* to the other as the switch inverted. This site consisted of a constant component that was located within *fimS* close to the P_{fimA} promoter and a variable component that was provided by DNA sequences outside *fimS* and adjacent either to IRL or to IRR. Only the combination of the P_{fimA}-proximal portion of the binding site with the portion associated with DNA sequences flanking IRL (phase-off switch) or IRR (phase-on switch) produced a fully functioning H-NS binding site. This arrangement provided yet another means by which the phase-on and phase-off forms of the *fim* switch could be distinguished.

Discussion

Those tyrosine integrase site-specific recombinases that have been characterized in most detail catalyse DNA integration and excision events such as Int-mediated bacteriophage lambda entry and exit from the chromosome of *E. coli* or DNA dimer resolution reactions such as those catalysed by XerC and XerD, or Cre, or Flp (Rajeev *et al.*, 2009). In the case of the Int recombinase, the Xis protein cofactor has been characterized as a directionality determinant, helping the lambda recombination system to distinguish between the integration and excision events (van Duyne, 2005; Radman-Livaja *et al.*, 2006). Such Recombination Directionality Factors (RDFs) have been described in the case of several other recombination systems and appear to be an important part of the machinery that ensures that an otherwise random event proceeds in a specific direction under a given set of physiological circumstances (Lewis and Hatfull, 2001; 2003; Lesic *et al.*, 2004).

The FimB integrase is unusual in that it promotes DNA inversion rather than integration/excision or DNA multimer resolution. In bacteria growing in mid-exponential phase at 37°C in L broth, the FimB protein inverts the *fimS* element from the on phase to the off phase and back again at approximately equal rates (McCusker *et al.*, 2008). However, when novobiocin is used to inhibit DNA gyrase activity, allowing DNA to become relaxed, the FimB-catalysed *fimS* inversion reaction adopts a pronounced bias in favour of the on orientation (Dove and Dorman, 1994; Kelly *et al.*, 2006; Müller *et al.*, 2009). Previously, the LRP NAP was identified as playing a role as an RDF in this reaction (Kelly *et al.*, 2006). Specifically, the LRP protein is required to maintain the *fimS* element in the on orientation and it does this by binding to the LRP-1 and LRP-2 sites within *fimS* (Fig. 1).

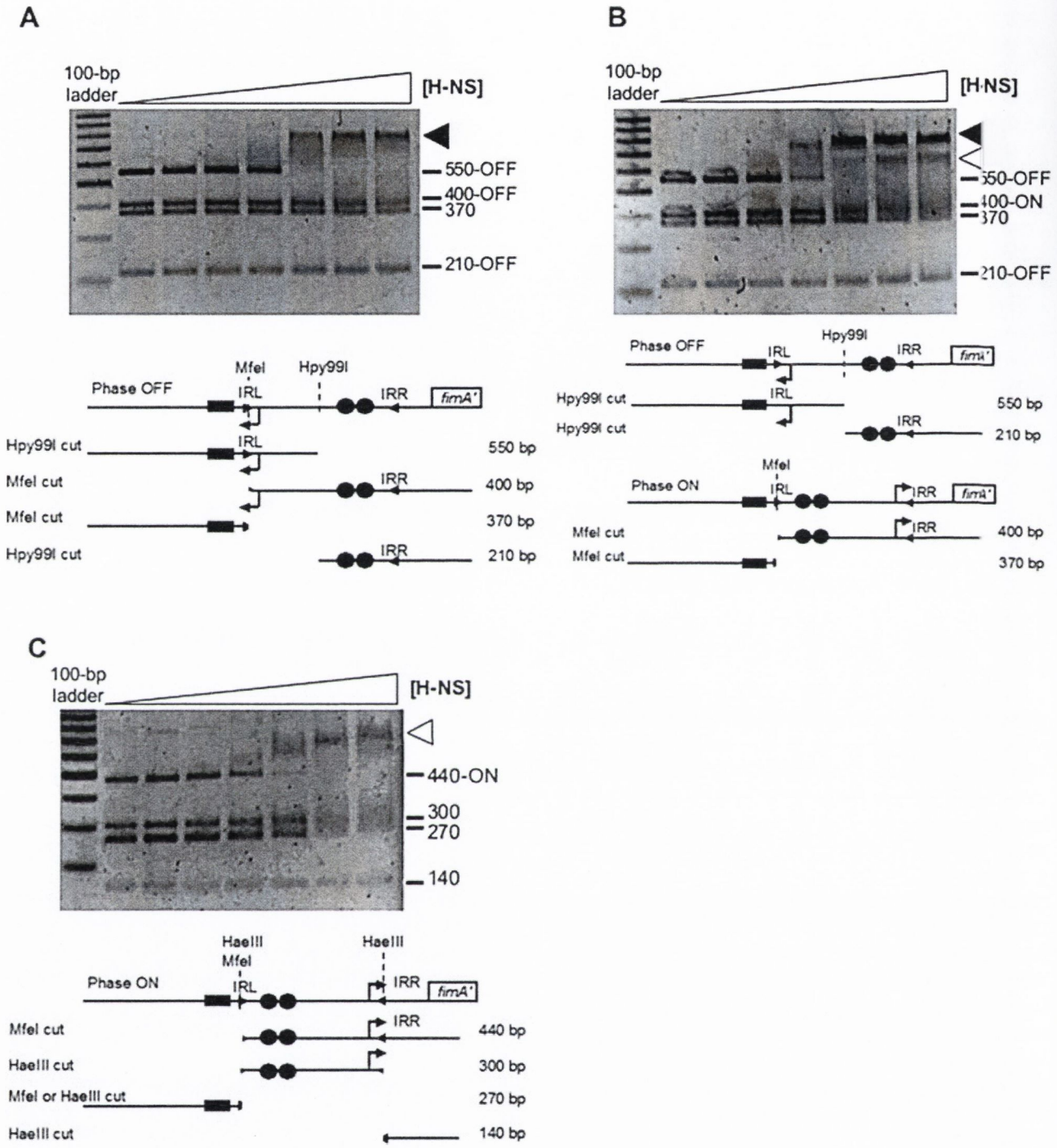


Fig. 5. Binding of H-NS to *fimS* is phase-variable. H-NS binding to the *fim* switch was assessed by a competitive *in vitro* binding assay. DNA fragments were amplified by PCR from phase-on-locked or phase-off-locked bacterial strains using a primer pair that gave a ~760 bp product. This PCR product was digested with restriction enzymes as shown schematically (vertical dashed lines) before being mixed and incubated with increasing amounts of purified H-NS protein (final concentrations used in each gel lane, left to right: 0, 7, 35, 70, 100, 135 and 165 nM). Each DNA fragment is identified on the agarose gels by size (bp) and *fimS* orientation (e.g. 550-OFF). DNA fragments that contain either only static (i.e. non-inverting DNA) or an invertible segment not attached to static DNA are indicated by size alone. Resolved complexes in which H-NS is bound at the phase-off IRL (A and B) are indicated by filled arrowheads while resolved complexes in which H-NS is bound at the phase-on IRR (B and C) are indicated by open arrowheads.

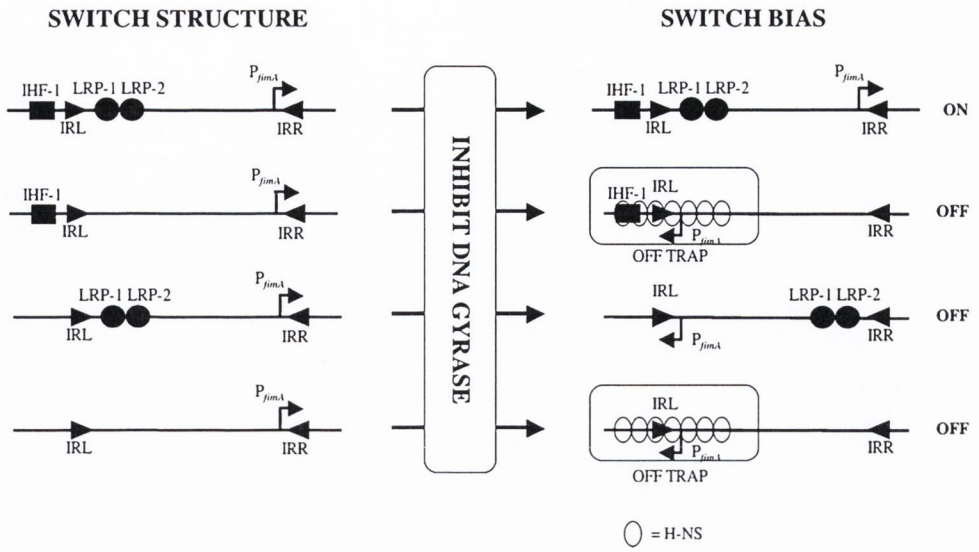


Fig. 6. H-NS, IHF, LRP and biasing of the *fimS* inversion mechanism. The structures of the wild-type *fim* switch and its mutant derivatives are summarized on the left of the figure. In each case the *fimS* element is in the on phase with the P_{fimA} promoter directed towards the IRR. The effect of DNA gyrase inhibition by novobiocin is summarized on the right of the figure for each form of the switch. The wild-type switch displays a bias towards the on phase whereas the three mutant forms of *fimS* show a bias towards off phase. However, in the cases of *fimS* lacking sites LRP-1 and LRP-2, or LRP-1, LRP-2 and IHF-1, this phase-off bias is dependent on the presence of the H-NS protein. In these two cases, H-NS is shown binding to a region that straddles IRL. Not to scale.

In the present study, the IHF protein was identified as a second RDF that is required to maintain *fimS* in the on orientation when gyrase activity is inhibited. The location of the IHF binding site in the non-inverting chromosomal DNA immediately adjacent to IRL creates the potential for differential interaction with LRP bound to sites LRP-1 and Lrp-2 within *fimS* because all three sites are only juxtaposed when the switch is in the on orientation (Fig. 1). It is proposed that the DNA bending activity of IHF and the DNA wrapping activity of LRP in combination with the relaxed nature of the *fim* DNA that obtains in the absence of adequate levels of gyrase activity create a nucleoprotein complex that prevents DNA inversion by FimB in the on-to-off direction (Fig. 6).

Previous work has suggested a role for the H-NS protein in influencing the FimB-mediated inversion of *fimS* (Higgins *et al.*, 1988; Kawula and Orndorff, 1991; O'Gara and Dorman, 2000). Careful examination of H-NS interactions with *fimS* and its flanking regions showed that H-NS interaction with the switch was contingent on *fimS* orientation (Figs 5 and 6). Our data show that the H-NS binding site in the *fimS* region is distributed across three incomplete sites. One of these is in the invertible *fimS* element and is adjacent to the P_{fimA} promoter. The others are located in the non-inverting parts of the chromosome immediately outside IRL and IRR. When the switch is in the on orientation, a complete site capable of binding H-NS in an EMSA is formed at the IRR end of the switch; when *fimS* is in the off orientation, a complete H-NS interaction site is formed only at the IRL end. The physiological significance of H-NS binding at IRR when *fimS* is in the on orientation is unknown. It may play a role in modulating the activity of the P_{fimA} promoter.

The data obtained in this study show that the H-NS–DNA complex involving IRL has a role to play in maintaining *fimS* in its off orientation following DNA relaxation. This is seen most clearly in the absence of LRP binding to *fimS*. In bacteria that express the H-NS protein, elimination of LRP binding to *fimS* sites LRP-1 and LRP-2 results in switching in the on-to-off direction when DNA gyrase is inhibited (Kelly *et al.*, 2006). This is not the case in the absence of H-NS: here bidirectional switching continues to be detected even though the LRP-1 and LRP-2 sites have been inactivated (Figs 4 and 6). Our results show that either H-NS or LRP can sustain the 'off' trap in relaxed DNA; only when both proteins fail to interact with *fimS* does the switch fail to become biased to the off orientation following gyrase inhibition. The results also show that IHF and LRP are not equivalent as RDF elements; prevention of IHF binding to the IRL-proximal site creates the 'off' trap regardless of the status of H-NS expression; abrogation of LRP binding to LRP-1 and LRP-2 only creates the 'off' trap if H-NS is present (Figs 4 and 6).

The data presented here cast new light on the issue of programmed versus stochastic phase variation of type 1 fimbriae in *E. coli*. The FimB site-specific recombinase acquires an orientational bias when DNA gyrase activity is impaired and DNA becomes relaxed. The maintenance of negatively supercoiled DNA depends on the activity of

Table 1. Bacterial strains, plasmids and bacteriophages.

Strain/plasmid/bacteriophage	Relevant details[a]	Reference/source
Strain		
CJD808	VL386 *fimB*::*kan*. Phase locked on	Dove and Dorman (1994)
CJD957	VL386 *fimB*::*kan*. Phase locked off	Dove and Dorman (1994)
CSH50	λ pL (209) *fimE*::IS*1*	Miller (1972)
VL386	Δ (*fimA–lacZ*)λ pL (209) *fimE*::IS*1*	Freitag *et al.* (1985)
VL386 IHF 1	VL386 *recD* IHF 1 × VL386	This study
VL386 Lrp 1/2	VL386 *recD* LRP1/2 × VL386	This study
VL386 IHF 1, Lrp 1/2	VL386 *recD* IHF 1, LRP1/2 × VL386	This study
VL386 *hns*$^-$	Linker- and DNA-binding-domain-encoding regions of *hns* replaced with *tetRA* element from Tn*10*	This study
VL386 *hns*$^-$, IHF 1	VL386 *hns*$^-$ × VL386 IHF 1	This study
VL386 *hns*$^-$, Lrp 1/2	VL386 *hns*$^-$ × VL386 LRP1/2	This study
VL386 *hns*$^-$, IHF 1, Lrp 1/2	VL386 *hns*$^-$ × VL386 IHF 1, LRP1/2	This study
VL386 *recD*	VL386 *recD*::Tn*10*	Smith and Dorman (1999)
VL386 *recD* IHF 1	Δ*fimB*::*kan*, 4 bp mutation in the IHF 1 site (see Fig. 2)	This study
VL386 *recD* LRP1/2	Δ*fimB*::*kan*, binding site mutations in LRP sites 1 and 2	Kelly *et al.* (2006)
VL386 *recD* IHF 1, LRP1/2	Δ*fimB*::*kan*, binding site mutations in IHF site 1 and LRP sites 1 and 2	This study
Plasmid		
pCPC2	700 bp (from *fimE*::IS*1* to *fimA–lacZ*) inserted in PstI site of pMMC106 (*fimS* phase off)	This study
pCPC2 IHF1	pCPC2 containing a 4 bp mutation in IHF1 site	This study
pKD46	Lambda genes *gam*, *bet* and *exo* under the control of the arabinose-inducible pBAD promoter (Amp^R)	Datsenko and Wanner (2000)
pMMC106	pMCL210 cut with XhoI, NheI and religated to delete the *lac* promoter.	McCusker *et al.* (2008)
pSGS501	*fimB*::*kan fimE*::IS*1 fimA–lacZ* cloned in pACYC184 (phase off)	Smith and Dorman (1999)
pSGS501 IHF1	Binding site mutations in IHF site 1 in pSGS501	This study
pSGS501 LRP1, 2	Binding site mutations in LRP sites 1 and 2 in pSGS501	Kelly *et al.* (2006)
pSGS501 IHF1, LRP1, 2	Binding site mutations in IHF site 1 and LRP sites 1 and 2 in pSGS501	This study
pSLD203	*fimB* cloned in pUC18	Dove and Dorman (1994)
Bacteriophage		
P1	*vir*	Lab stocks

a. Transductional crosses are represented as A × B, where A is the donor and B is the recipient.

DNA gyrase that is in turn dependent on the phosphorylation potential of the cell (van Workum *et al.*, 1996). Metabolically active bacteria typically adopt a motile or planktonic lifestyle and become sedentary when resources can no longer support this lifestyle. Attachment to and colonization of a surface represents a distinct lifestyle and the transition from motility to non-motility and attachment is known to involve complex signalling cascades and gene regulatory circuits (Müller *et al.*, 2009; Pesavento and Hengge, 2009).

DNA relaxation is a feature of the stationary phase of growth that correlates with a reduction in the negative supercoiling activity of DNA gyrase (Bordes *et al.*, 2003). It is tempting to speculate that the *fim* switch becomes biased towards the on phase when bacteria approach a state of low metabolic flux since this will promote type 1 fimbrial expression and assist attachment and colonization. The involvement of the LRP and IHF proteins in the establishment of a phase-on bias is also consistent with a need for physiological sensitivity on the part of *fim* gene expression. IHF protein levels peak at the transition from exponential growth to stationary phase (Ditto *et al.*, 1994) making this protein an attractive growth stage reporter. LRP has been established as a regulator of genes involved in the response to nutrient stress (Brinkman *et al.*, 2003; Yokoyama *et al.*, 2006) and LRP levels increase as bacteria approach stationary phase (Landgraf *et al.*, 1996). Variation in the supply or the activities of DNA gyrase and the NAPs described in this study offers the bacterium an excellent mechanism for fine-tuning the phase-variable expression of the *fim* structural genes to its physiological circumstances.

Experimental procedures

Growth media, growth conditions and genetic techniques

The bacterial strains used in this study were derivatives of *E. coli* K-12 and are listed in Table 1. Bacteria were routinely cultured aerobically in Luria–Bertani (LB) broth at 37°C or on L agar plates at 37°C for 12–16 h (Miller, 1992). Antibiotics were used at the following concentrations: carbenicillin, 100 μg ml^{-1}; kanamycin, 20 μg ml^{-1}; chloramphenicol, 25 μg ml^{-1}; tetracycline, 12.5 μg ml^{-1}. Plasmid DNA was introduced to bacterial cells by $CaCl_2$ transformation (Sambrook *et al.*, 1989) or by electroporation (Hanahan *et al.*, 1991)

Table 2. Oligonucleotides.

Primer name	Primer sequence (5′–3′)
hns.fw	CTC AAC AAA CCA CCC CAA TA
hns.rv	TGG CGG GAT TTT AAG CAA GT
hns.mut.fw[a]	TAA ACT GCA GCA ATA TCG CGA AAT GCT GAT CGC TGA CTA ATT AAG ACC CAC TTT CAC ATT
hns.mut.rv	TTG CTT GAT CAG GAA ATC GTC GAG GGA TTT ACC TTG CTC ACT AAG CAC TTG TCT CCT G
Ihf1.int.fw	GCA CCT CAA AAA CAC CAT CA
fimS.rv	GAC ATG GGC AGT CGT TCT GTA C
Ihf1.fw.pstI[b]	TTG GCG GCT GCA GGC AGG CCT ATC GCA TTA TTC GCG AT
fimS.rv.pstI[b]	GGC GGC TGC AGG ACA TGG GCA GTC GTT CTG TAC
Ihf1.mut.fw	GAG AAG AGG TTT GAT TTG GCT TAT CGA TAA TAA AGT TAA AAC AAC AAA TAA ATA CAA GAC
Ihf1.mut.rv	GTC TTG TAT TTA TTT GTT GTT TTA ACT TTA TTA TCG ATA AGC CAA ATC AAA CCT CTT CTC
Ihf1.bs.fw.BIO[c]	CTC ATA AAC GAA AAA TTA AAA AGA GAA G
Ihf1.bs.rv	ACG TAA CTT ATT TAT GAT ATG GAC AG
OL 4	CCG TAA CGC AGA CTC ATC CTC
OL 20	GAG TTT TAA TTT TCA TGC TGC TTT CC

a. The premature stop codon to be introduced into *hns* is underlined.
b. PstI restriction enzyme site used for cloning in pMMC106 is underlined.
c. This primer has a 5′ biotin tag.

using a Bio-Rad Gene Pulser. Transductions were carried out using bacteriophage P1 *vir* (Miller, 1992).

Assays involving the DNA gyrase inhibitor novobiocin were performed as follows: bacteria containing the *fimA–lacZ* transcriptional fusion were screened for their Lac phenotype on MacConkey lactose indicator medium as previously described (Dove and Dorman, 1994). Distinct phase-on (red) or phase-off (white) colonies were used to inoculate overnight cultures that were in turn used to inoculate 250 ml flasks containing 25 ml of LB and these were grown to an OD_{600} of ~0.1 at which point novobiocin (stock soloution 100 mg ml^{-1}) was added to a final concentration of 25, 50, 75 or 100 μg ml^{-1}. Cultures were further incubated overnight (20 h) before sampling for determination of *fim* switch orientation. Doubling times of VL386 pSLD203 (0, 25, 50, 75, 100 μg ml^{-1} novobiocin) and VL386*hns*$^-$ pSLD203 (0, 25, 50, 75 μg ml^{-1} novobiocin) were determined to be (in min) 41, 52, 221, 284, 276 and 54, 72, 149, 255 respectively. Wild-type cultures continued growing to saturation (OD_{600} ~ 5) whereas cultures of the *hns* mutant containing novobiocin achieved lower densities after overnight growth. Typical OD_{600} values for the *hns* mutant treated with 0, 25, 50 or 75 mg ml^{-1} novobiocin were 3.8, 3.8, 1.2 and 0.8 respectively. OD_{600} measurements were taken every 30 min, plotted as log OD_{600} against time and the gradient of the linear portion was used to plot the doubling time. Doubling times were the means of two independent experiments and replicates varied by less than 5%.

Molecular biological techniques

Plasmid DNA was isolated from bacterial cells using the HiYield plasmid mini-kit (RBC Biosciences). DNA fragments were purified from agarose gels using the HiYield gel/PCR DNA fragments extraction kit (RBC Biosciences). Restriction enzymes were purchased from New England Biolabs and used according to the manufacturer's directions. Automated DNA sequencing was carried out by GATC Biotech and oligonucleotide synthesis was performed by MWG Biotech. Phusion High-Fidelity DNA polymerase (Finnzymes) was used, according to manufacturer's directions, for all PCR.

Determination of fim *switch orientation*

The orientation of the *fim* switch was determined in cells expressing the FimB recombinase from the multicopy plasmid pSLD203 as previously described (Dove and Dorman, 1996; Smith and Dorman, 1999; Kelly *et al.*, 2006) with minor modifications described below. This method exploited a unique BstUI restriction site in the *fim* switch that results in restriction fragment length dimorphism among BstUI-digested PCR products that are characteristic of phase-on and phase-off switches.

Fifty microlitres of overnight culture was boiled for 5 min and oligonucleotides OL4 and OL20 (Table 2) were used with Phusion High-Fidelity DNA polymerase (Finnzymes) to amplify the *fim* switch region as a 726 bp DNA product. The PCR cycle involved denaturation at 98°C for 3 min, followed by 34 cycles of 98°C for 10 s, 62°C for 30 s and 72°C for 22 s, followed by a final extension for 10 min at 72°C. Ten microlitres of each PCR product was subsequently digested using BstUI and electrophoresed on a 2% (w/v) agarose/TAE gel. Phase-off switches yielded DNA fragments of 539 and 187 bp, while phase-on switches gave fragments of 433 and 293 bp. The proportion of phase-on- and phase-off-specific fragments in the population was determined by analysing DNA band intensities on the gel with Image J analysis software.

Creation of an hns *mutant allele*

Primers hns.mut.fw and hns.mut.rv (Table 2), which contain flanking sequences homologous to the *hns* open reading frame, were used to amplify the *tetRA* element from an *E. coli* strain containing transposon Tn*10*. The ~2 kb PCR product was subsequently integrated into the chromosome

using lambda *red*-mediated recombination as described previously (Datsenko and Wanner, 2000). Integration of the *tetRA* element introduced an in-frame stop codon (TAA) after 68 amino acids of the *hns* open reading frame and deleted the parts of the *hns* gene that encode the linker and nucleic acid-binding domains. The structure of the interrupted *hns* gene was confirmed by PCR using flanking primers hns.fw and hns.rv and DNA sequencing. The absence of the H-NS protein from the mutant was confirmed by Western blotting using (separately) mono- and polyclonal H-NS antibodies as described previously (Free *et al.*, 2001). This *hns* mutant strain also showed a *fimS* phase-locked-on phenotype on MacConkey lactose agar associated with rapid switching in the absence of *hns* (Higgins *et al.*, 1988). In all phenotypic tests this strain was indistinguishable from a congenic strain containing the *hns*206 allele (Dersch *et al.*, 1993).

Site-directed mutagenesis

Site-directed mutagenesis of the IHF 1 binding site was performed using the Quikchange II (Stratagene) kit. Plasmid pCPC2 was constructed by PCR amplification of a ~700 bp region of the *fim* locus encompassing the StuI site (located upstream of the IHF 1 site) and the BsrgI site (located downstream of *fimS*) using oligonucleotides pairs (Ihf1.fw.pstI and fim.rv.pstI) and subsequent cloning of this PCR product into pMMC106 (McCusker *et al.*, 2008). Ihf1.mut.fw and Ihf1.mut.rv were used to introduce a 4 bp mutation in the IHF 1 site (Fig. 2). The modifications made to the IHF 1 site introduced a ClaI restriction site. The mutated IHF 1 site was cloned into pSGS501 and into pSGS501-LRP 1/2 following digestion with StuI and MfeI. The presence of either LRP site 1 or LRP site 2 mutations introduced an XhoI restriction site (Kelly *et al.*, 2006) and the NAP binding site alterations were confirmed through PCR amplification of *fimS* and digestion using the novel restriction sites and DNA sequencing.

Allele replacement

Mutations in *fimS* were integrated into the chromosome as described previously (Smith and Dorman, 1999). Derivatives of plasmid pSGS501 were digested with EcoRV releasing an 8.5 kb fragment including the *fimB* gene (interrupted by a Kan^R cassette) and the *fimS* mutations. The digestion products were electrophoresed through 0.7% (w/v) agarose/TAE, to allow clear separation of the *fim* fragment from the linearized plasmid backbone (4.25 kb). The *fim* fragment was then extracted from the gel and precipitated using Pellet Paint co-precipitant (Novagen). Roughly 2 μg of this fragment was electroporated into strain VL386*recD* in a volume not exceeding 5 μl. Transformants were selected on L agar containing kanamycin and after single colony purification on MacConkey-lactose kanamycin agar, phase-locked-off colonies (white) were screened for the presence of the mutations by PCR and DNA restriction enzyme digestion at novel sites. DNA from the relevant portions of the *fim* locus in the integrants was then sequenced to ensure that the correct series of mutations was present.

Electrophoretic mobility shift assays

The binding of purified IHF protein (a kind gift from Professor Steven Goodman) to the wild type and mutant IHF 1 binding site was tested by EMSA as described previously (Yang and Nash, 1995). Biotinylated DNA probes (120 bp) containing either the wild type or mutant IHF 1 site were amplified using primers Ihf1.bs.fw.BIO and Ihf1.bs.rv with plasmids pCPC2 or pCPC2IHF-1$^-$ serving as templates. Probes (0.5 ng) were incubated in 0.5× TBE [45 mM Tris-borate (pH 8.3) containing 1.25 mM disodium EDTA], 5 μg ml^{-1} bovine serum albumin (BSA), 10% (v/v) glycerol, 60 mM KCl and 1 μg of poly(dI.dC) with increasing concentrations (0 to 30 nM) of purified IHF in a final reaction volume of 20 μl. Samples were incubated for 30 min at room temperature and loaded onto gels without the addition of loading dye. Protein–DNA complexes were resolved by electrophoresis through a 6% polyacrylamide gel for 1 h at room temperature. The gel was then transferred to Biodyne B membrane (Pall), UV cross-linked (GS GeneLinker, Bio-Rad) and developed using the EMSA kit (Pierce) as per manufacturer's recommendations.

To assess H-NS–DNA interactions, DNA fragments of ~760 bp (Fig. 5A and B) and ~730 bp (Fig. 5C) from the regions encompassing *fimS* were amplified from phase-locked substrates using primer pairs Ihf1.int.fw/fimS.rv and OL20/OL4 respectively. The binding of H-NS to defined restriction fragments of *fimS* (~1 μg of total DNA per reaction) was carried out in 20 μl of reaction mixtures containing increasing concentrations of purified H-NS protein (final concentrations: 0, 7, 35, 70, 100, 135 and 165 nM) in 20 mM Tris HCL, 1 mM EDTA, 100 μg ml^{-1} BSA, 1 mM DTT, 10% glycerol and 80 mM NaCl. The reactions were incubated at room temperature for 30 min before loading (without the addition of loading dye) onto a 3% (w/v) molecular screening agarose/TAE gel and electrophoresed at 90 V for 1 h. Gels were stained for 20 min in ethidium bromide (1 μg ml^{-1}) and visualized under UV light using an AlphaImager 2200 (AlphaInnotech) gel documentation system.

Acknowledgements

We thank Stephen Smith for helpful discussions and Steven Goodman for generously providing purified IHF protein. The work was supported by Grants 06/RFP/GEN017 and 07/IN1/B918 from Science Foundation Ireland.

References

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

Bates, A.D., and Maxwell, A. (2007) Energy coupling in type II topoisomerases: why do they hydrolyze ATP? *Biochemistry* **46:** 7929–7941.

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

Bordes, P., Conter, A., Morales, V., Bouvier, J., Kolb, A., and Gutierrez, C. (2003) DNA supercoiling contributes to discon-

nect sigmaS accumulation from sigmaS-dependent transcription in *Escherichia coli*. *Mol Microbiol* **48:** 561–571.
Brinkman, A.B., Ettema, T.J., de Vos, W.M., and van der Oost, J. (2003) The Lrp family of transcriptional regulators. *Mol Microbiol* **48:** 287–294.
Datsenko, K.A., and Wanner, B.L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci USA* **97:** 6640–6645.
Dersch, P., Schmidt, K., and Bremer, E. (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.
Ditto, M.D., Roberts, D., and Weisberg, R.A. (1994) Growth phase variation of integration host factor level in *Escherichia coli*. *J Bacteriol* **176:** 3738–3748.
Dorman, C.J. (2004) H-NS, a universal repressor for a dynamic genome. *Nat Rev Microbiol* **2:** 391–400.
Dorman, C.J., and Higgins, C.F. (1987) Fimbrial phase variation in *Escherichia coli*: dependence on integration host factor and homologies with other site-specific recombinases. *J Bacteriol* **169:** 3840–3843.
Dorman, C.J., Barr, G.C., Ní Bhriain, N., and Higgins, C.F. (1988) DNA supercoiling and the anaerobic and growth phase regulation of *tonB* gene expression. *J Bacteriol* **170:** 2816–2826.
Dove, S.L., and Dorman, C.J. (1994) The site-specific recombination system regulating expression of the type 1 fimbrial subunit gene of *Escherichia coli* is sensitive to changes in DNA supercoiling. *Mol Microbiol* **14:** 975–988.
Dove, S.L., and Dorman, C.J. (1996) Multicopy *fimB* gene expression in *Escherichia coli*: binding to inverted repeats *in vivo*, effect on *fimA* gene transcription and DNA inversion. *Mol Microbiol* **21:** 1161–1173.
van Duyne, G.D. (2005) Lambda integrase: armed for recombination. *Curr Biol* **15:** R658–R660.
Eisenstein, B.I. (1981) Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. *Science* **214:** 347–349.
Forterre, P., Gribaldo, S., Gadellem, D., and Serre, M.C. (2007) Origin and evolution of DNA topoisomerases. *Biochimie* **89:** 427–446.
Free, A., Porter, M.E., Deighan, P., and Dorman, C.J. (2001) Requirement for the molecular adapter function of StpA at the *Escherichia coli* bgl promoter depends upon the level of truncated H-NS proteins. *Mol Microbiol* **42:** 903–918.
Freitag, C.S., Abraham, J.M., Clements, J.R., and Eisenstein. B.I. (1985) Genetic analysis of the phase variation control of expression of type 1 fimbriae in *Escherichia coli*. *J Bacteriol* **162:** 668–675.
Hales, L.M., Gumport, R.I., and Gardner, J.F. (1994) Determining the DNA sequence elements required for binding integration host factor to two different target sites. *J Bacteriol* **176:** 2999–3006.
Hanahan, D., Jessee, J., and Bloom, F.R. (1991) Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol* **204:** 63–113.
Heddle, J.G., Barnard, F.M., Wentzell, L.M., and Maxwell, A. (2000) The interaction of drugs with DNA gyrase: a model for the molecular basis of quinolone action. *Nucleosides Nucleotides Nucleic Acids* **19:** 1249–1264.
Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., May, G., and Bremer, E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* **52:** 569–584.
Holden, N., Blomfield, I.C., Uhlin, B.E., Totsika, M., Kulasekara, D.H., and Gally, D.L. (2007) Comparative analysis of FimB and FimE recombinase activity. *Microbiology* **153:** 4138–4149.
Kawula, T.H., and Orndorff, P.E. (1991) Rapid site-specific DNA inversion in *Escherichia coli* mutants lacking the histonelike protein H-NS. *J Bacteriol* **173:** 4116–4123.
Kelly, A., Conway, C., Ó Cróinín, T., Smith, S.G.J., and Dorman, C.J. (2006) DNA supercoiling and the Lrp protein determine the directionality of *fim* switch DNA inversion in *Escherichia coli* K-12. *J Bacteriol* **188:** 5356–5363.
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.
Landgraf, J.R., Wu, J., and Calvo, J.M. (1996) Effects of nutrition and growth rate on Lrp levels in *Escherichia coli*. *J Bacteriol* **178:** 6930–6936.
Lesic, B., Bach, S., Ghigo, J.-M., Dobrindt, U., Hacker, J., and Carniel, E. (2004) Excision of the high-pathogenicity island of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. *Mol Microbiol* **52:** 1337–1348.
Lewis, J.A., and Hatfull, G.F. (2001) Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* **29:** 2205–2216.
Lewis, J.A., and Hatfull, G.F. (2003) Control of directionality in L5 integrase-mediated site-specific recombination. *J Mol Biol* **326:** 805–821.
McCusker, M.P., Turner, E.C., and Dorman, C.J. (2008) DNA sequence heterogeneity in Fim tyrosine-integrase recombinase-binding elements and functional motif asymmetries determine the directionality of the *fim* genetic switch in *Escherichia coli* K-12. *Mol Microbiol* **67:** 171–187.
McFarland, K.A., Lucchini, S., Hinton, J.C.D., and Dorman, C.J. (2008) The leucine-responsive regulatory protein, LRP, activates transcription of the *fim* operon in *Salmonella enterica* serovar Typhimurium via the *fimZ* regulatory gene. *J Bacteriol* **190:** 602–612.
Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Miller, J.H. (1992) *A Short Course in Bacterial Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
Müller, C.M., Aberg, A., Strasviçiene, J., Emody, L., Uhlin, B.E., and Balsalobre, C. (2009) Type 1 fimbriae, a colonization factor of uropathogenic *Escherichia coli*, are controlled by the metabolic sensor CRP-cAMP. *PLoS Pathog* **5:** e1000303.
O'Gara, J.P., and Dorman, C.J. (2000) Effects of local transcription and H-NS on inversion of the *fim* switch of *Escherichia coli*. *Mol Microbiol* **36:** 457–466.
Pesavento, C., and Hengge, R. (2009) Bacterial nucleotide-based second messengers. *Curr Opin Microbiol* **12:** 170–176.
Radman-Livaja, M., Biswas, T.T., Ellenberger, A., Landy, A., and Aihara, H. (2006) DNA arms do the legwork to ensure the directionality of 1 site-specific recombination. *Curr Opin Struct Biol* **16:** 42–50.

Rajeev, L., Malanowska, K., and Gardner, J.F. (2009) Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev* **73:** 300–309.

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

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

Wigley, D.B. (1995) Structure and mechanism of DNA topoisomerases. *Annu Rev Biophys Biomol Struct* **24:** 185–208.

van Workum, M., van Dooren, S.J., Oldenburg, N., Molenaar, D., Jensen, P.R., Snoep, J.L., and van Westerhoff, H. (1996) DNA supercoiling depends on the phosphorylation potential in *Escherichia coli. Mol Microbiol* **20:** 351–360.

Wright, K.J., Seed, P.C., and Hultgren, S.J. (2007) Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell Microbiol* **9:** 2230–2241.

Yang, S.W., and Nash, H.A. (1995) Comparison of protein binding to DNA *in vivo* and *in vitro*: defining an effective intracellular target. *EMBO J* **14:** 6292–6300.

Yokoyama, K., Ishijima, S.A., Clowney, L., Koike, H., Aramaki, H., Tanaka, C., *et al.* (2006) Feast/famine regulatory proteins (FFRPs): *Escherichia coli* Lrp, AsnC and related archaeal transcription factors. *FEMS Microbiol Rev* **30:** 89–108.

Appendix 2: Dorman C.J. and Corcoran C.P. (2009). Bacterial DNA topology and infectious disease. *Nucleic Acids Research* 37: 672-678.

672–678 Nucleic Acids Research, 2009, Vol. 37, No. 3 Published online 10 December 2008
doi:10.1093/nar/gkn996

SURVEY AND SUMMARY

Bacterial DNA topology and infectious disease

Charles J. Dorman* and Colin P. Corcoran

Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College, Dublin 2, Ireland

Received October 23, 2008; Revised November 20, 2008; Accepted November 25, 2008

ABSTRACT

The Gram-negative bacterium *Escherichia coli* and its close relative *Salmonella enterica* have made important contributions historically to our understanding of how bacteria control DNA supercoiling and of how supercoiling influences gene expression and vice versa. Now they are contributing again by providing examples where changes in DNA supercoiling affect the expression of virulence traits that are important for infectious disease. Available examples encompass both the earliest stages of pathogen–host interactions and the more intimate relationships in which the bacteria invade and proliferate within host cells. A key insight concerns the link between the physiological state of the bacterium and the activity of DNA gyrase, with downstream effects on the expression of genes with promoters that sense changes in DNA supercoiling. Thus the expression of virulence traits by a pathogen can be interpreted partly as a response to its own changing physiology. Knowledge of the molecular connections between physiology, DNA topology and gene expression offers new opportunities to fight infection.

INTRODUCTION

DNA gyrase was discovered in *Escherichia coli*, a bacterium that has played an important part in the foundation of modern molecular biology (1). DNA topoisomerase I was also discovered in *E. coli* (2), but the gene that encodes it, *topA*, was first identified as a suppressor of the *leu500* promoter mutation in *Salmonella enterica*, then called *S. typhimurium* (3). The genes that encode gyrase, *gyrA* and *gyrB*, have the interesting property of being up-regulated when DNA relaxes (4). In contrast, the *topA* gene is transcriptionally activated when DNA becomes more negatively supercoiled (5–7). This latter response is intuitively appealing: a promoter must open for transcription to begin and the energy of negative supercoiling can be used to bring about the necessary breakage of the hydrogen bonds between the paired bases (8). The molecular mechanism responsible for the DNA-relaxation-dependent activation of *gyrA* and *gyrB* has yet to be fully explained (9).

Reciprocal regulation of the transcription of the *topA* gene and the *gyrA* and *gyrB* genes by DNA negative supercoiling and relaxation, respectively, is consistent with the maintenance of a homeostatic balance of DNA supercoiling that benefits the cell (10–13). As DNA becomes more negatively supercoiled expression of the *topA* gene is enhanced, leading to a higher level of DNA topoisomerase I, a DNA relaxing enzyme. DNA relaxation has the opposite effect because it enhances the transcription of the genes coding for DNA gyrase which can then correct the supercoiled-relaxed balance to a value in keeping with the physiological needs of the cell (10–13).

This simple picture of topoisomerase gene regulation neglects a number of additional influences. For example, the Fis protein is a regulator of *topA*, *gyrA* and *gyrB* (14–16). This protein is the factor for inversion stimulation, hence the name 'Fis'. It was discovered originally as an important co-factor in the operation of invertible DNA switches that are catalyzed by members of the serine invertase family of site-specific recombinases (17). Fis is now known to play many regulatory roles in the cell, affecting the operation of several important DNA transactions such as bacteriophage integration and excision, expression of components of the translation machinery, DNA replication, and transposition (17,18). Fis represses the expression of its own gene, *fis*, and it has a highly characteristic expression pattern. The Fis protein is expressed to its maximum level in the early stage of exponential growth. Its intracellular concentration declines sharply thereafter and it is almost undetectable when the bacterial culture approaches the stationary phase of growth (19,20). This suggests that there is a window within which Fis-dependent molecular events can occur optimally. However, a straightforward correlation

*To whom correspondence should be addressed. Tel: +353 1 896 2013; Fax: +353 1 679 9294; Email: cjdorman@tcd.ie

between Fis concentration, growth phase and the operation of Fis-dependent systems has been difficult to obtain. The picture is made complicated by the fact that Fis is not essential for any of the processes to which it contributes and by the fact that the classic pattern of Fis protein expression can be overridden by the manipulation of growth conditions (21).

The Fis protein represses the transcription of both *gyrA* and *gyrB* and has a bi-functional relationship with *topA*: at high concentrations, Fis represses *topA* transcription and at low concentrations Fis is an activator (14–16). Fis regulates transcription positively by acting both as a conventional transcription factor that makes protein–protein contact with RNA polymerase and by creating a micro-domain of negatively supercoiled DNA in the vicinity of the target promoter (22–25). Like the genes coding for the main topoisomerases, the *fis* gene is regulated by changes in DNA supercoiling: increased negative DNA supercoiling stimulates the *fis* promoter (26). The contribution of Fis to global regulation of DNA transactions through changes in DNA supercoiling is best appreciated in the context of the impact of growth phase on DNA superhelicity. DNA is more negatively supercoiled in bacteria that are growing exponentially than in those where growth has slowed or ceased (27). Fis is thought to play a valuable role in offsetting the negative effects of DNA that is too relaxed or too negatively supercoiled by acting as a topological buffer. It creates micro-domains of DNA where the degree of DNA supercoiling is optimal for promoter function and preserves the integrity of these micro-domains regardless of changes to global supercoiling levels (28). It can perform this role throughout the genome because its DNA sequence requirements for DNA binding are non-stringent (29). Thus Fis acts in intimate association with gyrase and DNA topoisomerase I to set and reset DNA supercoiling levels in the cell.

The Fis protein is classified as a nucleoid-associated protein (NAP) and it is one of a number that belong to this group. Other abundant NAPs are HU and H-NS, two proteins that have the ability to constrain DNA supercoils (18). It is estimated that, during logarithmic growth, about half of the DNA in the bacterium is complexed with protein in ways that constrain negative supercoils (30,31). Thus the *effective* level of supercoiling, the portion that is available to do work in the cell and influence processes such as transcription, is only ~50% of the total detected when DNA superhelicity is measured with nucleic acid purified free of cellular components (32).

The physiological state of the cell is strongly influenced by the environment external to the bacterium (6,8). As the chemical and physical nature of the environment changes, the metabolic pathways of the microbe respond. DNA gyrase is intimately connected to these pathways by virtue of being an enzyme that requires ATP as an energy source and one that is inhibited by ADP: the ratio of the concentrations of ATP and ADP determines the level of gyrase activity (33,34). For this reason, shocks to the cell such as changes in osmolarity, temperature, pH, oxygen level, nutrient supply, etc. all potentially have an impact ultimately on the global level of DNA supercoiling (35–41). This is especially relevant in the cases of bacteria such as *E. coli* or *S. enterica* that can inhabit a wide range of environments. Thus, DNA supercoiling can be seen as a crude regulator of gene expression. It is variable in response to environmental signals and it has the potential to act widely within the genome (6,8). This leads to a model of global regulation in which the environment alters chromosome topology via topoisomerases and genes have evolved to respond to those environmentally determined changes (6). In addition to the influences of DNA supercoiling, further regulatory refinements are imposed by the multitude of locally acting transcription factors that are possessed by bacteria such as *E. coli* (42).

Pathogenic bacteria possess virulence genes that their commensal counterparts lack completely or they express virulence genes that are inactive due to mutation or cryptic in the commensal. The evolution of bacterial pathogens has involved the lateral transfer of virulence genes and their integration into the regulatory regime of the bacterium (43,44). Studies in a number of pathogens have provided evidence that the expression of many virulence genes is influenced by changes in DNA supercoiling (45–48). Given the impressive correspondence between the environmental stresses that pathogens must endure during infection, and the known impact of these stresses on the degree of DNA supercoiling in bacteria, this is perhaps unsurprising.

The infection process may be regarded as a series of relationships between the pathogen and the host of ever-deepening intimacy. Preliminary contact often involves attachment to the host by bacterial surface structures called fimbriae. The genes that encode these are often subject to complex regulation that includes a role for DNA supercoiling.

THE *fim* GENETIC SWITCH IN *E. coli* K-12

Type 1 fimbriae are important virulence factors in many bacterial species (49). They are expressed by most members of the *Enterobacteriaceae* and were the first bacterial fimbriae to be described (50,51). Type 1 fimbriae attach bacteria to mannosylated glycoproteins on a variety of eukaryotic cells. In *E. coli* K-12, these fimbriae are expressed phase-variably with bacterial populations containing fimbriate (phase-ON) and afimbriate (phase-OFF) members (Figure 1). Moreover, the two cell types are interchangeable. This is because the transcriptional promoter for the *fim* structural genes is part of an invertible DNA segment known as the *fim* switch, *fimS* (52). This 314 bp DNA segment is bounded by 9 bp perfect inverted repeats within which DNA cleavage and religation occur during the site-specific recombination reactions that invert the switch (53). Inversion is catalyzed by two tyrosine integrase site-specific recombinases that act independently and have distinct activities. FimB inverts the switch in both the ON-to-OFF and the OFF-to-ON directions with approximately equal efficiency and does this at a frequency of about 10^{-2} per cell per generation (54,55). The FimE protein has a marked preference for inverting the switch in the ON-to-OFF direction and its activity is dominant to the OFF-to-ON activity of FimB (53,55).

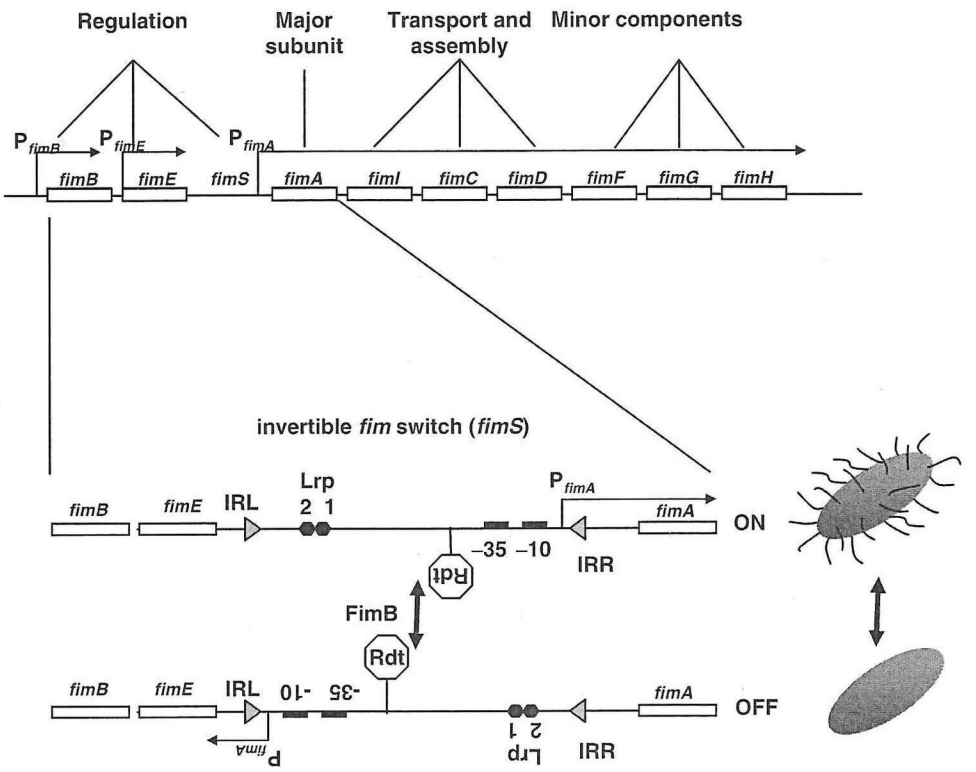


Figure 1. The invertible genetic switch in the *fim* operon of *E. coli*. The structure of the complete *fim* operon is summarized at the top of the figure. The positions and directions of transcription of each of the nine genes are shown, together with their functions. The positions of the transcription start sites associated with the three main promoters are represented by angled arrows. The invertible genetic element, *fimS*, that harbours the promoters for transcription of the structural genes, is shown in an expanded form in the centre and bottom of the figure. In the ON orientation, the P_{fimA} promoter is directed towards the *fimA* gene, resulting in an ON phenotype and a fimbriate bacterium. In the OFF orientation the *fimS* element has inverted and the P_{fimA} promoter has been disconnected from the *fimA* gene. This results in an OFF phenotype and an afimbriate bacterium. The 314 bp *fimS* element is bounded by 9 bp inverted repeats (grey arrowheads labelled IRL and IRR) that encompass the P_{fimA} promoter −10 and −35 boxes and transcription start site shown) and a Rho-dependent transcription terminator, Rdt. This terminator reduces the length and stability of the *fimE* transcript only when *fimS* is in the OFF orientation. In the ON orientation, the *fimE* gene reads across the switch into the *fimA* gene (56,57).

Many laboratory strains of *E. coli* K-12 lack an active *fimE* gene and invert the switch using FimB alone (54). Posttranscriptional control of *fimE* gene expression plays a key role in controlling *fimS* inversion in the complete wild-type *fim* operon. This is because the *fimS* element harbours a Rho-dependent terminator in addition to the promoter for *fim* structural gene transcription (56,57) (Figure 1).

Although the FimB integrase inverts the switch in a relatively unbiased manner, its activity becomes strongly biased in favour of the ON phase when DNA gyrase is inhibited (58). Inhibition of gyrase activity with the antibiotic novobiocin results in a clear dose-dependent preference for the ON orientation of *fimS* (58). This is not explained by changes in the expression of the *fimB* gene but is related to the quality of the FimB substrate. If the *topA* gene is inactivated by transposon insertion, the switch ceases to be invertible. It maintains thereafter the switch orientation (ON or OFF) that obtained at the moment that the *topA* gene was mutated. Again, this is not due to changes in the expression of the *fimB* gene or to global changes in DNA supercoiling. Instead it is due to a requirement for topoisomerase I activity in the immediate vicinity of the switch (58).

The simplest interpretation of the experimental data is that the switch becomes trapped in the ON orientation because this form of the switch is a poor substrate for FimB. This is not due to the creation of differentially supercoiled domains by the activity of the P_{fimA} promoter that might distinguish phase-ON from phase-OFF switches; complete inactivation of this promoter has no influence on switch biasing in the wake of DNA relaxation (59). Instead the trap is composed of a nucleoprotein complex that involves the left inverted repeat, two binding sites for the leucine-responsive regulatory protein within *fimS* and a reference site in the flanking, invariant DNA (Figure 1). Removal of the Lrp protein or abrogation of Lrp binding to the switch eliminates the OFF-to-ON bias that accompanies DNA relaxation; in fact, the switch now acquires a strong bias in the ON-to-OFF direction (59).

What is the physiological significance of inversion-biasing? DNA relaxation accompanies cessation of growth and a shift in the [ATP]/[ADP] ratio that that is unfavourable for DNA gyrase activity (27,33,34). In addition, the Lrp protein is a barometer of the metabolic status of the cell and an indicator of nutrient depletion (60). It is tempting to speculate that by evolving sensitivities to these factors, the cell has developed a mechanism to override the

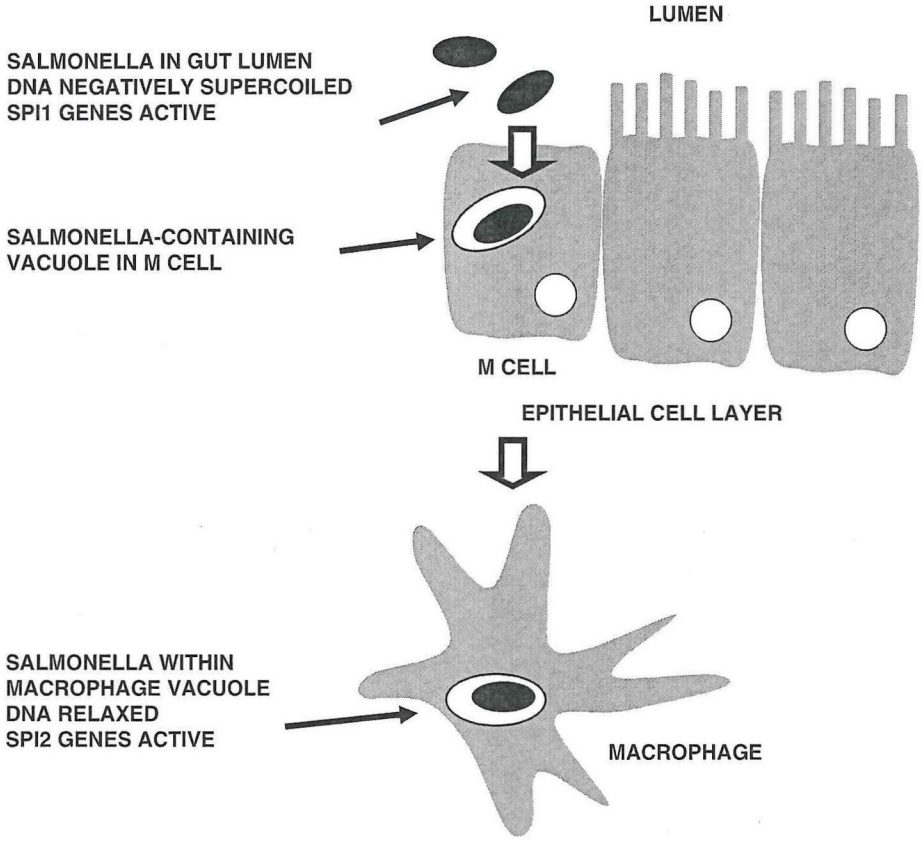


Figure 2. The intracellular life of *S. enterica* in the mammalian gut. A summary of the main steps in invasive disease in the murine gut caused by *S. enterica* serovar Typhimurium is presented. In the lumen of the gut the bacterium experiences environmental stresses that are known to result in a reduction in the linking number of its DNA. This increase in negative DNA supercoiling is part of the mechanism by which the SPI1 pathogenicity island genes are up-regulated. The bacteria traverse the antigen-sampling M cells in a *Salmonella*-containing vacuole. Following release on the basolateral surface, the bacterium may be engulfed by macrophage. *S.* Typhimurium undergoes DNA relaxation within the macrophage and this is part of the mechanism by which the SPI2 pathogenicity genes are activated. The products of these genes prevent the macrophage from killing the microbe, which is then able to establish a systemic disease.

stochastic DNA inversion behaviour of FimB in favour of a fimbriate phenotype. This may enhance the ability of the bacterium to participate in biofilm formation as a means to ride out physiologically unfavourable circumstances.

Type 1 fimbriae do not contribute exclusively to early phases of the host–pathogen interaction: they have been identified as important factors in the establishment of more intimate associations with the host during urinary tract infection by uropathogenic *E. coli* (61) and *Klebsiella pneumoniae* (62). Here, the fimbriae are expressed within bacterial communities living within epithelial cells of the bladder lining. The invertible *fim* switch in these bacteria is maintained in the ON phase, showing that DNA inversion in the ON-to-OFF direction is suppressed in this niche (63).

THE INTRACELLULAR LIFE OF *S. enterica*

Like *E. coli* K-12, *S. enterica* serovar Typhimurium (*S.* Typhimurium) uses type 1 fimbriae to interact with its host, although it controls their expression through mechanisms that are independent of DNA inversion (64). Unlike *E. coli* K-12, *S.* Typhimurium has the ability to invade mammalian epithelial cells and to survive engulfment by macrophage (Fig. 2). This is due to its possession of two separate type III secretion systems (TTSS) with separate sets of effector proteins that *S.* Typhimurium can use to modify the mammalian cells to its advantage (65–67). The TTSS that is encoded by the genes of the SPI1 pathogenicity island confer an invasive phenotype on the bacterium. The promoters of the SPI1 genes are up-regulated by negative DNA supercoiling (68). In this respect they resemble the TTSS genes of the dysentery bacillus *Shigella flexneri* (69). The TTSS that is encoded by the SPI2 pathogenicity island of *S.* Typhimurium is essential for the survival of the bacterium in the otherwise hostile environment of the macrophage. The effector proteins secreted via the SPI2 TTSS prevent phagolysosome fusion through modification of the macrophage vacuole that contains the engulfed bacterium (68). Interestingly, the promoters of the genes in the SPI2 island are up-regulated by DNA relaxation (70), which is the opposite of the SPI1 genes. This differential dependency on the state of DNA topology is likely to represent a key distinguishing factor between these two sets of virulence genes that ensures that each is active in the correct environment and repressed elsewhere. The lumen of the

mammalian gut exposes the bacterium to a range of stresses that have been shown to shift DNA supercoiling to more negative values (48). Indeed the recommended growth conditions for the induction of SPI1 genes in the laboratory involve low aeration and growth in a high-osmolarity medium (71). In contrast, SPI2 gene activation is favoured by a low-osmolarity growth regime (72). Measurements of plasmid topoisomer distributions have shown that bacterial DNA becomes more relaxed during growth of *S.* Typhimurium in the vacuole of cultured macrophage, which is consistent with SPI2 gene upregulation (70). Both SPI1 and SPI2 have a requirement for the Fis protein for optimal gene expression (20,70). This is in keeping with the role of Fis as a topological buffer (28). Fis is just one of the NAPs that has been shown to influence transcription within the major pathogenicity islands of *S.* Typhimurium. Like Fis, the HU NAP has a positive influence on SPI1 gene expression (73). In contrast, the H-NS protein represses the transcription of the genes of both SPI1 and SPI2 and it is assisted in this process by the Hha protein, a partial paralogue of H-NS (74).

The SPI1 and SPI2 pathogenicity islands possess genes coding for dedicated regulators of their own structural virulence genes (75). These operate in a regulatory environment in which DNA supercoiling and the NAPs set the regulatory background, in tune with signals coming from the external environment that modulate the metabolism of the bacterium.

CONCLUSIONS

DNA supercoiling has been identified as a factor that modulates the expression of virulence genes in pathogenic bacteria at different phases of the host–pathogen relationship. This is by no means confined to the four Gram-negative pathogens discussed above; DNA supercoiling has been identified as an important factor influencing gene expression in many other bacteria (45–48). It should also be emphasized that these effects on gene expression are not relevant only to pathogens but are also involved in the physiology of bacteria pursuing commensal or symbiotic lifestyles. The model that best describes the role of DNA supercoiling in bacterial gene regulation is one that takes a hierarchical view of the gene regulatory network of the cell. DNA supercoiling has a place at or near the apex of the hierarchy due to its potential to influence the activities of so many promoters simultaneously. The NAPs also have a high position in the hierarchy, but below that occupied by DNA supercoiling. Their widespread influences on transcription arise because each governs a large regulon of genes and the memberships of the different regulons overlap in ways that are conditional on environmental conditions. This form of flexible networking provides a backdrop for the activities of the conventional transcription factors, DNA binding proteins that regulate few, or possibly just one, promoters. As we come to appreciate the subtle sophistication of bacterial gene regulation and the complexity of its networks the task of intervening in infection by targeting the gene control programmes of the pathogen can indeed appear daunting. It is to be hoped that our ever-deepening knowledge of how bacteria manage their physiology at the level of gene expression will improve our position in this struggle.

FUNDING

This work was supported by Science Foundation Ireland. Funding for open access charges: Science Foundation Ireland.

Conflict of interest statement. None declared.

REFERENCES

1. Gellert,M., Mizuuchi,K., O'Dea,M.H. and Nash,H.A. (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc. Natl Acad. Sci. USA*, **73**, 3872–3876.
2. Wang,J.C. (1971) Interaction between DNA and an *Escherichia coli* protein omega. *J. Mol. Biol.*, **55**, 523–533.
3. Margolin,P., Zumstein,L., Sternglanz,R. and Wang,J.C. (1985) The *Escherichia coli supX* locus is *topA*, the structural gene for DNA topoisomerase I. *Proc. Natl Acad. Sci. USA*, **82**, 5437–5441.
4. Menzel,R. and Gellert,M. (1987) Fusions of the *Escherichia coli gyrA* and *gyrB* control regions to the galactokinase gene are inducible by coumermycin treatment. *J. Bacteriol.*, **169**, 1272–1278.
5. Pruss,G.J. and Drlica,K. (1989) DNA supercoiling and prokaryotic transcription. *Cell*, **56**, 521–523.
6. Dorman,C.J. (2006) DNA supercoiling and bacterial gene expression. *Sci. Prog.*, **89**, 151–166.
7. Blot,N., Mavathur,R., Geertz,M., Travers,A. and Muskhelishvili,G. (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Rep.*, **7**, 710–715.
8. Dorman,C.J. (2008) DNA supercoiling and bacterial gene regulation. In El-Sharoud,W.M. (ed.), *Microbial Physiology – A Molecular Approach*, Springer, Berlin, pp. 155–178.
9. Menzel,R. and Gellert,M. (1987) Modulation of transcription by DNA supercoiling: a deletion analysis of the *Escherichia coli gyrA* and *gyrB* promoters. *Proc. Natl Acad. Sci. USA*, **84**, 4185–4189.
10. DiNardo,S., Voelkel,K.A., Sternglanz,R., Reynolds,A.E. and Wright,A. (1982) *Escherichia coli* DNA topoisomerase I mutants have compensatory mutations in DNA gyrase genes. *Cell*, **31**, 43–51.
11. Pruss,G.J., Manes,S.H. and Drlica,K. (1982) *Escherichia coli* DNA topoisomerase I mutants: increased supercoiling is corrected by mutations near gyrase genes. *Cell*, **31**, 35–42.
12. Menzel,R. and Gellert,M. (1983) Regulation of the genes for *E. coli* DNA gyrase: homeostatic control of DNA supercoiling. *Cell*, **34**, 105–113.
13. Pruss,G.J., Franco,R.J., Chevalier,S.G., Manes,S.H. and Drlica,K. (1986) Effects of DNA gyrase inhibitors in *Escherichia coli* topoisomerase I mutants. *J. Bacteriol.*, **168**, 276–282.
14. Weinstein-Fischer,D. and Altuvia,S. (2007) Differential regulation of *Escherichia coli* topoisomerase I by Fis. *Mol. Microbiol.*, **63**, 1131–1144.
15. Schneider,R., Travers,A., Kutateladze,T. and Muskhelishvili,G. (1999) A DNA architectural protein couples cellular physiology and DNA topology in *Escherichia coli*. *Mol. Microbiol.*, **34**, 953–964.
16. Keane,O.M. and Dorman,C.J. (2003) The *gyr* genes of *Salmonella enterica* serovar Typhimurium are repressed by the factor for inversion stimulation, Fis. *Mol. Genet. Genomics*, **270**, 56–65.
17. Finkel,S.E. and Johnson,R.C. (1992) The Fis protein: it's not just for DNA inversion anymore. *Mol. Microbiol.*, **6**, 3257–3265.
18. Dorman,C.J. and Deighan,P. (2003) Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev.*, **13**, 179–184.
19. Bradley,M.D., Beach,M.B., de Koning,A.P., Pratt,T.S. and Osuna,R. (2007) Effects of Fis on *Escherichia coli* gene expression during different growth stages. *Microbiol.*, **153**, 2922–2940.
20. Kelly,A., Goldberg,M.D., Carroll,R.K., Danino,V., Hinton,J.C.D. and Dorman,C.J. (2004) A global role for Fis in the transcriptional

control of metabolic and type III secretion genes of *Salmonella enterica* serovar Typhimurium. *Microbiology*, **150**, 2037–2053.
21. Ó Cróinín,T. and Dorman,C.J. (2007) Expression of the Fis protein is sustained in late exponential and stationary phase cultures of *Salmonella enterica* serovar Typhimurium grown in the absence of aeration. *Mol. Microbiol.*, **66**, 237–251.
22. McLeod,S.M., Aiyar,S.E., Gourse,R.L. and Johnson,R.C. (2002) The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the *Escherichia coli proP* P2 promoter. *J. Mol. Biol.*, **316**, 517–529.
23. Schneider,D.A., Ross,W. and Gourse,R.L. (2003) Control of rRNA expression in *Escherichia coli*. *Curr. Opin. Microbiol.*, **6**, 151–156.
24. Auner,H., Buckle,M., Deufel,A., Kutateladze,T., Lazarus,L., Mavathur,R., Muskhelishvili,G., Pemberton,I., Schneider,R. and Travers,A. (2003) Mechanism of transcriptional activation by FIS: role of core promoter structure and DNA topology. *J. Mol. Biol.*, **331**, 331–344.
25. Travers,A., Schneider,R. and Muskhelishvili,G. (2001) DNA supercoiling and transcription in *Escherichia coli*: The FIS connection. *Biochimie*, **83**, 213–217.
26. Schneider,R., Travers,A. and Muskhelishvili,G. (2000) The expression of the *Escherichia coli fis* gene is strongly dependent on the superhelical density of DNA. *Mol. Microbiol.*, **38**, 167–175.
27. Bordes,P., Conter,A., Morales,V., Bouvier,J., Kolb,A. and Gutierrez,C. (2003) DNA supercoiling contributes to disconnect sigmaS accumulation from sigmaS-dependent transcription in *Escherichia coli*. *Mol. Microbiol.*, **48**, 561–571.
28. Rochman,M., Aviv,M., Glaser,G. and Muskhelishvili,G. (2002) Promoter protection by a transcription factor acting as a local topological homeostat. *EMBO Rep.*, **3**, 355–360.
29. Shao,Y., Feldman-Cohen,L.S. and Osuna,R. (2008) Biochemical identification of base and phosphate contacts between Fis and a high-affinity DNA binding site. *J. Mol. Biol.*, **380**, 327–339.
30. Bliska,J.B. and Cozzarelli,N.R. (1987) Use of site-specific recombination as a probe of DNA structure and metabolism *in vivo*. *J. Mol. Biol.*, **194**, 205–218.
31. Sinden,R.R. (1994) *DNA Structure and Function*, Academic Press, San Diego, pp. 343–345.
32. Travers,A. and Muskhelishvili,G. (2005) DNA supercoiling – a global transcriptional regulator for enterobacterial growth? *Nat. Rev. Microbiol.*, **3**, 157–169.
33. Westerhoff,H.V., O'Dea,M.H., Maxwell,A. and Gellert,M. (1988) DNA supercoiling by DNA gyrase: A static head analysis. *Cell Biophys.*, **12**, 157–181.
34. Snoep,J.L., van der Weijden,C.C., Andersen,H.W., Westerhoff,H.V. and Jensen,P.R. (2002) DNA supercoiling in *Escherichia coli* is under tight and subtle homeostatic control, involving gene-expression and metabolic regulation of both topoisomerase I and DNA gyrase. *Eur. J. Biochem.*, **269**, 1662–1669.
35. Higgins,C.F., Dorman,C.J., Stirling,D.A., Waddell,L., Booth,I.R., May,G. and Bremer,E. (1988) A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S.* Typhimurium and *E. coli*. *Cell*, **52**, 569–584.
36. Dorman,C.J., Barr,G.C., Ní Bhriain,N. and Higgins,C.F. (1988) DNA supercoiling and the anaerobic and growth phase regulation of *tonB* gene expression. *J. Bacteriol.*, **170**, 2816–2826.
37. Yamamoto,N. and Droffner,M.L. (1985) Mechanisms determining aerobic or anaerobic growth in the facultative anaerobe *Salmonella typhimurium*. *Proc. Natl Acad. Sci. USA*, **82**, 2077–2081.
38. Hsieh,L.S., Burger,R.M. and Drlica,K. (1991) Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *J. Mol. Biol.*, **19**, 443–450.
39. Hsieh,L.S., Rouvière-Yaniv,J. and Drlica,K. (1991) Bacterial DNA supercoiling and [ATP]/[ADP] ratio: changes associated with salt shock. *J. Bacteriol.*, **173**, 3914–3917.
40. Huo,Y.X., Rosenthal,A.Z. and Gralla,J.D. (2008) General stress response signalling: unwrapping transcription complexes by DNA relaxation via the sigma38 C-terminal domain. *Mol. Microbiol.*, **70**, 369–378.
41. Mojica,F.J., Charbonnier,F., Juez,G., Rodríguez-Valera,F. and Forterre,P. (1994) Effects of salt and temperature on plasmid topology in the halophilic archaeon *Haloferax volcanii*. *J. Bacteriol.*, **176**, 4966–4973.
42. Martínez-Antonio,A. and Collado-Vides,J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.*, **6**, 482–489.
43. Dobrindt,U., Hochhut,B., Hentschel,U. and Hacker,J. (2004) Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.*, **2**, 414–424.
44. Doyle,M., Fookes,M., Ivens,A., Mangan,M.W., Wain,J. and Dorman,C.J. (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science*, **315**, 251–252.
45. Ye,F., Brauer,T., Niehus,E., Drlica,K., Josenhans,C. and Suerbaum,S. (2007) Flagellar and global gene regulation in *Helicobacter pylori* modulated by changes in DNA supercoiling. *Int. J. Med. Microbiol.*, **297**, 65–81.
46. Fournier,B. and Klier,A. (2004) Protein A gene expression is regulated by DNA supercoiling which is modified by the ArlS-ArlR two-component system of *Staphylococcus aureus*. *Microbiology*, **150**, 3807–3819.
47. Rohde,J.R., Luan,X.S., Rohde,H., Fox,J.M. and Minnich,S.A. (1999) The *Yersinia enterocolitica* pYV virulence plasmid contains multiple intrinsic DNA bends which melt at 37 degrees C. *J. Bacteriol.*, **181**, 4198–4204.
48. Dorman,C.J. (1991) DNA supercoiling and environmental regulation of gene expression in pathogenic bacteria. *Infect. Immun.*, **59**, 745–749.
49. Thankavel,S.M., Shah,A.H., Cohen,M.S., Ikeda,T., Lorenz,R.G., Curtiss III,R. and Abraham,S.N. (1999) Molecular basis of the erythrocyte tropism exhibited by *Salmonella typhimurium* type-1 fimbriae. *J. Biol. Chem.*, **274**, 5797–5809.
50. Duguid,J.P., Smith,I.W., Dempster,G. and Edmunds,P.N. (1955) Non-flagellar filamentous appendices (fimbriae) and haemagglutinating activity in *Bacterium coli*. *J. Pathol. Bacteriol.*, **70**, 335–348.
51. Brinton,C.C. (1959) Non-flagellar appendices of bacteria. *Nature*, **183**, 782–786.
52. Abraham,J.M., Freitag,C.S., Clements,J.R. and Eisenstein,B.I. (1985) An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **82**, 5724–5727.
53. McCusker,M.P., Turner,E.C. and Dorman,C.J. (2008) DNA sequence heterogeneity in Fim tyrosine-integrase recombinase-binding elements and functional motif asymmetries determine the directionality of the *fim* genetic switch in *Escherichia coli* K-12. *Mol. Microbiol.*, **67**, 171–178.
54. Blomfield,I.C., McClain,M.S., Princ,J.A., Calie,P.J. and Eisenstein,B.I. (1991) Type 1 fimbriation and *fimE* mutants of *Escherichia coli* K-12. *J. Bacteriol.*, **173**, 5298–5230.
55. McClain,M.S., Blomfield,I.C. and Eisenstein,B.I. (1991) Roles of *fimB* and *fimE* in site-specific DNA inversion associated with phase variation of type 1 fimbriae in *Escherichia coli*. *J. Bacteriol.*, **173**, 5308–5314.
56. Joyce,S.A. and Dorman,C.J. (2002) A Rho-dependent phase-variable transcription terminator controls expression of the FimE recombinase in *Escherichia coli*. *Mol. Microbiol.*, **45**, 1107–1117.
57. Hinde,P., Deighan,P. and Dorman,C.J. (2005) Characterization of the detachable Rho-dependent transcription terminator of the fimE gene in Escherichia coli K-12. *J. Bacteriol.*, **187**, 8256–8266.
58. Dove,S.L. and Dorman,C.J. (1994) The site-specific recombination system regulating expression of the type 1 fimbrial subunit gene of Escherichia coli is sensitive to changes in DNA supercoiling. *Mol. Microbiol.*, **14**, 975–988.
59. Kelly,A., Conway,C., O Cróinín,T., Smith,S.G. and Dorman,C.J. (2006) DNA supercoiling and the Lrp protein determine the directionality of fim switch DNA inversion in Escherichia coli K-12. *J. Bacteriol.*, **188**, 5356–5363.
60. Brinkman,A.B., Ettema,T.J.G., de Vos,W.M. and van der Oost,J. (2003) The Lrp family of transcriptional regulators. *Mol. Microbiol.*, **48**, 287–294.
61. Wright,K.J., Seed,P.C. and Hultgren,S.J. (2007) Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cell. Microbiol.*, **9**, 2230–2241.
62. Rosen,D.A., Pinkner,J.S., Jones,J.M., Walker,J.N., Clegg,S. and Hultgren,S.J. (2008) Utilization of an intracellular bacterial community pathway in *Klebsiella pneumoniae* urinary tract infection and the effects of FimK on type 1 pilus expression. *Infect Immun.*, **76**, 3337–3345.

63. Hannan,T.J., Mysorekar,I.U., Chen,S.L., Walker,J.N., Jones,J.M., Pinkner,J.S., Hultgren,S.J. and Seed,P.C. (2008) LeuX tRNA-dependent and -independent mechanisms of *Escherichia coli* pathogenesis in acute cystitis. *Mol. Microbiol.*, **67**, 116–128.
64. McFarland,K.A., Lucchini,S., Hinton,J.C. and Dorman,C.J. (2008) The leucine-responsive regulatory protein, Lrp, activates transcription of the *fim* operon in *Salmonella enterica* serovar typhimurium via the *fimZ* regulatory gene. *J. Bacteriol.*, **190**, 602–612.
65. Groisman,E.A. and Ochman,H. (1997) . How *Salmonella* became a pathogen. *Trends Microbiol.*, **5**, 343–349.
66. Hensel,M. (2002) *Salmonella* pathogenicity island 2. *Mol. Microbiol.*, **36**, 1015–1023.
67. Galán,J.E. (2001) *Salmonella* interactions with host cells: type III secretion at work. *Annu. Rev. Cell. Dev. Biol.*, **17**, 53–86.
68. Galán,J.E. and Curtiss,R, III (1990) Expression of *Salmonella typhimurium* genes required for invasion is regulated by changes in DNA supercoiling. *Infect. Immun.*, **58**, 1879–1885.
69. Dorman,C.J. and Porter,M.E. (1998) The *Shigella* virulence gene regulatory cascade: a paradigm of bacterial gene control mechanisms. *Mol. Microbiol.*, **29**, 677–684.
70. Ó Cróinín,T., Carroll,R.K., Kelly,A. and Dorman,C.J. (2006) Roles for DNA supercoiling and the Fis protein in modulating expression of virulence genes during intracellular growth of *Salmonella enterica* serovar Typhimurium. *Mol. Microbiol.*, **62**, 869–882.
71. Lee,C.A. and Falkow,S. (1990) The ability of *Salmonella* to enter mammalian cells is affected by bacterial growth state. *Proc. Natl Acad. Sci. USA*, **87**, 4304–4308.
72. Garmendia,J., Beuzón,C.R., Ruiz-Albert,J. and Holden,D.W. (2003) The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the *Salmonella typhimurium* SPI-2 type III secretion system. *Microbiol.*, **149**, 2385–2396.
73. Schechter,L.M., Jain,S., Akbar,S. and Lee,C.A. (2003) The small nucleoid-binding proteins H-NS, HU, and Fis affect *hilA* expression in *Salmonella enterica* serovar Typhimurium. *Infect. Immun.*, **71**, 5432–5435.
74. Vivero,A., Baños,R.C., Mariscotti,J.F., Oliveros,J.C., García-del Portillo,F., Juárez,A. and Madrid,C. (2008) Modulation of horizontally acquired genes by the Hha-YdgT proteins in *Salmonella enterica* serovar Typhimurium. *J. Bacteriol.*, **190**, 1152–1156.
75. Rhen,M. and Dorman,C.J. (2005) Hierarchical gene regulators adapt *Salmonella enterica* to its host milieus. *Int. J. Med. Microbiol.*, **294**, 487–502.

Appendix 3: Corcoran C.P.. Cameron A.D.S and Dorman C.J. (2010). H-NS silences *gfp*, the Green Fluorescent Protein gene: *gfp*^{TCD} is a genetically remastered *gfp* gene with reduced susceptibility to H-NS-mediated transcription silencing and with enhanced translation. *Journal of Bacteriology* 192: 4790-4793.

JOURNAL OF BACTERIOLOGY, Sept. 2010, p. 4790–4793 Vol. 192, No. 18
0021-9193/10/$12.00 doi:10.1128/JB.00531-10

H-NS Silences *gfp*, the Green Fluorescent Protein Gene: *gfp*^{TCD} Is a Genetically Remastered *gfp* Gene with Reduced Susceptibility to H-NS-Mediated Transcription Silencing and with Enhanced Translation$^{\nabla}$§

Colin P. Corcoran, Andrew D. S. Cameron, and Charles J. Dorman*

Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin 2, Ireland

Received 10 May 2010/Accepted 24 June 2010

The bacterial nucleoid-associated protein H-NS, which preferentially targets and silences A+T-rich genes, binds the ubiquitous reporter gene *gfp* and dramatically reduces local transcription. We have redesigned *gfp* to reduce H-NS-mediated transcription silencing and simultaneously improve translation *in vivo* without altering the amino acid sequence of the GFP protein.

The green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* is widely used as a reporter of transcriptional activity and protein localization in both eukaryotes and prokaryotes (reviewed in reference 16). The DNA encoding *gfp* and many of its variants has a high A+T content (~60%) and strong predicted DNA curvature, both of which are key features of DNA that is targeted by the histone-like nucleoid structuring protein (H-NS) (10, 12, 14, 15, 22). H-NS is a global gene silencer that represses transcription by condensing chromosome and plasmid DNA in many Gram-negative bacteria. For example, H-NS binds and represses ~15% of the *Escherichia coli* and *Salmonella enterica* genomes (13), including A+T-rich genes that have been acquired horizontally (5). This raised the possiblity that H-NS may bind and repress the alien *gfp* gene and thus obscure measurements of promoter activity.

Many variants of GFP have been developed to optimize this versatile reporter for different applications. The alteration of these major characteristics of the protein have typically been achieved by making a very small number of amino acid substitutions involving only minor alterations to the DNA sequence of *gfp*. For example, the alteration of just one amino acid, arising from a single nucleotide substitution, was required to convert GFP to blue fluorescent protein (BFP) (8). Thus, many of the commonly used variants of GFP are encoded by genes that are ideal targets for H-NS and H-NS-like proteins.

The widely used and highly fluorescent *gfp*+ (17) has an A+T content of 59% and was predicted by the Bend.it algorithm (21) to contain several regions of curved DNA (Fig. 1A), making it a likely target for H-NS binding. The *gfp*+ gene was therefore redesigned to have an A+T content that was similar to that of the *E. coli* chromosome (49%). Gene Designer (20) was used both to reduce the A+T content of *gfp*+ and to optimize its codon usage for expression in *E. coli*. This approach produced iterative, equally optimized sequences that were screened for reduced intrinsic DNA curvature using Bend.it. The preferred sequence differed from *gfp*+ at 157 positions of the 717-base-pair gene, had 50% A+T content, and showed dramatically reduced predicted DNA curvature (Fig. 1A). This reengineered gene is referred to as *gfp*^{TCD} (TCD for Trinity College Dublin), and its DNA sequence is provided in Fig. S1 in the supplemental material.

Electrophoretic mobility shift assays (EMSA) were performed to determine the affinity of H-NS for *gfp*+ and *gfp*^{TCD} *in vitro*. H-NS bound to *gfp*+ with an apparent dissociation constant (K_d) of 4.9 ± 0.1 nM (Fig. 1B). A narrow protein concentration range (4.5 to 10.55 nM) was required for the transition from initial binding to the fully bound DNA probe that resolved as a single low-mobility complex. In contrast, H-NS displayed a relatively lower affinity for *gfp*^{TCD} (K_d, 7.5 ± 0.5 nM) (Fig. 1B), resulting in the formation of a poorly resolved H-NS–DNA complex over a wider range of protein concentrations (7.9 to 18.75 nM); the *gfp*^{TCD} probe resolved as a single H-NS–DNA complex only at protein concentrations above 25 nM.

The *E. coli proU* transcriptional regulatory region was used as a positive control for H-NS binding because it contains several well-characterized H-NS binding sites (1). As expected, the *proU* DNA probe was bound by H-NS (K_d, 6.2 ± 0.5 nM) in EMSA where it resolved as two separate H-NS–DNA complexes (Fig. 1C). H-NS has a relatively lower binding affinity for the *E. coli lacZ* gene (15), so this DNA was used as a negative control; EMSA analysis confirmed that *lacZ* was indeed more weakly bound by H-NS (K_d, 16.3 ± 1.8 nM) than was *proU* (Fig. 1C).

The *proU* operon was also used to test the effect of H-NS binding to *gfp*+ on transcription from a nearby promoter. Transcriptional repression of the *proU* operon in low-osmolarity conditions involves binding by H-NS to the A+T-rich, highly curved DNA located in the upstream and downstream regulatory elements (URE and DRE, respectively) that flank the *proU* promoter (1, 11, 15). The repressive complex formed

* Corresponding author. Mailing address: Department of Microbiology, Moyne Institute of Preventive Medicine, School of Genetics and Microbiology, Trinity College Dublin, Dublin 2, Ireland. Phone: 353 1 896 2013. Fax: 353 1 679 9294. E-mail: cjdorman@tcd.ie.

§ Supplemental material for this article may be found at http://jb.asm.org/.

$^{\nabla}$ Published ahead of print on 16 July 2010.

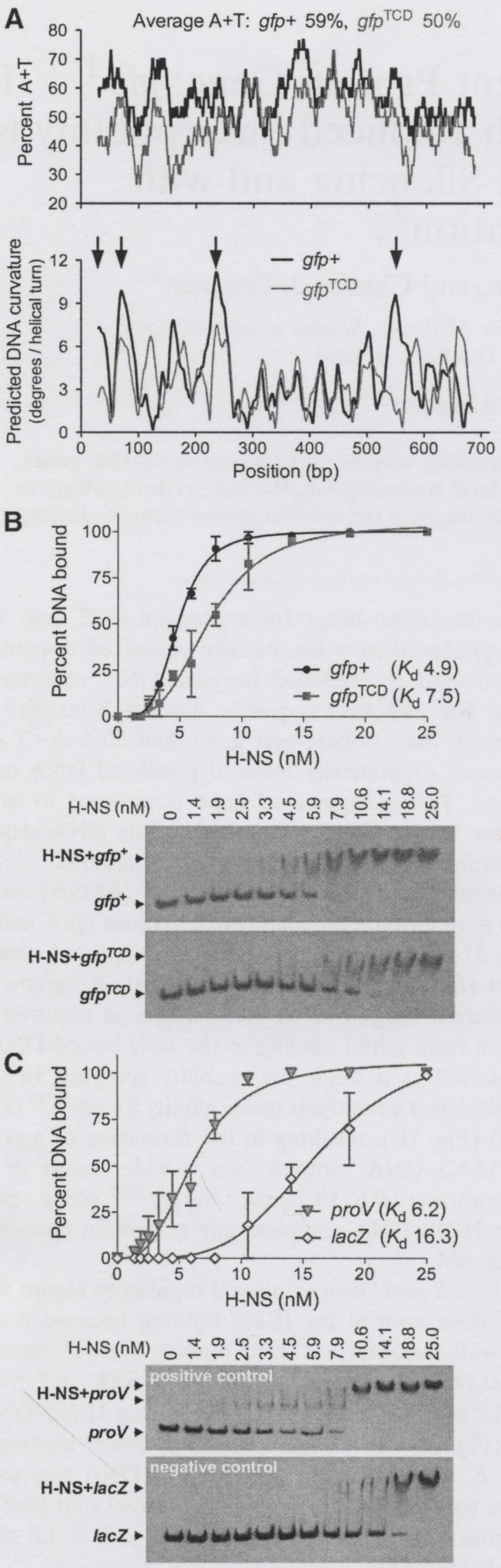
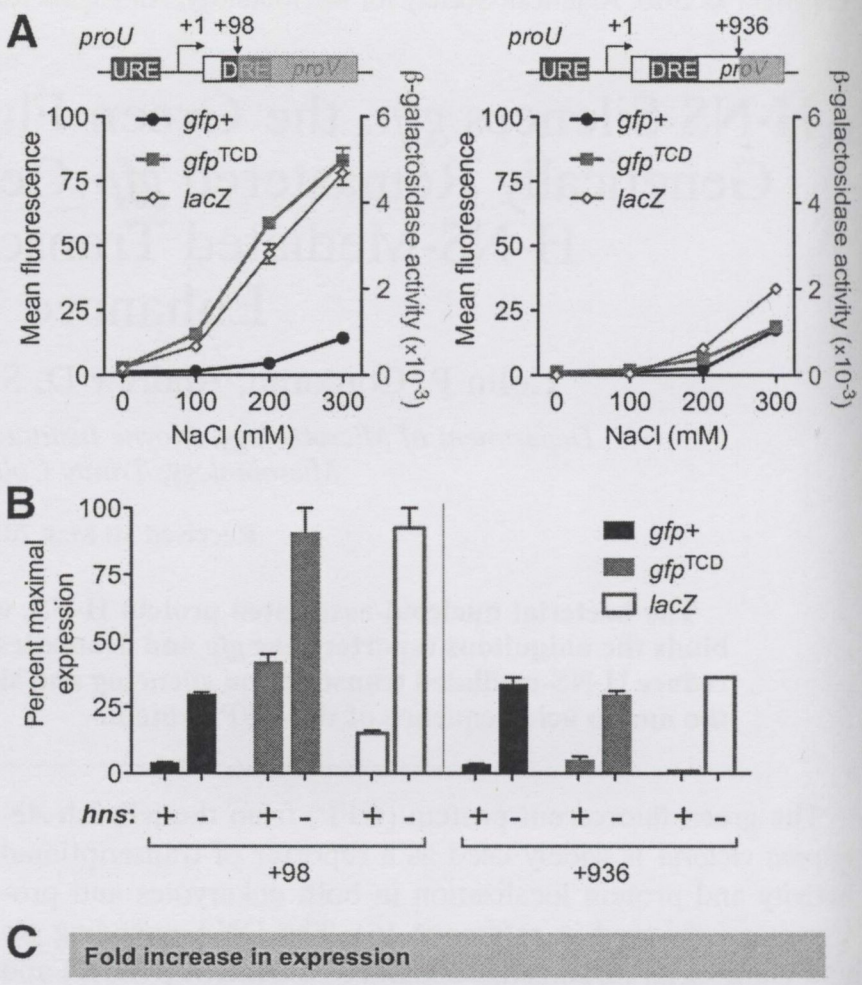


FIG. 1. DNA sequence features of the *gfp+* gene are conducive to H-NS binding. (A) A+T content (upper plot) and predicted intrinsic DNA curvature (lower plot) of *gfp+* and *gfp*^{TCD}. Arrows highlight regions of strong predicted curvature in *gfp+* that are reduced in *gfp*^{TCD}. (B and C) EMSA analysis of H-NS binding to *gfp+* and *gfp*^{TCD} (B) and *proU* and *lacZ* (C). Binding curves and K_d values are shown in the upper plots; mean values and ranges are plotted. Typical EMSA gels from which binding curves were derived are shown below. See supplemental material for a complete description of the experimental methods.



C

Fold increase in expression			
		gfp+	*gfp*^{TCD}
When DRE is replaced (+98 / +936)	wildtype	1.1	8.0
	Δ*hns*	0.9	3.1
When H-NS is removed (Δ*hns* / wildtype)	+98	7.3	2.2
	+936	9.4	5.6

FIG. 2. *gfp+* and *gfp*^{TCD} have different effects on transcription and translation. (A) The effect of increasing NaCl concentrations on the expression of reporter gene fusions inserted at +98 and +936 bp from the *proU* transcriptional start site. Mean values and ranges are plotted. Diagrams not to scale. (B) The effect of removing H-NS on fusion gene expression in 100 mM NaCl. Mean values and ranges are plotted. (C) Fold increases in fusion gene expression caused by changes in H-NS occupancy. See supplemental material for a complete description of the experimental methods.

by H-NS at *proU* is disrupted by increasing the osmolarity of the medium (7). While H-NS binding in the DRE is essential for osmoregulation of *proU*, it has previously been shown that the native DRE can be functionally replaced by an unrelated piece of DNA that is bound by H-NS (15). It was also shown that the *lacZ* gene cannot functionally replace the DRE and thus its presence in place of the DRE leads to derepression of *proU* in low-osmolarity conditions (100 mM NaCl) (15). Therefore, we tested the ability of *gfp+*, *gfp*^{TCD}, or *lacZ* to functionally replace the DRE by insertion of each gene within the DRE at +98 bp downstream from the transcriptional start site (Fig. 2A). Expression of the *proU-gfp+*(+98) fusion was repressed relative to expression of *proU-gfp*^{TCD}(+98), indicating that *gfp+*, but not *gfp*^{TCD}, functionally replaced the DRE as a binding site for H-NS and thus repressed *proU* expression. The *proU-lacZ*(+98) fusion was not repressed at low NaCl concentrations, consistent with earlier observations (15). Reporter fusions that did not disrupt the DRE were generated by insertion of each of the three reporter genes (*gfp+*, *gfp*^{TCD}, and

lacZ) at +936 bp downstream from the transcriptional start site (Fig. 2A). As predicted, expression of all three *proU*(+936) fusions was repressed by H-NS binding to the DRE at NaCl concentrations below 200 mM.

To confirm that H-NS binding to *gfp*+ accounted for repression of *proU-gfp*+(+98) *in vivo*, the expression of all *proU* fusions was tested in a Δ*hns* background. Cells were cultured in the repressive conditions of 100 mM NaCl, and the data were expressed as a percentage of maximal derepressed expression to facilitate comparisons between GFP fluorescence and β-galactosidase activity (Fig. 2B). For all three fusions, *proU* expression was elevated in the absence of H-NS. This revealed that even in the absence of the DRE, H-NS continued to bind the URE and repress the *proU-gfp*^{TCD}(+98) and *proU-lacZ*(+98) fusions. These data allowed an assessment of the relative effects of replacing only the DRE compared to those of eliminating H-NS protein from the cell (Fig. 2C). Replacing the DRE with *gfp*^{TCD} resulted in an 8-fold increase in expression relative to expression from the *proU-gfp*^{TCD}(+936) fusion; a similar comparison between *proU-gfp*+ at positions +98 and +936 showed that *gfp*+ inserted in the DRE maintained full repression. Consequently, eliminating H-NS genetically resulted in a 7.3- to 9.4-fold increase in expression from the *gfp*+ fusions while the expression of *proU-gfp*^{TCD}(+98) improved only 2.2-fold upon removal of H-NS (Fig. 2C). This also confirmed that H-NS-mediated repression of *proU* occurs primarily through binding to the DRE (Fig. 2C).

Even in the absence of H-NS, *proU-gfp*^{TCD}(+98) produced 3-fold-higher levels of GFP than *proU-gfp*+(+98) (Fig. 2C). The *gfp*^{TCD} gene is codon optimized for translation in *E. coli*, and these data suggest that *gfp*^{TCD} is translated 3-fold more efficiently than *gfp*+. Improved translation efficiency was confirmed by cloning *gfp*+ and *gfp*^{TCD} in the pPro vector under the control of the propionate-inducible *prpBCDE* promoter (9), for which there is no evidence of H-NS binding or repression (3). In pPro, *gfp*^{TCD} produced on average 3.5-fold more GFP in a wild-type background and 2.7-fold more GFP in a Δ*hns* background than *gfp*+ (Fig. 3A). The codon adaptation index (CAI) is a standard means to calculate the effects of species-specific codon biases on translation (18); Fig. 3B shows that *gfp*^{TCD} is predicted to have improved translation efficiency in both bacterial and eukaryotic model organisms. *gfp*^{TCD} therefore provides an ideal template for the directed evolution of novel *gfp* variants.

In many bacteria, including *E. coli*, H-NS selectively targets horizontally acquired genes of high A+T content preventing transcription of the potentially deleterious foreign DNA (5). The introduction of A+T-rich plasmid DNA can thus titrate H-NS away from native chromosomal locations, leading to a mild Δ*hns* phenotype with pleiotropic effects, including reduced virulence and motility (3, 6). We have shown that H-NS binds strongly in the *gfp*+ gene, perhaps explaining a recent report of reduced invasiveness of *Salmonella* due solely to the presence of a promoterless *gfp*+ gene on a multicopy plasmid (2). Over the years, many bacterial studies have relied on plasmid-based *gfp* fusions, and some of these studies may have experienced unrecognized pleiotropic effects due to H-NS titration. The *gfp* gene has been used in conjunction with plasmid-based high-throughput genetic screens in many landmark studies, including the identification of promoters involved in

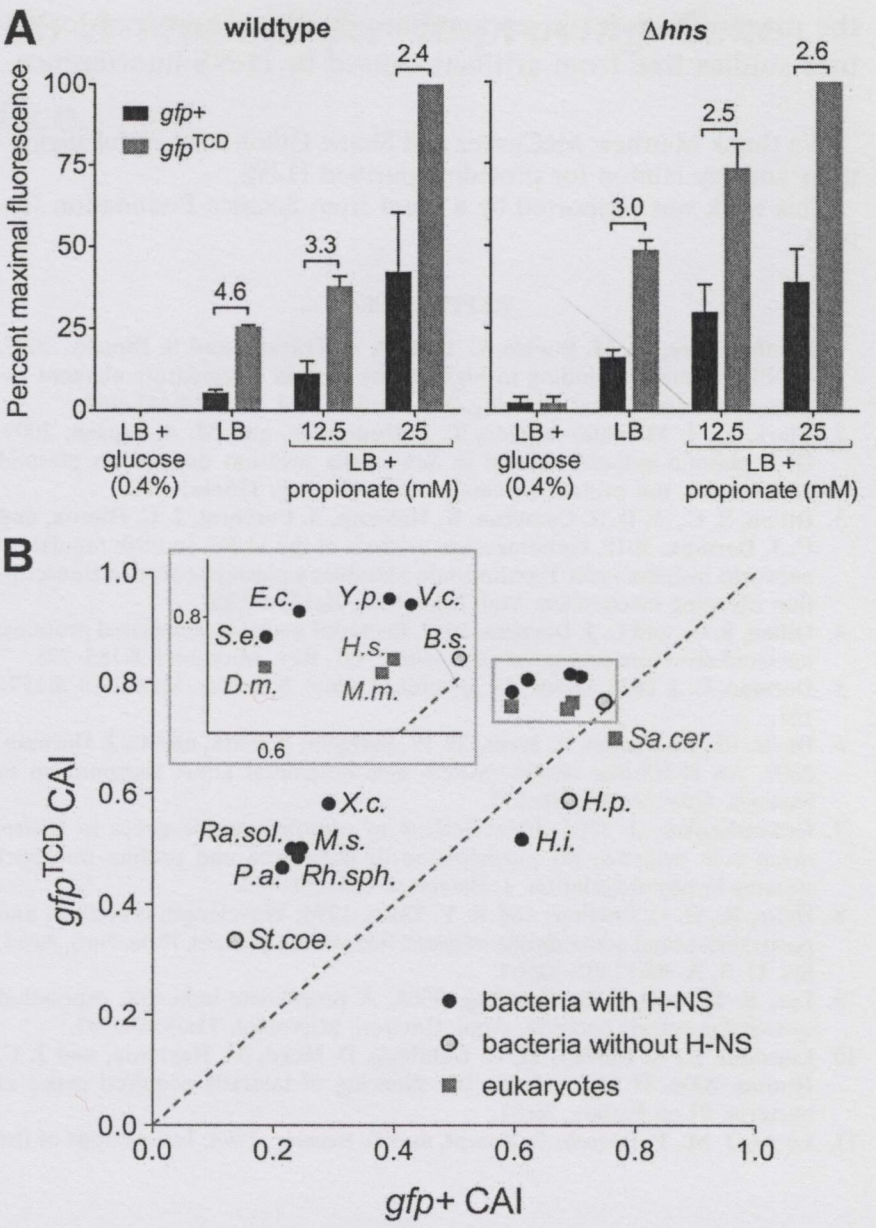


FIG. 3. *gfp*^{TCD} has improved translation efficiency. (A) Expression of *gfp*+ and *gfp*^{TCD} from the *prpBCDE* promoter on the plasmid pPro in wild-type and Δ*hns* cells. The *prpBCDE* promoter is repressed by glucose and induced by propionate. Mean values and ranges are plotted; fold differences between *gfp*+ and *gfp*^{TCD} fluorescence levels are shown above the bars. (B) Codon adaptation index (CAI) values for *gfp*^{TCD} and *gfp*+ in various organisms. Organisms listed above the dashed line will translate *gfp*^{TCD} more efficiently than *gfp*+ (and wild-type *gfp*). Organism classification and abbreviations are as follows. *Alphaproteobacteria*: *Rh.sph.*, *Rhodobacter sphaeroides*. *Betaproteobacteria*: *Ra.sol.*, *Ralstonia solanacearum*. *Epsilonproteobacteria*: *H.p.*, *Helicobacter pylori*. *Gammaproteobacteria*: *E.c.*, *Escherichia coli*; *H.i.*, *Haemophilus influenzae*; *P.a.*, *Pseudomonas aeruginosa*; *S.e.*, *Salmonella enterica*; *V.c.*, *Vibrio cholerae*; *X.c.*, *Xanthomonas campestris*; *Y.p.*, *Yersinia pestis*. *Actinobacteria*: *M.s.*, *Mycobacterium smegmatis*; *St.coe.*, *Streptomyces coelicolor*. *Firmicutes*: *B.s.*, *Bacillus subtilis*. *Eukaryota*: *D.m.*, *Drosophila melanogaster*; *H.s.*, *Homo sapiens*; *M.m.*, *Mus musculus*; *Sa.cer.*, *Saccharomyces cerevisiae*. See supplemental material for a complete description of the experimental methods.

the intracellular survival of *Salmonella enterica* serovar Typhimurium (19) and the adaptation by *E. coli* to metabolisms of different carbon sources (23).

Since nucleoid-associated proteins with a preference for A+T-rich DNA are found in most cell types, including archaea and eukaryotes (4), our discovery of H-NS binding and repression of *gfp* may have far-reaching implications for the use of *gfp* in all domains of life. The development of *gfp*^{TCD}, which is both a poor target for H-NS and a more accurate reporter of transcription that avoids H-NS-mediated downregulation of

the promoter under investigation, provides a new tool for future studies free from artifacts caused by H-NS interference.

We thank Matthew McCusker and Shane Dillon for helpful suggestions and Jay Hinton for providing purified H-NS.

This work was supported by a grant from Science Foundation Ireland.

REFERENCES

1. **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.
2. **Clark, L., I. Martinez-Argudo, T. J. Humphrey, and M. A. Jepson.** 2009. GFP plasmid-induced defects in *Salmonella* invasion depend on plasmid architecture, not protein expression. Microbiology **155:**461–467.
3. **Dillon, S. C., A. D. S. Cameron, K. Hokamp, S. Lucchini, J. C. Hinton, and C. J. Dorman.** 2010. Genome-wide analysis of the H-NS and Sfh regulatory networks in *Salmonella* Typhimurium identifies a plasmid-encoded transcription silencing mechanism. Mol. Microbiol. **76:**1250–1265.
4. **Dillon, S. C., and C. J. Dorman.** 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat. Rev. Microbiol. **8:**185–195.
5. **Dorman, C. J.** 2007. H-NS, the genome sentinel. Nat. Rev. Microbiol. **5:**157–161.
6. **Doyle, M., M. Fookes, A. Ivens, M. W. Mangan, J. Wain, and C. J. Dorman.** 2007. An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. Science **315:**251–252.
7. **Gowrishankar, J.** 1985. Identification of osmoresponsive genes in *Escherichia coli*: evidence for participation of potassium and proline transport systems in osmoregulation. J. Bacteriol. **164:**434–445.
8. **Heim, R., D. C. Prasher, and R. Y. Tsien.** 1994. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. U. S. A. **91:**12501–12504.
9. **Lee, S. K., and J. D. Keasling.** 2005. A propionate-inducible expression system for enteric bacteria. Appl. Environ. Microbiol. **71:**6856–6862.
10. **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.
11. **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.
12. **Navarre, W. W., S. Porwollik, Y. Wang, M. McClelland, H. Rosen, S. J. Libby, and F. C. Fang.** 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in *Salmonella*. Science **313:**236–238.
13. **Noom, M. C., W. W. Navarre, T. Oshima, G. J. Wuite, and R. T. Dame.** 2007. H-NS promotes looped domain formation in the bacterial chromosome. Curr. Biol. **17:**R913–914.
14. **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:**141–553.
15. **Owen-Hughes, T. A., G. D. Pavitt, D. S. Santos, J. M. Sidebotham, C. S. Hulton, J. C. Hinton, and C. F. Higgins.** 1992. The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. Cell **71:**255–265.
16. **Rizzo, M. A., M. W. Davidson, and D. W. Piston.** 1 December 2009, posting date. Fluorescent protein tracking and detection: fluorescent protein structure and color variants. Cold Spring Harb. Protoc. doi:10.1101/pdb.top63.
17. **Scholz, O., A. Thiel, W. Hillen, and M. Niederweis.** 2000. Quantitative analysis of gene expression with an improved green fluorescent protein. Eur. J. Biochem. **267:**1565–1570.
18. **Sharp, P. M., and W. H. Li.** 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. **15:**1281–1295.
19. **Valdivia, R. H., and S. Falkow.** 1997. Fluorescence-based isolation of bacterial genes expressed within host cells. Science **277:**2007–2011.
20. **Villalobos, A., J. E. Ness, C. Gustafsson, J. Minshull, and S. Govindarajan.** 2006. Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics **7:**285.
21. **Vlahovicek, K., L. Kajan, and S. Pongor.** 2003. DNA analysis servers: plot.it, bend.it, model.it and IS. Nucleic Acids Res. **31:**3686–3687.
22. **Yamada, H., S. Muramatsu, and T. Mizuno.** 1990. An *Escherichia coli* protein that preferentially binds to sharply curved DNA. J. Biochem. **108:**420–425.
23. **Zaslaver, A., A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M. G. Surette, and U. Alon.** 2006. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. Nat. Methods **3:**623–628.

Appendix 4: Dorman CJ and Corcoran CP (2010) Patent application PCT/EP2010/057856 *A Method for Improving Gene Expression*. European Patent Office.

"A METHOD FOR IMPROVING GENE EXPRESSION"

Field of the Invention

The present invention is directed to a method for improving and/or monitoring gene expression in a host cell comprising a protein encoding nucleic acid. The invention is also directed to a modified protein encoding nucleic acid, protein, expression system, plasmid vector and host cell. Preferably, the protein encoding nucleic acid is a fluorescent protein nucleic acid

Background to the Invention

Cell-based protein expression systems are well known and the aim of such expression systems is to overexpress a desired protein. Host cells include bacteria (such as *E.coli*, *B. subtilis* etc), yeast (such as *S.cerevisiae*) or eukaryotic cell lines. Conventional DNA sources and delivery mechanisms include viruses (such as baculovirus, retrovirus, adenovirus), plasmids, artificial chromosomes and bacteriophage (such as lambda). The expression system used will depend on the gene involved, for example *Saccharomyces cerevisiae* is often selected for proteins that require significant post-translational modification. Insect or mammalian cell lines may be used when human-like splicing of the mRNA is required. Additionally, bacterial expression has the advantage of easily producing large amounts of protein. *E. coli* is one of the most widely used expression hosts, and DNA is normally introduced in a plasmid expression vector.

It is common for genes from many organisms to be placed under the control of an inducible promoter in a fast growing organism such as *Saccharomyces cerevisae* or *E. coli*. This facilitates the expression of the gene product to very high levels (often >50% of total cell protein). The gene product can then be purified to homogeneity. An example of this is the expression of restriction endo-nucleases from multiple organisms in *E. coli*, which are then purified and used *in vitro* to digest DNA. For example the genes encoding the restriction endo-nucleases PstI, Hpy99I and PsiI from *Providencia stuartii* 164, *Helicobacter pylori* J99 and *Pseudomonas species* SE-G49, respectively, are all cloned into *E. coli* for over-expression and purification.

This type of process has been used extensively to manufacture medically-useful products since the 1970s when synthetic humanized insulin was developed by joining the insulin gene with a plasmid vector inserted into the *E. coli*. Insulin was the first FDA-licensed drug

produced through this recombinant DNA technology. Other bacterial recombinant proteins include recombinant human growth hormone, interleukin-2 lymphocyte growth factor and interferon, a cytokine. Vaccines can be manufactured in yeast for example. The first drug produced commercially by mammalian cell culture was tissue plasminogen activator (tPA), used to dissolve blood clots. Another recombinant protein produced by the mammalian CHO cells is glycoprotein factor VIII, a blood clotting factor. There are many other recombinant protein products manufactured in this way. Accordingly, improvements to such cell-based protein expression systems are always being developed.

Thus, the present invention is directed to improving gene expression in host cells.

The discovery of green fluorescent protein in the early 1960s started a new era in cell biology by enabling investigators to apply molecular cloning methods, fusing the fluorophore moiety to a wide variety of protein and enzyme targets, in order to monitor cellular processes in living systems using optical microscopy and related methodology. When coupled to recent technical advances in fluorescence and microscopy, including ultrafast low light level digital cameras and multitracking laser control systems, the green fluorescent protein and its colour-shifted genetic derivatives have demonstrated invaluable service in many thousands of live-cell imaging experiments.

The Green Fluorescent Protein (GFP) was originally isolated in the 1960's from the jelly fish *Aequorea victoria* and is encoded by the *gfp* gene and fluoresces green when exposed to blue light. GFP is composed of 238 amino acids (26.9 kDa). GFP has a typical beta barrel structure, consisting of one beta-sheet with alpha helix(s) containing the chromophore running through the center. Inward facing sidechains of the barrel induce specific cyclization reactions in the tripeptide Ser65–Tyr66–Gly67 that lead to chromophore formation. The hydrogen bonding network and electron stacking interactions with these sidechains influence the colour of wild type (wt) GFP and its numerous derivatives. The GFP from *A. victoria* has a major excitation peak at a wavelength of 395 nm and a minor one at 475 nm. Its emission peak is at 509 nm which is in the lower green portion of the visible spectrum.

However despite being isolated in the 1960's, its utility as a tool for molecular biologists was not realized until the 1990's when the cloning and nucleotide sequence of wild type *gfp* gene took place. At this time, it was determined that the GFP molecule folded and was fluorescent at room temperature, without the need for exogenous cofactors specific to the jellyfish. However, although this near-wild type GFP was fluorescent, it had several

drawbacks, including dual peaked excitation spectra, pH sensitivity, chloride sensitivity, poor fluorescence quantum yield, poor photostability and poor folding at 37°C.

Thus, GFP mutants were studied in order to find a protein with better characteristics. Indeed, the first reported crystal structure of a GFP was that of the S65T mutant (a single point mutation) by in the mid 1990's. This mutation dramatically improved the spectral characteristics of GFP, resulting in increased fluorescence, photostability and a shift of the major excitation peak to 488 nm with the peak emission kept at 509 nm. This matched the spectral characteristics of commonly available FITC filter sets, increasing the practicality of use by the general researcher.

In addition to the first single amino acid substitution, S65T, researchers have modified the GFP residues by directed and random mutagenesis to produce the wide variety of GFP derivatives in use today. For example, a 37°C folding efficiency (F64L) point mutant to this scaffold yielding enhanced GFP (EGFP) was discovered in 1995 by Ole Thastrup. EGFP allowed for the use of GFPs in mammalian cells. EGFP has an extinction coefficient of 55,000 $M^{-1}cm^{-1}$. Another mutation, superfolder GFP, related to a series of mutations that allow GFP to rapidly fold and mature even when fused to poorly folding peptides, was reported in 2006.

Many other mutations have been made, including colour mutants; in particular

- blue fluorescent protein (EBFP, EBFP2, Azurite, mKalama1);
- cyan fluorescent protein (ECFP, Cerulean, CyPet); and
- yellow fluorescent protein derivatives (YFP, Citrine, Venus, YPet). BFP derivatives (except mKalama1) contain the Y66H substitution.

The critical mutation in cyan derivatives is the Y66W substitution, which causes the chromophore to form with an indole rather than phenol component. Several additional compensatory mutations in the surrounding barrel are required to restore brightness to this modified chromophore due to the increased bulk of the indole group. The red-shifted wavelength of the YFP derivatives is accomplished by the T203Y mutation and is due to π-electron stacking interactions between the substituted tyrosine residue and the chromophore. These two classes of spectral variants are often employed for fluorescence resonance energy transfer (FRET) experiments.

There are many other mutants, including but not limited to the following:

- Semi-rational mutagenesis of a number of residues led to pH-sensitive mutants known as pHluorins, and later super-ecliptic pHluorins. By exploiting the rapid change in pH upon synaptic vesicle fusion, pHluorins tagged to synaptobrevin have been used to visualize synaptic activity in neurons.

- Redox sensitive versions of GFP (roGFP) were engineered by introduction of cysteines into the beta barrel structure. The redox state of the cysteines determines the fluorescent properties of roGFP.

- A variant combining mutations designed to improve folding at 37 °C (F99S, M153T, V163A) from GFPuv (Crameri et al., Nat Biotechnol 14 (1996), 315-9) were combined with mutations in the chromophore of GFPmut1 (F64L and S65T) (Cormack *et al.*, 1996) to create the GFP+ protein which was 320 times more fluorescent than the original GFP (Scholz et al. Eur. J. Biochem. 267 (2000), 1565-70). The GFPmut1 protein is identical with EGFP, which is commercially available, widely used even though only about 20% of EGFP/GFPmut1 is correctly folded at 37 °C (Tsien Annu Rev Biochem 67 (1998), 509-44). The development of GFP+ was necessary to allow the measurement of gene activity in the native location (Hautefort et al., App. Env. Microbiol. 69 (2003), 7480-7491).

It is know well-known that there are other fluorescent proteins isolated from different sources which may be used in the same manner as GFP. These include DsRed and the protein encoded by the *gfp* gene from the great star coral, *Montastraea cavernosa*.

DsRed is a recently cloned 28-kDa fluorescent protein isolated from the coral of the *Discoma* genus. DsRed has an emission maximum of 583-nm, which can be further extended to 602-nm by mutation of Lys-83 to Met. This emission spectrum makes it ideal as a fluorescent partner for GFP, as fluorescence of both proteins could be individually measured in a single cell (Baird et al., Proc. Natl. Acad. Sci. USA 97 (2000), 11984-11989). Similar to the early studies of GFP, a number of the limitations of this protein are presently being addressed through random mutagenesis of the coding sequence and screening for improved variants. One of the major issues with DsRed include the long maturation time before a red signal is detected and that it forms tetramers, both undesirable characteristics for a transcriptional/translational fusion.

Because of its easily detectable green fluorescence, GFP from *Aequorea* has been used widely to study gene expression and protein localization. Furthermore, GFP, like other fluorescent proteins, does not require a substrate or cofactor to fluoresce; hence, it is possible to directly express GFP and use it as a reporter in numerous species and in a wide variety of cells.

These fluorescent proteins and mutants or variants thereof can be introduced into organisms and maintained in their genome through breeding, injection with a viral vector, or cell transformation of either linear or circular DNA. For example, the GFP gene has been introduced and expressed in many bacteria, yeast and other fungi, plant, fly, and mammalian cells, including human.

Thus, GFP and other related fluorescent proteins and related mutants/variants as described above are now used routinely as reporters of gene expression in all types of cells. They are an essential tool for biologists and work is continually being carried out on developing improved variants to overcome inherent limitations of the native proteins.

Thus, the present invention is directed to improving the efficacy of fluorescent protein genes for use as a molecular tool, in particular when used in gene expression systems.

Statements of the Invention

According to a first general aspect of the invention, there is provided a method for improving gene expression in a host cell comprising a protein encoding nucleic acid comprising assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type protein encoding nucleic acid or mutant thereof; preparing an altered protein encoding nucleic acid by modifying the A and T nucleotide content of the wild type protein encoding nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell such that the intrinsic curvature of the altered protein encoding nucleic acid is reduced compared to the wild type protein encoding nucleic acid or mutant thereof; and using the altered protein encoding nucleic acid in host cell gene expression systems. Ideally, the modified A and T nucleotide content of the altered protein encoding nucleic acid results in reduced affinity, compared to the wild type protein encoding nucleic acid or mutant thereof, to host cell transcriptional repressor proteins.

According to a second general aspect of the invention, there is provided a method for improving gene and/or protein expression in a host cell comprising a fluorescent protein nucleic acid comprising assessing the A and T nucleotide content of the fluorescent protein nucleic acid; and modifying the A and T nucleotide content of the fluorescent protein nucleic acid to equal or lower the A and T nucleotide content of the host cell.

According to one embodiment of this second aspect of the invention, there is provided a method for improving gene and/or protein expression in a host cell comprising a fluorescent protein nucleic acid or mutant thereof comprising

assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type fluorescent protein nucleic acid or mutant thereof;

preparing an altered fluorescent protein nucleic acid by modifying the A and T nucleotide content of the wild type fluorescent protein nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell such that the intrinsic curvature of the altered fluorescent protein nucleic acid is reduced compared to the wild type fluorescent protein nucleic acid or mutant thereof; and

using the altered fluorescent protein nucleic acid in a host cell gene expression system.

According to a third aspect of the invention, there is provided a modified fluorescent protein nucleic acid comprising a sequence encoding a wild type fluorescent protein or mutant thereof with a lower A and T nucleotide content and/or reduced intrinsic curvature compared to the wild type fluorescent protein nucleic acid or mutant thereof. Ideally, the protein has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type fluorescent protein nucleic acid or mutant thereof.

According to a fourth aspect of the invention, there is provided a fluorescent protein encoded by the modified nucleic acid of the invention.

According to a fifth aspect of the invention, there is provided an expression system comprising the modified nucleic acid sequence or fluorescent protein according to the invention, preferably for use in a host cell such as *Escherichia coli*.

According to a sixth aspect of the invention, there is provided a plasmid vector comprising the modified nucleic acid sequence or fluorescent protein according to the invention, preferably for use in a host cell such as *Escherichia coli*.

According to a seventh aspect of the invention, there is provided a host cell comprising the modified nucleic acid sequence, fluorescent protein, plasmid vector, expression system of the invention.

According to a eighth aspect of the invention, there is provided a method of monitoring gene expression in a host cell using the modified nucleic acid, fluorescent protein, expression system or host cell of the invention.

Detailed Description of the Invention

In this specification, it will be understood that any of the percentage identities or homologies referred to in the specification are determined using available conventional methods over the entire/whole length of the sequence.

In this specification, it will be understood that the terminology "equal to or lower than the A and T nucleotide residue content of a host cell" covers an A and T residue content equal to or within +/- 15%, +/- 10%, +/- 9%, +/- 8%, +/- 7%, +/- 6%, preferably +/- 5%, +/- 4%, +/- 3%, +/- 2%, more preferably +/-1%, +/- 0.5%, +/- 0.1%, of the A and T residue content of the host cell.

DNA is made up of combinations of 4 nucleotides; Adenosine (A), Tyrosine (T), Guanosine (G) and Cytosine (C). Each species uses each of these nucleotides in differing ratios. A and T are grouped because they both contain only 2 hydrogen bonds while G and C both contain 3. It is known, for example, that approximately 50% of the genome of *Escherichia coli* (a Gram-negative bacterium), a typical host cell, is an A or a T nucleotide (A/T). In addition, it is known that, the A/T nucleotide content of the fluorescent protein gene, *gfp*+, for example is approximately 59%, i.e. it is "A/T rich". In this specification, we have used the term "A/T rich" or "A/T nucleotide rich" to convey an A/T nucleotide content higher than the average A/T residue content of any suitable host cell, including *E. coli* and many other host cells. Other proteins and fluorescent proteins will have different A/T nucleotide contents and may or may not be A/T rich compared to a host cell. As an example, the approximate A and T nucleotide content of the following Gram-positive and negative bacteria is given on the table below:

	Approximate A and T nucleotide content
Gram-negative	
E. coli	50%
Salmonella typhimurium	49%
Pseudomonas fluorescens	40%
Gram-positive	

Micrococcus luteus	25%
Streptomyces griseus	27%
Mycobacterium tuberculosis	33-35%
Bacillus subtilis	58%
Lactobacillus viridescens	60%
Staphylococcus aureus	67-68%
Clostridium perfringens	62%
Mycoplasma capricolum	75%

It will be understood that reducing the intrinsic curvature of the altered protein encoding nucleic acid ideally means lowering the curvature amplitude compared to the wild type protein encoding nucleic acid or mutant thereof.

Ideally, intrinsic curvature should be reduced to less than approximately 15° per helical turn, approximately 10° per helical turn, more preferably approximately 9° per helical turn, even more preferably approximately 7° per helical turn. For example, we found that all regions of intrinsic curvature greater than approximately 7° per helical turn in the fluorescent protein gfp should be reduced to approximately 7° per helical turn or less. For the fluorescent protein dsred we found that all regions of intrinsic curvature greater than approximately 9° per helical turn in the fluorescent protein dsred should be reduced to approximately 9° per helical turn or less.

We have found that codon optimization to increase the A/T content of a gene to match that of a G/C rich genome (<50% A/T) will make it a poor target for H-NS, similar codon optimization for an A/T rich genome may increase the affinity of H-NS for the target gene. Therefore, to avoid H-NS mediated complications when optimizing genes for expression in A/T rich genomes, it may be advantageous to optimize genes using a codon-table derived from highly expressed genes of the organism of interest that have a reduced A/T content compared to the average A/T content of the genome. The optimal level of A/T content required to avoid H-NS mediated complications in an A/T rich genome can then be determined empirically.

It will be understood that host cells are not limited to bacteria, and include all suitable prokaryote host cells, such as bacterial cells and yeast cells, and eukaryote host cells, such as mammalian cells (including human, primate and rodent cells). Furthermore, the host cells of the invention will be understood to express transcriptional repressor proteins,

such as nucleoid-associated transcriptional repressor proteins and typically H-NS or H-NS-like proteins.

Furthermore, the comments that follow in relation to H-NS are equally applicable to all transcriptional repressor proteins. The invention is also applicable to H-NS-like proteins or H-NS homologs. It will be understood that H-NS like proteins include but are not limited to the bacterial proteins Sfh, StpA, Hha, YdgT, MvaT, MvaU, Lsr2, BpH3. Such transcriptional repressor proteins and repressor-like proteins are also present in mammalian cells.

In this specification, it will be understood that the term "protein encoding nucleic acid" covers, but is not limited to fluorescent proteins. Non-fluorescent proteins which are over-expressed in host cells for purification are also contemplated. Such proteins include all suitable medical protein products (vaccines, hormones (insulin, growth hormone), clotting factors, cytokines etc) and enzymes for example. This is a non-exhaustive list.

In this specification, it will be understood that the term "fluorescent protein" covers, but is not limited to, fluorescent proteins such as GFP, YFP, CFP, BFP and DsRed and also covers any potential mutants or variants thereof. There are an extensive range of known variants and mutants of both GFP and DsRed including a wide variety of single or double amino acid substitutions. For convenience, where the term mutant is referred to in the following text, it will be understood that this term also covers fluorescent protein variants.

It will be understood that the following comments and teachings which although relate to a specific *gfp* variant tested, *gfp*+, are applicable to all fluorescent proteins and mutants and variants thereof in general. In the following description *gfp*T (or gfp^{T}) may also be referred to interchangeably as *gfp*TCD (or gfp^{TCD}).

According to a first general aspect of the invention, there is provided a method for improving gene expression in a host cell comprising a protein encoding nucleic acid comprising assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type protein encoding nucleic acid or mutant thereof; preparing an altered protein encoding nucleic acid by modifying the A and T nucleotide content of the wild type protein encoding nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell such that the intrinsic curvature of the altered protein encoding nucleic acid is reduced compared to the wild type protein encoding nucleic acid or mutant thereof; and using the altered protein encoding nucleic acid in host cell gene expression systems.

Ideally, the modified A and T nucleotide content of the altered protein encoding nucleic acid results in reduced affinity, compared to the wild type protein encoding nucleic acid or mutant thereof, to host cell transcriptional repressor proteins.

It will be understood that these teachings are applicable to any protein encoding nucleic acid with a high A and T nucleotide content (AT-rich) compared to a host cell nucleic acid content.

It will be understood that the protein encoding nucleic acid may be present in an extrachromosomal vector, such as a plasmid. Alternatively, the protein encoding nucleic acid may be integrated into the host cell genome.

We have surprisingly found that this method prevents the possible interference of transcriptional repressor proteins, such as H-NS proteins, with gene expression at the level of transcription. By reducing the affinity of transcriptional repressor proteins, such as H-NS and H-NS-like proteins, for target genes their rates of transcription should increase, leading to an increase in yield of the gene product. In this manner, the invention is applicable to any protein for over-expression and subsequent purification in a host cell. Such proteins include fluorescent proteins and non-fluorescent protein products.

The protein encoding nucleic acid may encode many medically useful protein products such as vaccines, hormones (insulin, growth hormone), clotting factors, cytokines etc and enzymes for example. In addition the protein encoding nucleic acid may encode various restriction endo-nucleases PstI, Hpy99I and PsiI from *Providencia stuartii* 164, *Helicobacter pylori* J99 and *Pseudomonas species* SE-G49. Many other protein products may be contemplated.

Helicobacter pylori J99 has an average genomic A/T content of 61%, which is the same A/T content as the fluorescent protein, *gfpmut2* (Figure 5B). This high average A/T content of *H. pylori* for example means it will have a large number of genes with a high A/T content Thus, due to a high AT content, it is to be expected that when genes are placed in *E. coli*, they are highly likely to be targeted by H-NS. The present invention provides a method for preventing this H-NS interaction by reducing the A/T nucleotide content to equal or lower than of the host cell, such as *E. coli*. These teachings are relevant to many different proteins and many different host cells.

Advantageously, the host cell is a bacterium, preferably a gram-negative bacterium, more preferably *Escherichia coli*. Other host cells may be contemplated including suitable prokaryotic and eukaryotic cells, such as yeast, insect, mammalian, primate and rodent cells.

According to another embodiment of this first aspect of the invention, there is provided a modified protein encoding nucleic acid comprising a sequence encoding a wild type protein or mutant thereof with an equal or lower A and T nucleotide content and/or reduced intrinsic curvature compared to the wild type protein encoding nucleic acid or mutant thereof. The resultant nucleic acid has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type protein encoding nucleic acid or mutant thereof.

The invention is also directed to a modified protein encoding nucleic acid of the invention and protein, expression system, plasmid vector and host cell comprising the modified protein encoding nucleic acid.

It will be understood that the following discussion on the second aspect of the invention relating to fluorescent proteins is equally applicable to this first aspect of the invention.

According to a second general aspect of the invention, there is provided method for improving gene and/or protein expression in a host cell comprising a fluorescent protein nucleic acid comprising assessing the A and T nucleotide content of the fluorescent protein nucleic acid or mutant thereof; and modifying the A and T nucleotide content of the fluorescent protein nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell.

In this manner, the A and T residue content of the fluorescent protein nucleic acid may be equal to or within +/- 15%, +/- 10%, +/- 9%, +/- 8%, +/- 7%, +/- 6%, preferably +/- 5%, +/- 4%, +/- 3%, +/- 2%, more preferably +/-1%, +/- 0.5%, +/- 0.1%, of the A and T residue content of the host cell. The modified fluorescent protein nucleic acid or mutant thereof can then be used as a molecular tool in gene expression systems and host cells.

According to one embodiment of this second aspect of the invention, there is provided a method for improving gene expression in a host cell comprising a fluorescent protein nucleic acid or mutant thereof comprising

assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type fluorescent protein nucleic acid or mutant thereof;

preparing an altered fluorescent protein nucleic acid by modifying the A and T nucleotide content of the wild type fluorescent protein nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell such that the intrinsic curvature of the altered fluorescent protein nucleic acid is reduced compared to the wild type fluorescent protein nucleic acid or mutant thereof; and

using the altered fluorescent protein nucleic acid in host cell gene expression systems.

Thus, the crucial aspect of this invention is to modify the fluorescent protein nucleic acid A/T nucleotide content and/or intrinsic curvature to be lower than the A/T nucleotide content of the host cell. Surprisingly, we found that the resultant fluorescent protein nucleic acid has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type fluorescent protein encoding nucleic acid or mutant thereof. Unexpectedly, this ensures improved fluorescent protein expression of a structurally identical fluorescent protein.

We have advantageously found that this method reduces the potential detrimental impact of a fluorescent protein nucleic acid or mutant thereof on gene and/or protein expression in a host cell comprising a fluorescent protein nucleic acid.

Specifically, the present invention is based on findings that transcriptional repressor proteins, such as nucleoid-associated transcriptional repressor proteins and typically H-NS or H-NS like proteins, targets fluorescent protein genes, such as the *gfp* gene or mutants or variants thereof. The transcriptional repressor proteins, namely the nucleoid-associated protein H-NS, are molecules that are abundant in *E. coli* and related organisms and are powerful repressors of transcription. H-NS can create bridges between different DNA molecules or between different parts of the same DNA molecule and can also nucleate along DNA. H-NS is essentially a protein with the ability to slow or block gene expression. Thus, these H-NS proteins interfere with gene expression at the level of transcription.

GFP is commonly used as a transcriptional reporter gene in host cells. We have now unexpectedly shown that H-NS interferes with the expression of the GFP protein in host cells such as bacteria. This discovery is unexpected and we postulate that many of the experiments carried out using GFP previously, in *E. coli* and related bacteria, are likely to have been complicated by the unsuspected participation of H-NS. The binding of the H-NS

protein to the *gfp*+ gene has the potential to undermine its fidelity as a reporter of gene expression. This can mislead investigators, perhaps causing them to underestimate the levels at which genes of interest are expressed or it may complicate their study by introducing a new and unsuspected layer of gene regulation and artefacts into the system under examination.

Importantly, we also postulate that H-NS binding and oligomerization causes the formation of new topological domains, which can lead to both local transcriptional repression and in addition, genome wide changes in gene expression. As most *gfp* transcriptional and translational fusions are located on multicopy plasmids, the possibility exists that global gene expression could be altered by the binding of H-NS in *gfp*, depleting the amount of free H-NS in the cell and thus leading to altered global gene expression. Thus, H-NS also has the potential to modify gene expression in a host cell by upregulating/downregulating various local/distal genes.

Considering that all of the GFP derivatives used today are highly related and contain minimal nucleotide substitutions, and given that H-NS binding is affected by general characteristics such as DNA curvature or intrinsic curvature (which typically requires large scale substitutions to alter), it is reasonable to extrapolate that the genes of all of the commonly used GFP variants, such as but not limited to GFPmut1, GFPmut2 and GFPmut3, are also bound by H-NS or other transcriptional repressor proteins. Thus, this problem is widespread for all GFP variants.

These unexpected findings have led to the development of the method for improving gene and/or protein expression in a host cell comprising a modified fluorescent protein nucleic acid.

It is known that, the gene encoding GFP has a high A and T nucleotide content and DNA with this nucleotide content is preferentially bound by the H-NS protein. This repression mechanism frequently involves cross-linking of DNA molecules by H-NS to either distant DNA molecules or different parts of the same molecule to form DNA-H-NS-DNA bridges. These DNA-H-NS-DNA bridges block the free movement of RNA polymerase and interfere with the process of transcription. By recognizing that H-NS binding is a problem with the *gfp*+ gene in particular, we have engineered a modified *gfp*+ gene that codes for a protein identical to the *gfp*+ protein, but is no longer bound by H-NS. Thus, advantageously, the present invention addresses all of the above-mentioned aspects to prevent H-NS binding and associated H-NS mediated transcriptional silencing to results in an improved

fluorescent protein for use in known expression systems/host cells. Specifically, the invention focuses on the optimization of *gfp* for expression in bacterial cells that express H-NS or related proteins. This modified fluorescent protein gene is thus modified to circumvent undesirable interference by H-NS.

Due to the degeneracy of the genetic code, we have modified the nucleotide sequence of the fluorescent gene without disturbing the order of the amino acids in the resultant protein product. Thus, the protein product encoded by the gene remains unchanged. Changing the nucleotide sequence in this manner enables the production of a modified fluorescent protein with altered A and T nucleotide content. Accordingly, we have been able to redesign the *gfp*+ gene such that the A and T nucleotide content matches/equals or lowers the average A and T nucleotide content from *E. coli* or other related bacterial and non-bacterial expression systems without altering the amino acid sequence or characteristics of the protein product, GFP. Specifically, we have utilized "codon optimization" involving modifying the *gfp*+ gene to abrogate binding of the H-NS protein to the *gfp* gene whilst maintaining the coding sequence to ensure that the protein that is expressed is identical at the amino acid sequence level to the GFP+ protein. In this way the modified gene continues to express a GFP protein that is structurally identical to GFP+.

The teachings of the invention in relation to altering A and T nucleotide content are widely applicable. These teachings are applicable to any fluorescent proteins and host cell transcriptional repressor proteins. It will also be understood the concept of the invention is equally applicable to a fluorescent (GFP) protein which is not identical in characteristics/properties to the fluorescent (GFP) protein or mutant/variant thereof that it is derived from.

Furthermore, it will be understood that these teachings are applicable to any fluorescent protein with a high A and T nucleotide content (AT-rich) compared to a host cell nucleic acid content.

In addition to the A and T nucleotide content, intrinsic curvature of the DNA is also important. DNA curvature or intrinsic curvature is related to the nucleotide content of DNA and refers to the curve of the DNA that is caused solely by the nucleotide content. This is influenced by the A/T content (since As and Ts have a higher internal bend angle leading to higher deflection from linear DNA) and also the sequence - there are certain combinations of nucleotides, including Gs and Cs, that can lead to increased bending of the DNA. The term "intrinsic curvature" is used to distinguish sequence determined bends

from bends that are introduced by DNA binding proteins. Promoter regions tend to be A/T rich to allow easy strand separation of the DNA. This often leads to promoter regions being intrinsically curved and, thus, good targets for the transcriptional repressor, H-NS. Thus, when the A and T nucleotide content of a nucleic acid is modified, the associated intrinsic curvature is similarly affected. Intrinsic curvature is a good measure/indicator of whether the modified nucleic acid does not have affinity to such transcriptional repressor proteins. An intrinsic curvature of low amplitude (for example below approximately 15° per helical turn, approximately 10° per helical turn, more preferably approximately 9° per helical turn, even more preferably approximately 7° per helical turn) is desirable.

Thus, the present invention involves modifying the A/T nucleotide residue content to equal or lower the average A and T nucleotide content of the host cell, carrying out codon optimization as needed and monitoring and modifying as necessary the intrinsic curvature of the nucleic acid to achieve an intrinsic curvature of low amplitude.

It will also be understood that the implications of these findings extend beyond host cells such as bacteria (*E. coli* and *Salmonella*) and yeasts. Thus, these teachings are applicable to a wide variety of host cells, whether prokaryotic or eukaryotic. Indeed, H-NS or H-NS-like proteins are widespread throughout many types of host cells including eukaryotic cells such as mammalian cells.

These findings have led to the present invention, which is directed to a method for improving gene and/or protein expression in a host cell. Specifically, we have developed a modified fluorescent protein gene which is impervious to interference by the transcriptional repressor proteins, such as H-NS, but which expresses an unaltered fluorescent protein. Hence, the fluorescent gene of the invention accurately reports transcriptional activity, due to the removal of the repressive effects of H-NS.

Advantageously, the new fluorescent gene of the invention can be used to monitor gene expression in bacteria such as *E. coli* without interference from H-NS binding to ensure that gene/protein expression in a host cell can now be conducted free from the undesirable complications that arise from transcriptional repressor proteins potentially binding to the fluorescent protein gene, with its associated bridging activity interfering with the faithful expression of that gene. Hence, the results should be more physiologically-relevant and free from artefacts caused by H-NS interference.

It is also postulated that lowering the A and T nucleotide content of a fluorescent protein in regions proximal to the promoter region and/or the ribosome binding site may improve gene/protein expression in a host cell.

Specifically, it is postulated that the high A/T content of the gfp+ gene may have had a direct role on transcription (independent of H-NS) since the location of A/T rich regions close to promoter regions can lead to reduced opening of the promoter region and thus reduced access for RNA polymerase and reduced transcription of the gene. This is because energy (usually supplied by underwound or "negatively supercoiled" DNA) is required to break the hydrogen bonds between the two DNA strands to allow access to the promoter region. The region requiring the lowest amount of energy separates first, using up the superhelical energy. Since A and T nucleotides only form 2 hydrogen bonds (G and C form 3), A/T rich regions become single stranded before regions with a lower A/T percentage. Therefore the presence of the A/T rich *gfp*+ gene proximal to a promoter region could influence the amount of superhelical energy (specifically, the amount of superhelical twist in the DNA) in the system and thus affect transcription of the gene of interest. This aspect is relevant to all A/T rich fluorescent proteins even if they weren't bound by H-NS. In this situation the A and T nucleotide content should be assessed and modified accordingly.

Further more specific embodiments of the first general aspect of the invention are given below.

According to one embodiment of the second aspect of the invention, the fluorescent protein nucleic acid is modified so that it is no longer A and T nucleotide rich (AT-rich) compared to the host cell nucleic acid average A and T nucleotide content. Ideally, the A and T nucleotide content of the fluorescent protein nucleic acid is modified to result in reduced affinity, compared to the wild type fluorescent protein nucleic acid or mutant thereof, to host cell transcriptional repressor proteins.

Preferably, the A and T nucleotide content of the fluorescent protein transcriptional repressor protein binding region nucleic acid is modified to equal or lower the A and T nucleotide content of the host cell transcriptional repressor protein nucleic acid binding region.

Preferably, the fluorescent protein nucleic acid has an intrinsic curvature of lower amplitude, for example below approximately 15° per helical turn, approximately 10° per

helical turn, more preferably approximately 9° per helical turn, even more preferably approximately 7° per helical turn)than the wild type fluorescent protein nucleic acid.

Advantageously, the host cell is a bacterium, preferably a gram-negative bacterium, more preferably *Escherichia coli*. Other host cells may be contemplated as described below.

According to another embodiment of this second aspect of the invention, the transcriptional repressor protein is a nucleoid-associated transcriptional repressor protein or repressor like protein, preferably H-NS or H-NS like proteins, such as Sfh and StpA.

The H-NS protein is a member of the family of nucleoid-associated proteins of bacteria. These proteins bind to DNA, regulate gene expression and organize the structure of the nucleoid (the part of the bacterial cell that contains the genetic material). H-NS is abundant (20,000 dimers per cell) and it binds to DNA sequences that have specific characteristics namely regions of intrinsic curvature and a high A and T content. Each H-NS dimer has two DNA binding domains and this facilitates the construction of DNA-H-NS-DNA. These bridges block the process of transcription (in which a gene is read by RNA polymerase) and silence the expression of genetic information. The H-NS protein binds to DNA sequences throughout the chromosomes of bacteria that express it, such as *Escherichia coli*, *Salmonella* and *Shigella*. The genes that it targets become silenced or down-regulated (Dorman, Nat. Rev. Microbiol. 5 (2007), 157-161).

H-NS is not the only DNA-protein-DNA bridge builder in biology. Such proteins are found in many cell types, including archaea and eukaryotes (Luijsterburg et al., Crit. Rev. Biochem. Mol. Biol. 43 (2008), 393-418). Importantly, a protein from the mouse has been shown to be able to substitute for H-NS in bacteria (Timchenko et al., EMBO J. 15 (1996), 3986-3992). Thus, the present invention relates to H-NS like proteins also.

According to another embodiment of this second aspect of the invention, the A and T nucleotide content of the entire fluorescent protein is modified compared to the wild type fluorescent protein nucleic acid or mutant thereof.

According to an alternative embodiment of this second aspect of the invention, the A and T nucleotide content of the fluorescent protein promoter region and/or ribosome binding site (RBS) nucleic acid is modified compared to the wild type fluorescent protein nucleic acid or mutant thereof.

Ideally, the A and T nucleotide content of the regions proximal to the fluorescent protein nucleic acid promoter region is modified such that the A and T nucleotide content to equal or lower the A and T nucleotide content of the host cell nucleic acid.

Advantageously, the fluorescent protein of the invention may be a green fluorescent protein (GFP), YFP, CFP, BFP or red fluorescent protein (DsRed) or a mutant or variant thereof.

According to a third general aspect of the invention, there is provided a modified fluorescent protein nucleic acid comprising a sequence encoding a wild type fluorescent protein or mutant thereof with an equal or lower A and T nucleotide content and/or reduced intrinsic curvature compared to the wild type fluorescent protein nucleic acid or mutant thereof. Ideally, the nucleic acid has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type fluorescent protein nucleic acid or mutant thereof.

According to one embodiment, the modified nucleic acid has reduced A and T nucleotide content across the entire length compared to the wild type fluorescent protein nucleic acid or mutant thereof.

According to an alternative embodiment, the modified nucleic acid has reduced A and T nucleotide content in the regions proximal to the promoter region and/or ribosome binding site (RBS) of the fluorescent protein nucleic acid compared to the same regions of the wild type fluorescent protein nucleic acid or mutant thereof.

Preferably, the modified nucleic acid has an A and T nucleotide content equal to or lower than a host cell average A and T nucleotide content.

Ideally, the host cell is a bacterium, preferably the Gram-negative bacterium *Escherichia coli* which has an A and T nucleotide content of approximately 50%.

According to another embodiment of the invention, the percentage of A and T nucleotides based on the full length modified nucleic acid sequence is from approximately 25% to 70% (for example the host cell *Mycobacterium tuberculosis* (a Gram-positive bacterium) has an approximate 35% A and T nucleotide content). Ideally, the A and T content is equal to or lower than that of the host cell.

According to a preferred embodiment of the invention, the nucleic acid comprises the nucleic acid sequence of Figures 4B or 6B or a sequence with at least 70%, preferably 80%, more preferably 85%, more preferably 90%, more preferably 95%, even more preferably 99% homology over the entire length to the nucleic acid sequence of Figures 4B or 6B. These are the "*gfp*T" and *DsRed*T genes of the following examples. The *gfp*T gene contains 157 nucleotide changes compared to the *gfp*+ coding sequence, without altering the amino acid sequence of the protein. Both *gfp* genes are 717 base pairs (bp) and thus vary at the nucleotide level by greater than approximately 20%.

Ideally, the nucleic acid of the invention has improved transcription compared to the wild type fluorescent protein nucleic acid or mutant thereof.

According to a fourth general aspect of the invention, there is provided a fluorescent protein encoded by the modified nucleic acid as described before.

According to a fifth general aspect of the invention, there is an expression system comprising the modified nucleic acid sequence or fluorescent protein as described above, preferably for use in a host cell such as *Escherichia coli*.

According to a sixth general aspect of the invention, there is provided a plasmid vector comprising the modified nucleic acid sequence or fluorescent protein as described above, preferably for use in a host cell such as *Escherichia coli*.

According to a seventh general aspect of the invention, there is provided a host cell comprising the modified nucleic acid sequence, fluorescent protein, plasmid vector, expression system as described above.

Ideally, the A and T nucleotide content of the modified fluorescent protein nucleic acid is equal to or lower than the A and T nucleotide content of the host cell nucleic acid.

Preferably, the fluorescent protein is a green fluorescent protein (GFP) or red fluorescent protein (DsRed). The green fluorescent protein mutant may be selected from the following; a spectral variant, a pHluorins, a variant with an altered Stokes shift, an oligomerization variant, a folding variant, a photoactivatable variant, a photoconversion variant, a photoswitchable variant, a redox sensitive variant and/or *gfp*$^{+}$.

The transcriptional repressor protein may be any nucleoid-associated repressor protein, preferably H-NS, repressor-like proteins, preferably H-NS like proteins such as Sfh and StpA.

According to a eighth aspect of the invention, there is provided a method of monitoring gene expression in a host cell using the modified nucleic acid, fluorescent protein, expression system or host cell of the invention.

The fluorescent gene of the invention may be used in many commercial applications. For example, it may be used in GFP-based kits for the study of gene expression.

Additionally, the modified nucleic acid of the invention may be inserted into a recombinant vector which may be any vector which may conveniently be subjected to recombinant DNA procedures. The choice of vector will often depend on the host cell into which it is to be introduced. Thus, the vector may be an autonomously replicating vector, i.e. a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, e.g. a plasmid. Alternatively, the vector may be one which, when introduced into a host cell, is integrated into the host cell genome and replicated together with the chromosome(s) into which it has been integrated.

The vector is preferably an expression vector in which the DNA sequence encoding the fluorescent protein of the invention is operably linked to additional segments required for transcription of the DNA. In general, the expression vector is derived from plasmid or viral DNA, or may contain elements of both. The term, "operably linked" indicates that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in a promoter and proceeds through the DNA sequence coding for the fluorescent protein of the invention.

The promoter may be any DNA sequence which shows transcriptional activity in the host cell of choice and may be derived from genes encoding proteins either homologous or heterologous to the host cell, including native *Aequorea* GFP genes.

Examples of suitable promoters for directing the transcription of the DNA sequence encoding the fluorescent protein of the invention in mammalian cells are the SV40 promoter (Subramani et al., Mol. Cell Biol. 1 (1981), 854-864), the MT-1 (metallothionein gene) promoter (Palmiter et al., Science 222 (1983), 809-814) or the adenovirus 2 major late promoter.

Specifically, the fluorescent protein gene of the invention may also be placed in plasmid vectors designed for simple cloning of e.g. the promoter of interest (transcriptional fusion) or the gene of interest (translational fusion) or both. While transcriptional fusions report promoter activity, translational fusions are often used to view the movement of the tagged protein by fluorescent microscopy.

Additionally, the fluorescent protein gene of the invention could also be integrated onto the chromosome to allow study of gene expression from its native location. One of the most effective ways to construct these fusions is using the lamda red mechanism, integrating the *gfp* gene or modified *gfp* (*gfp*T) gene after the end of the gene of interest's (GOI) coding sequence. While colonies can be screened for fluorescence using a FACs, fluorescent microscope, UV lamp or in some cases, by eye, we recommend using a linked selectable marker such as the kanamycin resistance cassette to allow initial selection of integrants.

Stilll additionally, a further version could also be developed for random insertion in bacterial genomes using transposon technology.

The number of available fluorescent reporter genes has increased in recent years as researchers have isolated genes encoding fluorescent proteins from an increasing variety of organisms and included the genes in cloning cassettes.

For example, fluorescent proteins from sea creatures have been used as reporter genes capable of integration into DNA via cloning cassettes. Products of these genes fluoresce under certain wavelengths of light, permitting the tracking of proteins in, e.g., heterologous cellls, such as dog and monkey cells. The most commonly used proteins of this nature fluoresce green, and were obtained from the jellyfish, *Aequorea victoria*, and sea pansy, *Renilla reniformis*. Additionally, a red fluorescent protein (RFP), known as drFP583, and a turquoise fluorescent protein, known as dsFP483, have been isolated from the IndoPacific mushroom corals (*Discosoma* sp. "red" and *Discosoma striata*, respectively). Both *Discosoma* and *Actinodiscus* are mushroom corals, soft bodied anthozoans that do not produce an external skeleton. It should be noted that the relationship between the genus *Discosoma* and the genus *Actinodiscus* is not well understood. Both *Actinodiscus* and *Discosoma* are members of the *Actinodiscidae* Family, which is a member of the *Corallimporpharia* (mushroom) Order. The taxonomy of the *Corallimporpharia* is poorly defined, and therefore, the nature of the relation of *Actinodiscus* to *Discosoma* is

uncertain. *Discosoma* and *Actinodiscus* are believed to be different genera of the same family, but they could be more closely or distantly related.

Finally, it will be understood that the teachings of this invention are equally applicable to other forms and mutations/variants of the *gfp* gene, such as those encoding proteins that fluoresce at different wavelengths (e.g. yellow), allowing improved versions of those genes also to be produced.

The Discoma red fluorescent protein (FP583, commercially known as DsRed) isolated by Matz in 1999 (Matz et al. Nat. Biotech. 17 (1999), 969-973), while providing a potential alternative to *gfp*+, contains the same intrinsic features make *gfp*+ a target for H-NS binding. The *dsred* gene is 55% A/T and is predicted to contain strong intrinsic curvature. Thus, the teachings of the present invention are also applicable to DsRed which may be optimized as described above to reduce H-NS affinity.

The invention will be more clearly understood by the following description of some embodiments thereof, given by way of example only, with reference to the accompanying drawings, in which:

Fig. 1 shows that H-NS binds to gfp^+ *in vivo* in accordance with Example 1. Chromatin immunoprecipitation (ChIP) using a H-NS specific monoclonal antibody was followed by quantitative PCR (qPCR). The Y-axis indicates fold enrichment relative to input DNA (DNA before addition of the H-NS specific antibody). Probe 2 showed over 12 fold enrichment over input DNA indicating H-NS binds in the *gfp*+ gene.

Figs. 2A and 2B relate to Example 2 and codon optimization. The new *gfp* gene (*gfp*T) of Example 2 contains reduced A/T content and DNA curvature compared to *gfp*+. The Bend.It program (Vlahovicek et al., Nucleic Acids. Res. 31 (2003), 3686-7) was used to determine the predicted curvature of the two *gfp* genes. Regions of strong intrinsic curvature in the *gfp*+ that are reduced in *gfp*T are indicated by filled arrows (Figure 2A). The new *gfp*T gene has reduced A/T content compared to the *gfp*+ gene (Figure 2B) While for the *gfp*+ gene the most A/T rich region is approaching 80% (average A/T content is 59%) the A/T content of some regions is reduced by over 20% in the new *gfp*T gene (average A/T content is 50%). The entire coding sequence (717bp) of both genes is shown in Figures 4A and 4B.

Fig. 3 shows the osmotic induction of *proU* fusions in accordance with Example 3.

Fig. 3A is a diagram of the downstream regulatory region (DRE) of *proV* containing H-NS binding sites essential for repression of *proU* in low osmolarity media. The

transcriptional start site of *proU* is indicated as +1. The positions of high-affinity H-NS binding sites and the points of insertion of the 3 reporter genes are indicated relative to +1.

Fig. 3B shows chromosomal *lacZ*, *gfp*+ and new *gfp*T fusions constructed at +98 bp (disrupts DRE) and Fig. 3C shows chromosomal *lacZ*, *gfp*+ and new *gfp*T fusions constructed at +936 bp (DRE intact) creating 3 reporter fusion strains each. β-galactosidase activity of the *proU-lacZ* fusions was monitored kinetically using a multiscan ascent plate reader (Thermo labsystems). The slope of at least ten samples was used along with the OD_{600} of the culture and volume of cells used to determine β-galactosidase activity. *gfp* fusions were measured using a flow cytometer (Beckman Coulter) and plotted by mean fluorescence values. Strains were grown overnight in LB with the indicated concentrations of NaCl as described previously (Lucht et al., J. Biol. Chem. 269 (1994), 6578-6586). Error bars show the standard error of the mean (SEM) of at least duplicate experiments.

Fig. 3(D) is a barchart showing the effect of removing H-NS on fusion gene expression in 100 mM NaCl. Mean values and ranges are plotted.

Fig. 3(E) is a table showing the fold-increases in fusion gene expression caused by changes in H-NS occupancy.

Figs. 4A and 4B shows the full coding sequences for the *gfp*+ gene and the new *gfp*T. *gfp*+ (Scholz et al. Eur. J. Biochem. 267 (2000), 1565-70) contains a number of mutations that improve the folding and emission spectrum of the GFP protein, optimizing it for use with flow assisted cell sorters (FACs). The sequence of *gfp*+ was obtained from pZep08 (Hautefort et al., App. Env. Microbiol. 69 (2003), 7480-7491). The new *gfp*T is the *gfp* made in accordance with Example 2.

Figure 5A is the full coding sequence for one of the most commonly used variants of the *gfp* gene, the *gfpmut2* gene. Recently, every potential promoter in *E. coli* was cloned into a low copy number vector containing the *gfpmut2* gene allowing for real-time monitoring of the transcriptional activity of the genome (Zaslaver et al. Nat. Meth. 3 (2006), 623-628).The *gfpmut2* differs to the *gfp*+ gene sequence by 48 nucleotides. The *gfpmut2* gene contains a higher A/T (61%) content than the *gfp*+ gene (59%) and also contains regions of high intrinsic curvature making it a strong H-NS target (Fig. 5B). 3 regions of strong intrinsic curvature are indicated by filled arrows.

Figure 6A is the full coding sequence for DsRed (Bartilson et al. Mol. Microbiol. 39 (2001), 126-135). Figure 6B is the full coding sequence for DsRedT. *gfp*T has reduced A/T content and intrinsic curvature and is predicted to be a poor target for H-NS. The DsRed coding sequence is 678 nucleotides in length, 55% of which are either an A or a T (Fig 6D). 2 regions of strong intrinsic curvature that are reduced in *dsred*T are indicated by arrows (Fig 6C). This high A/T content and high intrinsic curvature shows that DsRed

contains the same key determinants of H-NS binding affinity to a region as the *gfp*+ (Fig 2) gene and thus, is most likely bound by H-NS. Considering that DsRed is the basis for the most intensive research into fluorescent proteins since the *gfp* gene was isolated from *Aequorea victoria*, a wide range of derivatives of DsRed (Baird et al., Proc. Natl. Acad. Sci. USA 97 (2000), 11984-11989) may also be bound by H-NS.

Fig. 7A is a diagram of the the *fimA* promoter (P) is located in an invertible element (*fimS*). Site-specific recombination at the inverted repeats (IR), flanking *fimS*, results in inversion of the *fimA* promoter between phase ON and phase OFF.

Figs. 7B to D are the results of determining the response of *fimS* to novobiocin-induced DNA relaxation using a PCR based assay. Bands corresponding to phase ON and phase OFF are labelled. The percentage of phase ON and phase OFF switches in the population are shown below each gel. Fig 7B (Wild type) shows an increase in switching towards phase ON in response to increasing amounts of novobiocin. In contrast, in Fig 7C (*fimA-gfp*+), novobiocin causes an increase in switching towards phase OFF. Fig 7D (*fimA-gfp*T) behaves in a wild type manner, switching towards phase ON in response to novobiocin.

Fig 8 shows that H-NS binds to *gfp*+ with higher affinity than *gfp*T *in vitro*. Electrophoretic mobility shift assay (EMSA) analysis using purified H-NS and biotinylated *gfp*+ and *gfp*T probes (Fig 8A). Biotinylated *proU* and *lacZ* probes are used as positive and negative controls, respectively (Fig. 8B). H-NS binding to *gfp*+ and *gfpT* (Fig. 8C) in the presence of equal amounts of non biotinylated (unlabelled) probes. The concentration of purified H-NS used is indicated above each lane.

Fig. 9 shows that *gfp*T has improved translation efficiency.

Fig 9A is a graph showing the expression of *gfp*+ and *gfp*T from the *prpBCDE* promoter on the plasmid pPro in wildtype and *hns* mutant cells. The *prpBCDE* promoter is repressed by glucose and induced by propionate. Mean values and ranges are plotted; fold differences between *gfp*+ and *gfp*T fluorescence levels are shown above the bars.

Fig. 9B is a graph showing the codon adaptation index (CAI) values for *gfp*T and *gfp*+ in various organisms. Organisms above the dashed line will translate *gfp*T more efficiently than *gfp*+ (and wildtype *gfp*). [Organism classification] and abbreviations: [Alphaproteobacteria] *Rh.sph., Rhodobacter sphaeroides;* [Betaproteobacteria] *Ra.sol., Ralstonia solanacearum;* [Epsilonproteobacteria] *H.p., Helicobacter pylori;* [Gammaproteobacteria] *E.c., Escherichia coli; H.i., Haemophius influenzae; P.a., Pseudomonas aeruginosa; S.e., Salmonella enterica; V.c., Vibrio cholerae; X.c., Xanthomonas campestris; Y.p., Yersinia pestis;* [Actinobacteria] *M.s., Mycobacterium smegmatis; St.coe., Streptomyces coelicolor;* [Firmicutes] *B.s., Bacillus subtilis;*

[Eukaryota] *D.m., Drosophila melanogaster; H.s., Homo sapiens; M.m., Mus musculus; Sa.cer., Saccharomyces cerevisiae.*

EXAMPLES

GENERAL MATERIALS

gfp+ (Scholz et al. Eur. J. Biochem. 267 (2000), 1565-70), contains a number of mutations that improve the folding and emission spectrum of the GFP protein, optimizing it for use with flow assisted cell sorters (FACs).

- *gfp*^{T} is the *gfp* made in accordance with Example 2 of the invention which has been codon optomized.
- The following examples were carried out in the *E. coli* k12 strain, CSH50 (The Coli genetic stock centre).
- XL1 Blue was used as a cloning strain (Stratagene).
- pZep08 (Hautefort et al., App. Env. Microbiol. 69 (2003), 7480-7491) was used as the source of *gfp*+.
- All restriction emzymes used were from New England Biolabs (NEB).
- Phusion polymerase (Finnzymes) was used as per manufacturers instructions for all PCRs other than qPCR, where SYBR green qPCR mix (QuantiTect) was used.
- The plamid prep kit used was by RBC biosciences.
- All other standard reagents were from Sigma-Aldrich.
- Antibiotics, where needed, used in the following concentrations; Carbenicillin 100 μg/ml, Kanamycin 50 μg/ml and chloramphenicol 25 μg/ml.
- After integration of chromosomal cassettes and single colony pufirifaction under antiobiotic selection, the cassettes were considered stable and antibiotic selection unneccessary. Strains containing plasmids were maintained under selection to prevent plasmid loss

Example 1

H-NS binds to the *gfp*+ gene *in vivo.*

H-NS binding in *gfp*+ was confirmed *in vivo* by chromatin immunoprecipation using an H-NS specific monoclonal antibody and quantitative PCR.

Chromatin immunoprecipitation involves crosslinking proteins to DNA in live cells using formaldehyde, purifying the DNA-protein complexes and then using a specific antibody to

the protein of interest (in this case H-NS). This antibody-protein-DNA complex can then be purified, the protein removed from the DNA and the DNA quantified using quantitative real-time PCR. This identifies if a piece of DNA was bound by H-NS. The fold enrichment of the DNA is an indication of the affinity of the protein of interest for the DNA.

Materials and Methods

Chromatin immunoprecipitation (ChIP) using a monoclonal H-NS specific antibody was performed as described previously (Lucchini et al., PLoS Pathog. 2 (2006), e81). Quantitative PCR was performed on enriched DNA and unenriched (input) DNA using the Rotor-Gene 3000 real-time PCR machine (Corbett Research) and SYBR green (QuantiTect) as per manufacturers instructions. Input DNA was quantified using a Nanodrop 2000 (Thermo Scientific). The primers used were designed using Primer3 and are listed below.

ChIP primer list

gfp.chip.1.fw 5'-GGGTCATACCAACGTTCTGG-3'
gfp.chip.1.rv 5'-TTGTGCCCATTAACATCACC-3'
gfp.chip.2.fw 5'-TACAAGACGCGTGCTGAAGT-3'
gfp.chip.2.rv 5'-TGTGTCCGAGAATGTTTCCA-3'
gfp.chip.3.fw 5'-GGCATGGATGAGCTCTACAAA-3'
gfp.chip.3.rv 5'-TTTCCTTACGCGAAATACGG-3'
gfp.chip.4.fw 5'-TCACTACCGGGCGTATTTTT-3'
gfp.chip.4.rv 5'-TTGAGCAACTGACTGAAATGC-3'
gfp.chip.5.fw 5'-TGTCGGCAGAATGCTTAATG-3'
gfp.chip.5.rv 5'-CTGCCATTCATCCGCTTATT-3'

Each product was between 100 and 150 bp in length and encoded for only a single specific product (analyzed by agarose gel electrophoresis after a 40 cycle PCR). ChIP DNA samples were diluted 1 in 10 in AnalaR water (BDH). Each reaction was also performed using 20 ng of input DNA (4 ng/μl). The SYBR green PCR was set up in a 25 μl reaction containing 2.5 μl AnalaR water, 5 μl 1.5 μM forward and reverse primer mix, 12.5 μl SYBR green PCR mix and 5 μl of the DNA sample. The following thermocycle was used: 1. 95 °C for 15sec, 2. 52 °C for 60sec, 3. 72 °C for 15 sec. This was repeated for 40 cycles. Cycle threshold (C_T) values were extracted and the change in cycle threshold (ΔC_T) for each probe determined as follows; $\Delta C_T = C_T$ input sample - C_T ChIP sample.

The ΔC_T of the ChIP probe with the highest C_T value (i.e. the lowest amount of amplified DNA), was normalized to 0 by adding that ΔC_T to the ΔC_T of all of the ChIP probes. This ensures a positive value for the fold enrichment calculation.

Fold enrichment was calculated as follows; Fold enrichment= $(1+PCRyield)^{\Delta CT}$. The PCR yield refers to the efficiency of the amplification. Since PCR amplification is exponential, the amount of DNA is doubled in every cycle. Therefore, when the C_T values of 10 fold dilutions of input DNA are plotted against the log10 of the amount of input DNA per reaction, the slope should be -3.32 [$2^{3.32}$=10]. When the slope approaches this number the PCR yield=1.

Results and Conclusion

As shown in Figure 1, probe 2, located in the *gfp*+ gene, is strongly enriched for H-NS binding while the other probes are not. Probe 2 showed over 12 fold enrichment over input DNA indicating the presence of H-NS. This suggests that there is a strong H-NS binding site in this region (as supported by the DNA curvature and A/T content data, Figs. 2A and 2B). This data also infers that H-NS bound in this region is not nucleating along the DNA but instead, may be forming DNA-protein bridges with a second H-NS binding site This second site could be proximal to the *gfp*+ gene and lead to the transcriptional repression of a local gene (as described in Example 3) or distal, potentially forming a new topological domain and altering the expression of multiple genes.

Considering that most *gfp* transcriptional and translational fusions are located on multicopy plasmids, the possibility exists that global gene expression could be altered by the binding of H-NS in *gfp*, depleting the amount of free H-NS in the cell and thus leading to altered global gene expression.

H-NS titration from native genes was previously shown to occur upon the introduction of an of A/T rich plasmid DNA (Doyle et al., Science 315 (2007), 251-252).

This is further supported by the recent observation that the presence of a promoterless *gfp*+ gene on a multicopy plasmid caused a dramatic reduction in *Salmonella* invasiveness (Clark et al., Microbiology 155 (2009), 461-467).

It is also postulated that H-NS binding occurs in wild type *gfp* and it's derivatives as the minor modifications in the coding sequence required to alter the characteristics of the protein (often requiring only a single amino acid substation) are unlikely to alter the

determinants of H-NS binding and thus, the affinity of H-NS for the DNA. We have demonstrated in Example 2 how the 48 nucleotide difference between the *gfpmut2* and *gfp*+ genes is insufficient to significantly change the A/T content or curvature of the DNA. This instead requires a more thorough approach described in Example 2.

Example 2

Codon Optomization

As shown in Example 1, chromatin immunoprecipitation using a H-NS specific monoclonal antibody showed H-NS binding in the *gfp*+ gene (Fig. 1). This was supported by bioinformatic analysis of the *gfp*+ gene that showed high intrinsic curvature and high A/T content-typical features of H-NS bound regions (Fig. 2A).

Materials and Methods

To investigate if H-NS binding in *gfp*+ was affecting gene expression, we used the well-characterized H-NS regulated promoter *proU* and showed that *gfp*+ could functionaly replace a H-NS binding site in the *proU* DRE.

We then redesigned the *gfp*+ gene to be optimized for *E. coli* codon usage, which resulted in a drop in the A/T content and intrinsic curvature (Fig. 2B). This new *gfp*T gene was shown to be unable to replace the H-NS site in the *proU* DRE (described in the previous report), behaving similarly to the *lacZ* fusion (which has been shown to not bind H-NS). The characteristics of the *lacZ* gene were previously investigated (Owen-Hughes et a., Cell 71 (1992), 255-65) when differing results were obtained from *lacZ* and *luxAB* reporter fusions inserted in the *proU* DRE.

In vitro experiments exploiting the altered migration of curved DNA in agarose depending on the presence or absence of ethidium bromide showed that both *luxAB* and the *proU* DRE contained curved regions, while the migration of *lacZ* gene was not altered by the presence of ethidium bromide and thus, the *lacZ* gene was not intrinsically curved.

The affinity of purified H-NS for *luxAB*, *proU* DRE and *lacZ* was assessed using electrophoretic mobility shift assays (EMSAs). EMSA involves incubation of a purified protein with purified DNA followed by electrophoresis through a polyacrylamide gel. The DNA that is bound by the purified protein migrates slower than unbound DNA and thus can be monitored and quantified. The *lacZ* DNA was reproducibly shown to be bound with less efficiency (required more H-NS) than both the *luxAB* DNA and the *proU* DRE region

(Owen-Hughes et al., Cell 71 (1992), 255-65). This was consistent with their observation that H-NS preferentially binds to curved DNA.

As mentioned, the initial basis for their study was the differences in expression between the *luxAB* and the *lacZ* reporter fusions inserted in the *proU* DRE. While the *luxAB* fusion showed repression of *proU* of in low osmolarity media, the *lacZ* fusion showed derepression of *proU* under the same conditions.

The differences in transcription were due to H-NS binding in *luxAB* and the lack of H-NS binding in the *lacZ* reporter fusion (Forsberg et al., J. Bacteriol 176 (1994), 2128-32).

Analysis of DNA curvature was performed using the bend.it server as described in (Vlahovicek et al., Nucleic Acids. Res. 31 (2003), 3686-7).

The prediction of DNA curvature can be achieved using a number of different models. Rod models are the simplest form of DNA models and represent DNA as a cylindrical rod of constant diameter, made up of short cylindrical segments (e.g. the size of a base pair) and then to compute a given rod parameter (e.g. DNA curvature) on the basis of segment composition. Dinucleotide models define the segment as two adjacent base pairs while trinucleotide models define the unit around the central base pair of a given trinucleotide. Analysis of DNA curvature was performed using the bend.it server, which calculates the curvature of DNA molecules based on their DNA sequences using dinucleotides and trinucleotides models.

The genetic code uses 64 nucleotide triplets (codons) to encode 20 amino acids and stop, meaning each amino acid is encoded by on average 3 codons. The frequencies with which codons are used by different organisms varies significantly leading to variation in G/C content. The degeneracy of the genetic code enables many alternative nucleotide sequences to encode the same protein and allows for the codon optomization of a protein to a specific organism, without altering the protein at the amino acid level.

The simplest way to redesign a DNA sequence is to work from the amino acid sequence and use a 'one amino acid – one codon' approach where, for every amino acid, the most abundant codon for the organism of interest is used. A highly expressed gene designed as such would result in depletion of the transfer RNA pool for those codons, potentially allowing the incorrect incorporation of another tRNA leading to translation error. In contrast to this method, the program for optimization of *gfp*+ (Gene Designer) (Villalobos et al.,

BMC Bioinformatics 7 (2006), 285) optimizes genes for expression by using a codon usage table in which each codon is given a probability score based on the frequency distribution of the codons in the desired genome normalized for every amino acid. For the redesign of the *gfp*+ gene, we used the EColi_CII table that is derived from a collection of highly expressed *E. coli* genes. This approach avoids the use of rare codons which are strongly associated with low levels of protein expression due to ribosome stalling and abortive translation.

While this ensured that the new gfp gene (*gfp*T) would be highly expressed in *E. coli*, the most important factor in the codon optimization was the corresponding reduction in A/T content and reduction in curvature.

Since candidate sequences are generated *in silico* using a Monte Carlo based algorithm where each codon choice is an independent probabilistic event, the software can perform the optimization on multiple occasions finding a new and equally optimal solution each time. This allowed for screening of the potential new *gfp* genes for reduced curvature. The variant chosen (*gfp*T) was synthesized and used in Example 3.

We chose to optimize the *gfp*+ gene for *E. coli* to give it an A/T content of 50%. We chose this because the majority of bacterial research is conducted in *E. coli* and *Salmonella*. As mentioned earlier, H-NS targets horizontally acquired A/T rich DNA, so making the *gfp*+ gene a similar average A/T content to that of the *E. coli* genome (50%), was intended to allow it blend in and not be target by H-NS as an horizontally acquired gene. This method could be used for all other fluorescent proteins to prevent H-NS binding in the introduced DNA.

Results and Conclusions

Codon optomization reduced the A/T richness and intrinsic curvature of the *gfp*+ gene. Figure 2A shows the intrinsic curvature results for the *gfp*+ and *gfp*T genes and Figure 2B shows the A/T content results for the the *gfp*+ and new *gfp* ^{T} gene.Fig 2 shows that the new *gfp*T gene contains reduced A/T content and DNA curvature compared to *gfp*+.

The entire coding sequence (717 bp) of the two *gfp* genes was analysed using the Bend.It program that determines the predicted curvature (Figure 2A) and A/T content (Figure 2B) of DNA. Regions of strong intrinsic curvature in the *gfp*+ gene that are reduced in the *gfp*T gene are indicated by filled arrows.

While for the *gfp*+ gene the most G/C rich region is 0.6 (60%), the new *gfp*T gene has strongly G/C rich regions approaching 0.8 (80%). This reflects the overall drop in A/T content between the *gfp*+ gene (59%) and the *gfp*T gene (50%)

These plots indicate a general reduction in both A/T content and predicted curvature of the *gfp*+ gene, thus reducing two determinants of H-NS binding affinity. This data is a strong indicator that the intrinsic qualities of the DNA that allow H-NS to bind in the *gfp*+ gene are not present in the *gfp*T gene. The data also shows that the *gfpmut2* gene is A/T rich and contains strong predicted curvature, and thus is most likely also bound by H-NS. It is therefore probable that many of the *gfp* variants contain similar traits and are bound by H-NS.

Example 3

New gfp^{T} gene as a transcriptional reporter to a known H-NS regulated gene

Previous studies have shown H-NS binding in the downstream regulatory region (DRE) of *proU* and that H-NS binding in this region was responsible for repression of *proU* under low osmolarity conditions (Bouffartigues et al., Nat. Struct. Mol. Biol. 14 (2007), 441-8). In this example, we use the new *gfp*T gene (as developed in Example 2) as a transcriptional reporter to a known H-NS regulated gene.

Materials

Primer list:

fimA-gfp+-cat. fw

5'- GAT TGA TGC GGG TCA TAC CAA CGT TCT GGC TCT GCA GTA ATG AGC GTT CTA GAT TTA AGA- 3'

fimA-gfp+-cat. rv

5'- TCT GCA CAC CAA CGT TTG TTG CGC TAC CCG CAG CTG AAC TCT ACG AGA CAG CAC ATT AAC- 3'

kan.int.fw

5'-AAT GTC ATG ATA ATA ATG GTT TCT TAG ACG TCA GGT GGC GTG TAG GCT GGA GCT GCT TCG - 3'

kan.int.rv

5'-AGT TCA GCT GCG GGT AGC GCA ACA AAC GTT GGT GTG CAG ACA TAT GAA TAT CCT CCT TA - 3'

+98proU-gfp+.fw

5' - GGC AAT TAA ATT AGA AAT TAA AAA TCT TTA TAA AAT ATA ATG AGC GTT CTA GAT TTA AGA - 3'

+936proU-gfp+.fw

5' - CCG CCG GAC ACC GAA TGG CTT AAT TCG TAA AAC CCC TTA ATG AGC GTT CTA GAT TTA AGA - 3

+98proU-gfpT.fw

5' - GGC AAT TAA ATT AGA AAT TAA AAA TCT TTA TAA AAT ATG ATT GAT TAA GAA GGA GAT - 3'

+936proU-gfpT.fw

5' - CCG CCG GAC ACC GAA TGG CTT AAT TCG TAA AAC CCC TTG ATT GAT TAA GAA GGA GAT - 3'

proU.(stop).kan.rv

5' - GTC CGC CGC TGG CGT GGT ATC CCA CGG ATT ATT TTG ATC ACA TAT GAA TAT CCT CCT TA - 3'

+98proU-lacZ.fw

5' - GCA ATT AAA TTA GAA ATT AAA AAT CTT TAT AAA ATA TAA TGA GCG GAT AAC AAT TTC ACA - 3'

+936proU-lacZ.fw

5' - CGC CGG ACA CCG AAT GGC TTA ATT CGT AAA ACC CCT TAA TGA GCG GAT AAC AAT TTC ACA - 3'

proU.(stop).cat.rv

5' - GTC CGC CGC TGG CGT GGT ATC CCA CGG ATT ATT TTG ATC ACA TAT GAA TAT CCT CCT TAG - 3'

kan.XmaI.fw

5' – CAT AAC GAG CCC GGG TGT AGG CTG GAG CTG CTT C- 3'

kan.XmaI.rv

5'- CAT AAC GAG CCC GGG CAT ATG AAT ATC CTC CTT A- 3'

fimA.fw

5'- GGA AAG CAG CAT GAA AAT TAA AAC TC - 3'

fimA.rv

5'- GGT TAT TGA TAC TGA ACC TTG AAG G - 3'

proU.fw

5'- AGG GGT TGC CCT CAG ATT CTC - 3'

proU.rv

5'- GTC AGT CGG TGC AGT CGT C - 3'

Methods

Integration of linear (PCR amplified DNA) was performed as described by Datsenko and Wanner (Proc. Natl. Acad. Sci. USA 97 (2000), 6640-6645). The linear DNA was obtained as follows;

The *gfp*+-*cat* cassette from pZep08 was PCR amplified using primers *fimA-gfp*+-*cat*.fw and *fimA-gfp*+-*cat*.rv and integrated into the chromosome of the CSH50 in the fimA gene selecting for chloramphenicol resistance (encoded by the *cat* gene). The kanamycin resistance cassette (*kan*) was PCR amplified from pKD4 Datsenko and Wanner (Proc. Natl. Acad. Sci. USA 97 (2000), 6640-6645) using primers *kan*.int.fw and *kan*.int.rv that have 5' extensions allowing for chromosomal replacement of the cat gene in the *fimA-gfp*+-*cat* construct. The chromosomally located *gfp*+-*kan* cassette was then used as template for integration of *gfp*+-*kan* into the *proU* gene. PCR products amplified using either +98*proU-gfp*+.fw or +936*proU-gfp*+.fw along with the *proU*.(stop).*kan*.rv were integrated onto the chromosome creating *gfp*+, transcriptional fusions to the *proU* promoter.

The *gfp*T gene was synthesized by DNA2.0 (San Diego, CA, USA) and provided cloned into a custom vector (pJ204, pUC ori, encodes ampicillin resistance). The *kan* cassette from pKD4 was PCR amplified using primers kan.XmaI.fw and kan.XmaI.rv, which incorporated XmaI sites to allow for cloning into the XmaI site of pJ204 (located about 30bp downstream of the *gfp*T stop codon). The linear and plasmid DNA were digested with XmaI (NEB), ligated using T4 DNA ligase (Roche) as recommended by the manufacturer's instructions.The ligated DNA was then transformed by heat shock into the *E. coli* cloning strain XL1 Blue made competent for transformation using calcium chloride (These are standard laboratory techniques). Plasmids were purified (RBC biosciences) and used as template using primers; +98*proU-gfp*T.fw or +936*proU-gfp*T.fw along with *proU*.(stop).*kan*.rv. The amplification of a ~2 Kb product confirmed which colonies contained the *kan* gene in the correct orientation (promoter facing away from the *gfp*T coding sequence). These PCR products were then integrated into the chromosome creating *gfp*T transcriptional fusions to the *proU* promoter.

M1655 *lacZY-cat* (an *E. coli* strain containing the *cat* gene inserted in *lacA*; obtained from D. M. Stoebel, Department of Microbiology, School of Genetics and Microbiology, Trinity College, Dublin 2, Ireland.) using the primers +*98proU-lacZ*.fw or +*936proU-lacZ*.fw and the *proU*.(stop).*cat*.rv primer. These PCR products were then integrated into the chromosome creating *lacZ* transcriptional fusions to the *proU* promoter.

A stop codon (TAA or TGA) was included in the every forward primer so that it integrated in-frame and prevented formation of a translational fusion.

Presumptive integrants were screened for an increased size compared to the wild type (WT) gene using either *fimA*.fw and *fimA*.rv (that amplify a ~550 bp region of *fimA* in WT CSH50) or *proU*.fw and *proU*.rv (that amplify a ~1.4 Kb flanking region of the *proU* promoter in WT CSH50).

The presence of either *gfp+-cat* or *gfp+-kan* in *fimA* lead to an increase in size of ~2 Kb compared to WT since the integrations were made to insert at a single site.

Integration in *proU* causes a ~1.2Kb deletion when integrated at +98 bp from the *proU* transcriptional start site or a ~330 bp deletion when integrated at +936 bp from the *proU* transcriptional start site. The 3' integration event is directly before the stop codon in the *proV* (the first gene transcribed by the *proU* promoter).

Fusions of the correct predicted size (as analyzed by agarose gel electrophrophesis) were sequenced on both strands to ensure correct integration (GATC Biotech).

PD32 is an H-NS deficient strain containing the *bla* gene (that encodes resistance to the antibiotic ampicillin) inserted in the *hns* gene (Dersch et al., *Mol. Microbiol.* 8, (1993), 875-889). The mutant *hns* allele was transduced using phage P1vir (a standard technique in molecular biology) into the *proU* fusion strains to analyse expression from *proU* in the absence of H-NS.

Osmotic induction of proU fusions:

Single colonies were resuspended in 100 µl of L broth containing no salt (L0) and used to innoculte 3 ml broths containing increasing amounts of salt (NaCl). Cultures were incubated in aerated conditions overnight at 37 °C and 200rpm. 20 µl samlpes for flow cytometry (*gfp+*/*gfp*Tfusions) were added to 1ml of 2% (vol/vol) formaldehyde/phosphate buffer saline (PBS). Samples were left at 4 °C under tin-foil until measuring fluorescence using a flow cytometer (Beckan Coulter).

β–galactosidase activity of the *lacZ* fusions were measured using a described by Miller (1992) with minor differences. Reactions were performed in 96-well microtiter plates. The kinetics of substrate hydrolysis at 37 °C was measured for at least ten samples, at 30

second intervals after an initial 3 minute lag period, using a multiscan ascent plate reader (Thermo labsystems). The total volume of each reaction was maintained at 200 µl.

Beta-galactosidase activity was determined according to the following formula:

Slope (OD_{414}/time)/(OD_{600} x volume (µl) of cells used)

Sodium chloride (NaCl) was used to increase the osmolarity of the media as described previously (Lucht et al., J. Biol. Chem. 269 (1994), 6578-6586)

Results and Conclusion

Figure 3A shows the downstream regulatory region (DRE) of *proV* containing H-NS binding sites essential for repression of *proU* in low osmolarity media

Chromosomal *lacZ*, *gfp*+ and new *gfp*T fusions were constructed at +98 bp (disrupts DRE) (Fig. 3B) and +936 bp (DRE intact) (Fig. 3C) creating 6 reporter fusion strains.

We found that, as expected, insertion of a *lacZ* reporter gene at +98 bp causes derepression of the *proU* promoter at low salt. This is because H-NS does not bind in *lacZ* and its insertion at +98bp from the transcriptional start site deletes the H-NS binding site in the downstream regulatory element (DRE) (Fig 3A)

Insertion of the *gfp*+ gene at an identical site caused no derepression at low salt, demonstrating that H-NS binding in the *gfp*+ gene could functionally replace the well characterized, high affinity H-NS binding site in the DRE (Fig. 3B).

Insertion of the new *gfp*T gene disrupting the DRE leads to derepression of the *proU* promoter at low salt (similar to *lacZ*) indicating that H-NS does not bind in the new *gfp*T gene (Fig 3B). As a control, each reporter fusion was inserted at +936 bp leaving the DRE intact. As expected, these constructs showed repression of *proU* at low salt (Fig 3B).

To confirm that H-NS binding to *gfp*+ accounted for repression of *proU-gfp*+(+98) *in vivo*, the expression of all *proU* fusions was tested in an H-NS deficient background. Cells were cultured in the repressive conditions of 100 mM NaCl, and the data expressed as a percentage of maximal derepressed expression to facilitate comparisons between GFP fluorescence and β-galactosidase activity (Fig 3D). For all three fusions, *proU* expression was elevated in the absence of H-NS. This revealed that even in the absence of the DRE, H-NS continued to bind the URE and repress the *proU-gfp*T(+98) and *proU-lacZ*(+98)

fusions. These data allowed an assessment of the relative effects of replacing only the DRE compared to elimination of H-NS protein from the cell (Fig. 3E). Replacing the DRE with *gfp*T resulted in an 8-fold increase in expression relative to expression from the *proU-gfp*T(+936) fusion. A similar comparison between *proU-gfp*+ at position +98 and +936 showed that *gfp*+ inserted in the DRE maintained full repression. Consequently, eliminating H-NS genetically resulted in a 7.3 to 9.4-fold increase in expression from the *gfp*+ fusions while the expression of *proU-gfp*T(+98) improved only 2.2-fold upon removal of H-NS (Fig 3E).

These experiments were based on the previously described observation that H-NS binding in the DRE is required to maintain repression of the *proU* promoter in low osmolarity media. The derepression of the *proU* promoter upon the insertion of the *gfp*T gene at +98 bp indicates that the codon optimization process which reduced the A/T content and predicted curvature of the *gfp*T gene, was sufficient to prevent H-NS binding. The *gfp*T gene is shown here to be a faithful reporter of gene activity. Equally the *gfp*+ gene is shown to be an unfaithful reporter of gene activity, effectively replacing the native H-NS binding site in the DRE.

While these results suggest that fluorescence levels of the GFP protein have not been altered (Fig 3C), it is shown in Example 7 that the GFPT is translated more efficiently than GFP+ due to the codon optimization of GFPT (as described in Example 2).

Example 4

New *gfpT* gene as a transcriptional reporter to a second known H-NS regulated gene.

Summary

Control of type 1 fimbrial expression in *Escherichia coli* is controlled through inversion of a 314-bp invertible element (*fimS*), which contains the promoter for the major fimbrial subunit *fimA* (Fig 7A).

The site-specific recombinase FimB binds at the inverted repeats catalyzing the inversion of the *fimS* DNA segment leading to either fimbriate (phase ON) or afimbriate cells (phase OFF).

The inversion of *fimS* is sensitive to varying levels of DNA supercoiling. DNA supercoiling is controlled through the antagonistic actions of DNA gyrase (which tightly winds the DNA) and topoisomerase I (which relaxes the DNA). Under normal conditions, inversion from phase ON to phase OFF and from phase OFF to phase ON occurs at an equal rate. When the DNA is relaxed due to the addition of the DNA gyrase inhibiting drug novobiocin, switching from phase OFF to phase ON dramatically increases (Fig 7B).

We found that the introduction of the *gfp*+ gene led to a reversal in this phase ON switching bias, with the population instead biasing towards phase OFF in response to novobiocin (Fig 7C). Switching in an isogenic *fimA-gfp*T fusion strain biased towards phase ON in response to novobiocin (Fig 7D), indicating that the repressive effects on inversion of *fimS* caused by the presence of the *gfp*+ gene are absent in the new *gfp*T gene.

The negative effect of *gfp*+ on *fimS* inversion was most likely due to H-NS binding in *gfp*+ and interacting with H-NS bound in *fimS*, allowing new DNA bridges to form around *fimS*. This would alter local DNA topology and may affect the interaction of the FimB recombinase with the inverted repeats.

Materials and methods

Novobiocin sodium salt (Sigma) was prepared fresh in sterile water before use. The construction of the *fimA-gfp+-kan* fusion was described in Example 3. A *gfp*T-*kan* cassette was integrated into *fimA* using the primers

fimA-gfpT-kan:
fw 5' – GAT TGA TGC GGG TCA TAC CAA CGT TCT GGC TCT GCA GGA TTC ATT AAG AAG GAG AT– '3

fimA-gfpT-kan:
rv 5' – TCT GCA CAC CAA CGT TTG TTG CGC TAC CCG CAG CTG AAC TCA TAT GAA TAT CCT CCT TA– '3

Novobiocin induction was performed as follows. Briefly, single colonies were resuspended in 100 µl LB broth and used to innoculate 3 ml of LB in a 15 ml test tube, which were then incubated at 37 °C/ 200 rpm (aerobic conditions). Cultures were allowed grow until exponential phase (when the optical density at 600nM (OD_{600}) is between 0.2 to 0.4) at which point they were diluted and used to innoculate 3ml broths containing varying concentrations of novobiocin. Diltutions were calculated to allow 15 generations before cessation of growth at an OD_{600} of ~3, assuming each generation resulted in a doubling of the OD_{600} of the culture.

The orientation of *fimS* was determined using a PCR based assay. Primers OL20 (5' – CCG TAA CGC AGA CTC ATC CTC – '3) and OL4 (5' – GAC AGA ACA ACG ATT GCC AG – '3) were used to amplify from outside of the invertible region. The resulting PCR products were digested at a unique BstUI site, asymmetrically located within *fimS*, allowing for distinction between cells containing *fimS* in the phase ON or phase OFF orientation based on the size of the digested fragments (see Figs 7B-D). Digested fragments before electrophoresed through a 2% agarose/TAE gel containing ethidium bromide (1 µg/ml). Gels were visualized under ultraviolet light (AlphaImager 2200, AlphaInnotech). The bands corresponding to phase OFF and phase ON were measured by densitometry using Image J software, and the percentage of phase OFF and phase ON cells was calculated. Each gel shown is representative of at least 3 independent experiments.

Results and Conclusions

These results shown in Figures 7B to D show that the presence of the *gfp*+ gene can have a dramatic effect on the inversion of *fimS*, while the *gfp*T gene has no such effect. That a gene designed to act as a transcriptional reporter can have such a dramatic effect on the architecture of the DNA has serious implications for the results obtained from previous studies using *gfp*. If, as we have shown, *gfp*+ can affect a recombination event, it is possible that the presence of the *gfp* gene could affect other processes requiring recombination events such as chromosome replication, plasmid partitioning and phage

integration. Any alteration in these events could have a dramatic affect on the cell, possibly slowing growth, causing plasmid loss or have other unforeseen consequences, compromising the validity of the results obtained by using the reporter fusion.

Example 5

H-NS binds to *gfp*+ with higher affinity than *gfp*T *in vitro*

Summary

The regulatory regions flanking the *proU* promoter are A+T rich, highly intrinsically curved (Owen-Hughes et al., Cell 71 (1992), 255-65) and contain multiple high affinity H-NS binding sites (Bouffartigues et al., Nat. Struct. Mol. Biol. 14 (2007), 441-8). This DNA was therefore used as a positive control for H-NS binding (Figure 8). The *lacZ* reporter gene is a poor target for H-NS as it is relatively G+C rich and not intrinsically curved (Owen-Hughes et al., Cell 71 (1992), 255-65) and was used as a negative control for H-NS binding (Figure 8).

Electrophoretic mobility shift assays (EMSA) were performed to determine the affinity (K_{app}) of H-NS for *gfp*+ and *gfp*T *in vitro*. Since H-NS binds with low specificity and affinity and H-NS binding is highly co-operative, in order to assess H-NS binding affinity accurately for the two *gfp* genes a narrow range of protein concentrations was chosen. H-NS was found to bind *gfp*+ strongly with a K_{app} of 4.9 nM (Figure 8A). A further indication of the high affinity of H-NS for *gfp*+ is the narrow range of protein (4.5-10.55 nM) required for the transition from initial binding to fully bound probe, resulting in a single high molecular mass complex. This also illustrates the highly co-operative nature of H-NS binding. H-NS had a lower affinity for *gfp*T (Figure 8A; Kapp, 7.5 nM). The lower affinity of H-NS for *gfp*T resulted in smearing of the DNA over a wide range of protein concentrations (7.9-18.75nM) with the *gfp*T probe only resolving as a single bound complex at 25nM H-NS.

The *proU* regulatory region was used as a positive control for H-NS binding (Figure 8B). As expected, the *proU* probe was strongly bound by H-NS (K_{app} 6.2). The *proU* region contains a number of well characterized H-NS binding sites (Bouffartigues et al., Nat. Struct. Mol. Biol. 14 (2007), 441-8) and resolved as two separate high affinity complexes (arrowed). *lacZ* was poorly bound by H-NS (K_{app} 16.3) and the resolution of a single protein-DNA complex at high H-NS concentrations (25 nM) simply highlights the low specificity of H-NS, which at saturating concentrations binds independently of sequence to DNA (Tupper et al., EMBO J. 13 (1994) 258-268). Although the change in K_{app} between

gfp+ and *gfp*T measured *in vitro* was relatively small, the difference in binding affinity was highly significant *in vivo*. H-NS affinity for *gfp*+ and *gfp*T was also compared in the same reaction using biotinylated and unlabelled DNA in equal amounts (50 pM) (Figure 8C). These data showed that when both genes were present, H-NS bound specifically to *gfp*+ and only bound *gfp*T when all the *gfp*+ DNA had been bound (14.1 nM).

Materials and Methods

H-NS binding to each probe (0.4 ng DNA per reaction) was carried out in 20 ul reaction mixtures containing increasing concentrations of purified H-NS protein (final concentrations; 0-25nM) in 20mM Tris HCL, 1 mM EDTA, 100 ug/ml BSA, 1 mM DTT, 10% glycerol and 80 mM NaCl. Reactions were incubated at 4 °C for 30 min. 10 µl of each reaction was loaded (without the addition of loading dye) onto a 5% poly-acrylamide gel [5% acrylamide/bisacrylamide (30:1) (National Diagnostics), 2% glycerol, 0.5X TBE] and electrophoresed at 90 V for 2 h (4 °C) followed by electrophorectic transfer (30V for 1 hr) to Biodyne B 0.45 µM membrane (Pall). 0.5X TBE (45 mM Tris-borate [pH 8.3] containing 1.25 mM disodium EDTA), was used as both running and transfer buffers. The wet membrane was UV treated twice at 150 mJoule in a GS Gene Linker UV chamber (Biorad). The Chemiluminescent Nucleic Acid Detection Module (Pierce) was used as per manufacturers instructions followed by signal detection using developer and fixer solutions (Kodak) and Hyperfilm (Amersham Biosciences). Densitometric analysis was performed using Image J software.

PCR was used to amplify the entire coding sequence of *gfp*+ and *gfp*T and equivalently sized regions (717 bp) in *proU* and *lacZ*. The primers used are listed 5' to 3' below;

gfp+.bs.fw	ATG AGT AAA GGA GAA GAA CTT TTC
gfp+.bs.bio.rv	TTA TTT GTA GAG CTC ATC CAT G
gfpT.bs.fw	ATG AGC AAA GGC GAA GAG CT
gfpT.bs.bio.rv	TTA CTT ATA CAG TTC ATC CAT ACC G
proV.bs.fw	AGG GTG TTA TTT TCA AAA ATA TCA C
proV.bs.bio.rv	CAT ATG CGG CAT TAA GGC AA
lacZ.bs.fw	ATG ACC ATG ATT ACG GAT TCA CTG G
lacZ.bs.bio.rv	AGC GCG GCT GAA ATC ATC AT

Bio indicates primers that contained a 5' biotin tag that allows for visualization of the DNA. Primers with identical sequences to gfp+.bs.bio.rv and gfpT.bs.bio.rv but without 5' biotin tags were used as unlabelled DNA in Figure 8C.

Example 6

Optimization of a second fluorescent protein, DsRed, to reduce H-NS affinity

The same method of optimization as applied to *gfp*+ was used to alter the *dsred* gene *in silico*. The selected optimized gene, *dsred*T, differs from *dsred* by 155 nucleotide substitutions across the 678 base-pair gene has reduced A/T content (49%) and reduced predicted DNA curvature (Figure 6A to C). We predict that this optimized *dsred*T will have a a lower H-NS affinity than *dsred*.

Example 7

***gfp*T is translated more efficiently than *gfp*+**

Materials and methods

Cloning *gfp* variants in plasmid pPro

Blunt-ended PCR amplicons of the *gfp*+ and gfp^T open reading frames were generated using Phusion polymerase (NEB) and primers gfp+.pPro24-blunt.fw, gfp+.pPro24-PstI.rv, gfp^T.pPro24-blunt.fw, gfp^T.pPro24-PstI.rv (listed 5' to 3' below).

gfp+.pPro24-blunt.fw	AGT AAA GGA GAA GAA CTT TTC
gfp+.pPro24-PstI.rv	TCT ACT GCA GTT ATT TGT AGA GCT CAT CCA TG
gfpT.pPro24-blunt.fw	AGC AAA GGC GAA GAG CTG TTC
gfpT.pPro24-PstI.rv	TCT ACT GCA GTT ACT TAT ACA GTT CAT CC

PCR amplicons were first digested with PstI, then blunt ends were phosphorylated using T4 polynucleotide kinase in T4 ligase buffer (Roche Diagnostics, Mannheim, Germany) followed by purification with a HiYield gel/PCR DNA fragments extraction kit (RBC Biosciences). pPro24 (Lee & Keasling, 2005) was digested with SmaI and PstI, dephosphorylated using Antarctic phosphatase (NEB), and then ligated to PCR amplicons using a Rapid DNA ligation kit (Roche). Correct clones were confirmed by DNA sequencing.

The *prpBCDE* promoter in pPro24-*gfp* clones was induced with propionate as follows: streak-isolated colonies were used to inoculate 4 ml LB (86 mM NaCl) broth cultures and these were grown to an OD_{600} ~0.5. Cultures were then diluted 1/500 into fresh LB including glucose (to repress the *prpBCDE* promoter) or propionate (to induce it). Cultures

were grown overnight at 37 °C with shaking and samples were fixed and analysed by flow cytometry the following morning.

Calculation of Codon Adaptation Index (CAI) values.

CAI values were calculated using CAIcal (Puigbo et al., Biol Direct. 3, (2008), 38) from codon usage tables provided by the Codon Usage Database (Nakamura et al., Nucl. Acids Res. 28, (2000), 292). All codon usage tables used in this analysis were derived from whole genome sequences.

Results

Improved translation efficiency was confirmed by cloning *gfp*+ and *gfp*T in the *pPro* vector under control of the propionate-inducible *prpBCDE* promoter (Lee et al., Appl. Environ. Microbiol. 71 (2005), 6856- 6862.), for which there is no evidence of H-NS binding or repression (Dillon et al., Mol. Microbiol. (2010), doi:10.1111/j.1365-2958.2010.07173.x.). In *pPro*, gfpT produced on average 3.5-fold more GFP in a wildtype background and 2.7-fold more GFP in a *hns* mutant background compared to *gfp*+ (Figure 9A). The codon adaptation index (CAI) is a standard means to calculate the effects of species-specific codon biases on translation (Sharp et al., Nucl. Acids Res. 15, (1987), 1281-1295.). Figure 9B shows that *gfp*T is predicted to have improved translation efficiency in both bacterial and eukaryotic model organisms.

High levels of GFP, usually associated with expression from a multicopy plasmid, have previously been shown to be toxic to the host cell, leading to dramatic plasmid loss or mutation in order to reduce the amount of GFP being produced (Hautefort et al., App. Env. Microbiol. 69 (2003), 7480-7491). It is possible that a part of this toxicity is due to the sub-optimal codon usage of *gfp*+ leading to reduced translation efficiency. This would cause an increased burden on the translational machinery (including the ribosomal components), preventing efficient translation of essential host genes and thus, create toxic effects on the host.

GFPT is optimized for high expression in *E. coli* and thus may have reduced toxicity when highly expressed.

Another potential advantage may be increased fluorescence of GFPT vs GFP+ when the genes are weakly transcribed. This could arise due to the optimized coding sequence of GFPT allowing more efficient use of the host transfer RNA pool than GFP+, allowing faster translation and thus, a greater accumulation of fluorescent protein. This would result in

GFPT cells having a higher level of fluorescence than the GFP+ cells under identical conditions.

Therefore, GFPT potentially has a wider range of fluorescence than GFP+, without having the same detrimental effect on the host.

In the specification the terms "comprise, comprises, comprised and comprising" and the terms "include, includes, included and including" are all deemed totally interchangeable and should be afforded the widest possible interpretation.

The invention is in no way limited to the embodiment hereinbefore described which may be varied in both construction and detail within the scope of the claims.

CLAIMS

1. A method for improving gene expression in a host cell comprising a protein encoding nucleic acid comprising

assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type protein encoding nucleic acid or mutant thereof;

preparing an altered protein encoding nucleic acid by modifying the A and T nucleotide content of the wild type protein encoding nucleic acid or mutant thereof to equal or lower the A and T nucleotide content of the host cell such that the intrinsic curvature of the altered protein encoding nucleic acid is reduced compared to the wild type protein encoding nucleic acid or mutant thereof and the altered protein encoding nucleic acid has reduced affinity, compared to the wild type protein encoding nucleic acid or mutant thereof, to host cell transcriptional repressor proteins; and

using the altered protein encoding nucleic acid in a host cell gene expression system.

2. The method according to claim 1 wherein the host cell transcriptional repressor protein is a nucleoid-associated transcriptional repressor protein, including H-NS.

3. The method according to any of the preceding claims wherein the wild type protein encoding nucleic acid or mutant thereof is modified so that it is no longer A and T nucleotide rich (AT-rich) compared to the host cell nucleic acid average A and T nucleotide content.

4. The method according to any of the preceding claims wherein the modified A and T nucleotide content of the altered protein encoding transcriptional repressor protein binding region nucleic acid is equal to or lower than the A and T nucleotide content of the host cell transcriptional repressor protein nucleic acid binding region.

5. The method according to any of the preceding claims wherein the A and T nucleotide content of the altered protein encoding promoter region and/or ribosome binding site (RBS) nucleic acid is modified compared to the wild type protein encoding nucleic acid or mutant thereof.

6. The method according to claim 5 wherein the A and T nucleotide content of the regions proximal to the altered protein encoding nucleic acid promoter region are modified such

that the A and T nucleotide content is equal to or lower than the A and T nucleotide content of the host cell nucleic acid.

7. The method according to any of the preceding claims wherein the host cell is a bacterium, preferably *Escherichia coli* or *Salmonella*, or a yeast.

8. The method according to any of the preceding claims wherein the protein encoding nucleic acid is a fluorescent protein nucleic acid.

9. A modified nucleic acid comprising a sequence encoding a wild type protein or mutant thereof wherein the nucleic acid has an equal or lower A and T nucleotide content and/or reduced intrinsic curvature compared to the wild type protein encoding nucleic acid or mutant thereof characterised in that the protein has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type protein encoding nucleic acid or mutant thereof.

10. A modified fluorescent protein nucleic acid according to claim 9 comprising a sequence encoding a wild type fluorescent protein or mutant thereof, wherein the nucleic acid has an equal or lower A and T nucleotide content and/or reduced intrinsic curvature compared to the wild type fluorescent protein nucleic acid or mutant thereof characterised in that the protein has reduced affinity to one or more host cell transcriptional repressor proteins compared to the wild type fluorescent protein nucleic acid or mutant thereof.

11. The modified nucleic acid according to claim 10 with equal or lower A and T nucleotide content in the regions proximal to the promoter region and/or ribosome binding site (RBS) of the fluorescent protein nucleic acid compared to the wild type fluorescent protein nucleic acid or mutant thereof.

12. The modified nucleic acid according to claims 10 or 11 with improved transcription compared to the wild type fluorescent protein nucleic acid or mutant thereof.

13. The modified nucleic acid according any of claims 10 to 12 comprising the nucleic acid sequence of Figure 4B or 6B or a sequence with at least 70%, preferably 80%, more preferably 90%, more preferably 95%, even more preferably 99% homology over the entire length to the nucleic acid sequence of Figure 4B or 6B.

14. The modified nucleic acid according to any of claims 10 to 13 for use in a host cell expression system wherein the fluorescent protein has an A and T nucleotide content equal to or lower than the host cell average A and T nucleotide content.

15. The modified nucleic acid according to claim 14 wherein the host cell is a bacterium, preferably *Escherichia coli* or *Salmonella*, or a yeast.

16. A fluorescent protein encoded by the modified nucleic acid of any of claims 11 to 16.

17. A host cell expression system comprising the modified nucleic acid sequence or fluorescent protein according to any of claims 10 to 16, preferably for use in a host cell such as *Escherichia coli* or *Salmonella*.

18. A plasmid vector comprising the modified nucleic acid sequence or fluorescent protein according to any of claims 10 to 16, preferably for use in a host cell such as *Escherichia coli*.

19. A host cell comprising the modified nucleic acid sequence, fluorescent protein, plasmid vector, expression system according to any of claims 11 to 18.

20. A host cell according to claim 19 wherein the A and T nucleotide content of the modified fluorescent protein nucleic acid is equal to or lower than the A and T nucleotide content of the host cell nucleic acid.

21. The method, modified nucleic acid, fluorescent protein, expression system, plasmid vector or host cell according to any of claims 10 to 20 wherein the fluorescent protein is a green fluorescent protein (GFP), yellow fluorescent protein (YFP) cyan fluorescent protein (CFP), blue fluorescent protein (BFP) or red fluorescent protein (DsRed) or a mutant thereof.

22. The method, modified nucleic acid, fluorescent protein, expression system, plasmid vector or host cell according to any of claims 10 to 20 wherein the fluorescent protein is a green fluorescent protein mutant selected from the following a spectral variant, a pHluorins, a variant with an altered Stokes shift, an oligomerization variant, a folding variant, a photoactivatable variant, a photoconversion variant, a photoswitchable variant, a redox sensitive variant and/or gfp^{+}.

23. The modified nucleic acid, fluorescent protein, expression system, plasmid vector or host cell according to any of claims 10 to 20 wherein the host cell transcriptional repressor protein is a nucleoid-associated repressor protein, including H-NS.

24. A method of monitoring gene expression in a host cell comprising the use of the modified nucleic acid, fluorescent protein, expression system or host cell according to any of claims 10 to 23.

ABSTRACT

5

"A METHOD FOR IMPROVING GENE EXPRESSION"

10 The present invention is directed to a method for improving gene expression in a host cell comprising a modified protein encoding nucleic acid comprising the steps of assessing the A and T nucleotide content and/or the intrinsic curvature of a wild type protein encoding nucleic acid or mutant thereof, preparing an altered protein encoding nucleic acid with modified A and T nucleotide content and using the altered protein encoding nucleic acid in 15 host cell gene expression systems. The present invention is also directed to the modified nucleic acid sequence, protein, plasmid vector, expression system comprising the altered protein encoding nucleic acid.

20

25 31738WO.finalspec.3June2010

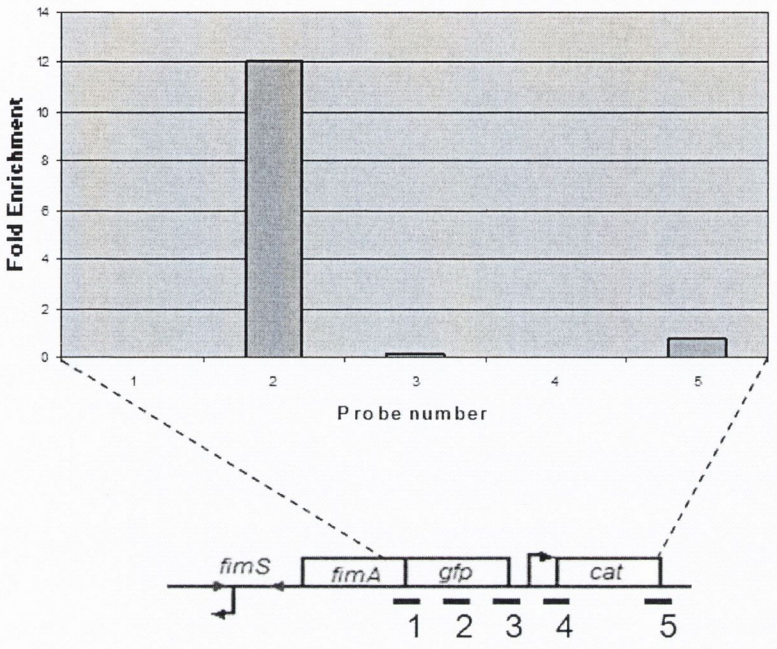

Fold Enrichment
14
12
10
8
6
4
2
0
1
2
3
4
5
Probe number
fimS
fimA
gfp
cat
1
2
3
4
5

FIGURE 1

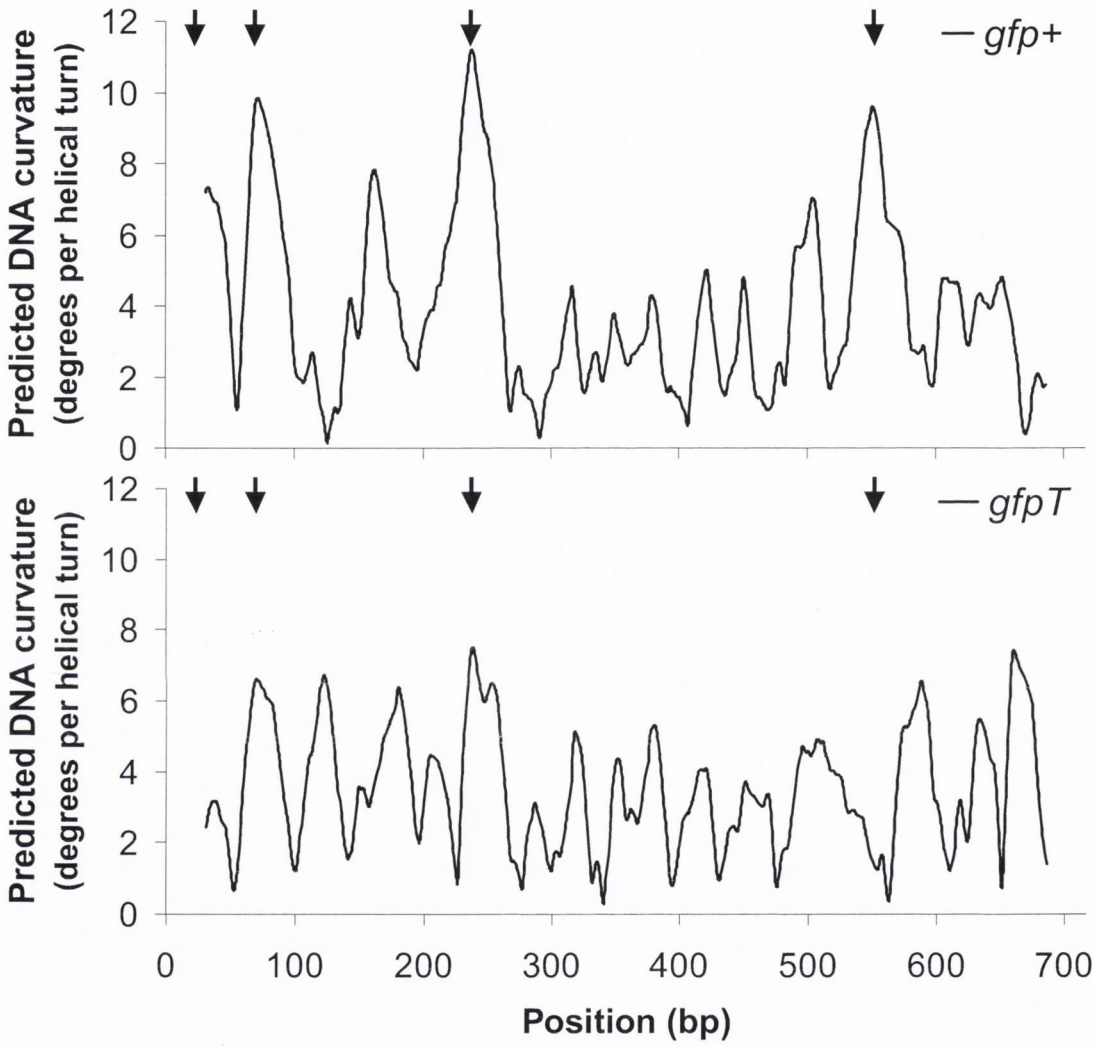

Predicted DNA curvature
(degrees per helical turn)
12
10
8
6
4
2
0
— gfp+
Predicted DNA curvature
(degrees per helical turn)
12
10
8
6
4
2
0
— gfpT
0
100
200
300
400
500
600
700
Position (bp)

FIGURE 2A

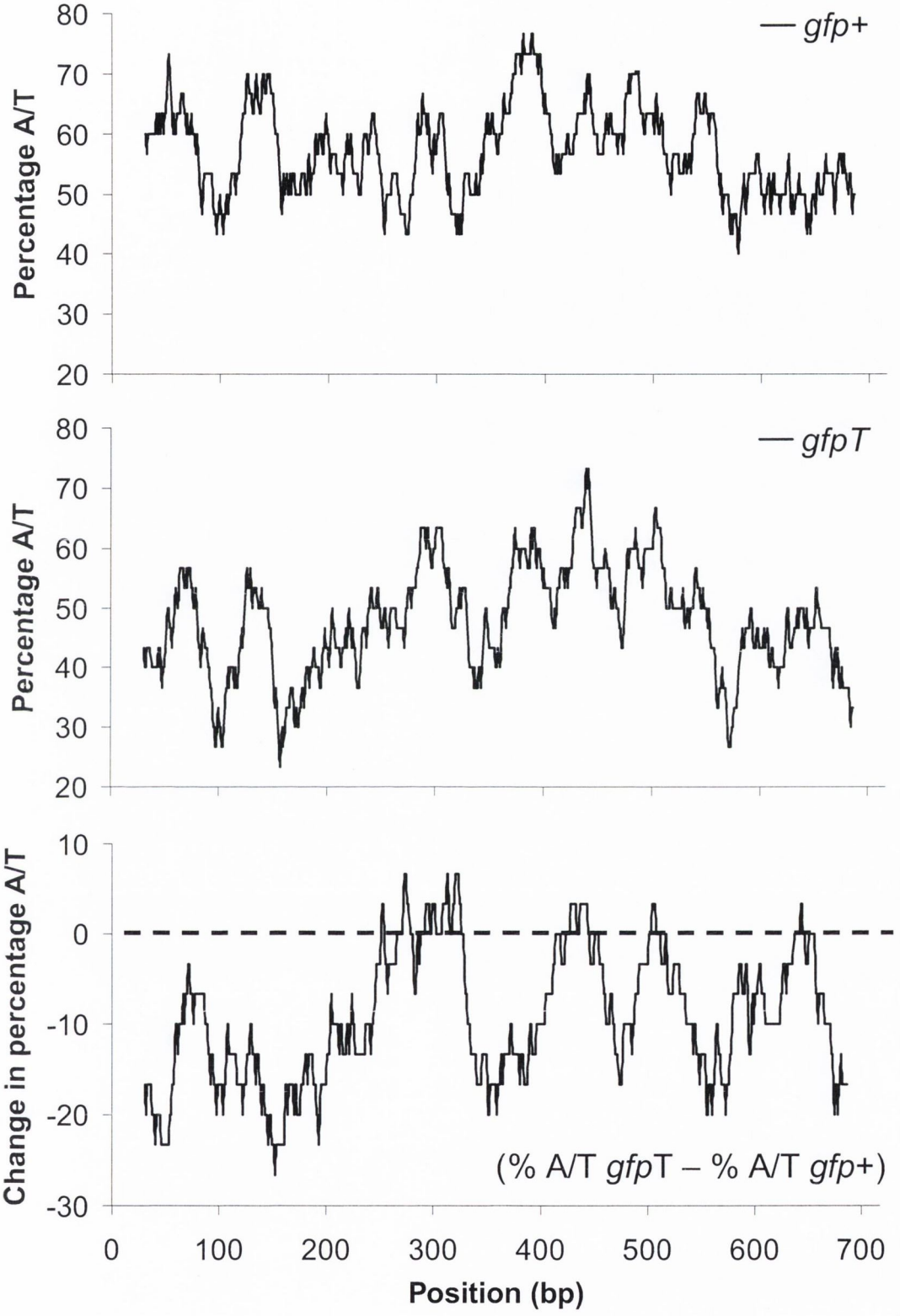

80
70
60
50
40
30
20
Percentage A/T
— gfp+
80
70
60
50
40
30
20
Percentage A/T
— gfpT
10
0
-10
-20
-30
Change in percentage A/T
(% A/T gfpT – % A/T gfp+)
0
100
200
300
400
500
600
700
Position (bp)

FIGURE 2B

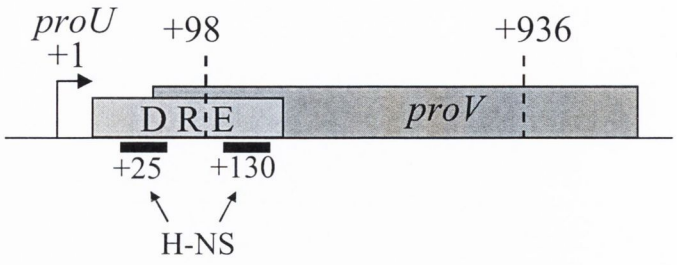


FIGURE 3A

***proU* fusions disrupting the DRE**

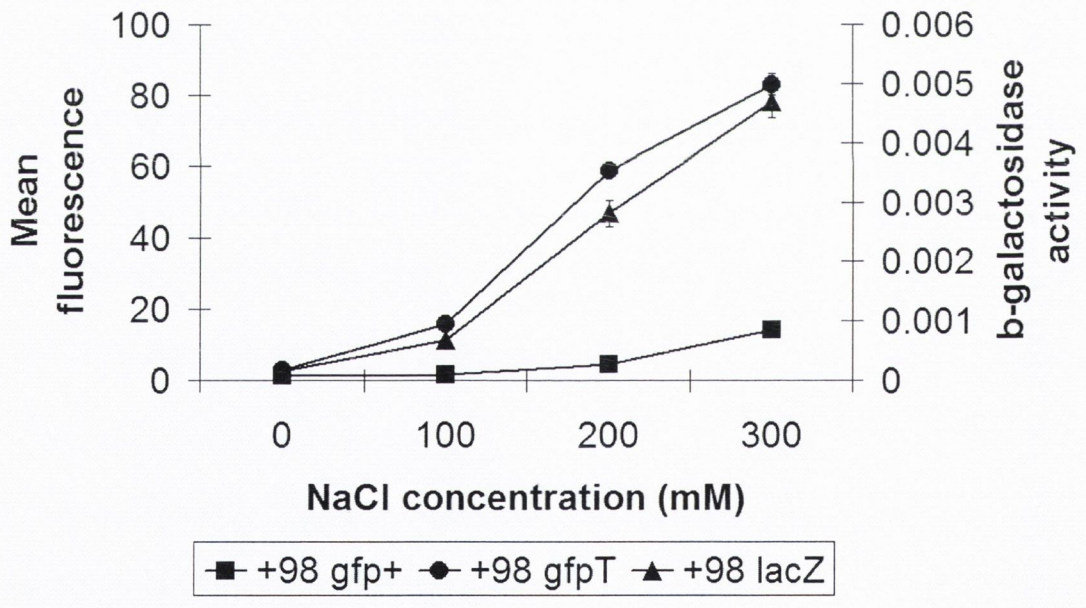


FIGURE 3B

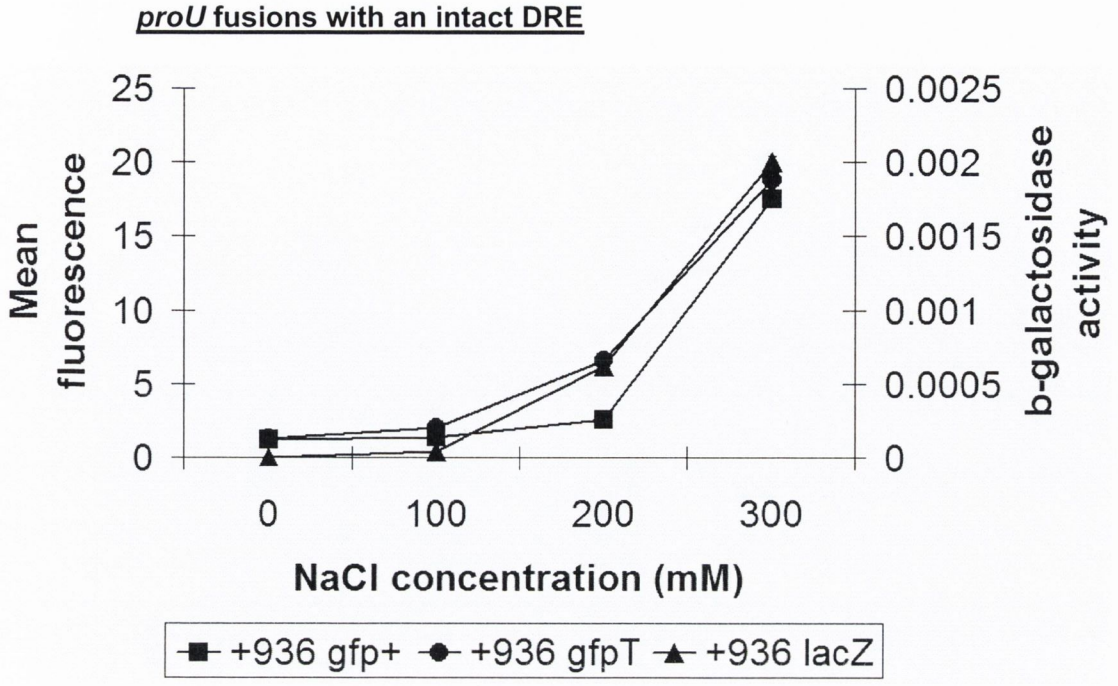


FIGURE 3C

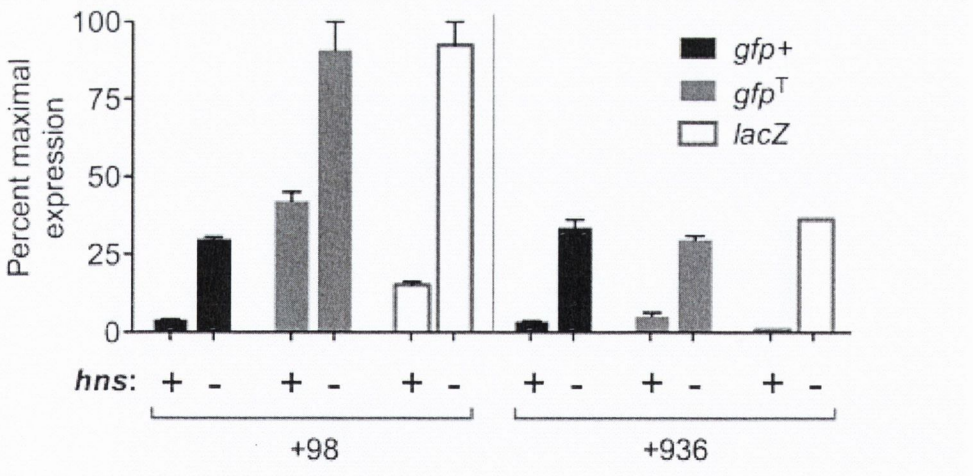


FIGURE 3D

Fold increase in expression			
		gfp+	*gfp*^{T}
When DRE is replaced (+98 / +936)	wildtype	1.1	8.0
	Δ*hns*	0.9	3.1
When H-NS is removed (Δ*hns* / wildtype)	+98	7.3	2.2
	+936	9.4	5.6

FIGURE 3E

gfp+ from pZep08

```
atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggt
 M  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G
gatgttaatgggcacaaattttctgtcagtggagagggtgaaggtgatgctacatacgga
 D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G
aagcttacccttaaatttatttgcactactggaaaactacctgttccatggccaacactt
 K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L
gtcactactttgacctatggtgttcaatgcttttcccgttatccggatcatatgaaacgg
 V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R
catgacttttcaagagtgccatgcccgaaggttatgtacaggaacgcactatatctttc
 H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  S  F
aaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgatacccttgtt
 K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
aatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaa
 N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K
ctcgagtacaactataactcacacaatgtatacatcacggcagacaaacaaaagaatgga
 L  E  Y  N  Y  N  S  H  N  V  Y  I  T  A  D  K  Q  K  N  G
atcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagac
 I  K  A  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D
cattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattac
 H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y
ctgtcgacacaatctgccctttcgaaagatcccaacgaaaagcgtgaccacatggtcctt
 L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L
cttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaataa
 L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  K  -
```

FIGURE 4A

Modified gfp gene (gfp^T)used in the Examples

```
atgagcaaaggcgaagagctgttcaccggtgttgtcccgattctggtggaactggacggt
 M  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G
gacgttaatggtcataagtttagcgtgagcggcgagggtgaaggtgacgccacgtacggc
 D  V  N  G  H  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G
aaactgaccctgaaattcatttgcaccaccggtaagctgccggtcccgtggccaactttg
 K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L
gtgaccaccctgacctacggtgtccagtgttttagccgttacccggaccacatgaagcgt
 V  T  T  L  T  Y  G  V  Q  C  F  S  R  Y  P  D  H  M  K  R
cacgacttttcaagagcgcaatgcctgagggctatgttcaagaacgtactatctcgttt
 H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  S  F
aaggatgatggcaattacaagacgcgtgccgaagttaagttcgagggcgacacgctggtt
 K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
aatcgcattgagctgaaaggcattgatttcaaagaggatggtaacatcctgggtcacaaa
 N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K
ctggagtataactacaactctcataatgtgtatatcacggcggacaaacagaagaacggt
 L  E  Y  N  Y  N  S  H  N  V  Y  I  T  A  D  K  Q  K  N  G
atcaaggcgaatttcaaaatccgtcacaacattgaagatggtagcgtgcaactggcagat
 I  K  A  N  F  K  I  R  H  N  I  E  D  G  S  V  Q  L  A  D
cattaccaacagaataccccgatcggcgacggtccggttttgctgccggacaatcattac
 H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y
ctgagcacgcaatccgcgctgagcaaagatccgaacgaaaaacgcgaccacatggtcttg
 L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  V  L
ctggaatttgtcaccgctgcgggcatcacccacggtatggatgaactgtataagtaa
 L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  K  -
```

FIGURE 4B

Gfpmut2 sequence

```
atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggc
 M  S  K  G  E  E  L  F  T  G  V  V  P  I  L  V  E  L  D  G
gatgttaatgggcaaaaattctctgtcagtggagagggtgaaggtgatgcaacatacgga
 D  V  N  G  Q  K  F  S  V  S  G  E  G  E  G  D  A  T  Y  G
aaacttacccttaaatttatttgcactactgggaagctacctgttccatggccaacactt
 K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L
gtcactactttcgcgtatggtcttcaatgctttgcgagatacccagatcatatgaaacag
 V  T  T  F  A  Y  G  L  Q  C  F  A  R  Y  P  D  H  M  K  Q
catgacttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatatttttac
 H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I  F  Y
aaagatgacgggaactacaagacacgtgctgaagtcaagtttgaaggtgataccttgtt
 K  D  D  G  N  Y  K  T  R  A  E  V  K  F  E  G  D  T  L  V
aatagaatcgagttaaaaggtattgattttaaagaagatggaaacattcttggacacaaa
 N  R  I  E  L  K  G  I  D  F  K  E  D  G  N  I  L  G  H  K
atggaatacaactataactcacataatgtatacatcatggcagacaaaccaaagaatgga
 M  E  Y  N  Y  N  S  H  N  V  Y  I  M  A  D  K  P  K  N  G
atcaaagttaacttcaaaattagacacaacattaaagatggaagcgttcaattagcagac
 I  K  V  N  F  K  I  R  H  N  I  K  D  G  S  V  Q  L  A  D
cattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattac
 H  Y  Q  Q  N  T  P  I  G  D  G  P  V  L  L  P  D  N  H  Y
ctgtccacacaatctgccctttccaaagatcccaacgaaaagagagatcacatgatcctt
 L  S  T  Q  S  A  L  S  K  D  P  N  E  K  R  D  H  M  I  L
cttgagtttgtaacagctgctgggattacacatggcatggatgaactatacaaataa
 L  E  F  V  T  A  A  G  I  T  H  G  M  D  E  L  Y  K  -
```

FIGURE 5A

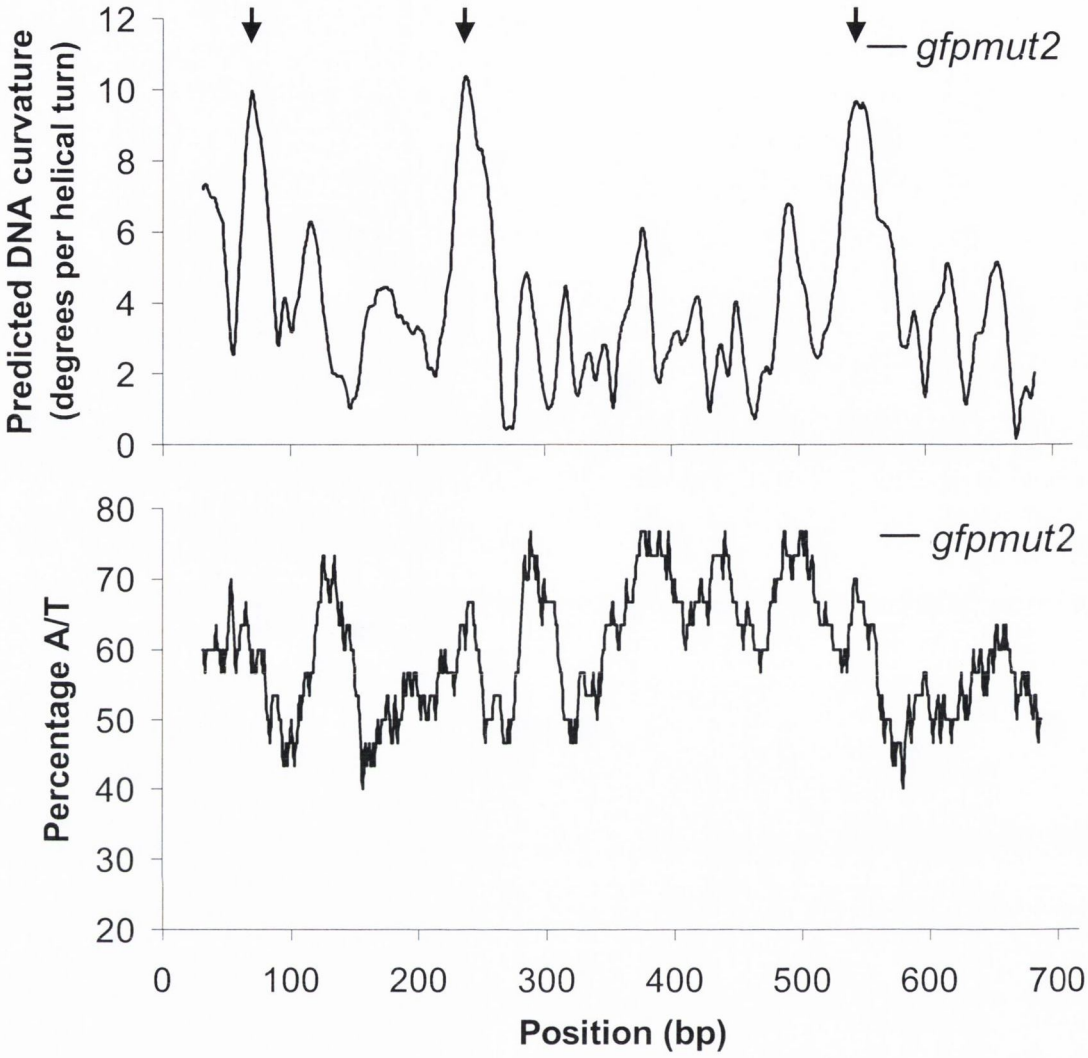

Predicted DNA curvature
(degrees per helical turn)
12
10
8
6
4
2
0
— gfpmut2
Percentage A/T
80
70
60
50
40
30
20
— gfpmut2
0
100
200
300
400
500
600
700
Position (bp)

FIGURE 5B

DsRed sequence

```
atgaggtcttccaagaatgttatcaaggagttcatgaggtttaaggttcgcatggaagga
 M  R  S  S  K  N  V  I  K  E  F  M  R  F  K  V  R  M  E  G
acggtcaatgggcacgagtttgaaatagaaggcgaaggagaggggaggccatacgaaggc
 T  V  N  G  H  E  F  E  I  E  G  E  G  E  G  R  P  Y  E  G
cacaataccgtaaagcttaaggtaaccaaggggggacctttgccatttgcttgggatatt
 H  N  T  V  K  L  K  V  T  K  G  G  P  L  P  F  A  W  D  I
ttgtcaccacaatttcagtatggaagcaaggtatatgtcaagcaccctgccgacatacca
 L  S  P  Q  F  Q  Y  G  S  K  V  Y  V  K  H  P  A  D  I  P
gactataaaaagctgtcatttcctgaaggatttaaatgggaaagggtcatgaactttgaa
 D  Y  K  K  L  S  F  P  E  G  F  K  W  E  R  V  M  N  F  E
gacggtggcgtcgttactgtaacccaggattccagtttgcaggatggctgtttcatctac
 D  G  G  V  V  T  V  T  Q  D  S  S  L  Q  D  G  C  F  I  Y
aaggtcaagttcattggcgtgaactttccttccgatggacctgttatgcaaaagaagaca
 K  V  K  F  I  G  V  N  F  P  S  D  G  P  V  M  Q  K  K  T
atgggctgggaagccagcactgagcgtttgtatcctcgtgatggcgtgttgaaaggagag
 M  G  W  E  A  S  T  E  R  L  Y  P  R  D  G  V  L  K  G  E
attcataaggctctgaagctgaaagacggtggtcattacctagttgaattcaaaagtatt
 I  H  K  A  L  K  L  K  D  G  G  H  Y  L  V  E  F  K  S  I
tacatggcaaagaagcctgtgcagctaccagggtactactatgttgactccaaactggat
 Y  M  A  K  K  P  V  Q  L  P  G  Y  Y  Y  V  D  S  K  L  D
ataacaagccacaacgaagactatacaatcgttgagcagtatgaaagaaccgagggacgc
 I  T  S  H  N  E  D  Y  T  I  V  E  Q  Y  E  R  T  E  G  R
caccatctgttcctttaa
 H  H  L  F  L  -
```

FIGURE 6A

DsRedT sequence

```
atgcgtagctctaaaaacgtgatcaaagaattcatgcgcttcaaggtacgtatggaaggt
 M  R  S  S  K  N  V  I  K  E  F  M  R  F  K  V  R  M  E  G
accgtgaacggtcacgaatttgagattgaaggcgagggcgagggccgcccgtacgaaggc
 T  V  N  G  H  E  F  E  I  E  G  E  G  E  G  R  P  Y  E  G
cacaataccgttaaactgaaagttaccaagggtggtccgctgcctttcgcatgggacatc
 H  N  T  V  K  L  K  V  T  K  G  G  P  L  P  F  A  W  D  I
ctgtctccacagtttcagtatggctctaaagtatacgtgaaacacccagccgatatcccg
 L  S  P  Q  F  Q  Y  G  S  K  V  Y  V  K  H  P  A  D  I  P
gactataagaaactgtcttttccggaaggtttcaaatgggaacgcgtaatgaacttcgag
 D  Y  K  K  L  S  F  P  E  G  F  K  W  E  R  V  M  N  F  E
gatggcggcgttgttaccgttactcaagactctagcctgcaggacggctgcttcatctat
 D  G  G  V  V  T  V  T  Q  D  S  S  L  Q  D  G  C  F  I  Y
aaggtgaaattcatcggtgttaacttccctagcgatggtccggtgatgcagaagaagact
 K  V  K  F  I  G  V  N  F  P  S  D  G  P  V  M  Q  K  K  T
atgggttgggaagcgtctaccgaacgtctgtacccgcgtgacggtgttctgaaaggcgaa
 M  G  W  E  A  S  T  E  R  L  Y  P  R  D  G  V  L  K  G  E
attcacaaagcgctgaaactgaaggacggtggccactacctggttgagtttaagagcatc
 I  H  K  A  L  K  L  K  D  G  G  H  Y  L  V  E  F  K  S  I
tatatggcaaagaaaccggtccaactgccgggttattactacgtggattctaaactggat
 Y  M  A  K  K  P  V  Q  L  P  G  Y  Y  Y  V  D  S  K  L  D
attaccagccataacgaagactacactatcgttgaacagtacgagcgcaccgaaggccgt
 I  T  S  H  N  E  D  Y  T  I  V  E  Q  Y  E  R  T  E  G  R
caccatctgtttctgtga
 H  H  L  F  L  -
```

FIGURE 6B

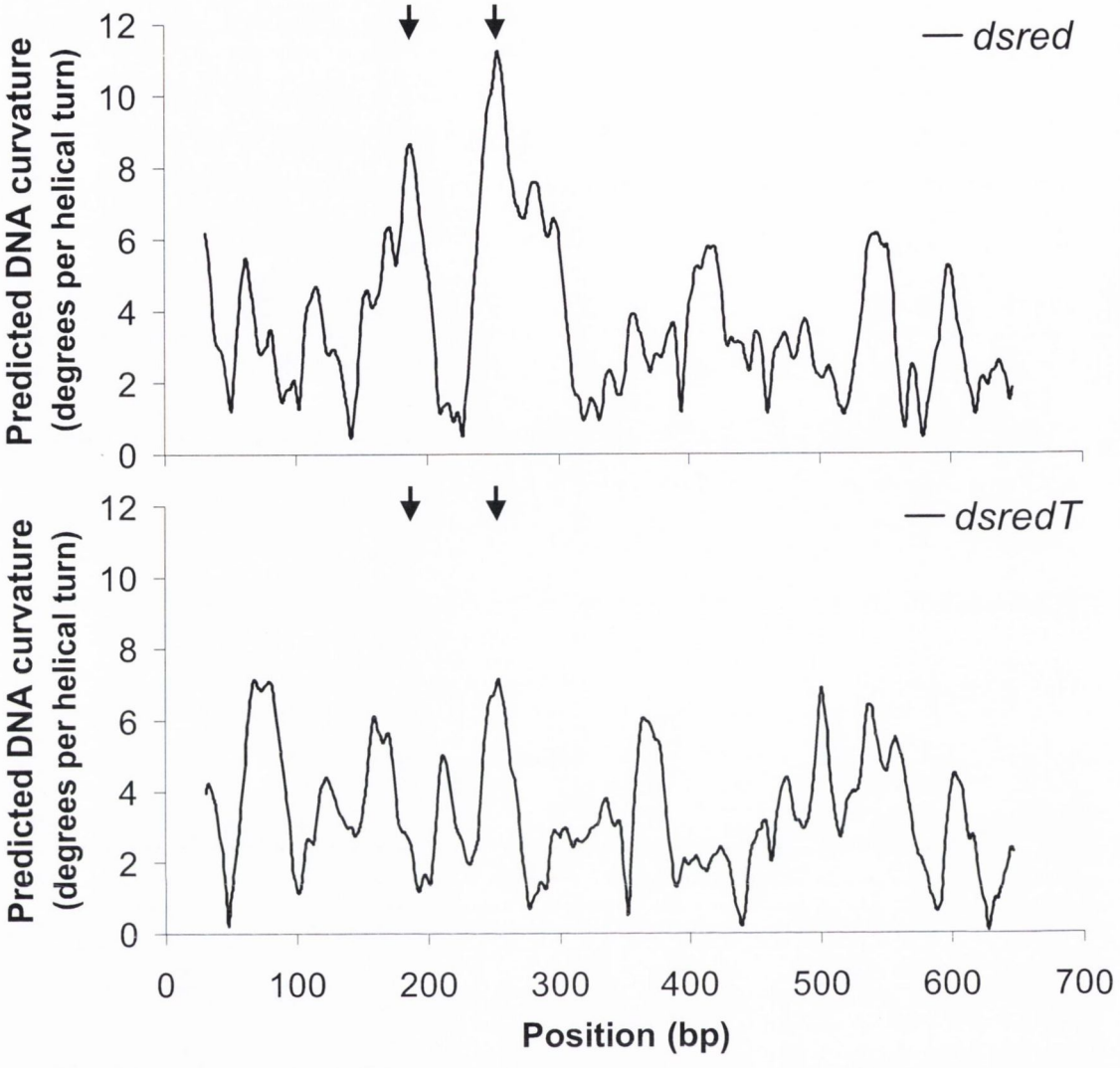

Predicted DNA curvature (degrees per helical turn)
12
10
8
6
4
2
0
— dsred
Predicted DNA curvature (degrees per helical turn)
12
10
8
6
4
2
0
— dsredT
0
100
200
300
400
500
600
700
Position (bp)

FIGURE 6C

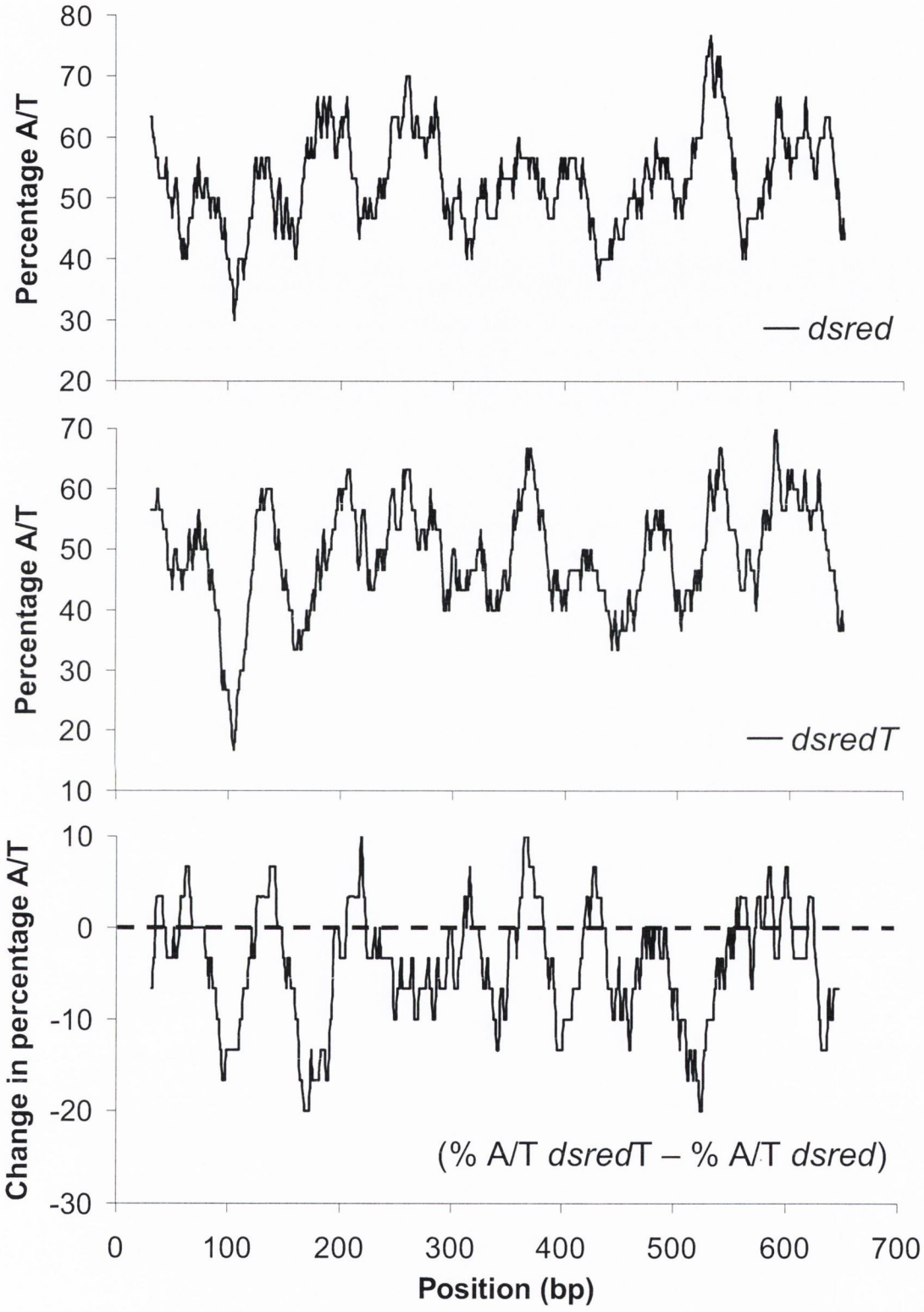

Percentage A/T
80
70
60
50
40
30
20
— dsred
Percentage A/T
70
60
50
40
30
20
10
— dsredT
Change in percentage A/T
10
0
-10
-20
-30
(% A/T dsredT – % A/T dsred)
0
100
200
300
400
500
600
700
Position (bp)

FIGURE 6D

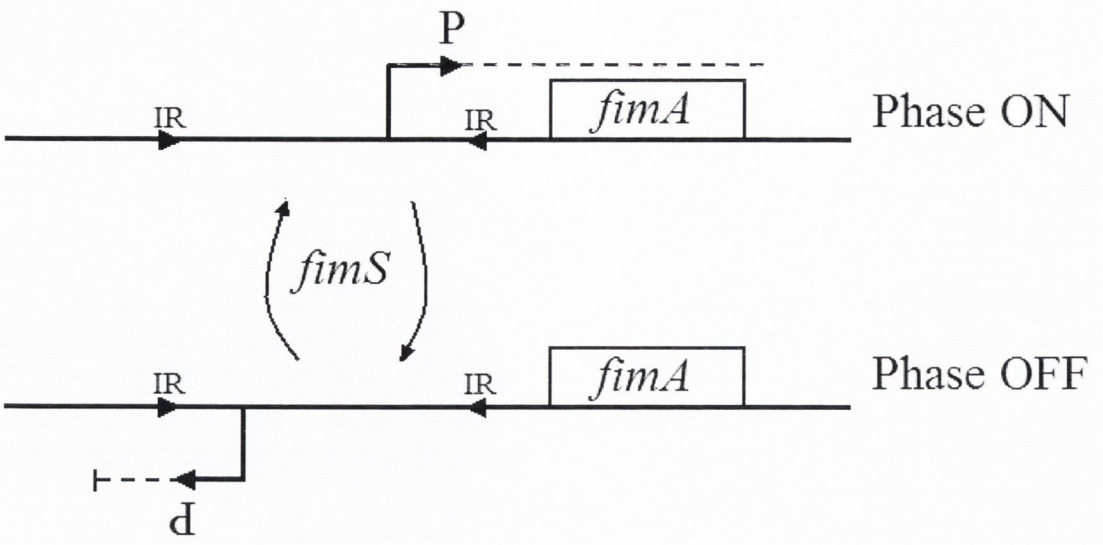


FIGURE 7A

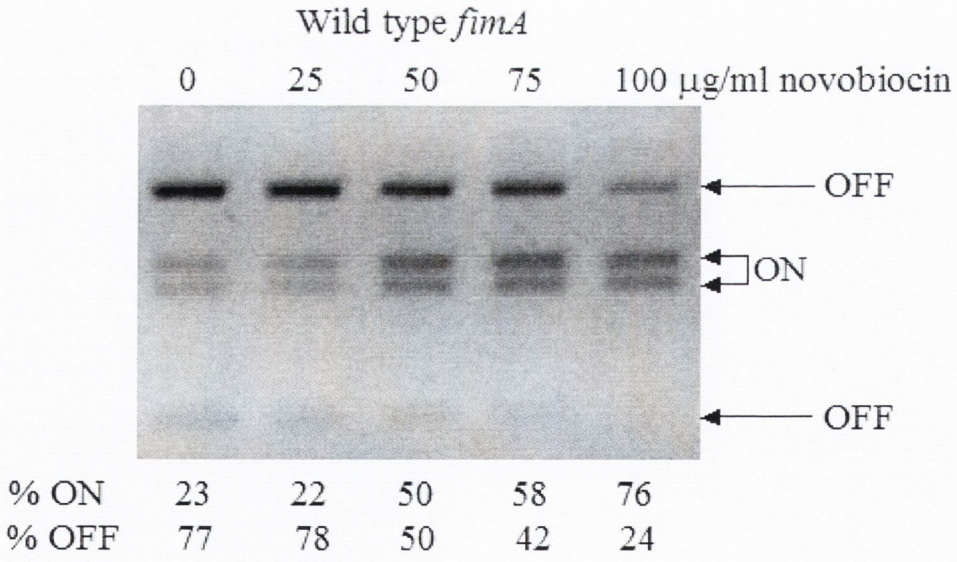


FIGURE 7B

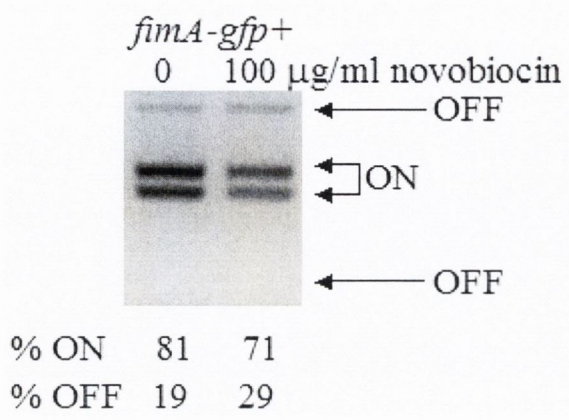


FIGURE 7C

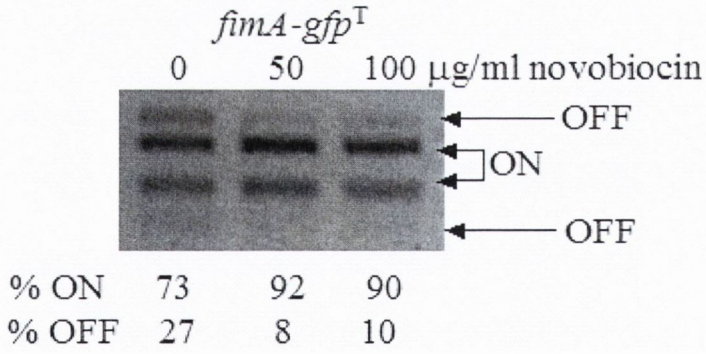


FIGURE 7D

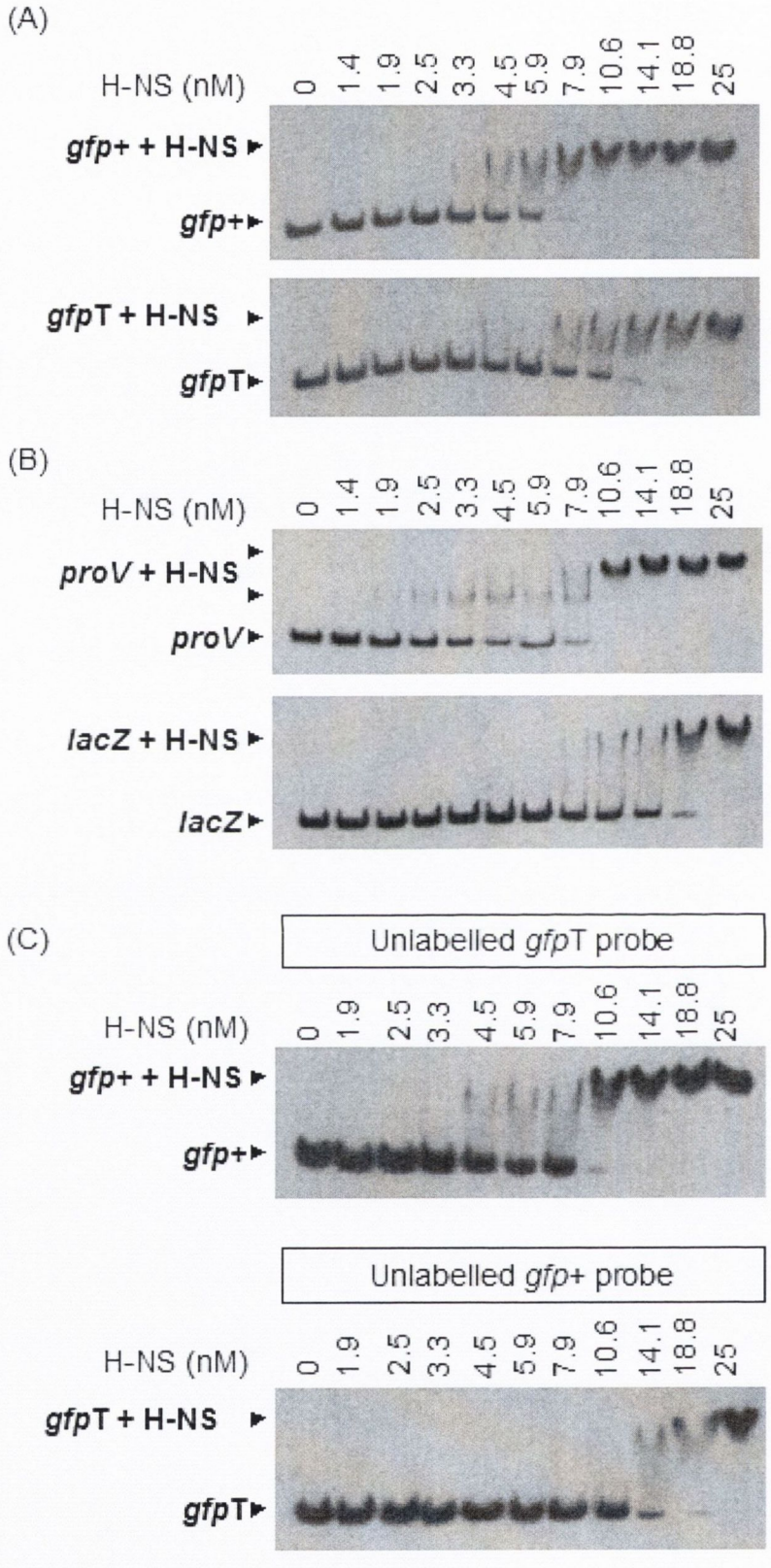

(A)
H-NS (nM)
0 1.4 1.9 2.5 3.3 4.5 5.9 7.9 10.6 14.1 18.8 25
gfp+ + H-NS
gfp+
gfpT + H-NS
gfpT
(B)
H-NS (nM)
0 1.4 1.9 2.5 3.3 4.5 5.9 7.9 10.6 14.1 18.8 25
proV + H-NS
proV
lacZ + H-NS
lacZ
(C)
Unlabelled gfpT probe
H-NS (nM)
0 1.9 2.5 3.3 4.5 5.9 7.9 10.6 14.1 18.8 25
gfp+ + H-NS
gfp+
Unlabelled gfp+ probe
H-NS (nM)
0 1.9 2.5 3.3 4.5 5.9 7.9 10.6 14.1 18.8 25
gfpT + H-NS
gfpT

FIGURE 8

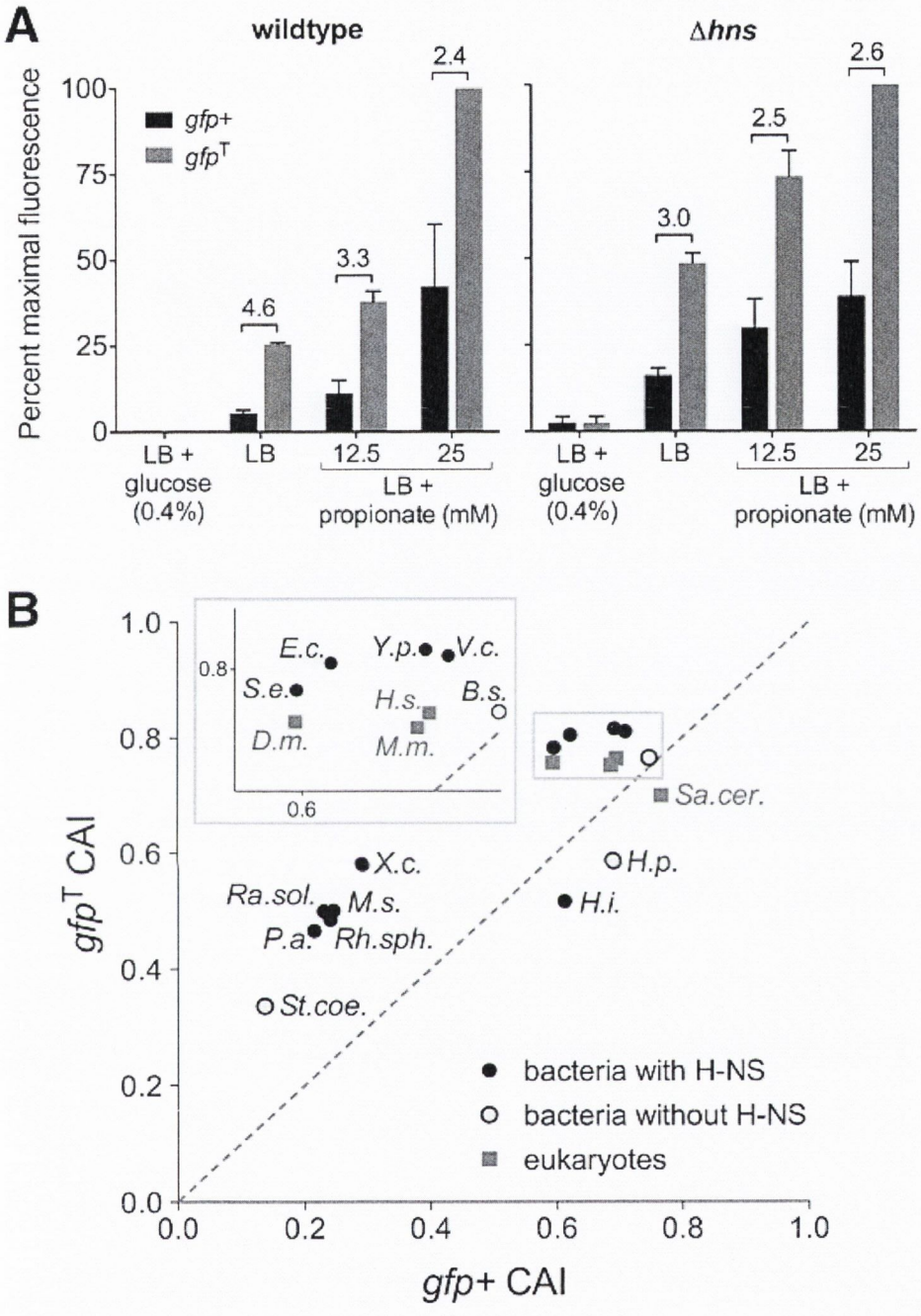

A
wildtype
Δhns
Percent maximal fluorescence
gfp+
gfp^T
2.4
4.6
3.3
3.0
2.5
2.6
LB + glucose (0.4%)
LB
12.5
25
LB + propionate (mM)
B
gfp^T CAI
gfp+ CAI
E.c.
Y.p.
V.c.
S.e.
H.s.
B.s.
D.m.
M.m.
Sa.cer.
X.c.
H.p.
Ra.sol.
M.s.
H.i.
P.a.
Rh.sph.
St.coe.
bacteria with H-NS
bacteria without H-NS
eukaryotes

FIGURE 9