THE MOLECULAR BASIS OF PYELONEPHRITIS-ASSOCIATED PILI PHASE VARIATION IN *ESCHERICHIA COLI*

by

Xiangwu Nou

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Experimental Pathology

Department of Pathology

The University of Utah

June 1996

Copyright © Xiangwu Nou 1996

All Rights Reserved

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

of a dissertation submitted by

Xiangwu Nou

This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.

Chair: David A. Low

John R. Roth

J. raves

H. Weis

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the dissertation of [redacted]Nou in its final form and have found that (1) its format, citations, and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

[redacted]
Date

[redacted]
David A. Low
Chair, Supervisory Committee

[redacted]
Walter Stevens
Chair/Dean

Approved for the Graduate Council

[redacted]
Ann W. Hart
Dean of The Graduate School

ABSTRACT

The expression of pyelonephritis-associated pili (Pap) in *Escherichia coli* is under a phase variation control mechanism in which individual cells alternate between pili$^+$ (ON) and pili$^-$ (OFF) states. This occurs through a process involving DNA methylation by deoxyadenosine methylase (Dam). Methylation of two GATC sites (GATC-I and GATC-II) within the *pap* regulatory region is differentially inhibited in phase ON and phase OFF cells. The GATC-I site of phase ON cells is nonmethylated and GATC-II site is fully methylated. Conversely, in phase OFF cells the GATC-I site is fully methylated whereas the GATC-II site is nonmethylated. Two transcription activators, Lrp and PapI, are required for this specific methylation inhibition. Low resolution DNA footprint analyses using nonmethylated *pap* DNA indicated that Lrp binds near the GATC-II, whereas PapI does not bind specifically to *pap* regulatory region. However, the addition of Lrp and PapI together resulted in an additional footprint around the GATC-I site, indicating that both Lrp and PapI are required for binding to the GATC-I region.

To define the role of Dam methylation in *pap* gene regulation, the GATC-I and GATC-II sites were mutated so that they could not be methylated, and the effects of these mutations on Pap phase variation were examined. The results indicated that methylation of GATC-I blocks formation of the phase ON state by inhibiting PapI-dependent Lrp binding to this DNA region. In contrast, methylation of GATC-II is required for the phase OFF to ON transition. Evidence suggests that this occurs by the inhibition of Lrp to sites overlapping the *papBA* promoter, which may occlude RNA polymerase.

The Lrp binding sites in the *pap* regulatory region were further defined by methylation protection analysis. Six Lrp binding sites were found, each separated by

about three helical turns of DNA. Lrp bound with highest affinity to three sites (1, 2, and 3) proximal to the *papBA* promoter. A mutational analysis indicated that the binding of Lrp to sites 2 and 3 inhibits *pap* transcription, which is consistent with the fact that Lrp binding site 3 is located between the -35 and -10 RNA polymerase binding region of *papBA* promoter. The addition of PapI decreased the affinity of Lrp for sites 1, 2, and 3 and increased its affinity for the distal Lrp binding sites 4 and 5. Mutations within Lrp binding sites 4 and 5 shut off *pap* transcription, indicating that the binding of Lrp to this *pap* region activated transcription. The *pap* GATC-I and GATC-II sites are located within Lrp binding sites 5 and 2, respectively, providing a mechanism by which Dam controls Lrp binding and Pap phase variation. A model for Pap phase variation is presented based on these results.

To Baihua, Aurora, and Angelina

TABLE OF CONTENTS

ABSTRACT iv

LIST OF TABLES ix

LIST OF FIGURES x

ACKNOWLEDGMENTS xii

Chapter

I. INTRODUCTION 1

Overview 2
The Pap Pili 3
DNA Methylation 13
Lrp and Gene Regulation 18
Pap Pili Phase Variation 26
Introduction to Work in This Thesis 34
References 35

II. REGULATION OF PYELONEPHRITIS-ASSOCIATED PILI PHASE-VARIATION IN *ESCHERICHIA COLI*: BINDING OF THE PAPI AND THE LRP REGULATORY PROTEINS IS CONTROLLED BY DNA METHYLATION 48

Summary 49
Introduction 49
Results 50
Discussion 53
Experimental Procedures 55
Acknowledgments 56
References 56

III. METHYLATION PATTERNS IN *PAP* REGULATORY DNA CONTROL PYELONEPHRITIS-ASSOCIATED PILI PHASE VARIATION IN *E. COLI* 58

Summary 59
Introduction 59
Results 61
Discussion 66
Experimental Procedures 68
Acknowledgments 69
References 69

IV. DIFFERENTIAL BINDING OF LRP TO TWO SETS OF *PAP* DNA BINDING SITES MEDIATED BY PAPI REGULATES PAP PHASE VARIATION IN *ESCHERICHIA COLI*..... 71

Introduction..... 72
Results..... 73
Discussion..... 76
Materials and Methods..... 83
Acknowledgments..... 83
References..... 83

V. SUMMARY AND DISCUSSION..... 85

The Role of Lrp in Pap Pili Phase Variation..... 86
The Role of PapI in Pap Pili Phase Variation..... 102
The Role of DNA Methylation in Pap Pili Phase Variation..... 107
The Roles of Other Regulatory Factors in Pap Pili Phase Variation..... 110
A Working Model for Pap Pili Phase Variation..... 112
Implications for Other Pili Phase Variation Systems..... 119
Concluding Remarks..... 121
References..... 122

LIST OF TABLES

Table		Page
1-1	Lrp regulon	21
2-1	The *papI* and *lrp* genes are required for transcription initiated at the *papBA* pilin promoter	50
3-1	E. coli strains, plasmids, and bacteriophage used in this study	60
3-2	The effects of Lrp and Dam on *pap* transcription in E. coli strains containing wild-type or mutant *pap* regulatory DNAs	62
3-3	The effects of PapI overproduction on methylation protection of the $GATC^{1028}$ site and *pap* transcription	65
4-1	Phenotypes of *pap* regulatory mutants	80
4-2	Effects of Lrp and PapI on the phenotypes of *pap* regulatory mutants	80
4-3	Effects of Dam methylase on the phenotypes of *pap* regulatory mutants	80
4-4	Escherichia coli strains, plasmids and bacteriophage used in this study	82

LIST OF FIGURES

Figure		Page
1-1	Organization of the *pap* genes	7
1-2	Mechanisms of pili phase variation in bacteria	29
2-1	Map of plasmid pDAL336	50
2-2	Gel retardation analysis	51
2-3	DNase I footprint analysis of fully methylated and non-methylated *pap* DNAs	52
2-4	Summary of DNase I footprint results using the *pap* regulatory region	53
2-5	DNase I footprint analysis of hemimethylated *pap* DNAs	54
3-1	The *pap* regulatory region	61
3-2	Analysis of LacZ phenotypes of E. coli strain MC4100 containing single-copy wild-type and regulatory mutant *pap'lacZ* fusions	61
3-3	Analysis of the GATC methylation states of *pap* DNA sequences isolated from wild-type and regulatory mutant E. coli strains	63
3-4	DNase I footprint analysis of wild-type and mutant *pap* DNAs	64
3-5	Pap phase variation model	67
4-1	The localization of Lrp binding sites in the *pap* regulatory region by methylation protection analysis	74
4-2	The location of Lrp binding sites in the *pap* regulatory region	75
4-3	A quantitative DNase I footprint analysis of nonmethylated wild-type and Lrp binding site mutant *pap* DNAs	77
4-4	Summary of the quantitative footprint analysis of Lrp binding to wild-type and mutant *pap* DNAs	79
4-5	Pap phase variation model	81
5-1	Comparison of Lrp footprints in *pap, ilvIH, lysU* and *ompC* regulatory regions	92

5-2 Proposed DNA conformations in phase ON and phase OFF cells..... 113

5-3 Model of Pap pili phase variation..... 116

ACKNOWLEDGMENTS

Formost, I am deeply indebted to my thesis advisor Dr. David Low. As a respected teacher, David patiently guided me throughout my work of this thesis with insights and enthusiasm. His guidance and instructions have been key to work presented in this thesis.

Past and current members of the Low laboratory: Bruce Braaten, Larry Blyn, Chris White-Ziegler, Majan van der Wood, Linda Kaltenbach, Brad Nicholen, Nate Wyand, and Brad Hale, all deserve my profond thanks for providing helps in various experiments, in English language, and in preparing this thesis, as well as for provinding encouragements and friendship and for sharing their insights and wisdom. I especially want to thank Dr. Chris Hill and Felix Vajdos for providing me with purified Lrp.

I would like to thank my supervisory committee members for their support and suggestions. I especially wish to thank Dr. John Roth for bringing my attention to the Molecular Biology Program at The University of Utah

I wish to thank Blackwell Science Ltd., Cell Press, and Oxford University Press for granting me permission to use copyrighted materials, which constitute Chapters II, III, and IV, of this thesis.

CHAPTER I

INTRODUCTION

Overview

Urinary tract infection (UTI) is a common health concern, especially among women. About one tenth of men and one fifth of women experience episodes of UTI during their lifetimes (74, 82, 96). UTI can be caused by a variety of pathogens, including bacteria, yeast and viruses. However, the most common etiologic agents of UTI are uropathogenic *Escherichia coli* strains (72, 130, 144). It is estimated that 60-85% of human UTI are caused by *E. coli* strains. The source of UTI-causing strains is believed to be the opportunistic pathogens in the fecal flora (46, 141). Those bacterial strains are capable of, upon contamination, colonizing and invading the lower urinary tract, causing cystitis, and further ascending to the upper urinary tract, causing pyelonephritis. In some cases, they invade the deep tissue, causing urosepsis (72, 124, 130, 139).

Uropathogenic and nonpathogenic *E. coli* strains differ in their abilities to colonize and to invade host tissues. Bacterial cells from a pathogenic strain can adhere to host epithelial cells by way of specific types of pili (also known as fimbriae)(130, 139). Such pili are usually a complex surface structure composed of specific pili subunits and adhesin molecules with different specificity to the components of host cell surface (72, 84). Most *E. coli* clinical isolates from pyelonephritis patients express a distinct type of pilus that enable the bacterial cells to attach to the epithelial cells lining the urinary tract (72). These pili are therefore referred to as **py**elonephritis-**a**ssociated **pi**li (Pap) (68, 118).

Production of pili confers to the bacterial cell the advantage to adhere tightly to host epithelial cell surfaces, a necessary step for colonization in environments like the urinary tract, where bacterial cells not attaching to the host cells are readily removed (72). Nevertheless, the presence of pili can sometime also be disadvantageous for a bacterial cell. First, the tight adhesion to epithelial cells mediated by pili reduces the mobility of the bacterial cells (pili are not structures of locomotion), which restricts the bacterial cells from moving to new locations. Second, the proteinacious pili may be an easy target for the host

immunological system, rendering piliated bacterial cells more vulnerable to eradication by the host (132). Third, the production of the pili also consumes a large quantity of energy.

Uropathogenic *E. coli* have developed a mechanism that allows them to maintain the advantages of pili production and avoid its disadvantages by altering the states of pili expression. This phenomenon is known as phase variation, by which bacterial cells alternate between phase ON (active expression) and phase OFF (no expression) states. In some other systems, such as *Neisseria gonorrhoeae* pili expression, the bacterial cells produce different types of pili with different antigenic characteristics, a phenomenon known as antigenic variation. Antigenic variation as well as phase variation occurs by different mechanisms in different cases, though in most cases reversible rearrangement of DNA is often involved (132). Understanding the mechanisms underlying different types of antigenic and phase variation is therefore important for understanding bacterial pathogenesis.

The Pap Pili

Characterization of Pap Pili

A uropathogenic *E. coli* strain can produce different types of pili, including Type 1 pili, Pap pili, S pili, and others (28, 72, 81). A given type of pilus is characterized by its morphology and adhesion properties but can exist as different serotypes. Pap pili produced by uropathogenic *E. coli* strains are hairlike proteinacious surface structures composed of a rigid shaft and a flexible tip (84). The shaft is composed of about one thousand major pilin monomers, which are arranged as a tightly packed right-handed helix that is 7 nm wide and 0.2-2 μm long with a 2.0-2.5 nm central helical hole (84, 85). The major pilin subunits account for 99.9% of the pilus mass and determine the antigenic specificity of the pilus (58, 104). The tip is composed of a variety of minor pilin subunits, which form a thin open helical structure (2 nm wide) termed the fimbrillum (84, 85). The tip determine the adhering specificity of the pilus (93, 103). A bacterial cell may produce hundreds to thousands of Pap pili on its surface.

Unlike the more common Type 1 pili, which bind to mannose residues of epithelial cells and therefore the binding to cells is inhibited by the presence of mannose (134), Pap pili can agglutinate human or guinea pig erythrocytes in the presence of mannose (145). This mannose-resistant hemagglutination is due to the presence of an adhesin in the pili with a different receptor specificity. *E. coli* strains expressing Pap pili, as well as purified pili from those strains, bind to erythrocytes expressing P blood group antigens (75, 76, 80), which are a family of globoseries glycolipid containing Gal(α1-4)Galβ moiety (Gal-Gal), including globotriaosylceramise (GbO3) and globoside (GbO4). Several lines of experiments confirmed that Pap pili adhesin recognized Gal-Gal disaccharide moiety of glycolipids on the cell surface (72). First, *E. coli* strains expressing Pap pili were shown to bind to cells carrying such glycolipids naturally or to cells or latex beads treated with such glycolipids (31, 89, 98, 142). Second, adherence of strains expressing Pap pili to cells carrying Gal-Gal containing glycolipids or treated with Gal-Gal containing glycolipids was blocked by substances containing the Gal-Gal moiety (75, 89, 142). Third, strains expressing Pap pili were shown to adhere to glycolipids containing Gal-Gal but not to those without Gal-Gal (17). In humans the globoseries glycolipids GbO3 and GbO4 are found predominantly within the kidney, the ureter, the red blood cells, and less abundantly within the gastrointestinal tract (24, 84). This provides an explanation for why Pap piliated *E. coli* strains colonize and invade the human urinary tract, causing UTI, whereas other tissues are relatively not infected. Recently Pap pili were found also to bind to immobilized fibronectin (157), providing a mechanism for Pap piliated bacterial cells to interact with the extracellular matrix (59).

Pap and UTI

A number of epidemiological studies have established the importance of the expression of Pap pili in bacteria caused urinary tract infections (72, 139). Marild *et al.* (106) showed that about 90% of children with first-time acute pyelonephritis carried *E. coli* expressing Pap pili. Sanderberg *et al.* (135) showed that the majority of adult women of

various ages with acute pyelonephritis carried Pap-piliated *E. coli* as causing agents. In the case of uncomplicated urosepsis, Johnson *et al.* (73) showed that up to 100% of the blood isolates expressed Pap pili. Overall, about 70% of *E. coli* isolates from pyelonephritis patients and 36% from cystitis patients expressed Pap pili, whereas only about 24% from asymptomatic bacteriuria patients and 19% from fecal flora of healthy individuals expressed Pap pili (72). However, in individual patients compromised by pregnancy, anatomic abnormalities, or instrumentation, the frequency of clinical isolates expressing Pap pili was substantially lower comparing with individual patients without such complicating factors (32, 72, 73, 99). These studies indicate that although nonpiliated strains are capable of infecting the lower urinary tract and upper urinary tract of the compromised individuals, Pap piliated strains are more likely to reach to the upper urinary tract of healthy individuals and cause pyelonephritis.

The importance of Pap pili in uropathogenesis was also demonstrated by using mice and monkeys as animal models (72). In both cases, inoculation with strains expressing Pap pili led to colonization of the urinary tract by the bacteria and to urinary tract infections (37, 54, 118, 131). Administration of substances containing Gal-Gal along with the bacteria inoculum protected the animals from urinary tract colonization by the bacteria (140). Immunization with purified Pap pili proteins also protected the animals from urinary tract infection following subsequent challenges with the *E. coli* strains expressing homologous Pap pili (55, 118, 131).

Urinary tract infections by Pap piliated *E. coli* cause local and systematic inflammatory responses which are responsible for the symptoms. Strains expressing Pap pili, either living or killed, elicit inflammatory responses when inoculated into mice (95). This inflammatory response is blocked by coadministration of a Gal-Gal containing substance (95). This observation suggests that the role of Pap pili in uropathogenesis is not restricted to allow the bacterial cells to adhere to the uroepithelial cells. It has been found that Pap piliated *E. coli* cells stimulate epithelial cells to release interleukin-6 and

interleukin-8 *in vivo* and *in vitro* (2, 3, 62, 63). The cytokines, in turn, activate the host inflammatory responses (139).

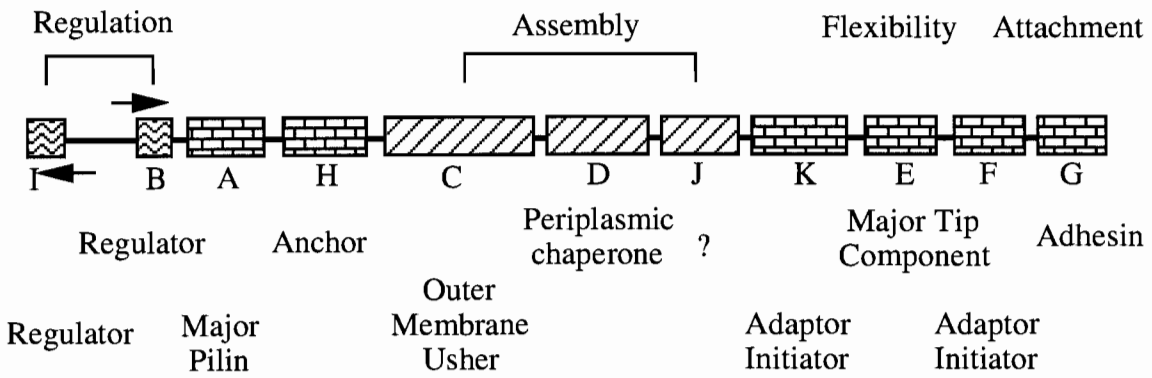
The *pap* Operon

The genes encoding the Pap pili are located on the chromosome as a cluster in the uropathogenic *E. coli* strains. It is believed that the uropathogenic *E. coli* strains have gained virulence determinant genes as a consequence of mobile element transposition and that different types of pili among a variety of enteric bacteria may have common origins (86). Eleven *pap* genes in the cluster have been identified and the function of the products studied (144). As shown in Figure 1-1, the *pap* genes are transcribed divergently, with *papI* being transcribed leftward as a monocistronic unit and the other 10 *pap* genes rightward as a polycistronic unit.

The products of *papI* and *papB* are positive regulatory proteins for the transcription of the *pap* operons (43). The *papA* gene encodes for the major pilin subunit of the Pap pili (9). PapA is the most heterogeneous gene product among different Pap pili producing isolates and determines the antigenic specificity of the Pap pili. However, PapA is not necessary for Gal-Gal adhesion. Mutants in *papA* gene produce cell adhesin which mediates Gal-Gal adhesion in rough strains but not in smooth strains (93, 150) suggesting that the pili shaft is important to extend the adhesin away from cell surface when the cell is coated with polysaccharides.

The genes *papK*, *papE*, *papF*, *papG,* and *papH* code for the minor components of Pap pili. PapG is the adhesin, present on the very tip of a pilus and determining the specific adhesion to the Gal-Gal moiety of glycolipids on the surfaces of epithelial cells (85). Mutants in the *papG* gene are unable to mediate mannose-resistant hemagglutination. The assembly of Pap pili starts with the assembly of the tip, and the subunits are sequentially added at the base so that the pilus "grows" from the base (69). PapF is a linker between the adhesin and the tip in the pilus. Its incorporation initiates the assembly of the pilus (71).

Figure 1-1. Organization of the *pap* genes. Eleven *pap* genes are organized as two divergently transcribed operons. The function of each gene product is indicated. The *papI-papB* intergenic region is the *pap* regulatory region.



Transcriptional regulatory protein

Pilin protein

Protein required for pili assembly

PapE subunits are then added to PapF. PapE is the most abundant minor pilin subunit and the only one known to form homopolymers. The presence of a string of PapE is believed to confer flexibility to the tip of the pilus (85). PapE also mediates the fibronectin-dependent adhesion (158). The incorporation of PapK into the pilus tip then initiates the assembly of PapA polymer (71), forming the long rigid shaft of the pilus structure. PapH serves both as a stopper of PapA polymerization and a molecular anchor for the pilus to the membrane. Mutants in *papH* produce nonanchored long pili and PapH overproducers produce shorter pili.(8).

The products of *papC* and *papD* are not present in the pilus but are required for Pap pili assembly (84). PapD is a periplasmic molecular chaperone that binds to the major and minor pilin subunits to prevent premature folding (94). Interaction with PapD also prevent the major and minor pilin subunits from being degraded by the periplasmic DegP proteinase (84). PapC, called outer membrane usher, is a specific outer membrane protein that determines the site and the order of pilus assembly (36). It may also serve as a deck and anchor for the assembly of Pap pili. The function of PapJ has not been well defined, though it has been suggested to facilitate the assembly of the PapA subunits (84).

Genetic Regulation of Pap Expression

The expression of Pap pili genes is affected by a variety of environmental factors, including carbon source, nitrogen source, temperature, and phase of growth (46). Generally, growth on rich medium inhibits the expression of pili genes and growth on minimal medium stimulates the expression. Similarly, growth at body temperature (37°C) induces the production of Pap pili whereas growth at room temperature (23°C) shuts off the pili production. It has been proposed that bacterial cells sense such environmental signals to determine whether they are inside a host (160). The genetic basis underlying such responses to environmental signals is complex. Many proteins, both operon specific factors encoded by the *pap* genes and global regulatory factors, have been shown to be involved in this process. The regulation of Pap pili expression occurs at both transcriptional and

posttranscriptional levels (102). In addition, the regulation of *pap* transcription involves a phase variation mechanism.

Transcriptional regulation

Both PapB and PapI are required for the transcription of the *papBA* operon (6). PapB is also autoregulated. Overexpression of PapB *in trans* leads to the down regulation of the transcription from the *papBA* operon (43). PapB has been shown to specifically bind to two sites in the intercistronic regulatory region as well as a site in the coding region of *papB* . Of those three binding sites, the most upstream one, centered at about -240 of the *papBA* transcription initiation point and about -90 of that of *papI*, has the highest affinity. A second binding site overlaps with the -10 region of the *papBA* promoter and a third one in the *papB* coding sequence at +100. Assuming the binding of PapB to the lower affinity sites at -10 and +100 interferes with the binding of RNA polymerase, this may explain the autoregulation of PapB. PapI does not bind to DNA, and recent evidence from our laboratory suggests it functions by affecting the DNA binding specificity of a major regulatory protein, Lrp (see further discussion later).

The global regulator CAP is also required for transcription and has been shown to bind in a cAMP-dependent manner to the regulatory region centered at 215.5 bp upstream of the *papBA* transcription initiation point and 115.5 bp upstream of the *papI* transcription initiation point, adjacent to the PapB high affinity binding site (52). Mutations introduced into this CAP binding site shut off the transcription from both operons. The involvement of CAP in transcription activation has been well studied. Depending on the properties of the promoter and the location of the binding site relative to the RNA polymerase binding site, CAP either directly interacts with RNA polymerase or induces DNA conformational change by bending DNA (25). The facts that the CAP binding site is far away from that of the RNA polymerase and that other regulatory factors are involved in the transcriptional activation suggest that *papBA* promoter is a class III CAP-activated promoter, in which direct CAP-RNA polymerase interaction has not been directly observed (25, 128). It has

been proposed that CAP, by interacting with PapB upon binding to the regulatory region, induces DNA conformational change that favors the formation of an open complex at the promoter regions (52).

The chromatin-associated protein H-NS may also play a regulatory role in the transcription of *pap* operons. The transcription of the *papBA* operon, when located on a multicopy plasmid, is stimulated twofold by mutation in *drdX,* a gene identical to *hns*. Moreover, cAMP-CAP and PapB are no longer required for *papBA* transcription in such mutants (53). It has been proposed that H-NS represses an intrinsically active *papBA* promoter and that cAMP-CAP and PapB are required as antirepressors to antagonize the silencing by H-NS (44). Recent work in this laboratory using chromosomally located single copy *papBA* operon indicates, however, that Pap phase variation occurs in an *hns* mutant background (146). This finding argues against the hypothesis that H-NS plays a central role in the regulation of *papBA* transcription.

Another chromosomal locus that is involved in the transcriptional regulation of *pap* operons was identified by screening for mutations giving a locked OFF phenotype following Tn-10 mutagenesis (20). This gene, originally named *mbf* for **M**ethylation **B**locking **F**actor, was later found to be identical to *lrp* (22), the gene encoding a global regulatory protein that has been implicated in the regulation of dozens of genes or operons (26). Lrp is absolutely required for the transcription of the *papBA* operon. Recent evidence suggests that Lrp play a central role in the regulation of *papBA* expression regulation and Pap pili phase variation (21).

Posttranscriptional regulation

In a single pilus, the shaft is composed of about a thousand of PapA major pilin subunits and the fimbrillum is composed of minor pilins present at much lower numbers (84). An obvious question then is how cells maintain the balance of those subunits. Such differential gene expression from a single polycistronic operon can be achieved by transcriptional attenuation or termination, by transcription from an internal promoter, or by

differential degradation of processed mRNA. A stem-loop terminator sequence is found between *papA* and *papH* genes. It is postulated that the transcription from the *papBA* promoter is attenuated at this sequence and only a small proportion of the transcription events can proceed passing this sequence and generate mRNA for the *papH* and downstream genes (6, 7, 116). Therefore, the predominant species of mRNA produced by the transcription from the *papBA* promoter is a dicistronic mRNA encompassing *papB* and *papA*. However, probing a northern blot with *papA* sequence revealed two mRNA species, one the expected size of the dicistronic mRNA of *papB* and *papA* and a smaller one that was not detected by probing with *papB* sequence (7). This is because the dicistronic mRNA is further processed so that a shorter mRNA species encompassing only the *papA* sequence is produced, while the part encompassing *papB* is rapidly degraded. This mRNA processing is dependent on the activity of RNase E (115), a RNase involved in the processing of a variety of mRNA precursors (111, 113, 114). The processed *papA* monocistronic mRNA is translationally active and markedly more stable than the dicistronic mRNA, with a half life of about 27 min in contrast to the 2.5 min half life of the *papB papA* dicistronic mRNA. It is believed that the *papA* mRNA is stabilized by the stem-loop terminator structure at the end of *papA* and another stem-loop structure formed at the head of the processed *papA* mRNA (115).

The production of Pap pili is shut off when cells are grown at 26°C or lower temperature. This thermoregulation is abolished by mutations in the gene *rimJ* (159), which encodes an N-terminal acetylase of ribosomal protein 5S (162). In such mutants, transcription from *papBA* promoter at lower temperature occurs as well as at 37°C, as determined by β-galactosidase activity of *papBA-lacZYA* fusion and by Northern blotting assays (160). The involvement of RimJ suggests that posttranscriptional events are, at least partly, responsible for Pap pili thermoregulation. It is hypothesized that the *pap* mRNA is destabilized at low temperature so that it is quickly degraded and ribosome or RimJ is involved in such a process (160). However, recent observations from our laboratory

suggest that H-NS may play an important role in thermoregulation of *pap* and that RimJ may affect the function of H-NS (C. White-Zigler and D. Low, unpublished data).

Phase variation

The expression of Pap pili is also controlled by a phase variation mechanism at the transcriptional level. This phase variation control involves differential DNA methylation and interaction with the global regulatory protein Lrp (15, 21). This will be further discussed in a later section.

DNA Methylation

Deoxyladenosine Methylase (Dam)

Deoxyladenosine methylase (Dam), also known as DNA adenine methyltransferase, is a 32 KDa protein that methylates the adenine at the N^6 position in a DNA sequence containing 5'-GATC-3' (107, 108). Dam is present in a variety of prokaryotic organisms, including *Escherichia coli, Salmonella typhimurium, Streptococcus pneumoniae,* and *Vibrio cholerae* (10, 107). In *E. coli*, mutation or overexpression of the *dam* gene leads to a hypermutable phenotype and uncoordinated DNA replication (108). The function of Dam has been shown to be involved in various cellular processes, such as mismatch repair, DNA replication initiation and transcriptional regulation (108).

Mismatch Repair

Mutants of the *dam* gene or Dam overproducer strains exhibit elevated frequencies of point mutations (64, 109). This is because Dam is an integral part of the methyl-directed mismatch repair system of bacterial cells. Mismatches occurring during DNA replication are first corrected by the proofreading function of the DNA polymerase. Those that are missed by the proofreading machinery are recognized by a mismatch repair system that includes MutS, MutL, and MutH, which subsequently replaces one of the mismatching nucleotides with one that matches (87, 112). Normally, DNA methylation trails DNA replication by a few minutes. Sometimes this lag can be up to approximately 10 min or one third of the cell

cycle due to the limited level of Dam in the cells . As a result, the newly synthesized strand is undermethylated immediately following the passage of DNA replication fork, leading to the hemimethylated GATC sites (108). The protein MutH cleaves DNA 5' to the G in the nonmethylated strand of a hemimethylated GATC site adjacent to a mismatch recognized by MutS and MutL (87, 112). The mismatch repair system is therefore able to discriminate the newly replicated strand from the old template strand. A stretch of DNA up to several Kb on the newly replicated, and therefore nonmethylated, strand is then excised and resynthesized by DNA polymerase III (87). Using λ phages containing mismatches in specific genes, it was shown that the repair of the mismatch following phage infection was biased in favor of the methylated strand when hemimethylated phage DNA was used, whereas the repair was random when nonmethylated phage DNA was used (123). Similar results were obtained using plasmid DNA carrying mismatches in a defined *in vitro* system (87). This is in agreement with the early observation that in *dam* mutants the mutation rate is much higher then in the wild-type. In the case of Dam overproduction, the lag between DNA replication and DNA methylation is greatly reduced, resulting in completely methylated DNA before the mismatches can be corrected. The methyl-directed mismatch repair cannot correct the mismatches when both strands are methylated, because a methylated GATC site is not recognized by the repair system (112).

Coordination of DNA Replication Initiation

In a rapidly growing population of bacterial cells, a new round of DNA replication occurs before the end of the previous one . As a result, multiple replication initiation sites (*oriC*) exist at the same time. In wild-type cells, the initiation at these sites is coordinated so that all the sites are used. Therefore, there are even numbers of initiation sites at any given time during the cell cycle (137). However, the initiation in the *dam* mutant cells is not coordinated so that odd numbers of initiation sites are present at any time (19). This finding suggests that Dam methylase is involved in the coordination of DNA replication initiation.

The distribution of GATC sequences in a DNA molecule is not even or random. There are 11 GATC sites in the 245 bp minimal *oriC* of *E. coli,* of which 8 are phylogenically conserved (163). Hemimethylated (but not completely methylated or nonmethylated) DNA containing *oriC* attaches to the cytoplasmic membrane *in vivo* and *in vitro* (119). Fully methylated plasmid DNA with *oriC* as a replication origin transforms *dam* mutants very poorly, and only hemimethylated plasmid DNA is recovered following such transformation. Mutants in *dam* can be transformed normally by nonmethylated plasmids with *oriC* as replication origin. In contrast, the wild-type cells are transformed normally by either methylated or nonmethylated plasmids with *oriC* as replication origin (133). These observations indicate that hemimethylated *oriC* sites are not effective for initiation of DNA replication in *E. coli.*

The GATC sites in the *oriC* region become hemimethylated after initiation and the passage of the replication fork. It has been suggested that such hemimethylated *oriC* regions are subsequently sequestered by binding to cytoplasmic membrane and are not accessible to methylation by Dam methylase or the DNA replication machinery for a new round of DNA replication (27). The sequestered *oriC* are released later, become methylated, and are subject to DNA replication initiation (27). In *dam* mutant cells, the *oriC* sites cannot be hemimethylated and therefore presumably cannot be sequestered. These unsequestered *oriC* sites are available for replication initiation at any time, resulting in uncoordinated replication initiation.

DnaA is a protein that determines the initiation of DNA replication (97). Interestingly, there are six GATC sites in the promoter region of *dnaA*, and the expression of *dnaA* is stimulated by DNA methylation (23, 83). The *dnaA* promoter may be sequestered after the passage of replication fork by the same mechanism as that of *oriC* (27). Therefore, DNA methylation can coordinate the initiation of DNA replication by linking the expression of the key factor in DNA replication initiation with the availability of the replication origin.

Transcriptional Modulation

DNA methylation has not been recognized as a general mechanism controlling gene expression in prokaryotes. There are only a few known cases in which the expression of a gene is affected by DNA methylation by Dam. In most of these cases, a mutation in the *dam* gene has a modest two- to sixfold effect on the transcription of the genes in question, either positively or negatively (25, 128).

Besides the positive regulation of *dnaA* by DNA methylation, the best known genes regulated by DNA methylation are those encoding transposon transposases. Expression of phage Mu *mom* gene, of which the promoter is downstream of a cluster of GATC sites, is reduced by 20-fold in a *dam* mutant host cell (60). The mechanism by which DNA methylation stimulates gene transcription is not completely understood. It is possible that the methylation at or near the promoter region causes DNA conformational change that precludes the binding of a negative transcriptional regulator. One such repressor gene, *momR*, was identified by screening mutants that render the transcription of *mom* independent of DNA methylation (18). The protein MomR, which is identical to the oxidative stress regulatory protein OxyR, binds to the *mom* regulatory region encompassing the cluster of GATC sites when the DNA is nonmethylated. Binding of MomR is not observed when DNA containing the *mom* regulatory region is methylated (18).

DNA methylation negatively affects the transcription of *sulA, trpA, trpS, glnS* and the transposase genes of some transposons, including Tn-10, Tn-5 and Tn-903 (11, 138). GATC sites are present within the promoter regions of all of these genes. The expression of those genes is stimulated in *dam* mutants by two- to tenfold. DNA methylation at the promoter region is believed to interfere by steric hindrance with the access of RNA polymerase to the promoter, causing reduced transcription (11, 138). Alternatively, the repression of these genes by DNA methylation can also be due to reduced affinity for the binding of a positive transcriptional regulatory protein.

In all of the systems described above, the methylation states of the GATC sequences are not regulated. Those GATC sequences are either fully methylated during most of the cell cycle or hemimethylated immediately following DNA replication. As a result, no variable patterns of DNA methylation corresponding to environmental signals are observed. Although such DNA methylation sensitive expression ties many cellular activities to the cell cycle, it is not an effective means for cells to respond to environmental signals.

Nonmethylated GATC Sites

It was believed that all the GATC sites in *E. coli* chromosome were completely methylated, except for a short period of temporal hemimethylation immediately following the DNA replication (108). However, using more sophisticated techniques, it has recently been shown that about 0.2% of the estimated 18,000 GATC sites in *E. coli* K-12 chromosome are completely nonmethylated (129). This finding raises the question why those GATC sites are resistant to methylation. One possibility is that some of these sites are located in stretches of non-B-form DNA sequences that are resistant to methylation or protected by some other types of conformational steric hindrance. Another possibility is that these sites are protected by binding of proteins that are involved in transcriptional regulation or chromosomal organization (129). The fact that the methylation states of many of these GATC sites depend on environmental signals, including nutrition and temperature, suggests that the methylation states are coupled with the transcriptional activity of some genes (56, 129).

Some of the nonmethylated sites have been identified by cloning the flanking sequences. Of the nine nonmethylated GATC containing sequences cloned by Wang and Church (152), seven were shown to match genes in *E. coli* database. All the seven nonmethylated GATC sites are located within the 5' noncoding regions of specific genes, suggesting the involvement of binding of transcription regulatory proteins in methylation protection of those GATC sites. Though some of the sequences containing nonmethylated

GATC sites exhibit strong consensus to the CAP binding sequence, protection by the binding of other proteins is also suggested.

Hale *et al.* (56) have similarly cloned 60 sequences containing nonmethylated GATC sites. An incomplete analysis of the 60 clones identified 10 different sequences that were located to the 5' noncoding regions of some known genes and some unidentified open reading frames. More importantly, the methylation states of these nonmethylated GATC sites were shown to be affected by stages of growth, the presence or absence of leucine, and the presence or absence of the global regulatory protein, Lrp (leucine responsive regulatory protein).

Lrp and Gene Regulation

Leucine Responsive Regulatory Protein.

Lrp was first identified by mutations that affected the transport of branched-chain amino acids (4) and later independently identified by researchers studying different gene regulation systems (5, 20, 70, 92). Purified Lrp is a 164 amino acid, 18.8 KDa protein with a pI of 9.3 that exists as a dimer in solution. It is moderately abundant, about 3,000 dimers per cell or about 0.1% of total cellular protein when cells are grown on a glucose-based minimal medium, as determined by antibody titration (161).

Genes encoding Lrp in *Salmonella typhimurium, Klebsiella aerogenes* and *Enterobacter aerogenes* have also been cloned and sequenced. They are about 90% identical to their *E. coli* counterpart and among themselves on the DNA sequence level. The amino acid sequences encoded by the *lrp* genes of the four different enteric bacteria are almost identical, differing only at two positions. Almost all of the mutations on the DNA sequence level are located at the third "wobble" position of codons, resulting in no change at the amino acid sequence level (45). The extremely high degree of conservation of Lrp among the four genera suggests that Lrp is a very important protein and is highly adapted for its function in enteric bacteria.

Lrp does not show extensive homology to any known protein, except to AsnC (79), *Pseudomonas putida* BkdR (105), and *Zymomonas mobilis* Grp (120), which are all involved in the metabolism of amino acids. AsnC is a positive transcriptional regulator of the asparagine synthetase encoding gene *asnA*; BkdR is a regulatory protein of the genes involved in the branched-chain amino acids degradation, and Grp is a protein involved in the regulation of glutamate uptake, respectively. Lrp shares 25% identity and at least 50% close similarity with AsnC, 36.5% identity and 55.8% similarity with BkdR, and 33% identity and 52% similarity to Grp, suggesting that they are evolutionally related. A plasmid carrying the *lrp* gene is capable of partially complementing *bkd* mutations in regulating the *bkdA1A2B-lpdV* operon (105). A plasmid carrying the *grp* gene has been shown to partially complement *lrp* mutations, as determined by activation of *ilvIH* transcription (120). Recently, the whole genome of *Haemophilus influenzae* was sequenced. Two *lrp*-like genes were found, which encode proteins carrying 77.2% identity and 86.7% similarity and 29.6% identity and 52.6% similarity to *E. coli* Lrp. However, the functions of these two gene products have not been reported.

A helix-turn-helix motif has been identified within Lrp (26, 121), suggesting a DNA-binding protein and consistent with its role as a transcription regulatory protein. Mutational analyses using the *ilvIH* system suggest that Lrp has three functional domains. Mutations affecting DNA binding occur within the N-terminal third of the protein, coincidental to the proposed helix-turn-helix motif, suggesting a DNA binding domain. Mutations within the C-terminal third of the protein affect only the responsiveness to leucine, suggesting a leucine responding domain. Mutations within the middle part of the protein result in reduced transcription without affecting DNA binding, suggesting a transcription activation domain (26, 121).

The expression of *lrp* is autoregulated by Lrp. Wang *et al.* (156) showed that the expression of Lrp was stimulated threefold in a *lrp* mutant and reduced tenfold in a Lrp overproducing strain. This autoregulation is not affected by leucine. The *lrp* gene is

maximally expressed when cells are grown on minimal medium. The expression is reduced by four- to 10-fold when the cells are grown in rich medium. However, this effect of rich medium on *lrp* expression is not mediated by Lrp, as the same effect is observed in *lrp* mutants (90). Therefore, there is another mechanism which regulates the expression of Lrp in responding to growth conditions.

Lrp Regulon

The gene encoding Lrp has been independently identified by a number of groups studying different gene regulation systems. These include the biosynthesis and the degradation of amino acids, the transportation of amino acid and peptide, the synthesis of tRNA, and the production of pili (26). Conceivably Lrp regulates the transcription of a large number of genes involved in different aspects of cellular activities. Systematic searches for Lrp regulated genes have identified a large number of genes or gene products of which the expression is regulated by Lrp (40, 90, 143). In one of these studies polypeptides whose production is affected by the presence or absence of Lrp were identified by two-dimensional gel electrophoresis. It has been shown that the expression of at least 30 polypeptides is affected by Lrp (40). Another study shows that in at least 66 clones carrying random λ p*lac*Mu, the expression of *lacZ* is affected by Lrp, suggesting that a large number of genes with these random insertions are regulated by Lrp (90). However, these studies do not distinguish genes directly or indirectly regulated by Lrp.

Table 1-1 is an incomplete list of members of Lrp regulon identified in various studies. Lrp-regulated genes (operons) are involved in a variety of aspects of cell activities (26). As shown in Table 1-1, these include, but are not limited to, biosynthesis of amino acids (*glnALG*, *gltBDF*, *ilvIH*, *leuABCD*, and *serA*) (40, 41, 90, 122), catabolism of amino acids (*sdaA*, *tdh*, *gcv*, and *glyA*) (90, 92), transport of amino acids and oligopeptides (*livJ*, *livKHMGF*, *oppABCD*, and *ompC*) (5, 40, 42, 57), catabolism of sugars (*malT, malEFG, malKlamBmalM*, CP8, [recently shown to be an allele of *gltD*])

Table 1-1. Lrp regulon. An incomplete list of genes or operons that are regulated by Lrp. A plus or minus sign in the "Effect of Lrp" column indicates that Lrp positively or negatively regulate the corresponding gene or operon. A plus sign in the "Effect of Leucine" column indicates that leucine is required for the function of Lrp in regulating the corresponding gene or operon, a minus sign indicates that leucine inhibits the function of Lrp; an equal sign indicates that leucine has little or no effect on the function of Lrp, a star indicates that the effect of leucine is not reported or is not clear.

Class	Operon	Function	Effect of Lrp	Effect of leucine	Ref.
Amino acid biosynthesis	*glnALG*	Glutamine biosynthesis	+	*	(40)
	gltBDF	Glutamate biosynthesis	+	-	(40, 41)
	ilvIH	Branched chain amino acid biosynthesis	+	-	(122)
	serA	Serine biosynthesisis	+	-	(92, 125)
	leuABCD	Leucine biosynthesis	+	-	(90)
Amino acid degradation	*sdaA*	Serine degradation	-	-	(92)
	tdh	Thereonine degradation	-	-	(92)
	gcv	Glycine degradation Generating 1C units	+	=	(90)
	glyA	Serine to glycine transfer	-	*	(30)
Amini acid transport	*livJ*	High affinity branched chain amino acid transport	-	+	(57)
	livKHMGF	High affinity branched chain amino acid transport	-	+	(57)
	sdaC	Serine transport	+	+	(136, 143)
	oppABCD	oligopeptide transport	-	*	(5)
	ompC	general transport	-	=	(42)
Sugar degradation	*malT*	Regulating maltose usage	+	*	(143)
	mal EFG	Maltose uptake	+	*	(143)
	malKlamB	Maltose uptake	+	*	(143)
	CP8 (*gltD*)	Pentose degradation	+	*	(30, 143)
Pili formation	*papBA*	Pap pili	+	=	(22)
	faeBC	K88pili	-	*	(65, 66)
	fanABC	K99 pili	+	-	(22)
	sfaBA	S pili	+	=	(147, 148)
	daa	F1845 pili	+	=	(147, 148)
	fimBE	Type 1 pili phase variation	+	+	(13)
Others	*lysU*	Lysyl-tRNA synthesis	-	-	(77, 90)
	pnt	DADPH synthesis	+	-	(30, 50)
	osmY	periplasmic protein, Function unknown	-	-	(88)
	lrp	Transcriptional regulation	-	=	(90, 156)

(90, 143), production of pili (*papBA, faeBC, fanABC, sfaBA, daaA*, and *fimA*) (12, 13, 15, 22, 66, 67, 147), and others (*lysU* [48, 91]; *pnt* [30] and *osmY* [88]). There are also a number of unidentified genes and proteins which appear to be regulated by Lrp (40, 90).

The most striking aspect of the Lrp regulon is not only the large number of its members but also the diverse patterns of the effects of Lrp on the expression of those genes. Lrp can be either a positive or a negative regulator of different operons, and leucine can either positively or negatively affect the function of Lrp on a particular regulatory system. Yet in some other cases, leucine has no effect on the function of Lrp. Therefore, there are six modes of regulation by Lrp (26). (i) Lrp activates transcription, and the activation is inhibited by leucine, such as the regulation of the *ilvIH* operon (122); (ii) Lrp activates transcription, and the activation is dependent or stimulated by leucine, such as in the case of regulation of *fimB* and *fimE* (13); (iii) Lrp activates transcription, and leucine has little or no effect on the activation, such as in the case of the regulation of *papBA* (22). (iv) Lrp represses transcription, and the repression is relieved by leucine, such as in the case of the regulation of *sdaA* (92); (v) Lrp represses transcription, and the repression requires leucine, such as in the case of the regulation of *livJ* and *livKHMGF* (57); and (vi) Lrp represses transcription, and leucine has little or no effect on the repression, such as in the case of the regulation of *ompC* (42).

It has been proposed that Lrp coordinates the response of bacterial cells to conditions of feast or famine (26). Lrp functions to regulate metabolic pathways in response to the availability of amino acids and nitrogen bases in the environments. In general, when cells are grown in a low nutrient environment, genes required for the biosynthesis of amino acids are activated by Lrp, and genes required for the degradation of amino acids are repressed by Lrp. Now cells are adjusted to a famine. Conversely, when cells are grown in a nutrient rich environment, the activation of amino acid biosynthesis genes and the repression of amino acid degradation genes are inhibited by the elevated level of leucine. This feast *vs.* famine response may also account for the transcriptional

activation of pili genes by Lrp, because special types of pili are often required for colonization in specific nutrient-deficient environments.

Besides its function as transcription regulatory protein for a variety of operons, Lrp may also have a function as a chromosomal organizer (30). This hypothesis is based on the observation that Lrp binds not only to the regulatory region of operons under its regulation but also to some other sequences. For example, Lrp has been reported to bind to 4 of the 10 HinFI fragments of the plasmid pBR322 (30), though the relative affinity of Lrp to those fragments in comparison with other Lrp-bound DNA sequences is not reported. Taking the binding of Lrp to pBR322 as an indication, there might be a huge number of low affinity Lrp binding sites on the chromosome. Lrp binding to those hypothetical sites would significantly affect the local structure and overall conformation of the chromosome, with a concomitant effect on the transcription of some operons.

Interaction of Lrp with DNA Molecules

Lrp binds specifically to DNA fragments containing the regulatory regions of the genes it regulates (47, 49, 66, 117, 122, 126, 153). In the case of *ilvIH*, a 331 bp region upstream of the transcription initiation point is sufficient for both transcription initiation and leucine-mediated repression of the operon (61). *In vitro*, Lrp binding to the *ilvIH* regulatory region is reduced by leucine, but not by isoleucine or valine (126). Lrp binds specifically to at least six sites over a 200 bp region in the regulatory region, as assayed by MPE footprinting (153). The binding to those sites is highly cooperative. The two sites with the highest affinity for Lrp binding, sites 2 and 4, contain dyad symmetrical sequence 5'-AGAATtttATTCT-3' and 5'-AGGATtttATCGT-3', respectively. The other sites, which have much lower affinity for Lrp binding, have sequences that conform to half of the symmetric sites. At low concentration, Lrp binds cooperatively to the upstream binding sites. At higher concentration, Lrp also cooperatively binds to the four downstream binding sites. It is the binding to these downstream sites that is required for Lrp activation of the *ilvIH* operon, since mutations in the upstream sites reduces the transcription slightly and

mutations in each of the downstream sites, except the one most proximal to the transcription initiation point, reduce the transcription significantly (153). The stoichiometry of Lrp binding to *ilvIH* regulatory region has been determined, corresponding to a dimer binding to each site (154).

Salmonella typhimurium has a cryptic *ilvIH* operon that is transcribed, but the translation is prematurely terminated due to a nonsense mutation near the start codon (127). The upstream region of *S. typhimurium ilvIH* operon is less than 60% identical to that of *E. coli* but still bound specifically by purified Lrp (155). MPE footprinting also revealed six Lrp binding sites, of which three were similar to those from *E. coli.* A consensus sequence, AgaATTTTATtcT, has been proposed as Lrp binding sequence, based on the analysis of the 12 Lrp binding sites identified in *E. coli* and *S. typhimurium ilvIH* regulatory regions (155).

An Lrp binding consensus sequence has also been drawn by selecting from a pool of random sequences the ones bound by Lrp. A pool of 50 bp DNA fragments with the middle 20 bp randomized was used as probe in Lrp mediated gel retardation. DNA molecules bound by Lrp were purified from the gel and amplified by PCR and subjected to a new round of selection. After a few rounds of selection-amplification, the selected sequences were cloned and sequenced. Comparison of the cloned sequences revealed a consensus sequence of YAGHAWATTWTDCTR, where Y = C or T, H = no G, W = A or T, D = no C, R = A or C (29). This consensus sequence is similar to the one drawn by comparing the Lrp binding sites of *E. coli* and *S. typhimurium ilvIH* regulatory regions. Surprisingly, the Lrp bound sequences selected both in the presence and in the absence of leucine are very similar, suggesting leucine does not affect the specificity of Lrp binding to DNA (29).

Another Lrp consensus sequence, TTTATTCtNaAT, has also been proposed based on the comparison of the regulatory regions of 11 Lrp regulated genes (125). However, further analysis has not been done on most of those sequences, and whether the sequences

corresponding to the proposed consensus sequence are bound by Lrp remains to be determined. It is possible that neither of the proposed consensus sequence accounts for the true Lrp-recognizing motif in all the systems regulated by Lrp, because such sequences do not exist in the regulatory regions of all the genes regulated by Lrp, and in some cases, such sequences, when present in the regulatory region, footprinting analyses fail to detect Lrp binding (30).

Lrp has also been shown to bend DNA containing Lrp binding sites from the *ilvIH* regulatory region. Circular permutation experiments showed that with a single Lrp binding site, the high affinity site number 4 of the *ilvIH* Lrp binding sites, Lrp binding induced a bending of 55°. When two adjacent sites were tested, the DNA was bent at least 130°. Considering the extensive footprinting caused by Lrp binding, it is most likely that the binding of Lrp induces a DNA conformational change such that a nucleoprotein complex is formed, either by DNA looping or wrapping around the protein molecules (154).

However, binding of Lrp to the regulatory region alone is not sufficient for its function as transcription activator, according to the functional analyses of a number of *lrp* activation mutants. The Lrp molecules produced by these activation mutants bind to the regulatory region of *ilvIH* normally but do not activate transcription (121). These same mutants Lrp also failed to activate *papBA* transcription, although the binding patterns to the *pap* regulatory region were similar to that of the wild-type Lrp (146). It has been proposed that Lrp interacts either directly with RNA polymerase or with another transcription activator, such as CAP, to activate transcription.

Pap Pili Phase Variation

E. coli isolates from patients with urinary tract infections produce Pap pili of different serotypes, which can be distinguished by specific antibodies to the major pilin subunit, PapA (33, 34). A number of *pap* operons have been cloned (9, 35, 149). In some cases a single clinical isolate expresses multiple antigenic types of Pap pili simultaneously. For example, a serotype O6 strain, C1212, contains two *pap* sequences that encode two

antigenically distinct Pap pili, antigenic type $F7_1$ and antigenic type $F7_2$ (100, 149). The type $F7_1$ pili are composed of pilin monomers of 17 KDa (Pap-17), whereas type $F7_2$ pili contain 21 KDa (Pap-21) pilin monomers. Although the majority of the cell (about 85%) in a single colony express the Pap-21, only a small portion, about 5%, of cells in the same colony express Pap-17. About another 10% of cells express neither Pap-17 nor Pap-21(101). The differential expression state of Pap pili among the cells in a single colony is due to phase variation, by which individual cells alternate the states of pili expression. The expression of Pap-17 becomes constitutive when the operon is cloned on a multicopy plasmid and transformed into *E. coli* K-12, suggesting that the presence of the excessive copies of *pap* genes abolishes the regulation of the expression. The expression of Pap-17 is restored to the wild-type level when the operon is subcloned on a single copy plasmid (101). Similarly, in another uropathogenic isolate, A55, which contains only the chromosome encoded Pap pili similar to Pap-17, about 6% of cells express the Pap pili (101). These observations indicate that the phase variation of Pap pili expression is delicately regulated and it can only be observed when the genes are present at single or very low copies.

The phase variation of Pap pili, examplified by the expression of Pap-17, was further studied by subcloning the *pap* sequence into a phage λ vector so that the *papBA* promoter controls the transcription of *lacZYA*. The λ construct containing the *pap* regulatory sequence and the *lacZYA* fusion was then integrated into the chromosome of *E. coli* K12 strain MC4100 at the *att* site. The expression of LacZYA in this strain is regulated in the same way as the expression of the Pap pili (16). Both Lac^+ and Lac^- colonies are observed at frequencies reflecting the piliated and nonpiliated cells from a parental colony. The expression of a Lac^+ phenotype is also repressed by growing in rich medium or in glucose based minimal medium. The expression state of LacZYA is heritable as phase ON (Lac^+) colonies give rise to predominantly phase ON colonies and phase OFF (Lac^-) give rise to predominantly phase OFF colonies. The frequencies of Pap pili phase variation of

cells grown on M9-glycerol minimal medium at 37°C are determined by plating thousands of cells from a single colony, assuming each parental Lac^+ cell gives rise to a Lac^+ colony and each Lac^- cell gives rise to a Lac^- colony. The phase ON to OFF switch occurs at a frequency of 2.6×10^{-2} per cell per generation and from phase OFF to phase ON at a frequency of 1.57×10^{-4} .

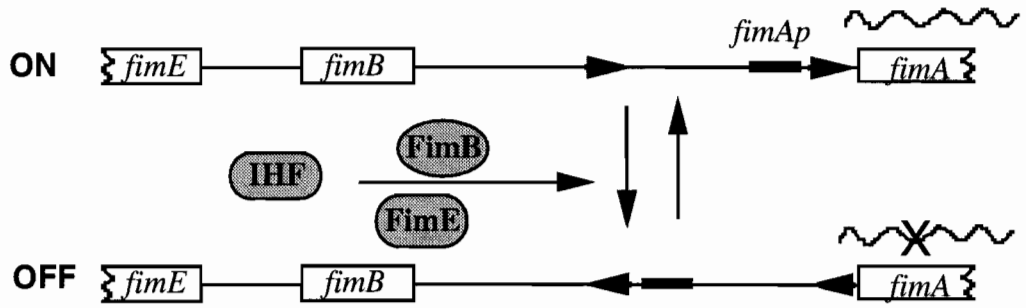
Mechanisms of Pili Phase Variation in Bacteria

A number of mechanisms of genetic variation (phase variation and antigenic variation) have been described (132) (see Figure 1-2). Type 1 pili phase variation occurs by RecA-independent site-specific recombination. This process requires the activities of FimB, FimE and IHF (39, 78). DNA sequencing indicates that a 314 bp DNA fragment, which contains the promoter region of the major pilin *fimA* gene, is differentially oriented in phase ON and phase OFF cells. In one orientation the promoter faces towards the *fimA* gene so that transcription can occur. In the other orientation the promoter faces away from *fimA* gene so that transcription is precluded (1, 38).

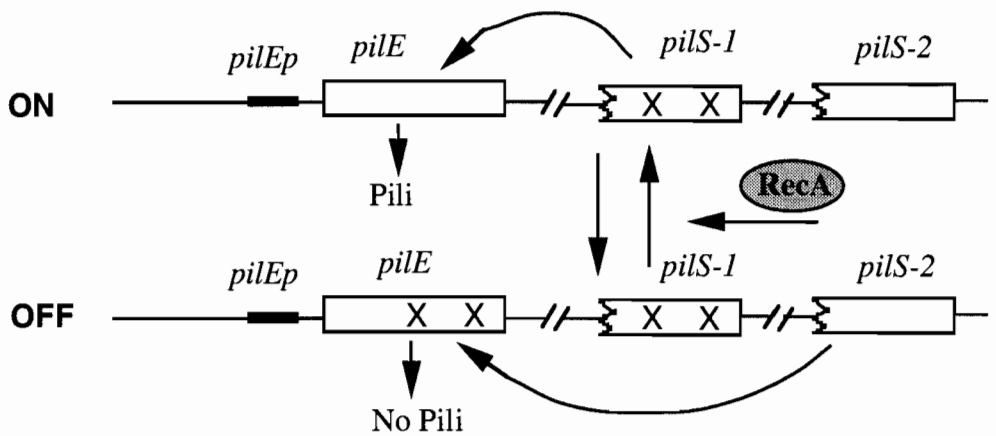
In contrast, the phase and antigenic variations of *N. gonorrhoeae* occur by RecA-dependent general homologous recombination (110). On the chromosome there are a *pilE* locus, where the major pilin gene *pil* is expressed, and *pilS* loci, where a number of silent copies of the *pil* gene encoding pilins of different antigenic specificity are located. The copies of the *pil* genes at the *pilS* locus are silent because they lack the amino terminal portion of the gene including the promoter. However, those silent copies of the *pil* gene can replace the copy at the *pilE* locus by a RecA-dependent unreciprocal recombination, leading to the expression of a pilin with different antigenic specificity (110). As *N. gonorrhoeae* cells are readily transformed by DNA they pick up in the medium, the donor DNA to the *pilE* locus can also come from sibling cells undergoing autolysis. (51). When the replacing copy of *pil* gene contains a nonsense mutation so that no pilin is expressed or a misense mutation so that the expressed pilin cannot undergo proper modification and assembly, no

Figure 1-2. Mechanisms of pili phase variation in bacteria. (A). RecA-independent site-specific recombination of *E. coli* Type 1 pili. (B). RecA-dependent nonreciprocal recombination of *N. gonorrhoeae* pili. (C). Short repeats variation of *H. influenza* pili. (D). Differential DNA methylation of *E. coli* Pap pili.

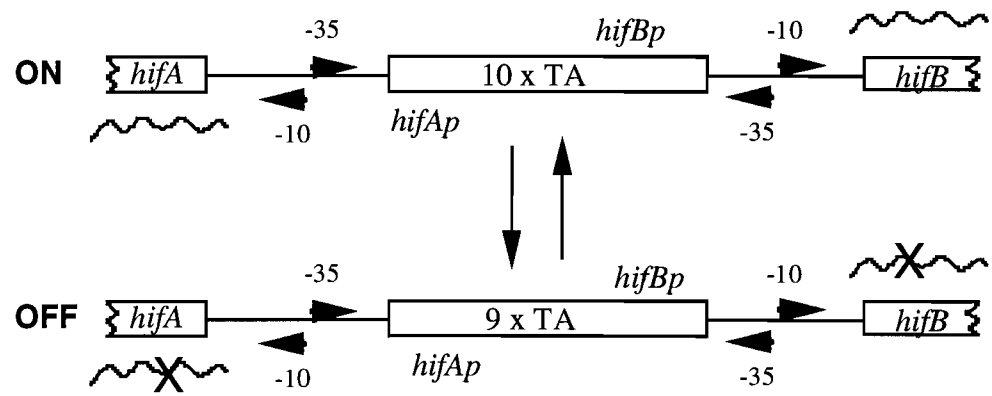
A. **RecA-independent site specific recombination. --*E. coli* Type 1 pili**



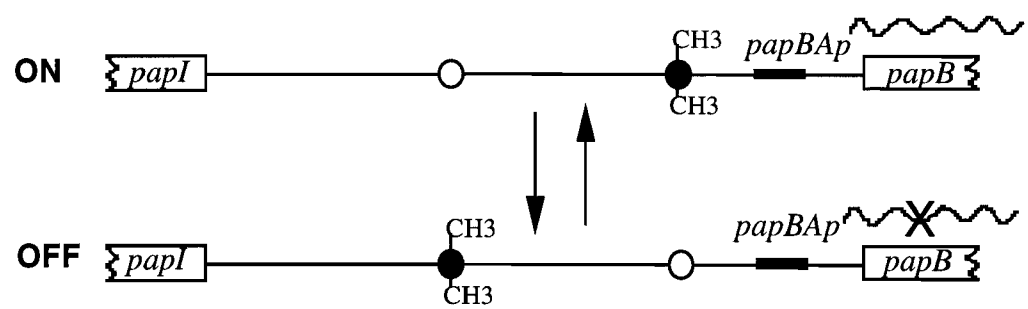
B. **RecA-dependent unreciprocal recombination. --*N. gonorrhoeae* pili**



C. **Short repeats variation. --*H. influenzae* pili**



D. **Differential DNA methylation --*E. coli* Pap pili**



pilus is produced, leading to a phase OFF state (110).

Another common mechanism of pili phase variation is the use of short repeated sequence. Such repeated sequences can be found in the coding or the regulatory region. Mispairing can happen during DNA replication or DNA recombination, resulting in deletion or insertion of the repeating units, which in turn changes the coding frame or the promoter activity, and ultimately the expression state of the pili (132). This mechanism is employed by the phase variation system of *Haemphilus influenzae* pili. There are 10 TA repeats between the -10 and -35 sequences of the overlapping promoters of the divergently transcribed pili genes hifA and hifB, which confers the maximal expression of the pili genes. When mispairing occurs during DNA replication so that one TA repeat is deleted or inserted, the expression of the pili genes is turned off or significantly reduced, resulting in a phase OFF state (151).

Mechanism of Pap Pili Phase Variation

Whereas most phase variation systems, such as that of *E. coli* Type 1 pili, *N. gonorrhoeae* pili, and *H. influenzae* pili, involve DNA sequence rearrangements or alteration, Pap pili phase variation does not. The *pap* regulatory region of DNA from populations of phase ON and phase OFF cells was sequenced following PCR amplifications. No changes at the DNA sequence level were detected in the phase ON and phase OFF populations. Moreover, Pap pili phase variation occurs independent of RecA. These and other results strongly indicate that Pap pili phase variation occurs by a mechanism that does not involve DNA rearrangement (16).

Interestingly, DNA sequence analysis of *pap* revealed two GATC sites, which are the Dam methylase targeting sequence, locate in two nearly perfect (24/27 bp identical) inverted repeats in the regulatory region of the *pap* DNA, 102 bp apart (15). DNA from phase ON and phase OFF populations were analyzed by Southern blot following digestion by a number of restriction enzymes. Though the restriction patterns by most restriction enzymes did not change, different patterns were observed when Dam methylation sensitive

enzymes were used. Using restriction enzymes MboI, which cut only at nonmethylated GATC, DpnI, which cut only at methylated GATC, and Sau3AI, which cuts at both methylated and nonmethylated GATC sequence, it was established that the GATC proximal to *papI* (denoted as GATC1028 or GATC-I) is nonmethylated in the phase ON cells and methylated in the phase OFF cells, and conversely, the GATC proximal to *papB* (GATC1130 or GATC-II) is methylated in the phase ON cells and nonmethylated in the phase OFF cells (15).

The differential DNA methylation patterns displayed by DNA from Pap phase ON and phase OFF cells suggest DNA methylation by Dam methylase is involved in regulating Pap phase variation. Examination of Pap phase variation in a Dam$^-$ background showed Dam methylase is indeed required for phase variation. In a mutant that carries a Tn-9 insertion in the *dam* gene, no phase OFF to phase ON transition was observed and the transcription from the *papBA* promoter was at the same level of the phase OFF population (15). However, complementation with plasmid carrying the *dam* gene did not restore the transcription and the phase variation. Nevertheless, replacing the mutant *dam* gene with a wild-type copy by transduction did restore the transcription and the phase variation to the wild-type level (15). These results indicate that high levels of Dam methylase also inhibit transcription from the *papBA* promoter and phase variation. Indeed, in the strain carrying the *dam* gene on a plasmid, the Dam level was fourfold higher and both GATC sites in the *pap* regulatory region were methylated (15).

The finding of nonmethylated GATC sites in the Pap regulatory region on the *E. coli* chromosome was unusual, because it was believed that virtually all the GATC sites on *E. coli* chromosome were methylated, except for a short period of hemimethylation immediately following DNA replication (108). However, a GATC site can remain nonmethylated if it is protected from methylation by the binding of a protein. Mutational analysis showed that PapI, PapB and CAP were not required for the protection of GATC-II from methylation, indicating another factor was involved in the protection of the GATC

sites from being methylated (20). Mutagenesis using mini Tn-10 identified a chromosomal locus unlinked to *pap* responsible for the observed protection. Mutation of this gene, *mbf* (for methylation blocking factor), resulted in the methylation of both GATC sites in the *pap* regulatory region. Also, the transcription of *pap* is shut off in *mbf* mutants (20). Mbf seems solely responsible for the methylation protection at GATC-II. However, both Mbf and PapI were required for the protection of GATC-I from being methylated (20). The gene *mbf* has been shown to be identical to *lrp* (22).

A model was proposed for the DNA methylation- and Lrp-dependent Pap phase variation (14). According to this model, the phase variation is determined by the interaction of the two GATC sites with Dam and Lrp or Lrp/PapI. For example, in a phase ON cell GATC-I is nonmethylated and GATC-II hemimethylated following DNA replication. Now Dam and Lrp or Lrp/PapI can compete for the two GATC sites. If Dam methylates GATC-II and Lrp/PapI bind to GATC-I, the cell remains phase ON. Conversely, if Dam methylates GATC-I and Lrp binds to GATC-II, the cell becomes phase OFF. A phase OFF to phase ON transition can be explained by a similar process.

Introduction to Work in This Thesis

Previous work has established that Pap pili phase variation is controlled by a mechanism involving the differential DNA methylation at two GATC sites in the regulatory region of the *pap* operons (14). This differential DNA methylation is dependent on the function of a global transcription regulatory protein, Lrp, which is hypothesized to protect the GATC sites from being methylated by binding to these sequences. A *pap* encoded protein, PapI, is also required for the methylation protection of the GATC-I from being methylated (20). Evidence will be presented in this thesis to show that (i) Lrp binds specifically to the *pap* regulatory region, (ii) interaction of PapI with Lrp alters the binding of Lrp to *pap* DNA, (iii) methylation states of the DNA greatly affect its interaction with Lrp and with Lrp/PapI, and (iv) cooperative binding of Lrp and Lrp/PapI to different sites

in the regulatory region helps to regulate Pap pili phase variation. The roles of Lrp, PapI, and DNA methylation in Pap pili phase variation will be discussed and a mechanistic model for Pap pili phase variation will be presented based on the experimental observations and some speculations in comparison with other gene regulation systems.

References

1. **Abraham, J. M., C. S. Freitag, J. R. Clements, and B. I. Eisenstein**. 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82:** 5724-5727.

2. **Agace, W., H. Hedges, M. Ceska, and C. Svanborg**. 1992. Interleukin-8 and the neutrophil response in mucosal Gram-negative infections. J. Clin. Invest. **92:** 780.

3. **Agace, W., S. Hedges, J. Anderson, U. Anderson, M. Ceska, and C. Svanborg**. 1993. Selective cytokine production by epithelial cells following exposure to *E. coli*. Infect. Immun. **61:** 602.

4. **Anderson, J. J., S. C. Quay, and D. L. Oxender**. 1976. Mapping of two loci affecting the regulation of branched-chain amino acid transport in *Escherichia coli* K-12. J. Bacteriol. **126:** 80-90.

5. **Andrews, J. C. and S. A. Short.** 1986. *opp-lac* fusions and transcriptional regulation of the *Escherichia coli* trp-linked oligopeptide permease. J. Bacteriol. **165:** 432-434

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

7. **Baga, M., M. Goransson, S. Normark, and B. E. Uhlin**. 1988. Processed mRNA with differential stability in the regulation of *E. coli* gene expression. Cell **52:** 197-206.

8. **Baga, M., M. Norgren, and S. Normark**. 1987. Biogenesis of *E. coli* Pap-pili: PapH, a minor pilin subunit involved in cell anchoring and length modulation. Cell **49:** 241-251.

9. **Baga, S., S. Normark, J. Hardy, P. O'Hanley, D. Lark, O. Olsson, G. Schoolnik, and S. Falkow**. 1984. Nucleotide sequence of the *papA* gene encoding the Pap pilus subunit of human uropathogenic *Escherichia coli*. J. Bacteriol. **157:** 330.

10. **Bandyopadhyay, R. and J. Das**. 1994. The DNA adenine methyltransferase-encoding gene (dam) of *Vibrio cholerae*. Gene **140:** 67-71.

11. **Barras, F. and M. G. Marinus**. 1989. The great GATC: DNA methylation in *E. coli*. Trends Genet. **5:** 139-143.

12. **Bilge, S. S., C. R. Clausen, W. Lau, and S. L. Moseley**. 1989. Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to Hep-2 cells. J. Bacteriol. **171:** 4281-4289.

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

14. **Blyn, L. B** 1991. The Regulation of Pap Pilin Phase Variation. Ph. D. Thesis. The University of Utah.

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

16. **Blyn, L. B., B. A. Braaten, C. A. White-Ziegler, D. A. Rolfson, and D. A. Low**. 1989. Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: Evidence for transcriptional regulation. EMBO J. **8:** 613-620.

17. **Bock, K., M. E. Breimer, A. Brignole, G. C. Hannsson, K. Karlsson, A., G. Larsson, H. Leffler, B. Samuelsson, E., N. Stromberg, C. Svanborg Eden, and J. Thurin**. 1985. Specificity of binding of a strain of uropathogenic *Escherichia coli* to Galα1-4Gal-containing glycosphingollipids. J. Biol. Chem. **260:** 8545-8551.

18. **Bolker, M. and R. Kahmann**. 1989. The *Escherichia coli* regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu *mom* promoter. EMBO J. **8:** 2403-2410.

19. **Boye, E. and A. Lobner-Olesen**. 1990. The role of *dam* methyltransferase in the control of DNA replication *in E. coli*. Cell **62:** 981-989.

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

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

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

23. **Braun, R. E. and A. Wright**. 1986. DNA methylation differentially enhances the expression of one of the two *E. coli dnaA* promoters in vivo and in vitro. Mol. Gen. Genet. **202:** 246-250.

24. **Breimer, M. E. and K. A. Karlsson**. 1983. Chemical and immunological identification of glycolipid-based blood group ABH and Lewis antigen in human kidney. Biochim. Biophys. Acta **755:** 170.

25. **Busby, S., A. Kolb, and H. Bug.** 1995. The *E. coli* cyclic AMP receptor protein. p. 177-191.*In* F. Eckstein and D. M. J. Lilley (ed), Nucleic Acids and Molecular Biology, Vol. 9. Springer-Verlag, Berlin.

26. **Calvo, J. M. and R. G. Matthews**. 1994. The Leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol. Rev. **58:** 466-490.

27. **Campbell, J. L. and N. Kleckner**. 1990. *E. coli oriC* and the *dnaA* gene promoter are sequestered from *dam* methyltransferase following the passage of the chromosomal replication fork. Cell **62:** 967-979.

28. **Clegg, S. and G. F. Gerlach**. 1987. Enterobacterial fimbriae. J. Bacteriol. **169:** 934-938.

29. **Cui, Y., Q. Wang, G. D. Strormo, and J. M. Calvo**. 1995. A consensus sequence for binding of Lrp to DNA. J. Bacteriol. **177:** 2872-4880.

30. **D'Ari, R., R. T. Lin, and E. B. Newman**. 1993. The leucine-responsive regulatory protein: more than a regulator. Trends Biochem. Sci. **18:** 260-263.

31. **de Man, P., B. Cedergren, S. Enerback, A. C. Larsson, H. Leffler, L. A. L., B. Nilsson, and C. Svanborg-Eden**. 1987. Receptor-specific agglutination tests for detection of bacteria that bind globoderies glycolipids. J. Clin. Microbiol. **25:** 401-406.

32. **de Man, P., I. Claeson, M. Johanson, U. Jodal, and C. Svanborg-Eden**. 1989. Bacterial attachment as a predictor of renal abnormalities in boys with urinary tract infection. J. Pediatr. **115:** 915-922.

33. **de Ree, J. M., P. Schwillens, and J. F. van den Borsh**. 1986. Monoclonal antibodies for serotyping the P fimbriae of uropathogenic *Escherichia coli*. J. Clin. Microbiol. **24:** 121-125.

34. **de Ree, J. M. and J. F. van den Boesh**. 1987. Serological response to the P fimbriae of uropathogenic *Escherichia coli* in pyelonephritis. Infect. Immun. **55:** 2204-2207.

35. **Denich, K., B. B. Blyn, A. Craiu, B. Braaten, J. Hardy, D. A. Low, and P. D. O'Hanley**. 1991. DNA sequences of three *papA* genes from uropathogenic *Escherichia coli*: evidence of structural and serological conservation. Infect. Immun. **59:** 3849-3858.

36. **Dodson, K. W., F. Jacob-Dubuisson, R. T. Striker, and S. J. Hultgren**. 1993. Outer membrane PapC usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. Proc. Natl. Acad. Sci. USA **90:** 3670.

37. **Dominque, G. J., R. Laucirica, P. Baglia, S. Covington, J. A. Robledo, and S. C. Li**. 1988. Virulence of wild-type *Escherichia coli* uroisolates in experimental pyelonephritis. Kidney Int. **34:** 761-765.

38. **Dorman, C. J. and C. F. Higgins**. 1987. Fimbrial phase variation in *Escherichia coli*: Dependence on integration host factor and homologies with other site-specific recombinases. J. Bacteriol. **169:** 3840-3843.

39. **Eisenstein, B. I., D. S. Sweet, V. Vaughn, and D. I. Friedman**. 1987. Integration host factor is required for the DNA inversion that controls phase variation in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84:** 6506-6510.

40. **Ernsting, B. R., M. R. Atkinson, A. J. Ninfa, and R. G. Matthews**. 1992. Characterization of the regulon controlled by the leucine-responsive regulatory protein in *Escherichia coli*. J. Bacteriol. **174:** 1109-1118.

41. **Ernsting, B. R., J. W. Denninger, R. W. Blumenthal, and R. G. Matthews**. 1993. Regulation of the *gltBDF* operon of *Escherichia coli:* how is a leucine-insensitive operon regulated by the leucine-responsive regulatory protein? J. Bacteriol. **175:** 7160-7169.

42. **Ferario, M., B. R. Ernsting, D. W. Borst, D. E. Wiese II, R. M. Blumenthal, and R. G. Matthews**. 1995. The leucine-responsive regulatory protein of *Escherichia coli* negatively regulates transcription of *ompC* and *micF* and positively regulates translation of *ompF*. J. Bacteriol. **177:** 103-113.

43. **Forsman, K., M. Gorannson, and B. E. Uhlin**. 1989. Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. EMBO J. **8:** 1271-1277.

44. **Forsman, K., B. Sonden, M. Goransson, and B. E. Uhlin**. 1992. Antirepression function in *Escherichia coli* for the cAMP-cAMP receptor protein transcriptional activator. Proc. Natl. Acad. Sci. USA **89:** 9880-9884.

45. **Friedberg, D., J. V. Platko, B. Tyler, and J. M. Calvo**. 1995. The amino acid sequence of Lrp is highly conserved in four enteric microorganisms. J. Bacteriol. **177:** 1624-1626.

46. **Gaastra, W. and F. K. De Graaf**. 1982. Host-Specific Fimbrial Adhesins of Noninvasive Enterotoxigenic *Escherichia coli* Strains. Microbiol. Rev. 129-161.

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

48. **Gazeau, M., F. Delort, P. Dessen, S. Blanquet, and P. Plateau**. 1992. *Escherichia coli* leucine-responsive regulatory protein Lrp controls lysyl-tRNA synthetase expression. FEBS Lett. **300:** 254-258.

49. **Gazeau, M., F. Delort, M. Fromant, P. Dessen, S. Blanquet, and P. Plateau**. 1994. Structure-function relationship of the Lrp-binding region upstream of lysU in *Escherichia coli*. J. Mol. Biol. **241:** 378-389.

50. **Gerolimatos, B. and R. L. Hansen**. 1979. Repression of *Escherichia coli* pyridine nucleotide transhydrogenase by leucine. J. Bacteriol. **134:** 394-400.

51. **Gibbs, C. P., B. Y. Remann, E. Schultz, A. Kaufmann, and T. F. Meyer**. 1989. Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. Nature (London) **338:** 651-652.

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

53. **Goransson, M., B. Sonden, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, and B. E. Uhlin**. 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli.* Nature (London) **344:** 682-685.

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

55. **Hagberg, L., H. Leffler, and C. Svanborg Eden**. 1985. Non-antibiotic prevention of urinary tract infection. Infection **13:** S196-S200.

56. **Hale, W. B., M. W. van der Woode, and D. A. Low**. 1994. Analysis of nonmethylated GATC sites in the Escherichia coli chromosome and identification of sites that are differentially methylated in response to environmental stimuli. J. Bacteriol. **176:** 3438-3441.

57. **Haney, S. A., J. V. Platko, D. L. Oxender, and J. M. Calvo**. 1992. Lrp, a leucine-responsive protein regulates branched chain amino acid transport genes in *Escherichia coli* . J. Bacteriol. **174:** 108-115.

58. **Hanley, J., I. E. Salit, and T. Hofmann**. 1985. Immunochemical characterization of P pili from invasive *Escherichia coli.* Infect. Immun. **49:** 581-586.

59. **Hasty, D. L., H. S. Courtney, E. V. Sokurenko, and I. Ofek.** 1994. Bacteria-extracellular interactions., p. 197-211. *In* P. Klemm (ed.), Fimbriae: Adhesion, Genetics, Biogenesis and Vaccines. CRC Press, Boca Raton.

60. **Hattman, S.** 1982. DNA methyltransferase-dependent transcription of the phage Mu *mom* gene. Proc. Natl. Acad. Sci. USA **79:** 5518-5521.

61. **Haughn, G. W., C. H. Squires, M. DeFelice, C. T. Lago, and J. M. Calvo**. 1985. Unusual organization of the *ilvIH* promoter of *Escherichia coli.* J Bacteriol. **163:** 186-198.

62. **Hedges, S., P. Anderson, G. Lidin-Jason, P. de Man, and C. Svanborg**. 1991. Interleukin-6 response to deliberate colonization of the human urinary tract with Gram-negative bacteria. Infect. Immun. **59:** 421.

63. **Hedges, S., M. Svensson, and C. Svanborg**. 1992. Interleukin-6 response of epithelial lines to bacterial stimulation *in vitro*. Infect. Immun. **60:** 1295.

64. **Herman, G. E. and P. Modrich**. 1981. *Escherichia coli* K-12 clones that overproduce *dam* methylase are hypermutable. J. Bacteriol. **145:** 367-380.

65. **Huisman, T. T.** 1995. Regulation of Expression of the K88 and K99 Fimbrial Operons in *Escherichia coli*. Ph. D. Thesis. Vrije University, The Netherland.

66. **Huisman, T. T., D. Bakker, P. Klaasen, and F. K. de Graaf**. 1994. Leucine-responsive regulatory protein, IS1 insertions, and the negative regulator FaeJA control the expression of the fae (K88) operon in *Escherichia coli*. Mol. Microbiol. **11:** 525-536.

67. **Huisman, T. T. and F. K. de Graf**. 1995. Negative control of *fae* (K88) expression by the 'global' regulator Lrp is modulated by the 'local' regulator FaeA and affected by DNA methylation. Mol. Microbiol. **16:** 943-953.

68. **Hull, R. A., R. E. Gill, B. Hsu, B. Minshew, and S. Falkow**. 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:** 933-938.

69. **Hultgren, S. J. and S. Normark**. 1991. Biogenesis of the bacterial pilus. Curr. Opin. Genet. Devel. **1:** 313-318.

70. **Ito, K., K. Kawakami, and Y. Nakamura**. 1993. Multiple control of *Escherichia coli* lysyl-tRNA synthetase expression involves a transcriptional repressor and a translational enhance element. Proc. Natl. Acad. Sci. USA **90:** 302-306.

71. **Jacob-Dubuisson, F., J. Heuser, K. Dodson, S. Normark, and S. Hultgren.** 1993. Initiation of assembly and association of the structural elements of a bacterial pilus depends on two specialized tip proteins. EMBO J. **12:** 837.

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

73. **Johnson, J. R., P. L. Roberts, and W. E. Stamm**. 1987. Fimbriae and other virulence factors in *Escherichia coli* urosepsis: Association with patients' characteristics. J. Infect. Dis. **156:** 225.

74. **Johnson, J. R. and W. E. Stamm**. 1989. Urinary tract infections in women: diagnosis and therapy. Ann. Intern. Med. **111:** 906-917.

75. **Kallenius, G., R. Mollby, B. Svenson, J. Winberg, and H. Hultberg**. 1980. Identification of a carbohydrate receptor recognized by uropathogenic *Escherichia coli*. Infection **8:** S288-S293.

76. **Kallenius, G., R. Mollby, S. B. Svenson, J. Winberg, A. Lundblad, S. Svensson, and B. Cedergren**. 1980. The P^k antigen as receptor for the haemagglutinin of pyelonephritic *Escherichia coli*. FEMS Microbiol. Lett. **7:** 297-302.

77. **Kawakami, K., K. Ito, and Y. Nakamura**. 1992. differential regulation of two genes encoding lysyl-tRNA synthetase in *Escherichia coli*: *lysU* constitutive mutants compensate for a *lysS* null mutation. Mol. Micobiol. **6:** 1739-1745.

78. **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.

79. **Kolling, R. and H. Lother**. 1985. AsnC, an autogenously regulated activator of asparagine synthetase A transcription in *Escherichia coli*. J. Bacteriol. **164:** 310-315.

80. **Korhonen, T. K., V. Vaisanen, H. Saxen, H. Hultber, and S. B. Svenson**. 1982. P-antigen-recognizing fimbriae from human uropathogenic *Escherichia coli* strains. Infect. Immun. **37:** 286-291.

81. **Korhonen, T. K., V. Vaisanen-Rhen, M. Rhen, A. Pere, J. Parkkinen, and J. Finne**. 1984. *Escherichia coli* fimbriae recognizing sialyl galactosides. J. Bacteriol. **159:** 762-766.

82. **Kraft, J. K. and T. A. Stamey**. 1977. The natural history of symptomatic recurrent bacteriuria in women. Medicine **56:** 55-60.

83. **Kucherer, C., H. Lother, R. Kolling, M. A. Schauzu, and W. Messer**. 1986. Regulation of transcription of the chromosomal *dnaA* gene of *Escherichia coli*. Mol. Gen. Genet. **205:** 115-121.

84. **Kuehn, M. J., D. Haslam, S. Normark, and S. J. Hultgren.** 1994. Structure, function, and biogenesis of *Escherichia coli* P pili, p. 37-51. *In* P. Klemm, (ed.), Fimbriae: Adhesion, Genetics, Biogenesis and Vaccines. CRC Press, Boca Raton.

85. **Kuehn, M. J., J. Heuser, S. Normark, and S. Hultgren**. 1992. P pili in uropathogenic *E coli* are composite fibbers with distinct fibrillar adhesive tips. Nature (London) **356:** 252.

86. **Kusters, J. G. and W. Gaastra.** 1994. Fimbrial operons and evolution., p. 179-196. *In* P. Klemm (ed.), Fimbriae: Adhesion, Genetics, Biogenesis and Vaccines. CRC Press, Boca Raton.

87. **Lahue, R. S., K. G. Au, and P. Modrich**. 1989. DNA mismatch correction in a defined system. Science **254:** 160-164.

88. **Lange, R., M. Barth, J. Bohringe, and R. Hengge-Aronis**. 1993. Complex transcriptional control of the (78)-dependent stationary phase induced and osmotically regulated osmY (csi-5) gene suggests novel roles for Lrp, IHF, and cAMP-Crp in the stationary phase response of *Escherichia coli*. J. Bacteriol. **175:** 7910-7917.

89. **Leffler, H. and C. Svanborg-Eden**. 1980. Glycolipid receptors for uropathogenic *Escherichia coli* on human erythrocytes and uroepithelial cells. Infect. Immun. **34:** 920-929.

90. **Lin, R., R. D'Ari, and E. B. Newman**. 1992. p*lac*Mu insertions in genes of the leucine regulon: extension of the regulon to genes not regulated by leucine. J. Bacteriol. **174:** 1948-1955.

91. **Lin, R., B. Ernsting, N. Hirshfield, R. G. Matthews, F. C. Neidhardt, R. L. Clark, and E. B. Newman**. 1992. The *lrp* gene product regulates expression of *lysU* in *Escherichia coli* K-12. J. Bacteriol. **174:** 2779-2784.

92. **Lin, R. T., R. D'Ari, and E. B. Newman**. 1990. The leucine regulon of *E. coli:* A mutation in *rblA* alters expression of L-leucine-dependent metabolic operons. J. Bacteriol. **172:** 4529-4535.

93. **Lindberg, F., B. Lund, L. Johansson, and S. Normark**. 1987. Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. Nature (London) **328:** 84-87.

94. **Lindberg, F., J. M. Tennent, S. J. Hultgren, B. Lund, and S. Normark**. 1989. PapD, a periplasmic transport protein in P pilus biogenesis. J. Bacteriol. **171:** 6052.

95. **Linder, H., I. Engerberg, I. Mattsby, K. Jann, and C. Svanborg Eden**. 1988. Induction of inflammation by *Escherichia coli* on the mucosal level: requirement for the adherence and endotoxin. Infect. Immun. **56:** 1309-1313.

96. **Lipsky, B. A.** 1989. Urinary tract infections in men: Epidemiology, pathophysiology, diagnosis, and treatment. Ann. Intern. Med. **110:** 138-150.

97. **Lobner-Olesen, A., K. Skarstad, F. G. Hansen, K. von Meyenboeg, and E. Boye**. 1989. The DnaA protein determines the initiation mass of *Escherichia coli* K-12. Cell **57:** 881-889.

98. **Lomberg, H., B. Cedergren, H. Leffler, B. Nilsson, A. C. Carlstrom, and C. Svanborg-Eden**. 1986. Influence of blood group on the availability of receptors for attachment of uropathogenic *Escherichia coli*. Infect. Immun. **51:** 916-926.

99. **Lomberg, H., H. M., U. Jodal, H. Leffler, K. Lincoln, and C. Svanborg-Eden**. 1984. Virulence-associated traits in *Escherichia coli* causing first and recurrent episodes of urinary tract infection in children with or without vesicoureteral reflux. J. Infect. Dis. **150:** 561-569.

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

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

102. **Low, D. A., B. Braaten, and M. van der Woude.** 1996. Fimbriae. p. 146-157. *In* F. Neidhardt (ed.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. ASM publications, Washington, D.C.

103. **Lund, B., F. Lindberg, B. I. Marklund, and S. Normark**. 1987. The PapG protein is the α-D-galactopyranosyl-(1-4)-β-D-galactopyranose-binding

adhesin of uropathogenic *Escherichia coli*. Proc. Natl Acad. Sci. USA **84:** 5898-5902.

104. **Lund, B., F. Lindberg, and S. Normark**. 1988. Structure and antigenic properties of the tip located P pilus proteins of uropathogenic *Escherichia coli*. J. Bacteriol. **170:** 1887-1894.

105. **Madhusudhan, K. T., D. Lorenz, and J. R. Sokatch**. 1993. The *bkdR* gene of Pseudomonas putida is required for expression of the *bkd* operon and encodes a protein related to Lrp of *Escherichia coli*. J. Bacteriol. **175:** 3934-3940.

106. **Marild, S., B. Wettergren, J. Hellstrom, U. Jodal, K. Lincoln, I. Orskov, F. Orskov, and C. Svanborg-Eden**. 1988. Bacterial virulence and inflammatory response in infants with febrile urinary tract infection or screening bacteriuria. J. Pediatr. **112:** 348.

107. **Marinus, M. G.** 1984. Methylation of prokaryotic DNA. p.81-111. *In* A. Razin,, H. Cedar and A. D. Riggs (ed.), DNA methylation; Biochemistry and biological significance. Springer-Verlag. New York.

108. **Marinus, M. G.** 1987. DNA methylation in *Escherichia coli*. Ann. Rev. Genet. **21:** 113-131.

109. **Marinus, M. G., A. Poteete, and J. A. Arraj**. 1984. Correlation of DNA adenine methylase activity with spontaneous mutability in *Escherichia coli* K-12. Gene **28:** 123-125.

110. **Meyer, T. F., C. P. Gibbs, and R. Haas**. 1990. Variation and control of protein expression in *Neisseria*. Ann. Rev. Microbiol. **44:** 451-477.

111. **Misra, T. K. and D. Apirion**. 1979. RNase E, an RNA processing enzyme from *Escherichia coli*. J. Biol. Chem. **254:** 11154-11159.

112. **Modrich, P.** 1991. Mechanisms and biological effects of mismatch repair. Ann. Rev. Genet. **25:** 229-253.

113. **Mudd, E. A., H. M. Krisch, and C. F. Higgins**. 1990. RNase E, an endoribonuclease, has a general role in the chemical decay of *Escherichia coli* mRNA: Evidence showing that *rne* and *arm* are the same genetic locus. Mol. Microbiol. **4:** 2127-2136.

114. **Mudd, E. A., P. Pretki, D. Belin, and H. M. Krisch**. 1988. Processing of unstable bacteriophage T4 gene 32 mRNA into a stable species requires *Escherichia coli* ribonuclease E. EMBO J. **7:** 3601-3607.

115. **Nilsson, P. and B. E. Uhlin.** 1991. Differential decay of a polycistronic *Escherichia coli* transcript is initiated by RNase-dependent endonucleolytic processing. Mol. Microbiol. **5:** 1791-1799.

116. **Norgren, M., M. Baga, J. M. Tennet, and S. Normark**. 1987. Nucleotide sequence, regulation and functional analysis of the *papC* gene required for cell surface localization of the Pap pili of uropathogenic *Escherichia coli*. Mol. Microbiol. **1:** 169.

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

118. **O'Hanley, P., D. Lark, S. Falkow, and G. Schoolnik**. 1985. Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. J. Clin. Invest. **75:**

119. **Ogden, G. B., M. J. Pratt, and M. Schaechter**. 1988. The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. Cell **54:** 127-135.

120. **Peekhaus, N., B. Tolner, B. Poolman, and R. Kramer**. 1995. The glutamate uptake regulatory protein (Grp) of *Zymomonas mobilis* and its relation to the global regulator Lrp of *Escherichia coli.* J. Bacteriol. **177:** 5140-5147.

121. **Platko, J. V. and J. M. Calvo**. 1993. Mutations affecting the ability of the *Escherichia coli* Lrp to bind to DNA, activate transcription, or respond to leucine. J. Bacteriol. **175:** 11110-11117.

122. **Platko, J. V., D. A. Willins, and J. M. Calvo**. 1990. The *ilvIH* operon of *Escherichia coli* is positively regulated. J. Bacteriol. **172:** 4563-4570.

123. **Pukkila, P., J. Peterson, G. Herman, P. Modrich, and M. Meselson**. 1983. Effects of high levels of DNA adenine methylation on methyl-directed mismatch repair in *E. coli.* Genetics **104:** 571-582.

124. **Reid, G. and J. D. Sobel**. 1987. Bacterial adherence in the pathogenesis of urinary tract infection: A review. Rev. Infect. Dis. **9:** 470-487.

125. **Rex, J. H., B. D. Aronson, and R. L. Somerville**. 1991. The *tdh* and *serA* operons of *E. coli*: Mutational analysis of the regulatory elements of leucine-responsive genes. J. Bacteriol. **173:** 5944-5953.

126. **Ricca, E., D. A. Aker, and J. M. Calvo**. 1989. A protein that binds to the regulatory region of the *ilvIH* operon of *Escherichia coli.* J. Bacteriol. **171:** 1658-1664.

127. **Ricca, E., C. T. Largo, M. Sacco, and M. De Felice**. 1991. Absence of acetohydroxy acid synthase III in *Salmonella typhimurium* is due to early termination of translation within the *ilvI* gene. Mol. Microbiol. **5:** 1741-1744.

128. **Richet, E. and L. Sogaad-Andersen**. 1994. CRP induces the repositioning of MalT at the *Escherichia coli malKp* promoter primarily through DNA bending. EMBO J. **13:** 4558-4567.

129. **Rinquist, S. and C. L. Smith**. 1992. The *Escherichia coli* chromosome contains specific, unmethylated Dam and Dcm sites. Proc. Natl. Acad. Sci. USA **89:** 4539-4543.

130. **Roberts, J. A.** 1991. Etiology and pathophysiology of pyelonephritis. Amm. J. Kidney Dis **17:** 1-9.

131. **Roberts, J. A., K. Hardaway, B. Kaack, E. N. Fussel, and G. Baskin**. 1984. Prevention of pyelonephritis by immunization with P-fimbriae. J. Urol. **131:** 602-607.

132. **Robertson, B. D. and T. F. Meyer**. 1992. Genetic variation in pathogenic bacteria. Trends Genet. **8:** 422-427.

133. **Russel, D. W. and N. D. Zinder**. 1987. Hemimethylation prevents DNA replication in *E. coli*. Cell **50:** 1071-1079.

134. **Salit, I. and E. Gotschlich**. 1977. Hemagglutination by purified type 1 *Escherichia coli* pili. J. Exp. Med. **146:** 1169-1181.

135. **Sandberg, T., B. Kaijser, G. Lidin-Janson, K. Lincoln, F. Orskov, I. Orskov, E. Stokland, and C. Svanberg-Eden**. 1988. Virulence of *Escherichia coli* in relation to host factors in women with symptomatic urinary tract infection. J. Clin. Microbiol. **26:** 1471.

136. **Shao, Z. Q., R. T. Lin, and E. B. Newman**. 1994. Sequencing and characterization of the *sdaC* gene and identification of the *sdaCB* operon in *Escherichia coli* K-12. Eur. J. Biochem. **222:** 901-907.

137. **Skarstad, K., E. Boye, and H. B. Steen**. 1986. Timing of initiation of chromosome replication in individual *Escherichia coli* cells. EMBO J. **5:** 1711-1717.

138. **Sternberg, N.** 1985. Evidence that adenine methylation influences DNA-protein interactions in *Escherichia coli*. J. Bacteriol. **164:** 490-493.

139. **Svanborg, C., F. Orskov, and I. Orskov.** 1994. Fimbriae and disease, p. 239-254. *In* P. Klemm (ed.), Fimbriae: Adhesion, Genetics, Biogenesis and Vaccines. CRC Press, Boca Raton.

140. **Svanborg Eden, C., R. Freter, L. Hagberg, R. Hull, S. Hull, H. Leffler, and G. Schoolnik**. 1982. Inhibition of experimental ascending urinary tract infection by an epithelial cell-surface receptor analog. Nature (London) **298:** 560-562.

141. **Svanborg-Eden, C., G. Lidin-Janson, and U. Lindberg**. 1979. Adhesiveness to urinary tract epithelial cells of fecal and urinary *Escherichia coli* isolates from patients with symptomatic urinary tract infections or asymptomatic bacteriuria of varying duration. J. Urol. **122:** 185.

142. **Svenson, S. B., H. Hultberg, G. Kallenius, T. K. Korhonen, R. Mollby, and J. Winberg**. 1983. P-fimbriae of pyelonephritogenic *Escherichia coli:* Identification and chemical characterization of receptors. Infection **11:** 61-67.

143. **Tchetina, E. and E. B. Newman**. 1995. Identification of Lrp-regulated genes by inverse PCR and sequencing: regulation of two mal operons of *Escherichia coli* by leucine-responsive regulatory protein. J. Bacteriol. **177:** 2679-2683.

144. **Tennent, J. M., S. Hultgren, B. Marklund, K. Forsman, M. Goransson, B. E. Uhlin and S. Normark.** 1990. Genetics of adhesion

expression in *Escherichia coli.*, p. 97-110. *In* B. W.Iglewski and V. L. Clark (ed.), Molecular Basis of Pathogenesis. Academic Press, San Diego.

145. **Van den Bosch, J. F., U. Verboom-Sohmer, P. Postma, J. De Graaff, and D. MacLaren**. 1980. Mannose-sensitive and mannose-resistant adherence to human epithelial cells and urinary virulence of *Escherichia coli*. Infect. Immun. **29:** 226-233.

146. **van der Woude, M., L. S. Kaltenbach, and D. A. Low**. 1995. Leucine-responsive regulatory protein plays dual roles as both an activator and a repressor of the *E. coli pap* fimbrial operon. Mol. Microbiol. **17**: 303-312

147. **van der Woude, M. and D. A. Low**. 1994. Leucine-responsive regulatory protein and deoxyadenosine methylase control the phase variation and expression of the *E. coli sfa* and *daa* pili operons. Mol. Microbiol. **11:** 605-618.

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

149. **van Die, I. and H. Bergmans**. 1984. Nucleotide sequence of the gene encoding the F72 fimbrial subunit of a uropathogenic *Escherichia coli* strain. Gene **32:** 83-90.

150. **van Die, I., E. Zuidweg, W. Hoekstra, and H. Bergmans**. 1986. The role of fimbriae of uropathogenic *Escherichia coli* as carriers of the adhesin involved in mannose-resistant hemagglutination. Microb. Pathol. **1:** 51-56.

151. **van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten**. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell **73:** 1187-1196.

152. **Wang, M. X. and G. M. Church**. 1992. A whole genome approach to *in vivo* DNA-protein interactions in *E. coli*. Nature (London) **360:** 606-610.

153. **Wang, Q. and J. M. Calvo**. 1993. Lrp, a global regulatory protein of *Escherichia coli*, binds co-operatively to multiple sites and activates transcription of *ilvIH*. J. Mol. Biol. **229:** 306-318.

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

155. **Wang, Q., M. Sacco, E. Ricca, C. T. Lago, M. DeFelice, and J. M. Calvo**. 1993. Organization of Lrp-binding sites upstream of *ilvIH* in *Salmonella typhimurium*. Mol. Microbiol. **7:** 883-891.

156. **Wang, Q., J. Wu, D. Friedberg, J. Platko, and J. Calvo**. 1994. Regulation of the *Escherichia coli lrp* gene. J. Bacteriol. **176:** 1831-1839.

157. **Westerlund, B., P. Kuusela, T. Vartio, I. van Die, and T. K. Korhonen**. 1989. A novel lectin-independent interaction of P fimbriae of *Escherichia coli* with immobilized fibronectin. FEBS Lett. **243:** 199.

158. **Westerlund, B., I. van Die, C. Kramer, P. Kuusela, H. Holthofer, A. M. Tarkkanen, R. Virkola, N. Riegman, H. Bergmans, W. Hoekstra, and T. K. Korhonen**. 1991. Multifunctional nature of P fimbriae of uropathogenic *Escherichia coli*: mutations in *fsoE* and *fsoF* influence fimbriae binding to renal tubuli and immobilized fibronectin. Mol. Microbiol. **5:** 2965.

159. **White-Ziegler, C. A. and D. A. Low**. 1992. Thermoregulation of the *pap* operon: evidence for the involvement of RimJ, the N-terminal acetylase of ribosomal protein S5. J. Bacteriol. **174:** 7003-7012

160. **White-Ziegler, C. A.** 1993. Thermoregulation of Pyelonephritis-Associated Pili Gene Expression. Ph. D. Thesis. The University of Utah.

161. **Willins, D. A., C. W. Ryan, J. V. Platko, and J. M. Calvo**. 1991. Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a global response to leucine. J. Biol. Chem. **266:** 10768-10774.

162. **Yoshikawa, A., S. Isono, A. Sheback, and K. Isono**. 1987. Cloning and nucleotide sequencing of the genes *rimI* and *rimJ* which encode enzymes acetylating ribosomal proteins S18 and S5 of *Escherichia coli* K12. Mol. Gen. Genet. **209:** 481-488.

163. **Zyskind, J. W. and D. W. Smith**. 1986. The bacterial origin of replication, *oriC*. Cell **46:** 489-490.

CHAPTER II

REGULATION OF PYELONEPHRITIS-ASSOCIATED PILI PHASE-VARIATION IN *ESCHERICHIA COLI*: BINDING OF THE PAPI AND THE LRP REGULATORY PROTEINS IS CONTROLLED BY DNA METHYLATION

This chapter is reprinted from *Molecular Microbiology* (1993) 7:545-553 with permission from coauthers and Blackwell Science Ltd.

Molecular Microbiology (1993) 7(4), 545–553

Regulation of pyelonephritis-associated pili phase-variation in *Escherichia coli*: binding of the PapI and the Lrp regulatory proteins is controlled by DNA methylation

Xiangwu Nou, Brett Skinner, Bruce Braaten, Lawrence Blyn,† Dwight Hirsch‡ and David Low*
Division of Cell Biology and Immunology, Department of Pathology, University of Utah Medical Center, Salt Lake City, Utah 84132, USA.

Summary

Expression of pyelonephritis-associated pili (Pap) in *Escherichia coli* is under a phase-variation control mechanism in which individual cells alternate between pili$^+$ (ON) and pili$^-$ (OFF) states through a process involving DNA methylation by deoxyadenosine methylase (Dam). Methylation of two GATC sites ($GATC_{1028}$ and $GATC_{1130}$) within the *pap* regulatory region is differentially inhibited in phase ON and phase OFF cells. The $GATC_{1028}$ site of phase ON cells is non-methylated and the $GATC_{1130}$ site is fully methylated. Conversely, in phase OFF cells the $GATC_{1028}$ site is fully methylated whereas the $GATC_{1130}$ site is non-methylated. Two transcriptional activators, PapI and Lrp (leucine-responsive regulatory protein), are required for this specific methylation inhibition. DNA footprint analysis using non-methylated *pap* DNAs indicates that Lrp binds to a region surrounding the $GATC_{1130}$ site, whereas PapI does not appear to bind to *pap* regulatory DNA. However, addition of Lrp and PapI together results in an additional DNaseI footprint around the $GATC_{1028}$ site. Moreover, Dam methylation inhibits binding of Lrp/PapI near the $GATC_{1028}$ site and alters binding of Lrp at the $GATC_{1130}$ site. Our results support a model in which Dam and Lrp/PapI compete for binding near the $GATC_{1028}$ site, regulating the methylation state of this GATC site and, consequently, the *pap* transcription state.

Received 6 August, 1992; revised 5 October, 1992; accepted 6 October 1992. †Present address: Department of Cellular and Developmental Biology, University of California, Irvine, California 94037, USA. ‡On sabbatical leave from the Department of Veterinary Microbiology and Immunology, University of California, Davis, California 95616, USA. *For correspondence. Tel. (801) 581 4901; Fax (801) 581 8946; Email Low@Bioscience.utah.edu.

Introduction

The *pap* operon in uropathogenic *Escherichia coli* encodes pili and adhesin proteins that play important roles in the attachment of *E. coli* to urinary tract epithelial cells (Normark *et al.*, 1986). The expression of Pap pili is under phase variation control: bacterial cells switch between pili$^+$ (ON) and pili$^-$ (OFF) states. Unlike other phase-variation systems examined, switching between ON and OFF Pap pili expression states occurs without *pap* DNA rearrangements or mutations (Blyn *et al.*, 1989) by a mechanism involving deoxyadenosine methylase (Dam) (Blyn *et al.*, 1990). Dam methylase plays a direct role in Pap phase-variation by methylating two GATC sites in the *pap* regulatory region denoted $GATC_{1028}$ and $GATC_{1130}$. Methylation analysis of the *pap* regulatory region showed that in the ON population the $GATC_{1028}$ site is non-methylated and the $GATC_{1130}$ site is methylated. Conversely, in the OFF population the $GATC_{1028}$ site is methylated but the $GATC_{1130}$ site is non-methylated (Blyn *et al.*, 1990).

We recently identified and cloned a gene denoted *mbf* (methylation blocking factor) which is necessary for inhibiting the methylation of the *pap* $GATC_{1028}$ and $GATC_{1130}$ sites (Braaten *et al.*, 1991). This gene is identical to the recently described *lrp* gene (leucine-responsive regulatory protein) (Braaten *et al.*, 1992) which appears to regulate a number of genes involved in cellular metabolism and will be denoted *lrp* here. Although methylation protection of the $GATC_{1130}$ site does not require any *pap*-encoded proteins, the PapI regulatory protein is necessary in conjunction with Lrp for methylation protection of the $GATC_{1028}$ site (Braaten *et al.*, 1991). Here we show that Lrp binds to the $GATC_{1130}$ region in the absence of PapI. Binding of Lrp to the $GATC_{1028}$ region is detected only after PapI addition using non-methylated and hemi-methylated DNAs, but is not detected when using fully methylated DNA. These results indicate that DNA methylation patterns control Pap phase-variation by modulating the binding of the Lrp and PapI regulatory proteins.

Table 1. The *papI* and *lrp* genes are required for transcription initiated at the *papBA* pilin promoter.

Strain	Relevant genotype	Strain description	β-galactosidase specific activity[a]
DL963	MC4100 (pNN387)	Single copy vector control	2 ± 0.8
DL967	MC4100 (pDAL287)	*papBAp–lacZ* in pNN387	98 ± 38
DL968	DL967 *lrp*$^-$	DL967 *mbf-20*::mTn*10*[b]	51 ± 19
DL969	DL967 (pDAL262)[c]	*papBAp–lacZ* + PapI *in trans*	1980 ± 134
DL970	DL969 *lrp*$^-$	DL969 *mbf-20*::mTn*10*	58 ± 22
DL971	MC4100 (pDAL288)	*papIp–lacZ* in pNN387	58 ± 24
DL972	DL971 *lrp*$^-$	DL971 *mbf-20*::mTn*10*	9 ± 3
DL973	DL971 (pDAL262)	*papIp–lacZ* + PapI *in trans*	139 ± 68
DL974	DL973 *lrp*$^-$	DL973 *mbf-20*::mTn*10*	8 ± 1

a. Values are the mean of four independent measurements ± 1 standard deviation from the mean. Units are as defined by Miller (1972).
b. The *mbf-20* allele contains a mTn*10* insertion within the *lrp* gene (Braaten *et al.*, 1991) and was transduced into *E. coli* strains by phage P1 transduction (Blyn *et al.*, 1990).
c. Plasmid pDAL262 expresses PapI constitutively in *E. coli* MC4100 since the *lacI* gene encoding *lac* repressor is deleted in this strain.

Results

PapI and Lrp are required for pap *transcription*

To determine the roles of PapI and Lrp in *pap* phase-variation, we examined the effects of these gene products on pilin transcription initiated from the divergent *papBA* and *papI* promoters. For this analysis, we used single-copy plasmids pDAL287 and pDAL288, which contain *papBA–lacZ* and *papI–lacZ* operon fusions, respectively, but lack intact PapB and PapI regulatory protein coding sequences. *E. coli* containing plasmid pDAL287 expressed about 100 Miller units of β-galactosidase (DL967, Table 1). This β-galactosidase level increased 20-fold in the presence of plasmid pDAL262, which produces PapI under the control of the *lac* promoter, indicating that PapI is a positive regulator of the *papBA* promoter (see strain DL969, Table 1). However, in the absence of Lrp, PapI did not have any effect on transcription initiated at *papBAp* (compare strains DL968 and DL970). Similarly, Lrp had less than a twofold effect on the *papBAp* promoter in the absence of PapI (compare strains DL967 and DL968, Table 1). These results indicate that both Lrp and PapI are required for stimulation of transcription from the *papBA* promoter.

E. coli containing plasmid pDAL288 (*papI–lacZ*) expressed about 60 units of β-galactosidase (DL971, Table 1). This level dropped sixfold in the *lrp*$^-$ strain DL972, indicating that Lrp is a positive regulator of *papI* transcription (compare these results to the twofold effect of Lrp on *papBA* transcription). PapI appeared to have a two- to threefold autostimulatory effect (strain DL973) which was dependent on the presence of Lrp (strain DL974). Together, these results indicate that Lrp and PapI work together to regulate transcription from both the *papBA* and *papI* promoters. In the absence of Lrp, PapI did not stimulate transcription from either of the *pap* promoters. In contrast, even in the absence of PapI, Lrp had a twofold stimulatory effect on transcription from *papBAp* and a sixfold stimulatory effect on transcription from *papIp*.

Analysis of the binding of Lrp and PapI to pap *regulatory DNA*

Our previous results showed that Lrp and PapI play roles in methylation inhibition of the *pap* $GATC_{1028}$ and $GATC_{1130}$ sites (Braaten *et al.*, 1991). Therefore, we determined if PapI and Lrp bind to *pap* regulatory DNA sequences by incubating a 292 bp *pap* DNA fragment derived from plasmid pDAL336 (Fig. 1), radiolabelled at one end, with protein extracts containing either Lrp, PapI, or both

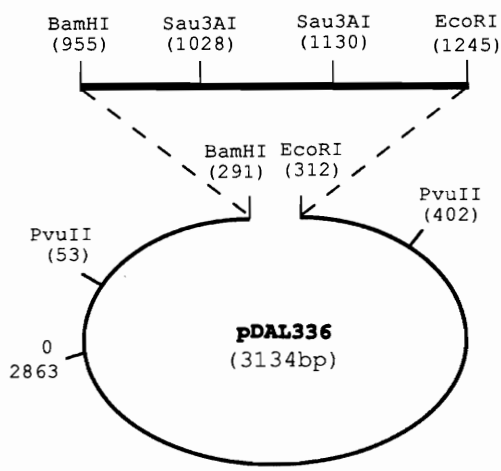


Fig. 1. Map of plasmid pDAL336. The 292 bp *Bam*HI–*Eco*RI DNA fragment of plasmid pDAL336 was used in gel retardation assays (Fig. 2) and DNase I footprinting analyses (Figs 3 and 5). The plasmid pTZ19U DNA sequence is shown by the light oval and the *pap* regulatory DNA sequence is shown by the dark line. The *pap* $GATC_{1028}$ and $GATC_{1130}$ sites are also indicated.

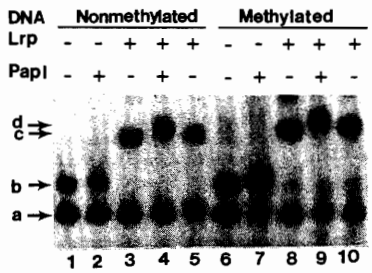


Fig. 2. Gel retardation analysis. Protein extracts containing Lrp and PapI were incubated with methylated and non-methylated *pap* regulatory DNA probes and analysed using high-ionic-strength PAGE (see the *Experimental procedures*). Additions of extracts containing Lrp and PapI are indicated above each lane. Shown on the left of the figure are DNA fragments: 'a', the 200 bp *Pvu*II–*Bam*HI DNA fragment of plasmid pDAL336B (used as an internal control); 'b', the unshifted 292 bp *Bam*HI–*Eco*RI *pap* DNA fragment; 'c', the shifted 292 bp DNA fragment resulting from Lrp binding; 'd', the shifted 292 bp DNA fragment resulting from Lrp- and PapI-binding.

proteins. DNA–protein complexes were separated by polyacrylamide gel electrophoresis (see the *Experimental procedures*). Addition of a protein extract containing PapI (Lrp^-) resulted in only a very slight band-shift of either fully methylated or non-methylated *pap* regulatory DNAs (Fig. 2, lanes 2 and 7). In contrast, addition of an extract containing Lrp ($PapI^-$) shifted band 'b' to the more slowly migrating band 'c' regardless of the *pap* methylation state (Fig. 2, lanes 3 and 8). These results indicate that Lrp binds to *pap* regulatory DNA sequences and that methylation of the $GATC_{1028}$ and $GATC_{1130}$ sites does not block Lrp binding.

Addition of PapI and Lrp together resulted in a further shift to band 'd' using methylated or non-methylated *pap* regulatory DNA sequences (Fig. 2, lanes 4 and 9). A control extract from a strain containing the *papI* gene in a transcriptional orientation opposite that of the phage T7 promoter and which is phenotypically $PapI^-$ did not result in any further shift of band 'c' (Fig. 2, lanes 5 and 10). Together, these results indicated that although PapI does not bind to *pap* DNA independently, it does appear to associate with the Lrp–*pap* DNA complex regardless of the *pap* methylation state.

Although the results shown in Fig. 2 indicated that PapI, in the presence of Lrp, bound to both methylated and non-methylated *pap* regulatory DNAs, they did not provide detailed information about the location of the DNA-binding sites or whether the interaction of PapI and Lrp with methylated DNA was the same as with non-methylated DNA. Therefore, we analysed the interactions of PapI and Lrp with *pap* regulatory DNA by DNaseI footprinting (Galas and Schmitz, 1978). PapI, in the absence of Lrp, did not alter the control DNaseI cleavage pattern using either non-methylated or methylated DNAs (Fig. 3, compare lanes 1, 2 and 6, 7, respectively), which is consistent with the results shown in Fig. 2 indicating that in the absence of Lrp, PapI does not bind to *pap* DNA. In contrast, addition of a partially purified Lrp extract resulted in an extended 120 bp footprint interrupted by DNaseI hypersensitive sites (Fig. 3, lanes 3 and 8; footprints are indicated by brackets 'c' through 'l'). These results suggest the possibility that Lrp induces bending of *pap* DNA (Hochschild, 1991). The methylation state of *pap* appeared to affect Lrp binding since the $GATC_{1130}$ site was hypersensitive to DNaseI cleavage in methylated, but not non-methylated, *pap* DNAs (Fig. 3, compare lanes 3 and 8; footprints 'i' and 'j').

Addition of PapI and Lrp to non-methylated *pap* DNA resulted in two additional protected regions of DNA near the $GATC_{1028}$ site (Fig. 3, lane 4; footprints 'a' and 'b'). The results presented in panel B (Fig. 3) show that footprint 'b' overlaps the $GATC_{1028}$ site. These results are consistent with our previous data showing that PapI is necessary for methylation protection of the $GATC_{1028}$ site (Blyn *et al.*, 1990). In addition to the new footprints 'a' and 'b' near the $GATC_{1028}$ site, a region near the $GATC_{1130}$ site became stronger and more extended in the presence of PapI (Fig. 3A, footprints 'i', 'j' and 'k'). These results, which are summarized in Fig. 4, show that although PapI does not bind to *pap* regulatory DNA sequences in the absence of Lrp, PapI does affect Lrp–*pap* DNA interactions.

Although addition of PapI and Lrp to non-methylated *pap* DNA resulted in footprinting of the $GATC_{1028}$ region, this region was not footprinted when fully methylated *pap* DNA was used (Fig. 3, compare lanes 4 and 9, footprints 'a' and 'b'). Similarly, footprints 'i', 'j', and 'k' near the $GATC_{1130}$ site were not altered after PapI addition when fully methylated DNA was used. These results show clearly that DNA methylation alters binding of Lrp–PapI to the *pap* regulatory region (see *Discussion*).

In our model for how *E. coli* alternates between phase ON and OFF methylation states, we hypothesized that after one round of DNA replication a fully methylated $GATC_{1028}$ site (phase OFF state) would give rise to a hemimethylated 'transition state'. If Lrp/PapI binds to this hemimethylated DNA and inhibits binding of Dam, then after one more round of DNA replication the $GATC_{1028}$ site would be non-methylated (phase ON state) (Blyn *et al.*, 1990). Although Lrp/PapI does not appear to bind to the $GATC_{1028}$ region of fully methylated *pap* DNA (Fig. 3), it is possible that binding to hemimethylated *pap* DNA might occur. Therefore, to test this hypothesis we constructed *pap* DNA fragments that contained a methylated top strand (orientation shown in Fig. 4) and non-methylated bottom strand (Hemi-1 DNA) as well as *pap* DNA containing a methylated bottom strand and non-methylated top strand (Hemi-2 DNA) (see the *Experimental procedures*).

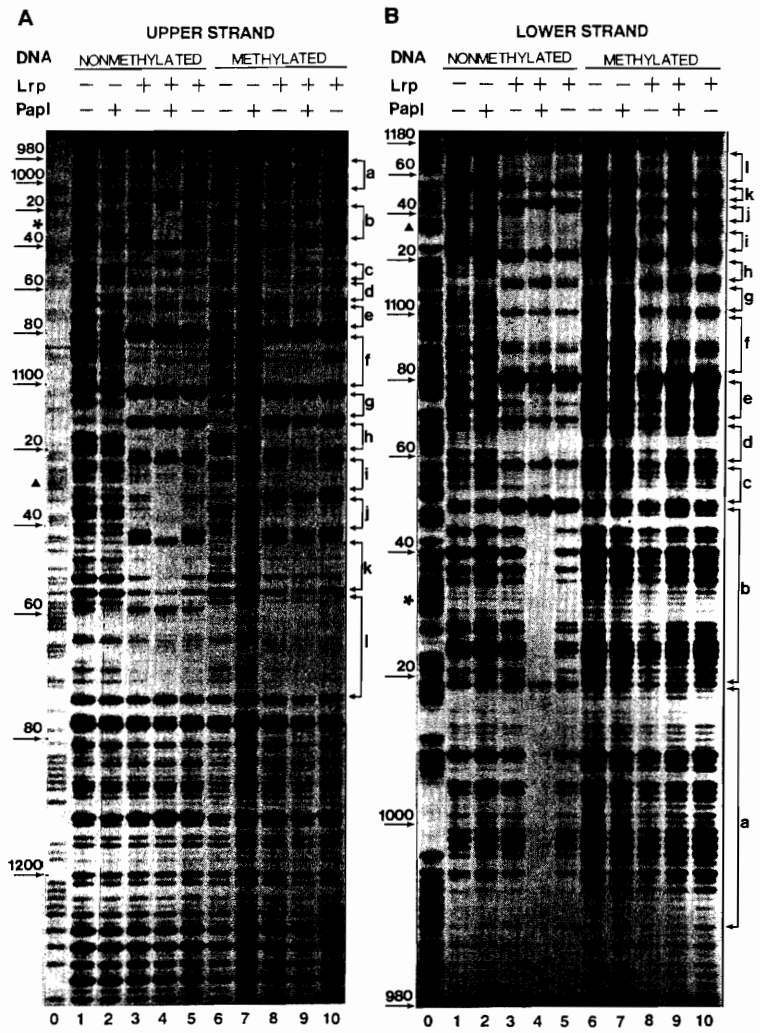


Fig. 3. DNaseI footprint analysis of fully methylated and non-methylated *pap* DNAs. Results obtained using the upper *pap* DNA strand (see Fig. 4 for orientation) are shown in (A) and results obtained using the lower DNA strand are shown in (B). Lane 0 is a G+A ladder of the upper (A) or lower (B) DNA strands. Additions of extracts containing Lrp and PapI are indicated above each lane. For lanes 3 and 8, no PapI extract was added; for lanes 5 and 10, extract from *E. coli* MC4100(pDAL288), which contains the *papI* gene in an orientation opposite to the T7 promoter and is phenotypically PapI$^-$, was added. Arrows and numbers on the left of each gel show the sequence co-ordinates corresponding to Fig. 4. Brackets and letters on the right of each gel show protected areas. The $GATC_{1028}$ site is marked by an asterisk and the $GATC_{1130}$ site is marked by a triangle.

We confirmed that these DNAs were hemimethylated by restriction enzyme digestion. Both the Hemi-1 and Hemi-2 DNAs were resistant to digestion with *Dpn*I, which cuts only fully methylated GATC sites, and *Mbo*I, which cuts only non-methylated GATC sites but were susceptible to digestion with *Sau*3A, which cuts at all GATC sites regardless of methylation state (data not shown).

DNaseI footprint analysis of the Hemi-1 and Hemi-2 DNAs was carried out under the same conditions as the analysis of non-methylated and fully methylated DNAs shown in Fig. 3. Addition of Lrp extract to both the Hemi-1 and Hemi-2 DNAs resulted in a DNA footprint pattern similar to that observed for fully methylated DNA (Fig. 5, lanes 1, 3 and 6, 8). The $GATC_{1130}$ site of fully methylated and hemimethylated DNAs is cleaved by DNaseI but is protected in non-methylated DNA (Figs 3 and 5). However, unlike results obtained with fully methylated *pap* DNA, addition of Lrp and PapI to Hemi-1 and Hemi-2 DNAs resulted in DNaseI protection of the $GATC_{1130}$ site. Thus, binding of Lrp and PapI to the Hemi-1 and Hemi-2 DNAs was different from that observed for non-methylated and fully methylated DNAs.

The $GATC_{1028}$ region of Hemi-1 and Hemi-2 DNAs was protected from DNaseI cleavage only after addition of Lrp and PapI together (Fig. 5). The Lrp/PapI footprint extended over the same base pairs as footprints 'a' and 'b' observed

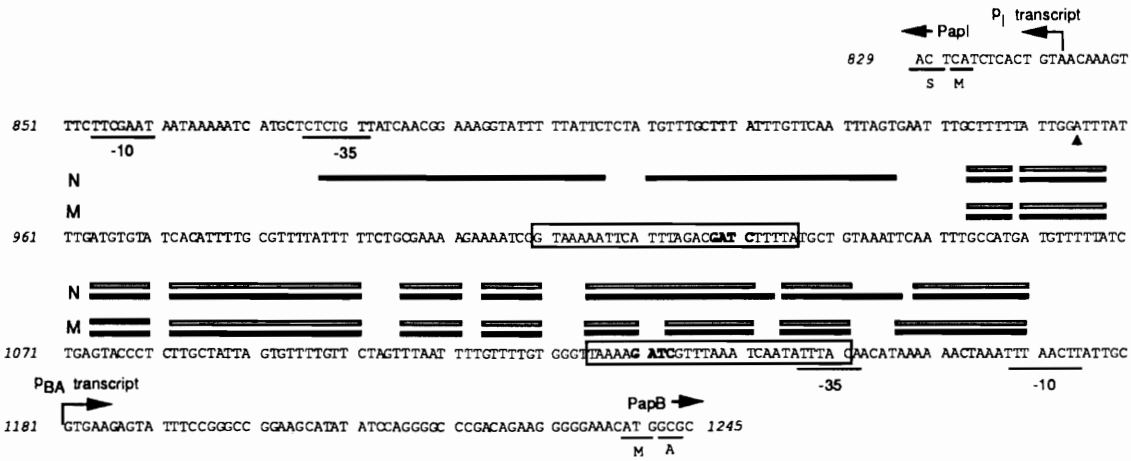


Fig. 4. Summary of DNaseI footprinting results using the *pap* regulatory region. The DNA sequence of the *pap* regulatory region is shown. The RNA polymerase-binding sites and both the transcription and translation start sites for *papI* and *papB* are indicated. The two 27 bp inverted repeats (three mismatches) containing the $GATC_{1028}$ and $GATC_{1130}$ sites (bold letters) are boxed. The bars over the DNA sequence represent areas protected from DNaseI cleavage. On the left side, the letters 'N' and 'M' indicate the results obtained using non-methylated and methylated DNAs respectively. Addition of Lrp extracts, in the absence of PapI, is indicated by an open bar. A solid bar indicates that PapI was added to the Lrp extract. The *pap* DNA sequence used for DNaseI footprinting was obtained from plasmid pDAL336B which contains DNA sequences from bp 953 (indicated by solid triangle) to bp 1244.

using non-methylated DNA (Fig. 3). However, it is apparent that both hemimethylated DNAs (Hemi-1 and Hemi-2) displayed a weaker footprint in the $GATC_{1028}$ region than was observed using non-methylated DNA after addition of the same levels of PapI and Lrp extracts. Similar results were obtained after analysis of the upper DNA strand (data not shown). Thus, Lrp/PapI appears to bind to the *pap* $GATC_{1028}$ region of hemimethylated *pap* DNAs with lower affinity than observed with non-methylated DNA. These results are in contrast to results obtained after addition of Lrp and PapI to fully methylated DNA in which binding to the $GATC_{1028}$ region was not detectable (Fig. 3).

Discussion

Recently we identified two regulatory proteins, Lrp and PapI, that are involved in methylation protection of the *pap* $GATC_{1028}$ and $GATC_{1130}$ sites (Braaten *et al.*, 1991). Here we show that Lrp binds near the $GATC_{1130}$ site in the absence of PapI (Fig. 3). However, after addition of both PapI and Lrp, binding near the $GATC_{1028}$ site as well as the $GATC_{1130}$ site occurred (Fig. 3). These results are consistent with our previous data showing that Lrp is required for methylation protection of the *pap* $GATC_{1028}$ and $GATC_{1130}$ sites whereas PapI is required in conjunction with Lrp for methylation protection of the $GATC_{1028}$ site (Braaten *et al.*, 1991). These data suggest the possibility that Lrp inhibits Dam methylation of the $GATC_{1130}$ site and Lrp/PapI inhibit methylation of the $GATC_{1028}$ site by binding near these sites and sterically blocking Dam binding.

In a previous report, we presented a general model for *pap* phase-variation which involved competition between Dam methylase and regulatory protein(s) for binding to the $GATC_{1028}$ site (Blyn *et al.*, 1990). Here we show that binding of Lrp and PapI to the *pap* regulatory region is modulated by the *pap* DNA methylation state. After the addition of Lrp and PapI to non-methylated *pap* DNAs we detected a 60 bp footprint surrounding the $GATC_{1028}$ site, which was not detected using fully methylated DNA (Fig. 3). We also detected binding of Lrp/PapI to the $GATC_{1028}$ region of hemimethylated *pap* DNAs, although the footprint was weaker than that observed using non-methylated DNA (compare Figs 3 and 5). Based on these results, the OFF methylation state would be maintained because a methylated $GATC_{1028}$ site prevents binding of PapI and Lrp. Phase switching from OFF to ON could occur after DNA replication if PapI/Lrp binds to the hemimethylated *pap* DNA intermediate, preventing Dam methylation of the $GATC_{1028}$ site on the newly synthesized DNA strand. If PapI/Lrp remains bound to this site, after one additional round of DNA replication the $GATC_{1028}$ site would be non-methylated and would remain in this ON methylation state until PapI/Lrp binding was disrupted. In support of this model, overproduction of Dam methylase blocks the OFF to ON transition (Blyn *et al.*, 1990), presumably owing to competition between PapI/Lrp and Dam for binding near the $GATC_{1028}$ site.

Binding of Lrp near the $GATC_{1130}$ site, in contrast to Lrp/PapI binding near the $GATC_{1028}$ site, is not blocked by DNA methylation (Fig. 3). However, the binding of Lrp to the $GATC_{1130}$ region, in the presence or absence of PapI,

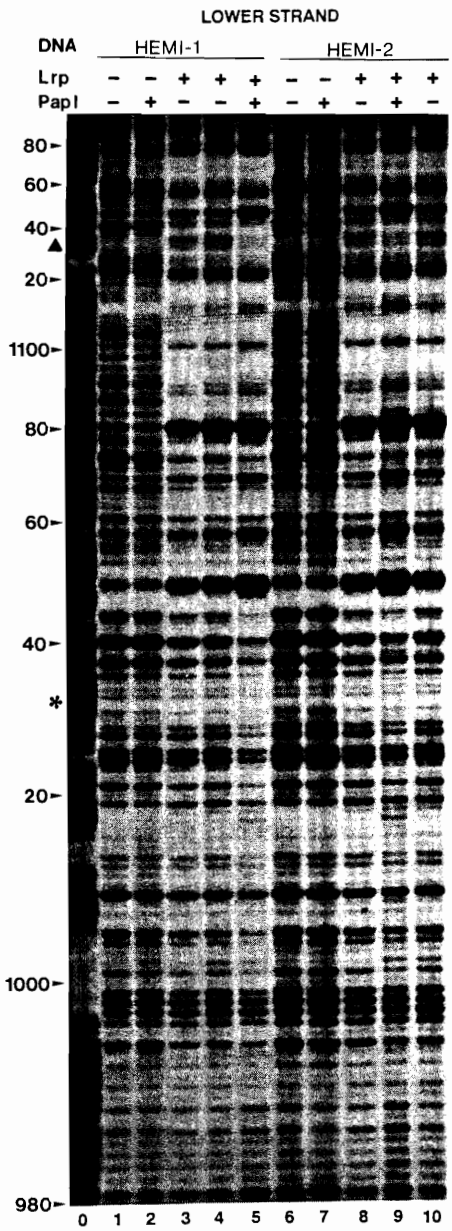


Fig. 5. DNaseI footprint analysis of hemimethylated *pap* DNAs. Hemimethylated *pap* DNAs were subjected to the same DNAseI footprint analysis as shown in Fig. 3 using equal levels of partially purified extracts containing Lrp and PapI. Hemi-1 DNA contains a methylated top strand and a non-methylated bottom strand, whereas Hemi-2 DNA has, conversely, a non-methylated top strand and a methylated bottom strand (see Fig. 4 for orientation). Results obtained using the lower DNA strand are shown. Additions of extracts containing Lrp and PapI are indicated above each lane and were carried out identically to the experiment shown in Fig. 3. Lane '0' shows a G+A ladder. Arrows and numbers on the left of each gel show the sequence co-ordinates corresponding to Fig. 4. The $GATC_{1028}$ site is marked by an asterisk and the $GATC_{1130}$ site is marked by a triangle.

was affected by DNA methylation since the $GATC_{1130}$ site from fully methylated DNAs was cleaved by DNaseI but was resistant to cleavage if non-methylated DNA was used (Figs 3 and 4). These results suggest the possibility that methylation of the $GATC_{1130}$ site may be important *in vivo* since in ON phase cells the $GATC_{1130}$ site is fully methylated but is non-methylated in OFF phase cells. In support of this hypothesis introduction of a *dam* null allele results in a OFF phase phenotype (Blyn *et al.*, 1990). Thus, although high Dam levels inhibit the OFF to ON transition, some Dam appears to be required for maintenance of the ON state.

PapI does not appear to bind specifically to *pap* DNA (Fig. 3), yet it is required along with Lrp for footprinting of the $GATC_{1028}$ region. PapI might function by forming a complex with Lrp, affecting the binding of Lrp to DNA. Alternatively, it is possible that PapI does bind to *pap* DNA near the $GATC_{1028}$ site but only in the presence of Lrp. In this regard, it is notable that the *pap* $GATC_{1028}$ and $GATC_{1130}$ sites are part of a near perfect (24/27 bp identity) inverted repeat (see boxed regions in Fig. 4). Thus, it is possible that Lrp binds to the $GATC_{1130}$ region in the absence of PapI but recognizes the similar $GATC_{1028}$ region in the presence of PapI. In any case, PapI appears to modulate protein binding at both *pap* GATC sites since addition of PapI resulted in an alteration of the Lrp–DNA footprint near the $GATC_{1130}$ site as well as a new footprint around the $GATC_{1028}$ site (Fig. 3).

Addition of PapI and Lrp to both non-methylated and methylated *pap* DNAs caused a similar shift in protein–DNA complex migration compared with addition of Lrp alone (compare bands 'c' and 'd', Fig. 2). However, DNaseI footprint analysis showed that these protein–DNA complexes were different. The protein–DNA complex obtained using non-methylated *pap* DNA was protected from DNaseI cleavage around the $GATC_{1028}$ site but the protein–DNA complex obtained using fully methylated DNA was not (Fig. 3). This could occur if PapI binds to Lrp but not directly to *pap* DNA (see above). In this case, PapI could interact with Lrp bound near the $GATC_{1130}$ site of methylated DNA without altering the Lrp–DNA footprint. Another explanation is that the additional shift in protein–DNA migration observed after addition of PapI and Lrp to methylated DNA (Fig. 2, lane 9) is the result of non-specific binding. In this regard, addition of PapI to *pap* DNA in the absence of Lrp resulted in a slight shift of the DNA (Fig. 2, lanes 2 and 7) yet no DNA footprint was detected (Fig. 3A and B, lanes 2 and 7). In these experiments a high level of non-specific DNA (10 μg) was added to prevent non-specific binding (see the *Experimental procedures*). However, it is still possible that non-specific binding occurred.

Previous results obtained using a multicopy plasmid system suggested that PapI is a positive regulator of *pap* transcription (Baga *et al.*, 1985). Our study, using single

copy plasmid pNN387, indicates that PapI is a strong positive regulator of transcription initiated from the p_{BA} promoter and is also weakly autoregulatory (Table 1). However, PapI addition did not have any effect on transcription initiated from either the p_I or p_{BA} promoters unless Lrp was present. These results along with DNA footprint data (Fig. 3) support the hypothesis that PapI and Lrp may interact to form an active transcriptional complex.

The results presented in Table 1 indicate that activation of transcription from the *papBAp* promoter can occur in the absence of the PapB regulatory protein if PapI is expressed from the *lacZp* promoter. It has been previously shown that PapB functions as a positive regulator of *pap* transcription but at high levels it represses *pap* transcription (Forsman *et al.*, 1989). Evidence indicates that binding of PapB to *pap* DNA sequences about 15 bp upstream of the proposed −35 region of the *papIp* promoter and adjacent to the Crp-binding site may stimulate transcription of *papI* (Forsman *et al.* 1989; Goransson *et al.*, 1989). Thus, PapB appears to play an indirect role in regulating *papBAp* transcription through stimulation of PapI production.

In the studies presented here we have not explored the possible role(s) of Crp in modulating binding of Lrp and PapI. Previous data suggest that Crp positively regulates the *papIp* promoter in conjunction with PapB and may also be directly required for transcription from the *papBAp* promoter (Goransson *et al.*, 1989). The Crp-binding site is located about 50 bp from the $GATC_{1028}$ site (towards *papI*) and about 180 bp from the *papBAp* promoter. Based on these data and a comparative analysis of the conserved 'GATC box' region of the *fae*, *dae*, and *sfa* operons we recently speculated that Crp might interact with RNA polymerase via a DNA loop between the $GATC_{1028}$ and $GATC_{1130}$ sites (van der Woude *et al.*, 1992). It is also possible that Crp-induced DNA bending plays a role in transcriptional activation.

Methylation of GATC sites by Dam has been shown to play an important role in a number of important regulatory processes in *E. coli* such as transposition (Roberts *et al.*, 1985; Yin and Rezikoff, 1988), mismatch repair (Modrich, 1989) and chromosome segregation (Ogden *et al.*, 1988). Each of these processes relies on DNA replication to generate a hemimethylated DNA intermediate that has altered interactions with DNA-binding proteins compared with its fully methylated precursor. This is an important way in which different biological events are co-ordinated with DNA replication in *E. coli*. Pap phase-variation, like previously described Dam-regulated systems, also involves alteration of regulatory protein–DNA interactions by DNA methylation (Fig. 3). However, the Pap regulatory system differs from other systems described since stable non-methylated GATC sites are formed, possibly as a result of binding of Lrp/PapI to DNA regions overlapping the $GATC_{1028}$ and $GATC_{1130}$ sites. It will be important to determine how DNA methylation modulates the interaction of Lrp and PapI with *pap* regulatory DNA sites and to ascertain the mechanism by which these regulatory proteins initiate formation of an active transcriptional complex.

Experimental procedures

Bacterial strains and plasmids

E. coli strain MC4100 (Casadaban, 1976) was used as a transformation recipient for all plasmids used in this study. Plasmid pDAL262B was constructed by cloning the 271 bp *Taq*I–*Sph*I DNA fragment (containing *papI* coding sequence but lacking the *papI* promoter) into *Sph*I–*Acc*I-digested plasmid pTZ18R. This places the *papI* gene under *lac* promoter control.

Plasmid pDAL283 was constructed by insertion of a *Hin*dIII–*Eco*RI DNA fragment containing *papI* from plasmid pDAL262B into plasmid pT7-5 (Tabor and Richardson, 1985) in the orientation in which the phage T7 promoter controls *papI* transcription.

Plasmid pDAL287 was constructed by insertion of the 489 bp *Hha*I DNA fragment derived from plasmid pDAL292 (Blyn *et al.*, 1990) into single-copy vector pNN387 (Elledge *et al.*, 1989) in the orientation in which the *papBA* promoter controls *lacZ* transcription. Plasmid pDAL288 was constructed similarly except that the 489 bp *Hha*I DNA fragment was inserted into vector pNN387 in the orientation in which the *papI* promoter controls *lacZ* transcription.

The 292 bp *pap* regulatory DNA fragment used for gel retardation and DNA footprint analysis was obtained from plasmid pDAL336 (Fig. 1). Plasmid pDAL336 was constructed by digesting plasmid pDAL292 (Blyn *et al.*, 1990) with *Bam*HI followed by digestion with Bal31 exonuclease to produce random deletions. After end-filling, the DNA was cut with *Eco*RI and recloned into *Sma*I–*Eco*RI-digested plasmid pTZ19U (US Biochemicals). The extent of *pap* deletion was determined by DNA sequence analysis. One subclone (plasmid pDAL336), which contained only *pap* DNA sequences to the right of bp #953, was used to generate DNA probes for gel retardation and DNaseI footprint analysis (see Fig. 4).

Preparation of cell extracts containing PapI and Lrp

A crude extract containing PapI (no Lrp) was prepared as described previously by Ausubel *et al.* (1989) using Lrp^- strain DL1393 containing plasmid pDAL283 (see above). Expression of PapI was induced by a shift from 30°C to 40°C as described by Tabor and Richardson (1985). A control extract lacking PapI was also prepared using strain DL1394, which contains the *papI* gene in the opposite transcriptional orientation to the T7 promoter and does not express PapI after thermal induction (used in Figs 2 and 3, lanes 5 and 10).

A partially purified extract containing Lrp was prepared from *E. coli* strain DL1173, which expresses about fivefold elevated levels of Lrp owing to the presence of plasmid pCV180 (Platko *et al.*, 1990) but lacks any *pap* DNA sequences. Bacteria were suspended in extract buffer (50 mM Tris-Cl at pH 7.5, 1 mM EDTA, 1 mM DTT, 1 mM PMSF), sonicated to break cells, and centrifuged at

11000 × *g* for 30 min to remove debris. Lrp was partially purified by ammonium sulphate precipitation and heparin agarose fractionation. Based on SDS–PAGE and Coomassie brilliant blue staining, Lrp comprised about 15% of total protein in the pooled peak fractions.

DNA probes

Fully methylated and non-methylated DNA probes used in gel retardation and DNA footprint experiments were prepared as follows. Plasmid pDAL336 was transformed into both Dam$^-$ strain DL738 (MC4100 dam-16 (del; kan^R)) and Lrp$^-$ strain DL844 (MC4100 *mbf-20*::mTn*10*) (Blyn *et al.*, 1990; Braaten *et al.*, 1991). Plasmid DNA isolated from DL738 was non-methylated whereas DNA from strain DL844 was methylated based on analysis using restriction enzymes *Mbo*I and *Dpn*I which cleave only non-methylated and methylated DNAs respectively. Radiolabelled probes were prepared by end-filling the *Bam*HI site (this labels the upper strand, as shown in Fig. 4) or the *Eco*RI site (this labels the lower strand) and digesting with *Eco*RI and *Bam*HI respectively to generate 292bp DNA probes labelled only one strand.

Hemimethylated DNA probes were prepared as follows. The 619bp *Pvu*II DNA fragments containing the *pap* regulatory region were isolated from both fully methylated and non-methylated plasmid pDAL336 DNAs. These DNA fragments were denatured at 90°C for 2 min in 60% DMSO containing 0.1% xylene cyanole FF, 0.1% bromophenol blue, and 2mM EDTA. DNA strands were separated by acrylamide gel electrophoresis as described previously (Maniatis *et al.*, 1982) and purified by brief electrophoresis into a 1% agarose gel. Gel slices containing DNA were placed into Spin-X columns at −20°C for 30 min, thawed at room temperature and then DNA was collected by centrifugation followed by ethanol precipitation. Hemimethylated DNA probes were prepared by mixing equimolar amounts of methylated and non-methylated single-strand DNAs together in annealing buffer (70mM Tris-HCl at pH 7.6, 10mM $MgCl_2$, 5mM dithiothreitol, 0.1mM EDTA), heating at 65°C for 10 min followed by a 30 min incubation at 37°C. Duplex hemimethylated DNA was separated from single strands by agarose gel electrophoresis.

Gel retardation and DNaseI footprint analyses

DNA-binding reactions for gel retardation were prepared as described by Ausubel *et al.* (1989). Each binding reaction contained 20000 c.p.m. of [^{32}P]-dATP-labelled *pap* DNA fragment (0.04ng to 0.1ng), 10μg of sonicated herring-sperm DNA, and 4μl of cell extract dialysed in binding buffer (60mM Tris-HCl pH 7.5, 40mM KCl, 100mM NaCl, 1mM EDTA, 1mM DTT). The total volume was adjusted to 10μl with binding buffer. The reaction mixture was incubated at room temperature for 15 min and then loaded onto a high-ionic-strength acrylamide gel as described by Ausubel *et al.* (1989). Electrophoresis was carried out at 400V for 30–90 min and exposed to film.

DNaseI footprinting (Galas and Schmitz, 1978) was carried out as follows. DNA probes, labelled on either the upper or lower strand (see above), were incubated with partially purified Lrp and PapI fractions and then incubated with DNaseI, as described by Ausubel *et al.* (1989). The protein–DNA binding reactions were similar to that described above for gel retardation except that 10^5 c.p.m. of ^{32}P-labelled DNA probe was used. After incubation for 15 min at room temperature, 2μl of DNaseI (2ng ml^{-1} in 25mM $CaCl_2$, 25mM $MgCl_2$) was added. The reaction was terminated after 1 min using 6μl of 50% glycerol containing EDTA (50mM) and bromophenol blue. Reactions were loaded onto high-ionic-strength polyacrylamide gels, as described above for gel retardation analysis, and bands containing DNA were excised from the acrylamide gel and eluted overnight at room temperature in 0.5ml of a solution containing 0.5M ammonium acetate, 1mM EDTA and 0.1% SDS. DNA was extracted once with an equal volume of chloroform and 0.5ml of 95% ethanol was added. The DNA was then precipitated by centrifugation at 15000 × *g* for 2 min, DNA pellets were suspended in 0.5ml of TE buffer and the DNA was reprecipitated as described above. Samples were analysed on 6% polyacrylamide–urea gels.

Acknowledgements

We are grateful to M. van der Woude and D. Stillman for helpful discussions and comments. This work was supported by National Institutes of Health predoctoral fellowship 5T32-GM07464 to L.B. and National Institutes of Health grants 2R01 AI23348 and 5K04 AI881 and National Science Foundations grant DMB-9019992 to D.L.

References

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1989) *Current Protocols in Molecular Biology*. New York: Green Publishing Associates/ Wiley Interscience.

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

Blyn, L.B., Braaten, B.A., White-Ziegler, C.A., Rolfson, D.A., and Low, D.A. (1989) Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. *EMBO J* **8:** 613–620.

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

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

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

Casadaban, M. (1976) Transposition and fusion of the *lac* genes to selected promoters in *E. coli* using bacteriophage lambda and Mu. *J Mol Biol* **104:** 541–555.

Elledge, S.J., Sugiono, P., Guarente, L., and David, R.W. (1989) Genetic selection for genes encoding sequence-specific DNA-binding proteins. *Proc Natl Acad Sci USA* **86:** 3689–3693.

Forsman, K., Goransson, M., and Uhlin, B.E. (1989) Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. *EMBO J* **8:** 1271–1277.

Galas, D., and Schmitz, A. (1978) DNAase footprinting: a simple method for the detection of protein–DNA binding specificity. *Nucl Acids Res* **5:** 3157–3170.

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

Hochschild, A. (1991) Detecting cooperative protein-DNA interactions and DNA loop formation by footprinting. *Meth Enzymol* **208:** 343–361.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Miller, J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Modrich, P. (1989) Methyl-directed DNA mismatch correction. *J Biol Chem* **264:** 6597–6600.

Normark, S., Båga, M., Goransson, M., Lindberg, F.P., Lund, B., Norgren, M., and Uhlin, B.E. (1986) Genetics and biogenesis of *Escherichia coli* adhesins. In *Microbial Lectins and Agglutinins; Properties and Biological Activity*. Mirelman, D. (ed.). New York: Wiley Interscience, pp. 113–143.

Ogden,G.B., Pratt, M.J., and Schaechter, M. (1988) The replicative origin of the *E. coli* chromosome binds to cell membranes only when hemimethylated. *Cell* **54:** 127–135.

Platko, J.V., Willins, D.A., and Calvo, J.M. (1990) The *ilvIH* operon of *Escherichia coli* is positively regulated. *J Bacteriol* **172:** 4563–4570.

Roberts, D., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1985) IS10 transposition is regulated by DNA adenine methylation. *Cell* **43:** 117–130.

Tabor, S., and Richardson, C.C. (1985) A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* **82:** 1074–1078.

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

Yin, J.C.P., and Reznikoff, W.S. (1988) The effect of *dam* methylation on Tn*5* transposition. *J Mol Biol* **199:** 35–45.

CHAPTER III

METHYLATION PATTERNS IN *PAP* REGULATORY DNA CONTROL PYELONEPHRITIS-ASSOCIATED PILI PHASE VARIATION IN *E. COLI*

This chapter is reprinted from *Cell* (1994) 76:577-588
with permission from coauthers and Cell Press

Methylation Patterns in *pap* Regulatory DNA Control Pyelonephritis-Associated Pili Phase Variation in E. coli

Bruce A. Braaten, Xiangwu Nou, Linda S. Kaltenbach, and David A. Low
Division of Cell Biology and Immunology
Department of Pathology
University of Utah Health Sciences Center
Salt Lake City, Utah 84132

Summary

We have examined the roles of pap DNA methylation patterns in the regulation of the switch between phase ON and OFF pyelonephritis-associated pili (Pap) expression states in E. coli. Two Dam methyltransferase sites, GATC^{1028} and GATC^{1130}, were shown previously to be differentially methylated in phase ON versus phase OFF cells. In work presented here, these sites were mutated so that they could not be methylated, and the effects of these mutations on Pap phase variation were examined. Our results show that methylation of GATC^{1028} blocks formation of the ON state by inhibiting the binding of Lrp and PapI regulatory proteins to this site. Conversely, methylation of GATC^{1130} is required for the ON state. Evidence indicates that this occurs by the inhibition of binding of Lrp to sites overlapping the pilin promoter. A model describing how the transition between the phase ON and OFF methylation states might occur is presented.

Introduction

In Escherichia coli, methylation of 5′-GATC-3′ DNA sequences by deoxyadenosine methylase (Dam) plays an important role in the timing of initiation of DNA replication (Bakker and Smith, 1989; Boye and Lobner-Olesen, 1990) as well as in coordinating cellular events such as Tn*10* transposition (Roberts et al., 1985) and mismatch repair (Modrich, 1989) with DNA replication. This occurs as a result of the formation of transiently hemimethylated GATC sites following passage of the replication fork. Because of the low intracellular Dam level, GATC sites remain in a hemimethylated state for up to 10%–20% of the cell cycle, providing a time window in which events tied to DNA replication can occur. For example, the MutH protein preferentially binds to hemi- or nonmethylated GATC sites and makes single strand nicks to the 5′ side of GATC on the nonmethylated strand. In this way, newly synthesized DNA is distinguished from template DNA by the mismatch repair system (Modrich, 1989).

Until recently it was thought that all GATC sites in E. coli are either fully methylated or, for a brief period following passage of the replication fork, hemimethylated. However, analysis of the pyelonephritis-associated pili (*pap*) operon showed that two GATC sites within the transcriptional regulatory region of this operon were nonmethylated under certain conditions, forming specific methylation patterns (Blyn et al., 1990). Many uropathogenic strains of Escherichia coli contain the *pap* operon (O'Hanley et al., 1985b), which codes for proteins involved in the synthesis and assembly of Pap pili at the surface of the bacterium (Hultgren and Normark, 1991). Pap pili are important virulence factors, because they mediate E. coli adherence to host uroepithelial cells and facilitate bacterial colonization of the urinary tract (O'Hanley et al., 1985a; Roberts et al., 1989). The expression of Pap pili by uropathogenic E. coli is subject to phase variation in which some bacteria within a single colony express Pap pili (phase ON) while other cells do not (phase OFF) (Blyn et al., 1989; Low et al., 1987).

Pap pili phase variation is controlled at the transcriptional level (Blyn et al., 1989). Two pap regulatory DNA Dam sites (GATC^{1028} and GATC^{1130}) near the *pap* pilin transcription start site are differentially methylated (Blyn et al., 1990). The GATC^{1028} site is nonmethylated in phase ON cells but methylated in phase OFF cells. Conversely, the GATC^{1130} site is methylated in phase ON cells but nonmethylated in phase OFF cells (Blyn et al., 1990). Evidence that these GATC methylation patterns might control *pap* gene expression was obtained by analysis of E. coli *dam* mutants. In both Dam$^-$ cells and cells that overexpress Dam, Pap gene expression is shut down (Blyn et al., 1990).

Further work showed that the global regulatory gene *lrp* (leucine-responsive regulatory protein) is required for specific methylation protection of both the GATC^{1028} and GATC^{1130} sites (Braaten et al., 1991; Braaten et al., 1992). Footprint analyses of nonmethylated *pap* DNAs showed that Lrp binds to the GATC^{1130} region; however, Lrp requires the regulatory protein PapI to bind to the GATC^{1028} region (Nou et al., 1993). Furthermore, both Lrp and PapI are required for activation of *pap* transcription (Nou et al., 1993). Thus, it seems likely that the GATC^{1028} site in phase ON cells is nonmethylated as a result of binding of Lrp/PapI to DNA sequences near this site, sterically inhibiting Dam methylation. Binding of Lrp/PapI to the GATC^{1028} region was not detected using fully methylated *pap* DNA, indicating that the *pap* DNA methylation state controls the binding of the Lrp and PapI regulatory proteins (Nou et al., 1993).

The results discussed above suggest that *pap* DNA methylation patterns might control *pap* transcription. However, because *dam* mutations have pleiotropic effects such as weak induction of the SOS response owing to inhibition of DNA repair (Peterson et al., 1985), it is possible that Dam affects *pap* transcription indirectly. To determine whether methylation of the *pap* GATC sites directly regulates *pap* transcription, we mutated these GATC sites so that they are nonmethylatable and observed the effects of these mutations on *pap* transcription. Based on DNA footprint analysis, these mutations did not disrupt binding of Lrp and PapI to *pap* regulatory DNA sequences. An isolate containing the GCTC^{1028} site mutation displayed a locked ON phenotype, even at high Dam levels, indicating that methylation of this site is required for transition to the phase OFF state. In contrast, cells containing the

Table 1. E. coli Strains, Plasmids, and Bacteriophage Used in This Study

E. coli Strain, Plasmid, or Bacteriophage	Description	Reference or Source
E. coli		
MC4100	F⁻*araD139 Δ(lacIPOZYA-argF) U169 rpsL thi-1*	Casadaban (1976)
DL1504	MC4100 λ354 lysogen	This study
DL1505	MC4100 λ354-1 lysogen	This study
DL1506	MC4100 λ354-2 lysogen	This study
DL1608	MC4100 λ354-3 lysogen	This study
DL842	MC4100 mTn*10::mbf-2*	This study
DL1609	DL1504 *mbf-2*	This study
DL1612	DL1505 *mbf-2*	This study
DL1615	DL1506 *mbf-2*	This study
DL1618	DL1608 *mbf-2*	This study
DL845	MC4100 mTn*10::mbf-7*	Braaten et al. (1991)
GM2929	F⁻*dam-13::* Tn*9 dcm6 hsdR2 recF143 merA⁻ merB⁻*	Marinus, unpublished data
DL1611	DL1504 *dam-13::* Tn*9* (cam^R)	This study
DL1614	DL1505 *dam-13::* Tn*9* (cam^R)	This study
DL1617	DL1506 *dam-13::* Tn*9* (cam^R)	This study
DL1620	DL1608 *dam-13::* Tn*9* (cam^R)	This study
DL1650	DL1504 (pPY1025, pTP166)	This study
DL1651	DL1505 (pPY1025, pTP166)	This study
DL1652	DL1506 (pPY1025, pTP166)	This study
DL1653	DL1608 (pPY1025, pTP166)	This study
DL738	MC4100 *dam-16* (Del; Kan^R)	This study
DL1544	DL738 (pDAL337)	This study
DL1711	DL738 (pDAL337-1)	This study
DL1712	DL738 (pDAL337-2)	This study
DL1713	DL738 (pDAL337-3)	This study
DL844	MC4100 *mbf-20*	This study
DL1547	DL844 (pDAL337)	This study
DL1545	DL844 (pDAL337-1)	This study
DL1546	DL844 (pDAL337-2)	This study
DL1761	DL1506 (pPY1025, pTZ18R)	This study
DL1762	DL1506 (pPY1025, pDAL262)	This study
DL1763	DL1608 (pPY1025, pTZ18R)	This study
DL1764	DL1608 (pPY1025, pDAL262)	This study
Plasmids		
pTZ19U	*amp* pMB1 replicon	U.S. Biochemical
pTZ18R	*amp* pMB1 replicon	U.S. Biochemical
pDAL262	pTZ18R containing a 271 bp TaqI–SphI PapI DNA fragment	Nou et al. (1993)
pDAL337	pTZ19U containing a 1.76 kb *papIB* regulatory sequence	This study
pDAL337-1	pDAL337 with *pap* GATC^{1028} site changed to GCTC	This study
pDAL337-2	pDAL337 with *pap* GATC^{1130} site changed to GCTC	This study
pDAL337-3	pDAL337 with *pap* GATCs$^{1028/1130}$ changed to GCTCs	This study
pRS550	*amp-kan-lacZYA* pMB1 replicon	Simons et al. (1987)
pDAL354	pRS550 containing a 1.76 kb *papIB* regulatory sequence	This study
pDAL354-1	pDAL354 with *pap* GATC^{1028} site changed to GCTC	This study
pDAL354-2	pDAL354 with *pap* GATC^{1130} site changed to GCTC	This study
pDAL354-3	pDAL354 with *pap* GATCs$^{1028/1130}$ changed to GCTCs	This study
pPY1025	pGB2 containing *lacI*^{q}	Youderian
pTP166	pBR322 containing p*tac-dam*	Marinus et al. (1984)
Bacteriophage		
λRS45	*amp'-lacZYA imm*^{21}	Simons et al. (1987)
λ354	λRS45-pDAL354 recombinant phage	This study
λ354-1	λRS45-pDAL354-1 recombinant phage	This study
λ354-2	λRS45-pDAL354-2 recombinant phage	This study
λ354-3	λRS45-pDAL354-3 recombinant phage	This study

GCTC^{1130} site mutation, alone or in combination with the GCTC^{1028} mutation, displayed a locked OFF phenotype. These results indicate that methylation of the GATC^{1130} site is required for *pap* transcription. This conclusion is supported by the finding that the GCTC^{1028} mutant displays a locked OFF phenotype in a *dam*⁻ background, indicating that Dam activity is necessary for *pap* transcription. Thus, the transcriptional state of the *pap* operon is constrained by the methylation states of the GATC^{1028} and GATC^{1130} sites; transcription only occurs if both of these sites are

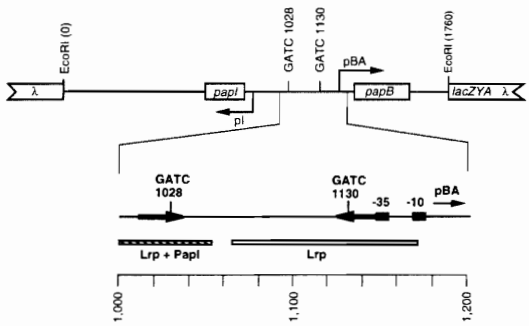


Figure 1. The *pap* Regulatory Region

The 1.76 kb EcoRI DNA fragment containing the *pap* regulatory region is shown at the top of the figure. Numbers in parentheses refer to the distance (bp) from the EcoRI (0) site. The pBA and pI promoters, controlling transcription of the *papB* and *papI* genes, respectively, are also shown. The lower part of the figure shows a detailed view of the GATC sites and pBA promoter. The GATC^{1028} and GATC^{1130} sites are contained within inverted repeats of 27 bp containing three mismatches, depicted by inverted arrows. The −35 and −10 RNA polymerase binding sites of the pBA promoter and the *pap* DNA regions bound by Lrp in the presence and absence of PapI (Nou et al., 1993) are also shown.

in the appropriate methylation configuration. A possible mechanism by which the reversible switch between ON and OFF methylation states might occur is presented in a model.

Results

Mutation of the *pap* GATC^{1028} and GATC^{1130} Sites Inhibits Pap Phase Variation

Previous work showed that *pap* transcription was greatly reduced in *dam*$^-$ and Dam overproducer E. coli isolates (Blyn et al., 1990). These results suggested that Dam plays an important role in *pap* transcription, but did not address whether this effect was a direct result of methylation of the *pap* GATC sites. Also, these conditions did not allow examination of the individual roles of the *pap* GATC^{1028} and GATC^{1130} sites in the regulation of *pap* transcription. Therefore, we constructed three *pap* regulatory DNA mutants where GATC sequence was changed to GCTC at either the 1028 site alone, the 1130 site alone, or both sites together (see Experimental Procedures and Table 1). The effects of these mutations on *pap* transcription were examined using single-copy *pap BAp-lacZYA* operon fusions in which β-galactosidase production is controlled by *pap* regulatory DNA from the *papBA* promoter (Figure 1).

Analysis of both single colonies and β-galactosidase assays of cultures in exponential growth show that the wild-type construct switches between the OFF and ON states, the 1028 mutant is locked in the ON state, and the 1130 and double-site mutants are locked in the OFF state (Figure 2 and Table 2). Thus, mutations that prevent methylation of the *pap* GATC sites lock cells in a single phase state. Moreover, the 1130 site mutation is epistatic to the

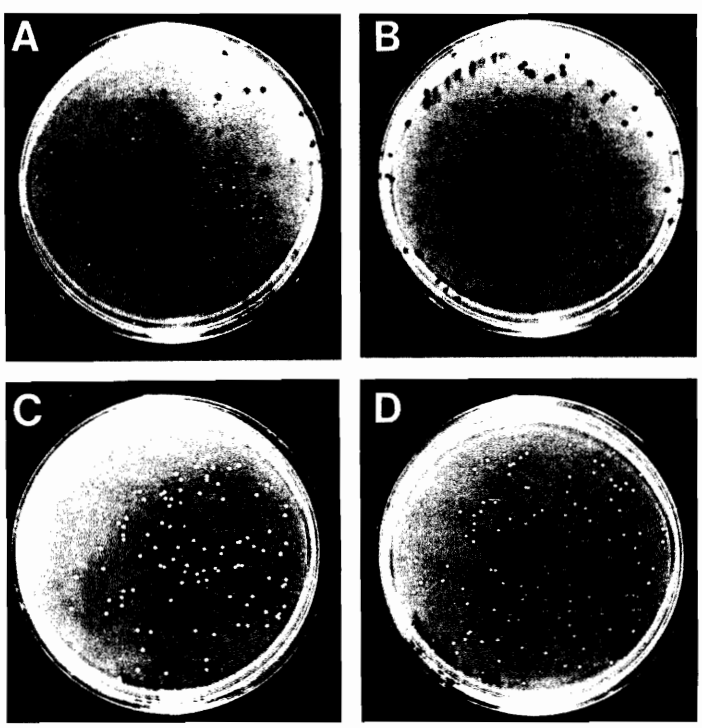


Figure 2. Analysis of LacZ Phenotypes of E. coli Strain MC4100 Containing Single-Copy Wild-Type or Mutant *pap'-lacZYA* Fusions

A single colony of each of the E. coli strains listed below was inoculated onto M9-glycerol medium containing the β-galactosidase indicator X-gal as described in Experimental Procedures. (A), DL1504 (wild-type GATC sites, phase variation phenotype); (B), DL1505 (GCTC^{1028} mutation, locked ON phenotype); (C), DL1506 (GCTC^{1130} mutation, locked OFF phenotype); (D), DL1608 (GCTC^{1028} and GCTC^{1130} mutations, locked OFF phenotype).

Table 2. The Effects of Lrp and Dam on *pap* Transcription in E. coli Strains Containing Wild-Type or Mutant *pap* Regulatory DNAs

Relevant E. coli Genetic Background	E. coli Strain	*pap* Site Mutated from GATC to GCTC	β-Galactosidase Activity (Miller Units)[a]
Wild-type	DL1504[b] phase OFF	None	55 ± 11
	DL1504[b] phase ON	None	1630 ± 199
	DL1505	1028	4177 ± 78
	DL1506	1130	7 ± 1
	DL1608	1028 and 1130	25 ± 1
lrp⁻ (mbf-2)	DL1609	None	6 ± 0
	DL1612	1028	7 ± 1
	DL1615	1130	6 ± 1
	DL1618	1028 and 1130	6 ± 1
dam⁻	DL1611	None	84 ± 6
	DL1614	1028	24 ± 0
	DL1617	1130	235 ± 27
	DL1620	1028 and 1130	36 ± 3
dam⁺ (4-fold higher)[c] (pPY1025/ pTP166; no IPTG in medium)	DL1650	None	56 ± 20
	DL1651	1028	5584 ± 63
	DL1652	1130	5 ± 0
	DL1653	1028 and 1130	26 ± 2
dam⁺ (69-fold higher)[c] (pPY1025/ pTP166; 1 mM IPTG in medium)	DL1650	None	70 ± 34
	DL1651	1028	5429 ± 545
	DL1652	1130	5 ± 0
	DL1653	1028 and 1130	24 ± 1

[a] β-Galactosidase-specific activities were measured by the method of Miller (1972).
[b] DL1504 displays a phase variation phenotype.
[c] Dam methylase activities were previously determined (Blyn et al., 1990).

1028 site mutation, suggesting that methylation of the $GATC^{1130}$ site is essential for *pap* transcription.

The wild-type E. coli lysogen DL1504 switches phase states, since both blue (phase ON) and white (phase OFF) colonies are observed (Figure 2A). The DL1504 phase OFF wild-type lysogen produced about 30-fold less β-galactosidase than its phase ON counterpart (Table 2). This result agrees with our previous work on a similar construct, DL379, where we found a 33-fold difference between phase OFF and phase ON β-galactosidase production (Blyn et al., 1989).

E. coli strain DL1505, which has the $GATC^{1028}$ site mutated to GCTC, produces only blue colonies (Figure 2B); therefore, the Pap switch is locked ON. We found that these locked ON cells expressed 2.5-fold more β-galactosidase than the wild-type phase ON strain DL1504. This result is expected, because in the wild-type population, cells are switching OFF at a rate that is about 60-fold higher than the ON switch rate. Consequently, after 24 generations, an ON cell produces a colony containing only 25% ON cells (Blyn et al., 1989). Because all cells in the locked ON strain DL1505 population are producing β-galactosidase, the locked ON population will express higher levels of the enzyme than a phase ON population of cells.

Mutation of the $GATC^{1130}$ site to GCTC, either alone or in combination with the 1028 site mutation, also prevents Pap phase variation, but in this case, only white colonies are observed, and the Pap switch is locked OFF (Figures 2C and 2D). The locked OFF 1130 site and double-site mutants expressed only 7 U and 25 U, respectively, of β-galactosidase, compared with 55 U for the phase OFF DL1504 strain. Because the phase OFF population contains about 0.9% ON cells as a result of low frequency switching to phase ON, one would expect to find higher levels of β-galactosidase activity in the phase OFF population versus the locked OFF population.

The 1028-Site Mutant Is Transcriptionally OFF in a *dam⁻* Background but Remains Locked ON under Conditions of Dam Overproduction

Previous results showed that *pap* transcription is inhibited under conditions of Dam overexpression and in the absence of Dam (Blyn et al., 1990). Similar data are shown in Table 2 for comparison. These effects of aberrant Dam levels on *pap* transcription could be due to alterations in the methylation of the $GATC^{1028}$ and $GATC^{1130}$ sites or, alternatively, might be caused by pleiotropic effects known to occur under these conditions. To distinguish between these possibilities, we measured the effects of alterations of Dam levels on *pap* transcription using the *pap* GCTC 1028, 1130, and double-site mutants.

The 1028 site mutant maintained a locked ON phenotype in the presence of 4-fold higher Dam levels and produced even more β-galactosidase than in the wild-type background (5584 U, compared with 4177 U in wild-type cells; Table 2). Similar results were obtained in the presence of 69-fold higher Dam levels. Because *pap* transcription in the 1028 site mutant was not reduced by Dam overproduction, these results indicate that methylation of the $GATC^{1028}$ site inhibits *pap* transcription. Moreover, the fact that the 1028 site mutant expressed significantly more β-galactosidase under high Dam levels suggests the pos-

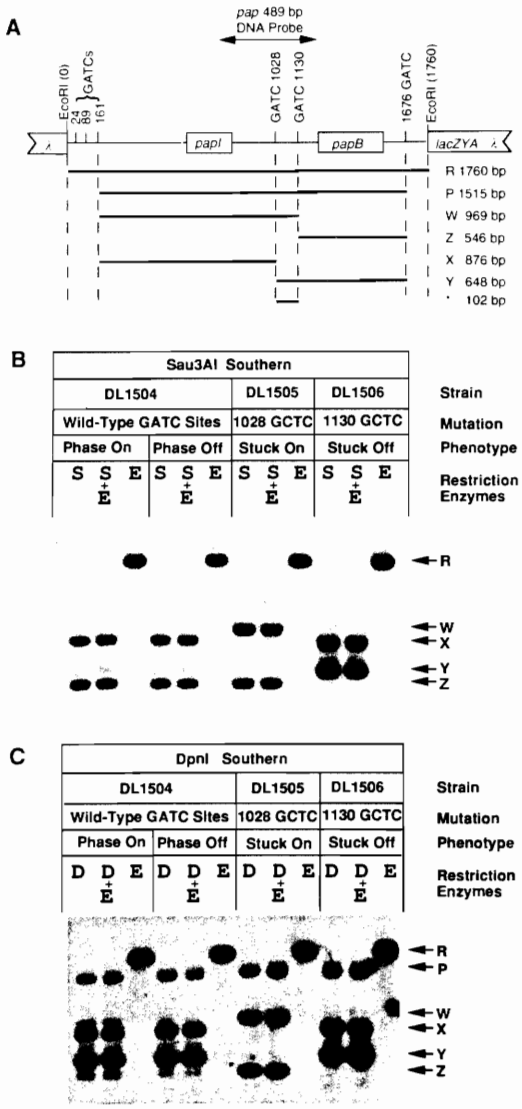


Figure 3. Analysis of the GATC Methylation States of *pap* DNA Sequences Isolated from Wild-Type and Regulatory Mutant E. coli Strains

Chromosomal DNAs were isolated from E. coli strains and digested with Sau3AI (S) and EcoRI (E), as shown in (B), or with DpnI (D) and EcoRI (E), as shown in (C). DNA fragments were separated by electrophoresis on 1.3% agarose gels, transferred to nylon membranes, and probed using a 489 bp *pap* regulatory DNA fragment (A) as described in Experimental Procedures. Each band in the Southern blots shown in (B) and (C) has been given a letter designation that corresponds to the DNA fragments shown in (A). Also shown in (A) is the 102 bp DNA fragment generated by Sau3AI cleavage at both the 1028 and 1130 GATC sites, which is not seen in the Sau3AI Southern in (B).

sibility that methylation of the 1130 site enhances *pap* transcription. As expected, both the 1130 site mutant and the double-site mutant also had low β-galactosidase activities in the Dam-overproducing cell, similar to that observed in cells expressing normal levels of Dam (Table 2).

In contrast with the the results obtained under conditions of Dam overproduction, in the absence of Dam the 1028 site mutant displayed a phase OFF phenotype (4177 U in the wild-type background, compared with 24 U in *dam*⁻ cells, Table 2). Thus, the locked ON phenotype of the 1028 site mutant is dependent upon the presence of Dam. This result, combined with the finding that the 1130 site mutation is epistatic to the 1028 site mutation (Table 2), strongly indicates that methylation of the 1130 site is required for *pap* transcription. Both the GCTC^{1130} and double mutant remained in the phase OFF state in the *dam*⁻ background, although the β-galactosidase level measured for the GCTC^{1130} mutant was higher.

Analysis of the In Vivo Methylation States of *pap* Regulatory DNA GATC Sites in Both Wild-Type and Mutant E. coli Strains

As discussed above, mutation of the *pap* GATC^{1028} and GATC^{1130} sites caused cells to be locked in a transcriptionally active (ON) or inactive (OFF) state, respectively (Table 2). These results suggest the possibility that these mutant strains are also locked in a single methylation state. To test this hypothesis, we analyzed the methylation states of the *pap* GATC sites by Southern blot analysis (see Experimental Procedures). Our results, shown in Figure 3, indicate that the nonmutated GATC^{1130} co-site of the 1028 site locked ON mutant is methylated, similar to that observed in phase ON cells. Conversely, the nonmutated GATC^{1028} co-site of the 1130 site locked OFF mutant is methylated, similar to that of phase OFF cells. Figure 3B shows control digestions with Sau3AI, which cuts GATC sites regardless of methylation state. The Sau3AI digest of the phase ON and phase OFF wild-type *pap* DNAs produced the expected DNA fragments X and Z (Figure 3A). The 102 bp band (asterisk) is not seen in the Southern shown here; however, we have observed this fragment using a 1.5% agarose gel (data not shown). Digestion of DL1505 DNA produced fragments W and Z, owing to mutation of the 1028 site, whereas digestion of DL1506 DNA produced fragments X and Y, owing to mutation of the 1130 site. To confirm that the GATC sites cut by Sau3AI are within the *pap* DNA fragment, we digested with EcoRI, which cuts at the borders of the *pap* DNA insert. Addition of EcoRI did not affect the Sau3AI digest pattern (Figure 3B).

Analysis of the methylation states of the *pap* regulatory DNA GATC sites was carried out using restriction enzyme DpnI, which cuts only fully methylated GATC sites. As shown in Figure 3C, most DNAs from phase OFF DL1504 cells are fully methylated at the GATC^{1028} site, as evidenced by the appearance of DNA fragments X and Y. In contrast, in phase ON DL1504 cells, DNA containing a fully methylated GATC^{1130} site was observed as indicated by DNA fragments W and Z. DNA fragments X and Y were also present, since the phase ON population contained 17% phase ON cells and 83% phase OFF cells (see above). Most DNAs isolated from the locked ON GCTC^{1028} mutant contained a fully methylated GATC^{1130} site (bands W and Z). Hence, the locked ON mutant retains the 1130 site methylation state found in phase ON cells. Conversely, DNA from the locked OFF GCTC^{1130} mutant con-

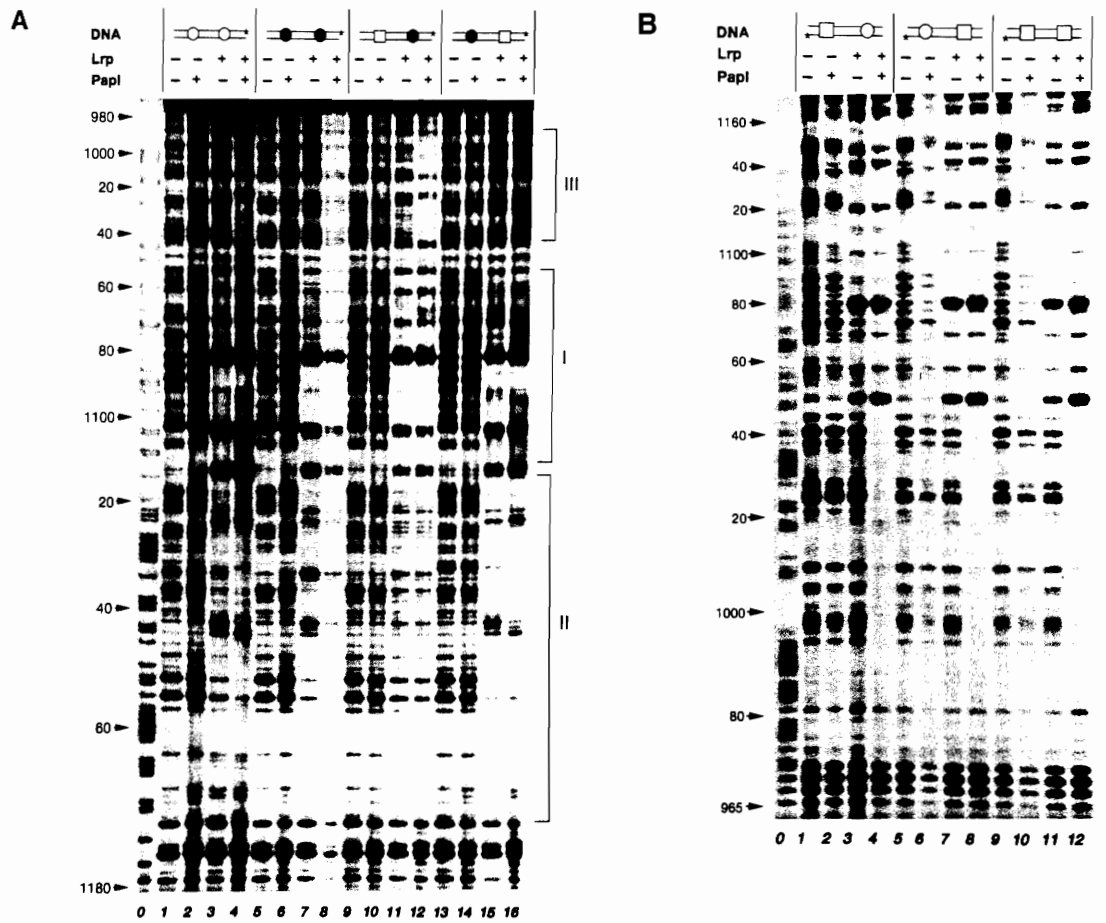


Figure 4. DNAase I Footprint Analysis of Wild-Type and Mutant *pap* DNAs

Results obtained using the upper *pap* DNA strand, with GATC[1028] on the left and GATC[1130] on the right, are shown in (A), and results obtained using the lower DNA strand are shown in (B). At the top of each figure, the methylation states of each GATC site are indicated by a black circle (methylated) or a white circle (nonmethylated), and the labeled DNA strand is marked with a star. GATC sites that have been mutated to GCTC sequence are indicated by a white square. Additions of Lrp and PapI are indicated above each lane. Arrows and numbers to the left of each panel show the base pair locations of *pap* regulatory DNA sequence. The brackets shown at the right in (A) indicate footprints detected after addition of Lrp (I and II) and Lrp/PapI (I, II, and III). Lane 0 is a G + A ladder of the upper (A) or lower (B) DNA strands.

tains a fully methylated GATC[1028] site (bands X and Y). Thus, the locked OFF mutant retains the 1028 site methylation state present in phase OFF cells. The presence of DNA fragment P, detected in DpnI digests from both the locked ON and locked OFF mutants, indicates that a small fraction of GATC co-sites were either hemimethylated as a result of DNA replication or nonmethylated.

Further analysis of the methylation states of *pap* DNAs was carried out by double digestion with restriction enzymes EcoRI and MboI to detect nonmethylated GATC sites that are cut by MboI. These results (data not shown) were consistent with the methylation state analysis of the *pap* GATC[1028] and GATC[1130] sites described above using DpnI. Together, these results show that the GATC site mutations lock cells in one methylation state as well as one transcription state.

The 1028 Site and 1130 Site GCTC Mutations Do Not Disrupt Binding of Lrp and PapI to *pap* Regulatory DNAs

Our analysis of the *pap* GATC site mutants described above is based on the assumption that mutation of the *pap* GATC sites to GCTC sequence does not cause aberrant regulation of *pap* transcription owing to, for example, inhibition of the binding of proteins to these sites. To determine whether this assumption is valid, we analyzed the binding of Lrp and PapI to both wild-type and mutant *pap* DNAs by DNAase I footprinting. As shown in Figure 4A, Lrp binds to both methylated and nonmethylated wild-type DNAs in the GATC[1130] site region, as evidenced by alternating protected and DNAase I-hypersensitive sequences (footprints I and II, lanes 3 and 7). Although PapI alone does not bind to *pap* DNA (Figure 4A, lanes 2 and 6), addition

Table 3. The Effects of PapI Overproduction on Methylation Protection of the GATC[1028] Site and *pap* Transcription in the GCTC 1130 Site Mutant

E. coli Strain	Plasmids Present[a]	IPTG[b]	GATC[1028] Sites Protected from Methylation (%)[c]	β-Galactosidase Activity (Miller Units)[d]
DL1761	pTZ18R (vector control)	None	11	9.6 ± 2.5
DL1761	pTZ18R (vector control)	100 μM	11	12.9 ± 3.3
DL1762	pDAL262 (PapI$^+$)	None	26	118.8 ± 6.6
DL1762	pDAL262 (PapI$^+$)	100 μM	68	372.7 ± 30.9

[a] Plasmid pPY1025 (*lacI*^{q}) is also present in each strain to repress transcription of the *papI* gene, which is under *lac* promoter control.
[b] IPTG was added to cultures as indicated to induce *papI* transcription.
[c] The percent of GATC[1028] sites that was cleavable by restriction enzyme MboI was measured by Southern blot analysis as described in Experimental Procedures.
[d] The specific β-galactosidase activities, measured using the method of Miller (1972), are shown.

of both PapI and Lrp results in a footprint around the GATC[1028] site (footprint III, lane 4). As we have shown previously (Nou et al., 1993), binding of Lrp and PapI near the 1028 site is methylation-sensitive, since we did not detect footprint III using fully methylated DNA (Figure 4A, lane 8).

Analysis of the 1028 site mutant DNA showed that Lrp addition resulted in a large footprint around the GATC[1130] site when this site was nonmethylated (Figure 4B, lanes 3 and 4). Notably, methylation of the 1130 site reduced the Lrp footprint around GATC[1130] between base pairs 1100 and 1150 (compare lanes 3 and 11, Figure 4A). Binding of Lrp/PapI to the GATC[1028] region was observed using DNAs containing either a methylated 1130 site (Figure 4A, lane 12) or a nonmethylated 1130 site (Figure 4B, lane 4). Thus, the 1028 site mutation does not inhibit binding of Lrp to the GATC[1130] region, nor does it inhibit Lrp/PapI binding to the GCTC[1028] region. These results indicate that the locked ON phenotype displayed by the 1028 site mutant is not due to aberrant protein–DNA interactions. This conclusion is supported by the finding that *pap* transcription in the 1028 site mutant is shut off in an *lrp*$^-$ background (Table 2). Thus, transcription in the locked ON mutant is still dependent on Lrp, suggesting that the normal regulatory mechanism is intact.

We argue above that the locked OFF phenotype of the 1130 site mutant is caused by a lack of methylation of the mutated GCTC[1130] site. However, a locked OFF phenotype would also result if the GCTC[1130] mutation inhibited binding of Lrp and PapI to the *pap* regulatory region. This latter possibility was tested and ruled out, since we found that Lrp binds around the GCTC[1130] region, and Lrp/PapI binds near the GATC[1028] region of the 1130 site mutant DNA (Figure 4A, lanes 15 and 16; Figure 4B, lanes 7 and 8). Binding of Lrp/PapI to the GATC[1028] region of the 1130 site mutant DNA was observed using nonmethylated DNA but was not detected using methylated DNA, similar to results obtained using wild-type *pap* DNAs (compare lane 8 of Figure 4B with lane 16 of Figure 4A). Analysis of the locked OFF GCTC double-site mutant also showed that binding of Lrp and PapI was identical to that observed using wild-type *pap* DNA (Figure 4B, lanes 11 and 12). Thus, the locked OFF phenotype of the 1130 site and double-site mutants is not caused by the inhibition of binding of Lrp and PapI to *pap* regulatory DNA.

Overproduction of PapI Partially Restores *pap* Transcription in the 1130 Site Mutant

Analysis of DNA from the GCTC[1130] locked OFF mutant cells indicates that most GATC[1028] sites are fully methylated in vivo (see Figure 3), suggesting that binding of Lrp/PapI near the GATC[1028] site rarely occurs in this mutant. However, DNA footprint analysis of the locked OFF 1130 site mutant indicates that Lrp and PapI can bind to the 1028 site region if GATC[1028] is nonmethylated, similar to results obtained using nonmutated *pap* DNA (Figure 4B, lane 8). What could account for this apparent discrepancy between the in vivo and in vitro data? One possibility is that the PapI level present in cells containing a single-copy *pap* operon is much lower than the level of PapI used for in vitro footprint analysis. If this is the case, then high intracellular levels of PapI within the locked OFF mutants should result in occupancy of the 1028 site region by Lrp/PapI and might also induce transition to the phase ON state. To test this hypothesis, we introduced plasmid pDAL262 (vector pTZ18R containing *papI* under *lac* promoter control) into the GCTC[1130] mutant strain. Plasmid pPY1025, expressing Lac repressor, was also introduced to allow regulation of *papI* transcription by addition of isopropyl β-D-thiogalactopyranoside (IPTG).

Analysis of the methylation state of the GATC[1028] site in the vector control strain DL1761 (the 1130 site mutant strain containing plasmid pTZ18R) showed that only 11% of the GATC[1028] sites were protected from methylation in the presence or absence of IPTG (Table 3). These results indicate that a small fraction of the 1130 site mutant DNAs contain Lrp–PapI bound to the GATC[1028] site, protecting this site from Dam methylation. In contrast, a much higher level of methylation protection of the GATC[1028] site was observed in strain DL1762, which contains the PapI expression plasmid pDAL262 (26% protection in the absence of IPTG and 68% protection with 100 μM IPTG). Thus, overproduction of PapI in the locked OFF 1130 site mutant results in significant methylation protection of the GATC[1028] site, similar to that observed in phase ON cells. These results support our hypothesis above that the reason for the discrepancy between in vitro footprint analysis and in vivo methylation state analysis is that there are limiting levels of PapI in the phase OFF mutant under normal physiological conditions.

Overproduction of PapI resulted in a high degree of

methylation protection of the GATC^{1028} site of the GCTC^{1130} locked OFF mutant, suggesting the possibility that the locked OFF phenotype could be reversed by high PapI levels. Measurement of β-galactosidase activity showed that there was a 29-fold increase in *pap* transcription at 100 μM IPTG compared with the vector control strain (Table 3). Even in the absence of IPTG, we observed a 12-fold increase in *pap* transcription, indicating that PapI expression was not fully repressed by plasmid pPY1025. Similar results were obtained using the 1028/1130 double-site mutant (data not shown). Thus, although methylation of the GATC^{1130} site is necessary for *pap* transcription under normal physiological conditions, this requirement can be partially overridden by a high intracellular level of PapI. Nevertheless, the level of β-galactosidase attained under maximal PapI induction in the 1130 site mutant was still significantly lower than that observed in wild-type phase ON cells (373 U versus 1630 U, respectively). Together, these results indicate that binding of Lrp/PapI to the GATC^{1028} site is necessary but not sufficient for optimal *pap* transcription and support our conclusion that methylation of the GATC^{1130} site is also necessary for the phase ON state.

Discussion

The data presented in this study show that the methylation states of both the *pap* GATC^{1028} and GATC^{1130} sites control Pap pili phase variation. Our previous work indicated that in Pap phase ON cells the *pap* GATC^{1028} site is nonmethylated and the GATC^{1130} site is methylated, whereas the converse methylation states are present in phase OFF cells (Blyn et al., 1990). Here we show that methylation of the *pap* GATC^{1028} site inhibits the phase ON state, whereas methylation of the GATC^{1130} site is required for the ON state (Table 2). These differential effects of Dam methylation on *pap* gene regulation were identified using E. coli isolates with mutations in the GATC^{1028} and GATC^{1130} sites that prevent their methylation. Notably, these GATC site mutations do not inhibit the binding of the Lrp and PapI regulatory proteins to *pap* DNA target sequences in vitro (Figure 4). Therefore, the effects of these mutations on *pap* gene regulation are due to changes in the methylation states of the *pap* GATC sites.

Methylation of the *pap* GATC^{1028} Site Inhibits the Phase ON State

Our results show that mutation of the 1028 site to GCTC results in a locked ON phenotype (Figure 2) and high expression of β-galactosidase (Table 2). Thus, methylation of the 1028 site appears to be required for transition to the phase OFF state. Previously, we found that Dam overproduction locks *pap* transcription in the OFF state (Blyn et al., 1990). However, it was not possible to determine whether this was due to methylation of the GATC^{1028} site, the GATC^{1130} site, or both sites. Here we show that methylation of the 1028 site, but not the 1130 site, prevents *pap* transcription, since the 1028 site mutant remains in the phase ON state under conditions in which Dam is overexpressed (Table 2).

How does methylation of the *pap* GATC^{1028} site inhibit *pap* transcription? It seems likely that this is due to inhibition of binding of Lrp and PapI to the GATC^{1028} region, since these proteins are required for *pap* transcription (Braaten et al., 1991). We recently showed that binding of Lrp/PapI to the GATC^{1028} region was detected using nonmethylated but not fully methylated *pap* DNAs (Nou et al., 1993), suggesting that methylation of one or perhaps both of the *pap* GATC sites blocks this binding. DNA footprint analysis of the 1028 site mutant DNA carried out here shows that even when the 1130 site is methylated, binding of Lrp/PapI to the 1028 region occurs (Figure 4A, lane 12). These results indicate that methylation of the 1028 site, but not the 1130 site, inhibits binding of Lrp/PapI to the 1028 site region.

Methylation of the *pap* GATC^{1130} Site Is Required for the Phase ON State

In contrast with the negative effect of methylation of the 1028 site on *pap* transcription, our results indicate that methylation of the 1130 site is necessary for transcription. First, cells containing the 1130 site GCTC mutation display a locked OFF phenotype (Figure 2C). This does not appear to be caused by inhibition of binding of Lrp or PapI, since binding of these proteins to the GCTC^{1130} mutant *pap* DNA was not inhibited (Figure 4). Second, the GCTC^{1028} locked ON mutant displayed a phase OFF phenotype in a Dam$^-$ background (Table 2), showing that Dam activity is essential for *pap* transcription. Third, the 1130 site mutation is epistatic to the 1028 site mutation, based on analysis of the double-site mutant (Table 2). Together, these results strongly indicate that methylation of the 1130 site is essential for *pap* transcription. This conclusion is supported by our finding that the 1028 site mutant showed a higher *pap* transcription level under conditions of Dam overexpression compared with the wild-type Dam level (30% increase, Table 2). We reason that this effect is due to a higher efficiency of methylation of GATC^{1130} sites in the 1028 site mutant DNA population.

In vitro DNAase I footprint analysis showed that addition of Lrp and PapI to DNA containing the 1130 site GCTC mutation resulted in a footprint around the 1028 site (Figure 4B, lane 8). However, in vivo, the methylation state analysis showed that almost all of the GATC^{1028} sites in cells containing the GCTC^{1130} site mutation are fully methylated, indicating that Lrp/PapI binding to this site occurs only rarely under physiological conditions (Figure 3). These results suggested the possibility that, in vivo, a much lower level of PapI is present in GCTC^{1130} mutant cells compared with the amount added for in vitro binding analysis. Our data support this hypothesis, since the in vivo overexpression of PapI in the GCTC^{1130} mutant resulted in an increased methylation protection of the GATC^{1028} site from 11% to 68% (Table 3). It seems likely that this increased methylation protection was due to enhanced binding of Lrp/PapI to the GATC^{1028} region, since we have shown previously that both PapI and Lrp are necessary for methylation protection of GATC^{1028} in vivo (Braaten et al., 1991), and the 1028 site DNA region is

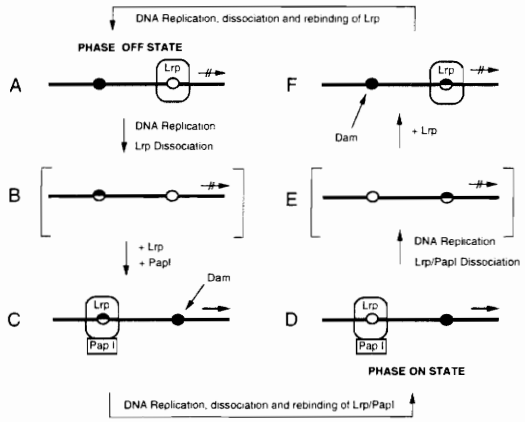


Figure 5. Pap Phase Variation Model

The *pap* DNA phase OFF state is shown in (A) with a methylated GATC^{1028} site on the left (black circle) and a nonmethylated GATC^{1130} site on the right (white circle). The *pap* DNA phase ON state shown in (D) has the converse methylation pattern, with a nonmethylated GATC^{1028} site on the left (white circle) and a methylated GATC^{1130} site on the right (black circle). Hemimethylated GATC sites generated as a result of DNA replication are indicated as half-filled circles. Transcription from the *papBAp* promoter is shown by arrows: a blocked arrow indicates that transcription does not occur, whereas an open arrow indicates transcription occurs. The brackets indicate transient intermediate states.

footprinted only in the presence of both of these proteins (Nou et al., 1993).

Overproduction of PapI partially reversed the locked OFF phenotype of the 1130 site mutant (Table 3). However, β-galactosidase expression did not correlate with the fraction of GATC^{1028} sites that were protected from methylation. For example, in a wild-type phase ON population containing 17% ON cells and expressing about 1600 U of β-galactosidase (Table 2), we found that only about 12% of the GATC^{1028} sites were nonmethylated (data not shown). In contrast, under conditions of PapI overproduction, the 1130 site mutant expressed only 372 U of β-galactosidase, but 68% of the *pap* GATC^{1028} sites were nonmethylated (Table 3). Thus, binding of Lrp/PapI to the GATC^{1028} region is necessary but not sufficient for maximal *pap* transcription. High levels of *pap* transcription also require methylation of the GATC^{1130} site.

Why is methylation of the 1130 site required for the phase ON state? A comparison of the binding of Lrp to nonmethylated *pap* DNA (Figure 4A, lane 3) and the 1028 site mutant DNA containing a methylated 1130 site (Figure 4A, lane 11) indicates that methylation of the GATC^{1130} site weakens the Lrp footprint and reduces DNAase I hypersensitivity between base pairs 1100 and 1150. Since the Lrp and RNA polymerase binding sites overlap at base pairs 1140–1165 (Figure 1), it seems likely that binding of Lrp near the 1130 site inhibits *pap* transcription. Therefore, methylation of the 1130 site might function to prevent repression of *pap* transcription caused by binding of Lrp near the *papBA* promoter.

A Model for Pap Phase Variation

A central question in *pap* gene regulation is how the transition between the phase OFF and ON states occurs. Based on methylation analysis carried out previously (Blyn et al., 1990) and the data shown here, in phase ON cells Lrp and PapI are bound near the 1028 site, protecting this site from methylation. In contrast, the 1130 site is methylated and unoccupied (Figure 5). In phase OFF cells the converse methylation pattern is found, indicating that the 1028 site is methylated and unoccupied, whereas Lrp is bound to the 1130 site, protecting this site from methylation. Our data, discussed above, indicate that the methylation patterns of the phase OFF and ON cells prevent transition to the alternate state by regulating the binding of the Lrp and PapI proteins. How then can cells switch between phase states? We hypothesize that phase switching requires DNA replication based on the model outlined in Figure 5. Following passage of the replication fork, the methylated GATC^{1028} site of phase OFF DNA is converted to a hemimethylated state, and Lrp is displaced from the 1130 site by the replication machinery (Figures 5A and 5B). Since Lrp/PapI binds to DNA containing a hemimethylated GATC^{1028} site with low affinity (Nou et al., 1993), this provides an opportunity for Lrp, facilitated by PapI, to bind near the 1028 site. In addition, the unoccupied 1130 site is now free to be methylated by Dam (Figure 5C). After an additional round of DNA replication, one DNA duplex will contain a nonmethylated GATC^{1028} site that if bound by Lrp/PapI will be in the Phase ON state (Figure 5D). Transition from the phase ON state to the phase OFF state can occur after DNA replication if the Lrp/PapI complex does not reassociate with the 1028 site region after displacement by the replication fork (Figure 5E). In this case, the 1028 site becomes unprotected and can be methylated by Dam, preventing further association of Lrp/PapI with this site (Figure 5F). In the absence of bound Lrp/PapI at the GATC^{1028} site, Lrp binds to the hemimethylated GATC^{1130} site, protecting this site from further methylation by Dam. After an additional round of DNA replication, *pap* DNA in the phase OFF state is generated in which the 1028 site is methylated and the 1130 site is nonmethylated and occupied by Lrp (Figure 5A).

It is likely that the binding transition between phase OFF and ON is limited mainly by the PapI level, since Lrp is relatively abundant (about 3000 Lrp dimers per cell) (Willins et al., 1991). Previous results have shown that *papI* transcription is positively regulated by cAMP (Görannson et al., 1989). Thus, under conditions that increase cAMP levels, the PapI level will also increase, which should lead to a higher OFF to ON switch rate according to our model. This appears to occur, since the OFF to ON switch rate of cells growing in minimal medium with glycerol as sole carbon source (high intracellular cAMP) is 35-fold higher than cells using glucose as carbon source (low intracellular cAMP) (Blyn et al., 1989).

How does PapI facilitate binding of Lrp to the GATC^{1028} region? Our recent data indicate that PapI forms a complex with Lrp, which could alter the binding specificity of Lrp (unpublished data). PapI might lower the affinity of Lrp binding at the 1130 site, increase the affinity of Lrp for the

1028 site, or both. In any case, binding of Lrp and PapI at the GATC^{1028} site would inhibit methylation of this site and must facilitate methylation of the 1130 site, since this site is methylated in phase ON DNA (Figure 5C). This could occur if binding of Lrp/PapI to the 1028 site reduced the affinity of Lrp for the 1130 site or increased the accessibility of the 1130 site to Dam. In addition, binding of Lrp/PapI at GATC^{1028} activates *pap* transcription, generating PapB regulatory protein from transcripts initiated at the *papBA* promoter (Figure 1). Since PapB is a positive regulator of *papI* transcription (Baga et al., 1985), this should raise the level of PapI and increase the probability that PapI/Lrp will bind back to the 1028 site after a second round of DNA replication, leading to the phase ON state (Figure 5D).

How does binding of Lrp and PapI to the 1028 site region activate *pap* transcription? One possibility is that transcription is activated indirectly as a result of methylation of the 1130 site and inhibition of binding of Lrp near the *papBA* promoter. This does not appear to be the case, however, since *pap* transcription is shut off in the absence of Lrp (Table 2). Another possibility is that Lrp bends the DNA, allowing Lrp, PapI, or both to contact RNA polymerase. In support of this hypothesis, Lrp has recently been shown to bend *ilvIH* DNA (Wang and Calvo, 1993a) and may bend *pap* DNA, based on footprint analysis showing periodic DNAase I hypersensitive sites between the 1028 and 1130 sites (Figure 4). Alternatively, it has been reported that *pap* transcription is activated in cells deficient for the histonelike protein H-NS (Göransson et al., 1990). Therefore, it is possible that binding of Lrp and PapI near GATC^{1028} interferes with the binding of H-NS, alleviating H-NS-mediated repression of *pap* gene expression by competitive inhibition.

The model presented in Figure 5 is only a basic outline of the events that might occur during Pap phase variation. We have depicted a single Lrp molecule binding to either the 1028 site or the 1130 site. However, our recent unpublished data indicate that Lrp binds to multiple *pap* DNA sites in a cooperative manner, resulting in the large DNAase I footprints seen in Figure 4. Such cooperativity was recently reported by Wang and Calvo (1993b) for binding of Lrp to the *ilvIH* operon. A comparative study of four fimbrial operons sharing DNA sequence similarity with *pap* shows that the spacing (102–103 bp) between the GATC sites is conserved, although the inter-GATC DNA sequence is not (van der Woude et al., 1992). Therefore, it is possible that spacing between Lrp binding sites is critical for proper orientation of DNA-bound Lrp. We are now testing this hypothesis by determining the effects of alterations in the spacing between the two *pap* GATC sites.

Similarities to Eukaryotic Methylation Patterns

Some eukaryotic cells, including those from mammals and plants, express cytosine methyltransferase (MTase), which methylates CpG doublets. Certain regions of the genome, known as CpG islands, are totally nonmethylated and appear to be closely associated with genes that are actively transcribed (Antequera and Bird, 1993). Evidence indicates that CpG methylation patterns might play diverse roles in tissue-specific gene expression (Cedar, 1988), X chromosome inactivation (Riggs and Pfeifer, 1992; Wolf et al., 1984), transposition of the *Spm* element in maize (Federoff et al., 1989), genomic imprinting (Swain et al., 1987), and epigenetic inheritance (Holliday, 1993). Although MTase appears to be essential for embryonic development (Li et al., 1992), the mechanisms by which DNA methylation patterns are formed and control cellular processes are not known.

It was previously thought that prokaryotes lacked methylation patterns and therefore could not provide a useful paradigm for eukaryotic systems. However, it is now clear that there are a number of nonmethylated GATC sites in the E. coli chromosome within *pap* and other fimbrial operons (van der Woude et al., 1992) as well as genes such as *mtl*, *gut*, and *fep* (Wang and Church, 1992). Our work here indicates that *pap* DNA methylation patterns control gene expression by locking the protein binding configurations of cells in the phase ON or phase OFF states, while protein binding, in turn, controls the methylation patterns. These regulatory interactions constitute an epigenetic switch. We speculate that similar mechanisms may exist in eukaryotes and that the study of *pap* gene regulation as well as other prokaryotic systems may prove useful in the study of the mechanisms by which DNA methylation patterns control gene regulation in eukaryotic cells.

Experimental Procedures

Bacterial Strains, Plasmids, and Bacteriophage

The E. coli strains, plasmids, and bacteriophage used in this study are shown in Table 1. E. coli λ lysogens, containing either a wild-type or mutant 1.76 kb *papIB* regulatory DNA sequence in single copy, were made by using pRS550 and λRS45, following the protocol of Simons et al. (1987). Using the P1 transduction method described by Miller (1972), congenic *dam*$^-$ and congenic *lrp*$^-$ E.coli strains were constructed from phage P1 lysates of GM2929 (dam$^-$) and DL842 (*mbf*-2). Overproducing *dam*$^+$ E. coli strains were made via the transformation procedure of Hanahan (1983), using plasmid pPY1025 (a gift of P. Youderian) and plasmid pTP166 (a gift of M. G. Marinus).

Media, Antibiotics, Restriction Endonucleases, and Radioactive Nucleotides

Luria-Bertani (LB) broth, LB agar, M9 minimal broth (M9), and M9 agar were prepared according to Miller (1972). The carbon source of M9 media was glycerol (Sigma) at a final concentration of 0.2% v/v. If used, the chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) (Gold Biotechnology) was at 40 μg/ml of medium. Antibiotics, when used, were at the following final concentrations: ampicillin (amp), 100 μg/ml; chloramphenicol (cam), 24 μg/ml; kanamycin (kan), 25 μg/ml; spectinomycin (spc), 100 μg/ml; and tetracycline (tet), 15 μg/ml. Amp, kan, and tet were obtained from Boehringer Mannheim Corporation, while cam and spc were from Sigma Chemical Company. The following restriction endonucleases used in this study were purchased from New England Biolabs, Incorporated: ApaI, BspHI, DdeI, DpnI, EcoRI, HhaI, MboI, and PvuII. NspV was obtained from GIBCO/BRL. All radioactive nucleotides were obtained from Amersham Corporation.

Site-Specific Mutagenesis and Construction of the Double-Site Mutant

The site-directed mutagenesis protocol of Kunkel (1985) was employed to mutate specific *pap* regulatory DNA GATC sites to GCTC sequences. Mutagenic primers were made using an Applied Biosystems 380 B DNA Synthesizer. The synthetic oligonucleotide 5′-TCATTTAGACGCTCTTTTATGC-3′ was used to mutate the GATC^{1028} site, whereas the mutagenic primer 5′-TGGGTTAAAAGCTCGTTTAAAT-3′ was used

for the GATC^{1130} site. To facilitate the mutagenesis procedure, 2 μg of T4 Gene 32 Protein (US Biochemical Corporation) was added to the post-annealed template/primer reaction mixture and incubated for 5 min at room temperature, followed by a 5 min incubation on ice before proceeding with the extension reaction.

The *pap* regulatory DNA double GCTC $^{1028/1130}$ mutant was constructed from the two types of single GATC to GCTC mutants. First, a 360 bp *pap* DNA regulatory fragment that contained both GATC sites (one site mutated to GCTC, the other not) was isolated from the DNA of each type of single site mutant by cutting their respective DNAs with ApaI and NspV. Next, each of the isolated and purified 360 bp DNAs was cut with DdeI, which resulted in the mutated 1028 site being located on a 215 bp DNA fragment, while the mutated 1130 site was on a 145 bp fragment. Finally, the DNA fragments containing the mutant GCTC sites were separated on a 1.2% agarose gel and purified using Prep-A-Gene (Bio-Rad Laboratories), and the two mutant DNAs were placed back into the *pap* regulatory region using established recombinant DNA technology (Sambrook et al., 1989).

All site-specific mutations were confirmed by sequencing the *pap* regulatory DNA from each type of mutant with the dideoxy chain termination method of Sanger et al. (1977), using the Sequenase Version 2.0 DNA sequencing kit (US Biochemical Corporation) and [α-^{35}S]dATP.

Southern Blot Analysis for the Methylation Pattern of Wild-Type and Mutant *papIB* DNAs

The large-scale CsCl preparation of bacterial genomic DNA described by Ausubel et al. (1989) was used to isolate chromosomal DNAs from the pertinent E. coli strains. Three samples from each chromosome preparation were doubly digested with EcoRI and one other of the following restriction endonucleases: Sau3AI, DpnI, or MboI. Chromosomal single digests were also made using each of the above endonucleases. The DNA fragments were separated by electrophoresis on 1.3% agarose gels and transferred to Hybond-N membranes (Amersham Corporation), using a PosiBlot pressure blotter (Stratagene). DNA hybridization was performed in 50% formamide at 42°C for 20 hr with a 489 bp HhaI *pap* DNA probe that had been radiolabeled with [α-^{32}P]dCTP using a random primed DNA labeling kit (Boehringer Mannheim). Quantitation of radioactivity was carried out using a Bio-Rad model GS-250 imager system.

DNAase I Footprint Analyses of Wild-Type and Mutant *papIB* DNAs in Various GATC Methylation States

Wild-type or mutant *papIB* regulatory DNAs were cloned into plasmid pTZ19U to construct the pDAL337 series of plasmids (Table 1). Nonmethylated GATC site plasmid DNAs of the pDAL337 series were obtained by transforming the *dam*$^-$ E. coli strain DL738 with each type of plasmid, and, after a period of growth, the plasmids were reisolated from their respective DL738 hosts (Table 1). Fully methylated GATC site DNAs from each of the pDAL337 series of plasmids (with the obvious exception of the double GATC to GCTC mutant) were obtained after their transformation into and reisolation from the *dam*$^+$ but *mbf*$^-$ E. coli strain DL844 (Table 1). Before footprinting, an aliquot of each of the above plasmid DNAs was cut with EcoRI, and the 1.76 kb *papIB* DNA fragment from each was isolated after electrophoresis from an agarose gel using Prep-A-Gene (Bio-Rad Laboratories). The GATC methylation state pattern of each of the *papIB* DNAs was then verified by agarose gel electrophoresis after incubating each DNA isolate with DpnI or MboI, which cut methylated or nonmethylated GATC DNA sites, respectively.

Wild-type and mutant DNA probes containing the pertinent *pap* GATC sites in a given methylation state were radiolabeled on either the upper or lower DNA strand as follows. The top DNA strand was labeled after cutting each plasmid with NspV and BspHI and purifying the resulting 392 bp *pap* fragment, which was then tagged with [α-^{32}P]dATP by filling the ends with Klenow enzyme (Boehringer Mannheim Corporation). The bottom strand was labeled after cutting each plasmid with NspV and PvuII (which cuts the pTZ19U vector DNA) and isolating the 1171 bp-sized fragment, which was radiolabeled with [α-^{32}P]dCTP by filling the ends with Klenow enzyme. The 1171 bp fragment was then cut with BspHI, yielding a 392 bp DNA bottom strand–labeled *pap* probe that was purified for use. DNAase I footprinting analyses (Galas and Schmitz, 1978) of the probes were carried out as previously described by Nou et al. (1993).

Assay for β-Galactosidase Activity

E. coli strains were grown at 37°C to log phase (OD_{600} between 0.3 and 0.9) in 5 ml tube cultures of M9-glycerol broth containing the appropriate antibiotics. Each tube culture was assayed in triplicate for β-galactosidase activity as described by Miller (1972), and a minimum of two colonies from each E. coli strain were tested. In the case of the Dam overproducer E. coli strains, two different sets of tube cultures per bacterial colony picked were assayed for β-galactosidase. The first set of culture tubes contained M9-glycerol broth with amp, kan, spc, and 1 mM IPTG (Research Organics Incorporated); the second set of culture tubes contained the same medium but no IPTG.

Acknowledgments

We wish to thank Marjan van der Woude and John Roth for helpful discussions, and the Protein–DNA Core Facility, Cancer Center, University of Utah for production of the oligonucleotides used in this study. The facility is supported by National Cancer Institute grant 5P30 CA42014–07. This work was supported by National Institutes of Health grants 2R01 AI23348 and 5K04 AI881 to D. A. L.

Received July 26, 1993; revised November 2, 1993.

References

Antequera, F., and Bird, A. (1993). CpG islands. In DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz, eds. (Basel: Birkhäuser Verlag), pp. 169–185.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1989). Current Protocols in Molecular Biology (New York: Wiley Interscience).

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

Bakker, A., and Smith, D. W. (1989). Methylation of GATC sites is required for precise timing between rounds of DNA replication in *E. coli*. J. Bacteriol. *171*, 5738–5742.

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

Blyn, L. B., Braaten, B. A., White-Ziegler, C. A., Rolfson, D. H., and Low, D. A. (1989). Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. EMBO J. *8*, 613–620.

Boye, E., and Lobner-Olesen, A. (1990). The role of Dam methyltransferase in the control of DNA replication in E. coli. Cell *62*, 981–989.

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

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

Casadaban, M. (1976). Transposition and fusion of the *lac* genes to selected promoters in E. coli using bacteriophage lambda and mu. J. Mol. Biol. *104*, 541–555.

Cedar, H. (1988). DNA methylation and gene activity. Cell *53*, 3–4.

Federoff, N., Masson, P., and Banks, J. A. (1989). Mutations, epimutations, and the developmental programming of the maize suppressor-mutator transposable element. BioEssays *10*, 139–144.

Galas, D., and Schmitz, A. (1978). DNase footprinting: a simple method for the detection of protein–DNA binding specificity. Nucl. Acids Res. *5*, 3157–3170.

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

Göransson, M., Sonden, B., Nilsson, P., Dagberg, B., Forsman, K., Emanuelsson, K., and Uhlin, B. E. (1990). Transcriptional silencing

and thermoregulation of gene expression in *Escherichia coli*. Nature *344*, 682–685.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. *166*, 557–580.

Holliday, R. (1993). Epigenetic inheritance based on DNA methylation. In DNA Methylation: Molecular Biology and Biological Significance. J. P. Jost and H. P. Saluz, eds. (Basel: Birkhäuser Verlag), pp. 452–468.

Hultgren, S. J., and Normark, S. (1991). Biogenesis of the bacterial pilus. Curr. Opin. Genet. Dev. *1*, 313–318.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA *82*, 488–492.

Li, E., Bestor, T. H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

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

Marinus, M. G., Poteete, A., and Arraj, J. A. (1984). Correlation of DNA adenine methlyase activity with spontaneous mutability in *Escherichia coli* K-12. Gene *28*, 123–125.

Miller, J. H. (1972). Experiments in Molecular Genetics (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Modrich, P. (1989). Methyl-directed DNA mismatch repair. J. Biol. Chem. *264*, 6597–6600.

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

O'Hanley, P., Lark, D., Falkow, S., and Schoolnik, G. (1985a). Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice. J. Clin. Invest. *75*, 347–360.

O'Hanley, P., Low, D., Romero, I., Lark, D., Vosti, K., Falkow, S., and Schoolnik, G. (1985b). Gal–Gal binding and hemolysin phenotypes and genotypes associated with uropathogenic *Escherichia coli*. N. Engl. J. Med. *313*, 414–447.

Peterson, K. R., Wertman, K. F., Mount, D. W., and Marinus, M. G. (1985). Viability of *Escherichia coli* K-12 DNA adenine methylase (*dam*) mutants requires increased expression of specific genes in the SOS regulon. Mol. Gen. Genet. *201*, 14–19.

Riggs, A. D., and Pfeifer, G. P. (1992). X-chromosome inactivation and cell memory. Trends Genet. *8*, 169–174.

Roberts, D., Hoopes, B. C., McClure, W. R., and Kleckner, N. (1985). IS*10* transposition is regulated by DNA adenine methylation. Cell *43*, 117–130.

Roberts, J. A., Kaack, M. B., and Fussell, E. N. (1989). Bacterial adherence in urinary tract infections: preliminary studies in a primate model. Infection *17*, 401–404.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, Second Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Sanger, F. S., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Simons, R. W., Houman, F., and Kleckner, N. (1987). Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene *53*, 85–96.

Swain, J. L., Stewart, T. A., and Leder, P. (1987). Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. Cell *50*, 719–727.

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

Wang, M. X., and Church, G. M. (1992). A whole genome approach to in vivo DNA–protein interactions in *E. coli*. Nature *360*, 606–610.

Wang, Q., and Calvo, J. M. (1993a). Lrp, a major regulatory protein in *Escherichia coli*, bends DNA and can organize the assembly of a higher order nucleoprotein structure. EMBO J. *6*, 2495–2501.

Wang, Q., and Calvo, J. M. (1993b). Lrp, a global regulatory protein of *Escherichia coli*, binds cooperatively to multiple sites and activates transcription of *ilvIH*. J. Mol. Biol. *229*, 306–318.

Willins, D. A., Ryan, C. W., Platko, J. V., and Calvo, J. M. (1991). Characterization of Lrp, an *Escherichia coli* regulatory protein that mediates a global response to leucine. J. Biol. Chem. *266*, 10768–10774.

Wolf, S. F., Jolly, D. J., Lunnen, K. D., Friedmann, T., and Migeon, B. R. (1984). Methylation of the hypoxanthine phosphoribosyl transferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc. Natl. Acad. Sci. USA *81*, 2806–2810.

CHAPTER IV

DIFFERENTIAL BINDING OF LRP TO TWO SETS OF *PAP* DNA BINDING SITES MEDIATED BY PAPI REGULATES PAP PHASE VARIATION IN *ESCHERICHIA COLI*

This chapter is reprinted from *The EMBO Journal* (1995) 14:5785:5797 with permission from coauthers and Oxford University Press

Differential binding of Lrp to two sets of *pap* DNA binding sites mediated by PapI regulates Pap phase variation in *Escherichia coli*

Xiangwu Nou, Bruce Braaten, Linda Kaltenbach and David A.Low[1]

Division of Cell Biology and Immunology, Department of Pathology, University of Utah, Health Sciences Center, Salt Lake City, UT 84132, USA

[1]Corresponding author

Pyelonephritis-associated pili (Pap) expression in *Escherichia coli* is subject to a phase variation control mechanism that is regulated by the leucine-responsive regulatory protein (Lrp), PapI, and deoxyadenosine methylase (Dam). In previous work, we found that the differential Dam methylation of two target sites in *pap* regulatory DNA, GATC-I and GATC-II, is essential for the transition between active and inactive *pap* transcriptional states. Here, we identify six Lrp binding sites within the *pap* regulatory DNA, each separated by about three helical turns. Lrp binds with highest affinity to three sites (1, 2 and 3) proximal to the *papBAp* promoter. A mutational analysis indicates that the binding of Lrp to sites 2 and 3 inhibits *pap* transcription, which is consistent with the fact that Lrp binding site 3 is located between the –35 and –10 RNA polymerase binding region of *papBAp*. The addition of PapI decreases the affinity of Lrp for sites 1, 2 and 3 and increases its affinity for the distal Lrp binding sites 4 and 5. Mutations within Lrp binding sites 4 and 5 shut off *pap* transcription, indicating that the binding of Lrp to this *pap* region activates *pap* transcription. The *pap* GATC-I and GATC-II methylation sites are located within Lrp binding sites 5 and 2, respectively, providing a mechanism by which Dam controls Lrp binding and Pap phase variation.

Keywords: DNA–protein interaction/Lrp/*pap* operon/phase variation

Introduction

The regulation of pyelonephritis-associated pili (pap) expression in *Escherichia coli* is under a complex phase variation control mechanism which involves leucine-responsive regulatory protein (Lrp), PapI, and deoxyadenosine methylase (Dam; Blyn *et al.*, 1990; Braaten *et al.*, 1992, 1994; van der Woude *et al.*, 1992). Lrp is a 19 kDa global regulator which binds to DNA as a dimer (Platko and Calvo, 1993; Wang and Calvo, 1993). Lrp activates some genes and represses others and, in many cases, leucine modulates Lrp activity. However, the *pap* operon is not responsive to leucine. Moreover, Lrp requires PapI as a cofactor to activate *pap* transcription, whereas Lrp appears to act by itself to regulate other members of the Lrp regulon such as *ilvIH* (Nou *et al.*, 1993; Calvo and Matthews, 1994).

Two DNA GATC sites which are targets for Dam methylase are located within the *papI–B* regulatory region. The GATC-II site is 50 bp upstream of the *pBAp* transcription start site, whereas GATC-I is 102 bp upstream of GATC-II (Figure 2A). We have found previously that the GATC-II site of phase OFF cells is protected from methylation, whereas GATC-I is fully methylated (Blyn *et al.*, 1990). Methylation protection of GATC-II requires Lrp but not PapI (Braaten *et al.*, 1991). *In vitro* footprint analyses have shown that in the absence of PapI, Lrp bound close to the GATC-II site (Nou *et al.*, 1993). Together, these data suggest that in phase OFF cells, Lrp binds near GATC-II, thus sterically blocking its methylation. Our recent results indicate that the binding of Lrp near GATC-II blocks basal transcription from the *papBAp* promoter (van der Woude *et al.*, 1995).

Phase ON cells have a methylation pattern that is the converse of phase OFF cells: GATC-I is protected from methylation whereas GATC-II is fully methylated. An *in vitro* footprint analysis showed that the addition of Lrp and PapI to non-methylated *pap* DNA resulted in footprinting of the GATC-I region as well as the GATC-II region (Nou *et al.*, 1993). These results are consistent with the finding that both Lrp and PapI are required for the protection of GATC-I from methylation *in vivo* (Braaten *et al.*, 1991). Recent data suggest that the binding of Lrp, in the presence of PapI, to the GATC-I region activates transcription ~8-fold over the basal transcription level. Lrp binding to the GATC-I region is not sufficient for transcription activation because Lrp mutants which bind to *pap* with the same affinity as wild-type Lrp fail to induce transcription (Platko and Calvo, 1993; van der Woude *et al.*, 1995). These results indicate that Lrp interacts with other parts of the transcriptional apparatus, such as RNA polymerase, when it is bound at the GATC-I region.

Methylation appears to play a direct role in controlling *pap* transcription. First, transcription is shut off in the absence of Dam or in the presence of excess Dam (4-fold above wild-type levels; Blyn *et al.*, 1990). Second, the mutation of the *pap* GATC-I site to the GCTC sequence results in a locked ON phenotype, even in the presence of high Dam levels (Braaten *et al.*, 1994). These results indicate that the methylation of GATC-I shuts off *pap* transcription. A DNA footprint analysis showed that the binding of Lrp to the GATC-I region, facilitated by PapI, was blocked by DNA methylation (Nou *et al.*, 1993). Based on these data, we hypothesized that Lrp and Dam methylase compete for binding to the GATC-I region, thereby controlling the switch between transcriptionally active and inactive states (van der Woude *et al.*, 1992).

Although the methylation of GATC-I inhibits *pap* transcription, the methylation of GATC-II appears to be required for *pap* transcription. Both wild-type and phase

© Oxford University Press

locked ON GCTC-I mutant *E.coli* strains are transcriptionally inactive in a *dam*$^-$ background (Braaten *et al.*, 1994). Moreover, the mutation of GATC-II to the GCTC sequence results in a locked OFF phenotype. This phenotype was not caused by the inhibition of binding of Lrp to the GCTC-II region, but appeared to be a direct result of the inability of this site to be methylated by Dam. Preliminary data suggested that the methylation of GATC-II might reduce the affinity of Lrp for this site, aiding in the binding transition of Lrp from GATC-II to GATC-I and the activation of *pap* transcription (Braaten *et al.*, 1994).

Previous analyses using DNase I footprinting indicated that Lrp bound to a 115 bp region around GATC-II in the absence of PapI, and footprinted an additional 60 bp region around GATC-I after PapI addition (Nou *et al.*, 1993). These footprints showed periodic regions of DNase I protection and hypersensitivity, characteristic of bent DNA. Because the cleavage of DNA by DNase I is affected not only by protein binding but also by the DNA structure, it was not possible to determine the regions of *pap* DNA contacted by Lrp. Further, the roles of Lrp binding sites in regulating *pap* phase variation were not explored. Here, we have used a methylation protection analysis to identify six *pap* Lrp binding sites within the *pap* regulatory region, each spaced about three helical turns apart. The GATC-I and GATC-II sites are each located within an Lrp binding site. A mutational analysis indicated that the binding of Lrp to sites 2 and 3 proximal to the *papBAp* promoter inhibited transcription, whereas the binding of Lrp to the promoter distal sites 4 and 5 activated transcription. Our results show that PapI facilitates the transition from phase OFF to ON by reducing the affinity of Lrp for the *papBAp* promoter proximal binding sites and increasing Lrp affinity for the distal *pap* binding sites.

Results

Identification of Lrp binding sites in the pap regulatory region.

We showed previously that Lrp protected a 115 bp region around the GATC-II site of *pap* DNA from DNase I cleavage. PapI and Lrp together footprinted an additional 60 bp around the GATC-I site (Nou *et al.*, 1993). These large footprints consisted of periodical protected and hypersensitive DNA regions, which may have resulted from DNA bending as well as the specific binding of Lrp (Hochschild, 1991). To localize further each of the *pap* Lrp binding sites, we used a methylation protection analysis with dimethyl sulfate (DMS) to identify the guanosine and adenosine residues in close contact with Lrp (Wissman and Hillen, 1991). As shown in Figure 1, guanosines in both the top strand (designated G_n) and bottom strand (designated C_n for the complementary guanosine) of the *pap* regulatory region were protected from DMS methylation by Lrp addition. These and other methylation protection data (results not shown), summarized in Figure 2A and B, indicate that the *pap* regulatory region contains six Lrp binding sites. Each of these sites is spaced 30–33 bp apart, center to center, or about three turns of a DNA helix. At lower Lrp levels, the binding of Lrp to sites 1, 2 and 3 occurred (designated here by [1,2,3]). At higher Lrp levels, sites 4, 5 and 6, designated [4,5,6], were also occupied (Figure 1). The addition of PapI did not result in any new protected bases, although PapI did alter the affinity of Lrp for the six sites identified (see below; Figures 3 and 4). All six Lrp binding sites contain the sequence 'Gnn(n)TTTt' (where 'n' is any nucleotide and 't' indicates that a fourth thymidine was present in all sites except site 4). The 'G' of the consensus binding site was protected from methylation in all six binding sites. Additional guanosines were protected in all sites except site 6. Four adenosines complementary to the consensus thymidines were also protected from methylation in sites [1,2,3]. We were unable to determine if the adenosines in sites [4,5,6] were protected because piperidine cleavage does not occur as efficiently at methylated adenosines compared with methylated guanosines (Wissman and Hillen, 1991). As shown in Figure 2C, the 'G' of the consensus binding site and the 'A' tract are positioned opposite one another across the DNA major groove.

Mutational analysis of the roles of Lrp binding sites in pap transcription.

To determine the roles of the Lrp binding sites identified above in controlling Pap phase variation, we introduced DNA base-pair replacements containing *Nhe*I restriction sites throughout the *pap* regulatory region (Figure 2B). Our data, summarized in Figure 2A, show that mutations within Lrp binding site 4 or 5 result in locked OFF *pap* transcription phenotypes, whereas mutations within site 2 or 3 lock cells in the phase ON state. All mutations in or close to Lrp binding site 2 or 3 resulted in an increased OFF to ON switch frequency or a locked ON phenotype based on an analysis of strains containing a single-copy *papBAp–lac* fusion (Table I). *E.coli* containing mutation #13 (site 3) displayed a locked ON transcription phenotype, as did *E.coli* containing mutation #24 (adjacent to site 2). Both of these mutant strains expressed high levels of β-galactosidase (~5800 Miller units). This level of transcription is ~3-fold higher than that observed for a wild-type population containing 30% phase ON cells (Table I), which is consistent with the locked ON phenotypes of these mutants. Mutants #16 and #17, which contain single base-pair substitutions of G1163 within site 3, also expressed high levels of β-galactosidase and a locked ON phenotype (Figure 2B and Table I). These data suggest that the 'G' in the consensus sequence may be important for Lrp binding (confirmed below). *E.coli* containing mutation #12 (site 2) or #11 (adjacent to site 1) displayed altered switching phenotypes with an ~8-fold higher OFF to ON switch rate compared with wild type. Mutant #11 also had a 2-fold lower ON to OFF switch rate, which resulted in a higher *papBA* transcript level as well as a higher fraction of cells in the phase ON state (Table I). Together, these results show that mutations within or near Lrp binding sites [1,2,3] resulted in either locked ON phenotypes or increased rates of switching from OFF to ON.

In contrast to mutations near or within binding sites [1,2,3], mutations #56 (site 4), #14 (site 5) and #10 (between sites 6 and 1) resulted in locked OFF phenotypes and very low transcript levels (Figure 2A and Table I). Mutation #7, a G1028 to T1028 transversion within the

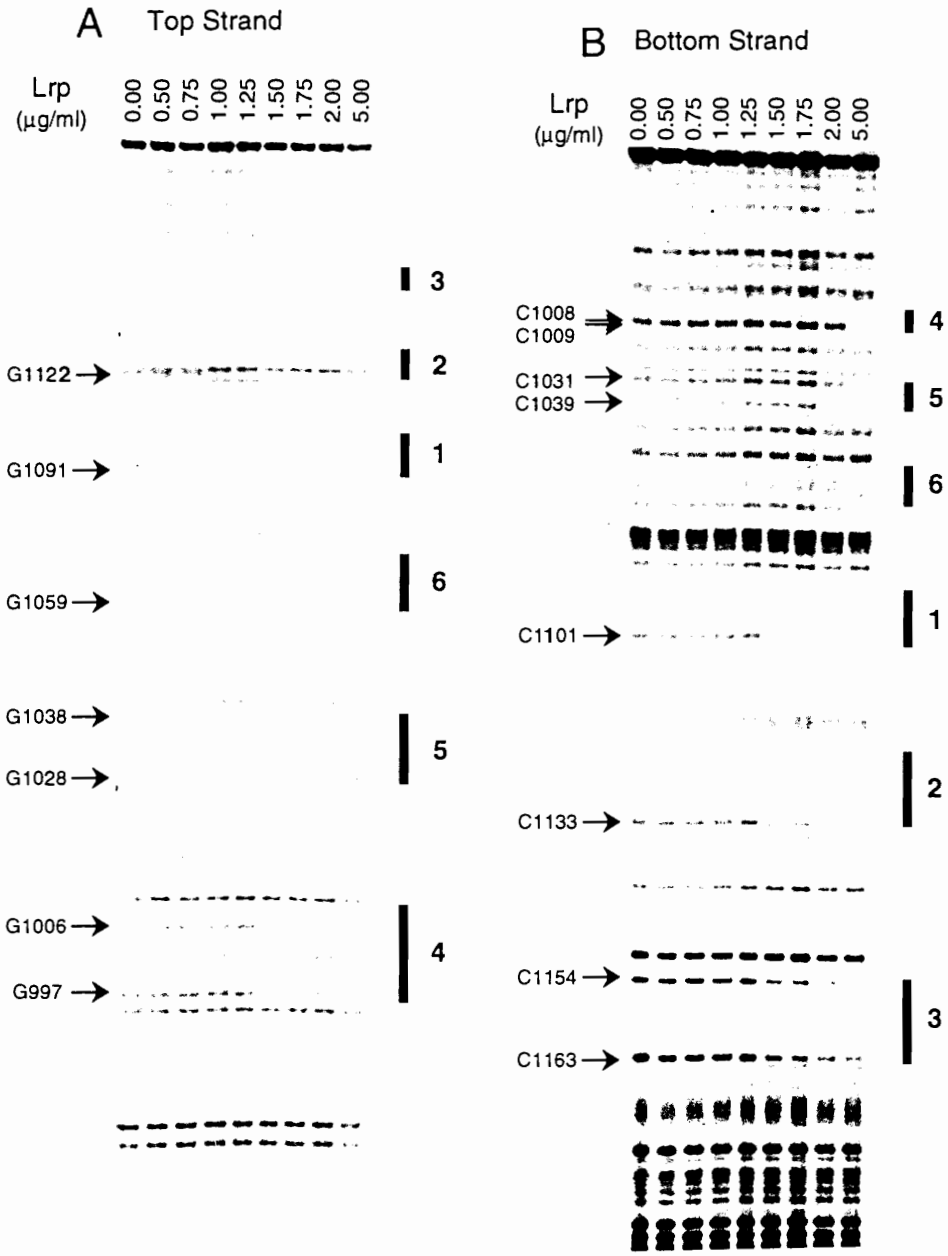


Fig. 1. The localization of Lrp binding sites in the *pap* regulatory region by methylation protection analysis. The amounts of Lrp shown were added to *pap* regulatory DNA, end-labeled with ^{32}P on the top (A) or bottom (B) DNA strand. A methylation protection analysis using DMS was carried out as described in Materials and methods, and samples were analyzed on polyacrylamide gels. The locations of *pap* top strand guanosines (see *pap* DNA orientation in Figure 2B) protected by Lrp addition are shown in (A) and the positions of *pap* bottom strand guanosines protected by Lrp addition are shown in (B). Bottom strand guanosines are designated as C_n based on the top strand position of the complementary cytosine. The locations of the six Lrp binding sites in the *pap* regulatory region, based on these and additional methylation protection data (results not shown), are indicated to the right.

GATC-I sequence of site 5, also gave a locked OFF phenotype (Figure 2A and B). Mutation #9, located adjacent to site 6, did not significantly affect *pap* switch rates or transcript levels, whereas the complete substitution of site 6 (mutant #57) resulted in only a 2-fold reduction in the phase OFF to ON rate (Table I). Thus, site 6 does not appear to play a significant role in the regulation of *pap* transcription. Taken together, these results indicate that the binding of Lrp to sites [4,5] has a positive regulatory function. In addition, the *pap* DNA sequence between sites 1 and 6 (defined by mutation #10) also plays a positive role (see Discussion).

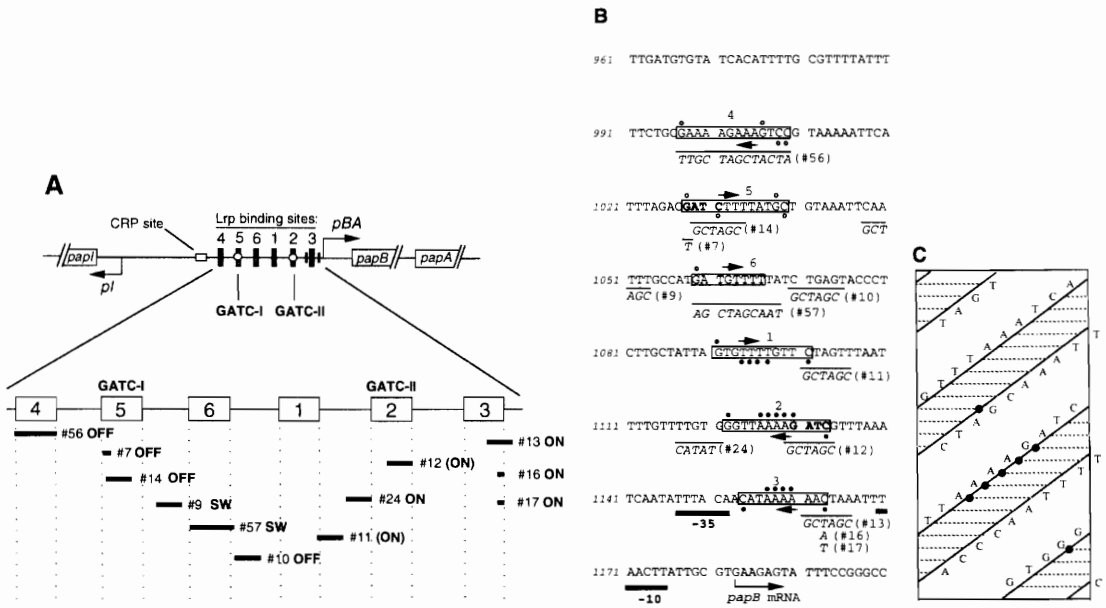


Fig. 2. The location of Lrp binding sites in the *pap* regulatory region. (**A**) The *papIBA* regulatory region and a summary of the mutation analyses. In the top half, the six Lrp binding sites identified in this study are shown as rectangles. The *pap* GATC-I and GATC-II sequences within sites 5 and 2, respectively, are shown as open circles. The –35 and –10 RNA polymerase binding region for the *papBAp* promoter is shown as two dark boxes around Lrp binding site 3. The bottom half shows the positions of each of the *pap* mutations and their transcriptional phenotypes: OFF, phase locked off; SW, reversible switching between phase OFF and ON states occurs; ON, phase locked ON; (ON), high OFF to ON rate. (**B**) The locations of Lrp binding sites and mutations within the *pap* regulatory region. Filled circles show guanosines and adenosines that are protected from DMS methylation after the addition of a low level (1.25 μg/ml) of Lrp (sites [1,2,3]). Filled circles above the sequence show the methylation protection of the top strand, whereas circles below the sequence show the methylation protection of the bottom strand. Open circles show guanosines and adenosines that are protected from methylation at a higher level of Lrp (5 μg/ml). Boxed regions show the six Lrp binding sites, each containing the consensus sequence Gnn(n)TTT(t), where 'n' is any nucleotide. Lrp site 6 is shown as a light box and extends only through the binding site consensus sequence because only a single guanosine was protected from methylation in this region. The arrows indicate the relative orientation of each Lrp binding site. The positions of substitution, transversion and transition mutations are shown below the DNA sequence by bars, with the substituted (mutated) base pairs in italics. The –35 and –10 RNA polymerase binding region for the *papBAp* promoter is shown, as well as the start site for transcription of the *papB* gene. (**C**) A planar representation of the DNA region around Lrp binding site 2 which contains GATC-II. The positions of bases are projected onto the surface of a cylinder that is then unrolled onto a flat surface. Closed circles show guanosines and adenosines that are protected from methylation due to Lrp binding.

To determine the relative dominance of sites [1,2,3] compared with sites [4,5] on *pap* transcription, we constructed a mutant (#47) containing mutations #14 (locked OFF) and #13 (locked ON). This double mutant displayed a locked ON phenotype with a β-galactosidase level similar to mutant #13 (Table I). These results indicate that mutation #13 is dominant to mutation #14 (see Discussion).

The binding of Lrp near GATC-I and GATC-II appears to inhibit the methylation of these sites by Dam (Nou *et al.*, 1993). Therefore, the methylation patterns of these sites provide information regarding the binding of Lrp to *pap in vivo*. An analysis of the GATC site methylation patterns of *pap* switch mutants #13 and #24 showed that the GATC-I sites of these locked ON mutants were almost completely protected from methylation, whereas the GATC-II site was fully methylated. These results suggest that Lrp is bound near GATC-I in these mutants, similar to phase ON wild-type cells (Blyn *et al.*, 1990; Nou *et al.*, 1993; Braaten *et al.*, 1994). Mutant #11, which had a higher fraction of phase ON cells because of a lower ON to OFF switch rate, showed a proportional increase in the methylation protection of GATC-I compared with wild-type cells (Table I). In contrast, the locked OFF mutants #10 and #14 showed the almost complete protection of GATC-II from methylation, whereas the GATC-I site of mutant #10 was fully methylated (mutant #14 lacks a GATC-I site). This methylation pattern suggests that Lrp binds close to GATC-II, similar to phase OFF cells.

The effects of Dam, Lrp and PapI on transcription from mutant pap constructs

Recent evidence indicates that Dam methylase plays a dual a role in regulating the *pap* transcriptional switch (Braaten *et al.*, 1994). The methylation of GATC-I inhibits the formation of the phase ON state, whereas the methylation of GATC-II is required for the phase ON state. Thus, *pap* transcription is shut down in a *dam*⁻ or a Dam overproducer (~4- to 69-fold) background (Blyn *et al.*, 1990; Table III). The introduction of mutant *dam* alleles into the *pap* switch mutants showed that mutant #13 remained locked ON in the absence of Dam or under high Dam levels, and thus is totally Dam independent. Mutant #12, which showed a higher rate of switching compared with wild-type cells under normal Dam levels (Table I), became locked ON in the absence of Dam. However, *pap*

transcription in mutant #12 was shut off at high Dam levels, similar to wild-type *pap* (Table III). A methylation analysis showed that at high Dam levels both the GATC-I and GATC-II sites of wild-type and mutant #13 DNAs were fully methylated and the GATC-I site of mutant #12 DNA was also fully methylated (Table III). An analysis of mutant #14 (locked OFF) showed that it remained locked OFF in the absence of Dam or under high Dam levels.

Mutations #12 and #13 are in close proximity to the −35 and −10 RNA polymerase consensus binding sites of the *papBAp* promoter, raising the possibility that the aberrant phase variation phenotypes of *E.coli* containing these mutations might be caused by an alteration in promoter activity rather than the disruption of Lrp binding. If this was so, then it is likely that *pap* transcription would no longer be dependent on Lrp (Braaten *et al.*, 1991). To determine if *pap* switch mutants #12 and #13 are dependent on Lrp, we introduced *lrp* null mutations into both wild-type and regulatory mutant cells. The transcription of *pap* in mutants #12 and #13 was shut off in the absence of Lrp, showing that these regulatory mutants remain dependent upon Lrp for *pap* transcription (Table II).

PapI is required for the methylation protection of the GATC-I site as well as *pap* transcription in *E.coli* containing a wild-type *pap* sequence (Braaten *et al.*, 1991; Nou *et al.*, 1993). Similar to wild-type *pap*, mutant #12 cells were dependent on PapI for transcription and showed little, if any, methylation protection of GATC-I in a *papI*⁻ background, consistent with their locked OFF phenotypes (Table II). However, mutant #13 remained locked ON in the absence of PapI, even though a methylation analysis showed that the GATC-I site was not protected from methylation under these conditions (Table II). Together, these results show that *pap* transcription in mutant #12, like wild-type cells, is dependent on both Lrp and PapI, whereas transcription in mutant #13 requires Lrp but is independent of PapI.

Determination of Lrp binding affinities for wild-type and mutant pap DNAs

To more fully understand the effects of mutations in the *pap* regulatory region, we analyzed the binding of Lrp, in the presence and absence of PapI, to both wild-type and mutant DNAs *in vitro* using quantitative DNase I footprinting (Figure 3). Lrp bound with highest affinity to sites [1,2,3] of wild-type *pap* DNA and with lower affinity to sites [4,5,6] (Figures 3A and 4A). The addition of PapI altered the affinity of Lrp for its DNA binding sites. With PapI, Lrp had the highest affinity for sites [4,5]. In contrast, the affinity of Lrp for the remaining four sites was reduced. A quantitative analysis showed that Lrp has an ~3-fold higher affinity for the GATC-II DNA region (sites [1,2,3]) than it does for the GATC-I region (sites [4,5]; Figure 4A). The addition of PapI increased the affinity of Lrp for the GATC-I region by ~4-fold and decreased the affinity of Lrp for the GATC-II region by almost 2-fold. Thus, the net effect of PapI is to shift the binding of Lrp from the GATC-II region (sites [1,2,3]) to the GATC-I region (sites [4,5]).

An analysis of mutant #12 (Figures 3B and 4B), which has an altered GATC-II site, showed that the affinity of Lrp for the GATC-II region (sites [2,3]) was reduced ~3-fold compared with wild-type *pap*. In contrast, the affinity of Lrp for sites [4,5] increased ~1.5-fold over the wild-type level. PapI increased the affinity of Lrp for sites [4,5] and reduced the affinity of Lrp for site 1, similar to wild-type *pap* DNA. These results are consistent with the increased OFF to ON switch rate of this mutant because the binding of Lrp to the GATC-II DNA region, which appears to act negatively on *pap* transcription, was significantly reduced. Moreover, the high affinity binding of Lrp to sites [4,5] around GATC-I, which appears to activate *pap* transcription, still required PapI, consistent with the PapI-dependent phenotype of this mutant.

An analysis of mutant #13, which contains a 6 bp replacement within site 3, showed that the binding of Lrp to sites [2,3] was reduced by ~3-fold, similar to that observed for mutant #12. However, Lrp had a higher affinity for sites [5,6] of mutant #13 DNA compared with either wild-type or mutant #12 DNAs (Figure 3C). Moreover, PapI did not affect the affinity of Lrp for site 5, which contains GATC-I, of mutant #13 DNA. Together, these results are consistent with the PapI-independent phenotype of this mutant (Table II) because the high affinity binding of Lrp to GATC-I (site 5) occurred in the absence of PapI. Similar footprinting results were obtained for mutant #24, which contains a 6 bp replacement within Lrp site 2, and mutants #16 and #17, which contain point mutations within site 3, all displaying locked ON phenotypes (results not shown).

Mutant #14 contains a 6 bp replacement of the GATC-I site, resulting in a locked OFF transcription phenotype (Table I). As shown in Figure 3D, the mutation of the GATC-I site did not reduce the affinity of Lrp by itself for sites 4 or 5, although the affinity of Lrp for sites [6,1,2,3] were reduced to the level that Lrp has for wild-type *pap* DNA in the presence of PapI. Strikingly, PapI did not increase the affinity of Lrp for sites [4,5] around the mutant #14 GATC-I site, as occurred with the wild-type *pap* GATC-I site. Thus, the binding transition of Lrp from the GATC-II to the GATC-I region, mediated by PapI, is blocked in mutant #14 and cells remain in the transcriptionally inactive state. These results indicate that the GATC-I sequence is necessary for the binding of Lrp/PapI to sites [4,5].

Discussion

The data presented here show that the transition between active and inactive *pap* transcriptional states is regulated by the binding of Lrp to two sets of *pap* DNA binding sites. In the absence of PapI, Lrp binds with highest affinity to sites [1,2,3] around GATC-II of the *pap* regulatory DNA. The binding of Lrp at sites [2,3] appears to inhibit *pap* transcription because mutations within or near these sites resulted in locked ON phenotypes or higher OFF to ON switch rates (Table I). In the presence of PapI, the affinity of Lrp for sites [4,5] increased ~4-fold, and the affinity of Lrp for sites [1,2,3] decreased ~2-fold (Figure 4A), resulting in a translocation of Lrp from the GATC-II to the GATC-I region. Recent data have shown that PapI binds specifically to the Lrp moiety of Lrp–*pap* DNA complexes (Kaltenbach *et al.*, 1995). Thus, it is possible that the affinity of Lrp for sites [1,2,3,4,5] is altered by a conformation change induced by the binding of PapI. It seems likely that the binding of Lrp/PapI to the GATC-I

region activates *pap* transcription because the mutation of the GATC-I site inhibits the binding of Lrp/PapI to sites [4,5] and prevents the formation of the phase ON transcription state (Figure 4D, mutant #14).

Previously, we have shown that methylation plays dual roles in *pap* transcriptional regulation (Braaten *et al.*, 1994). The methylation of the GATC-I site inhibits the binding of Lrp/PapI and shuts off transcription. In contrast, the methylation of GATC-II is required for *pap* transcription, possibly by reducing the affinity of Lrp for this DNA region. These data are consistent with recent evidence which indicates that the binding of Lrp at the GATC-II region inhibits the basal transcription level of the *papBAp* promoter, whereas the binding of Lrp near GATC-I activates *pap* transcription ~8-fold above the basal level (van der Woude *et al.*, 1995). Therefore Lrp acts as a repressor in the absence of PapI and as an activator in the presence of PapI, because PapI is required to shift the binding of Lrp from GATC-II to GATC-I (Figures 3 and 4). Here, our data indicate that Lrp binds to sites [1,2,3] near GATC-II in the phase OFF state (Figures 3A and 4A). Because Lrp binding site 3 is located between the −35 and −10 RNA polymerase consensus binding region (Figure 2A and B), this provides a mechanism (via steric hindrance) by which the binding of Lrp might directly inhibit *pap* transcription. It is less clear how the binding of Lrp near GATC-I activates *pap* transcription. Based on the analysis of Lrp activation mutants, the binding of Lrp is not sufficient for the activation of *pap* transcription, suggesting that Lrp interacts with another component of the regulatory machinery such as RNA polymerase or catabolite gene activator protein (van der Woude *et al.*, 1995).

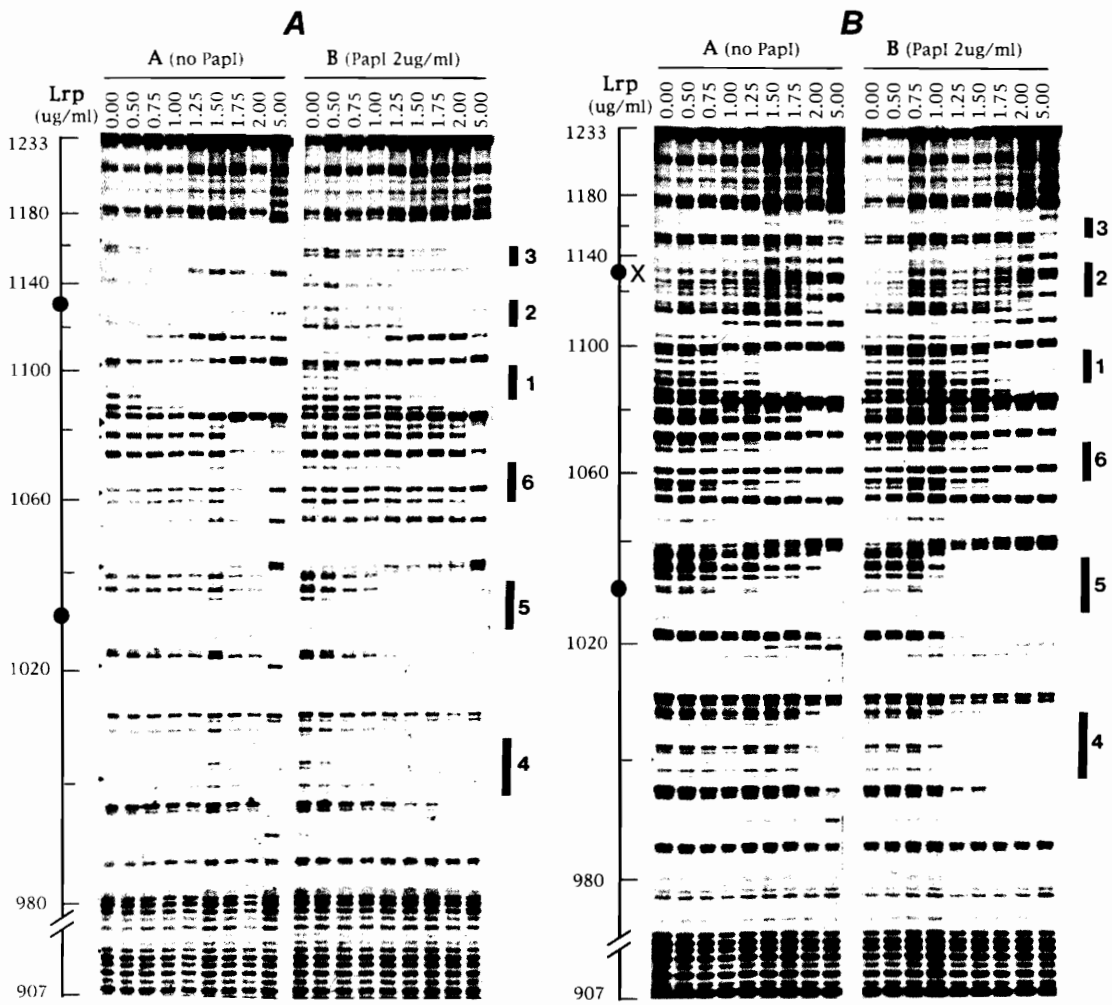
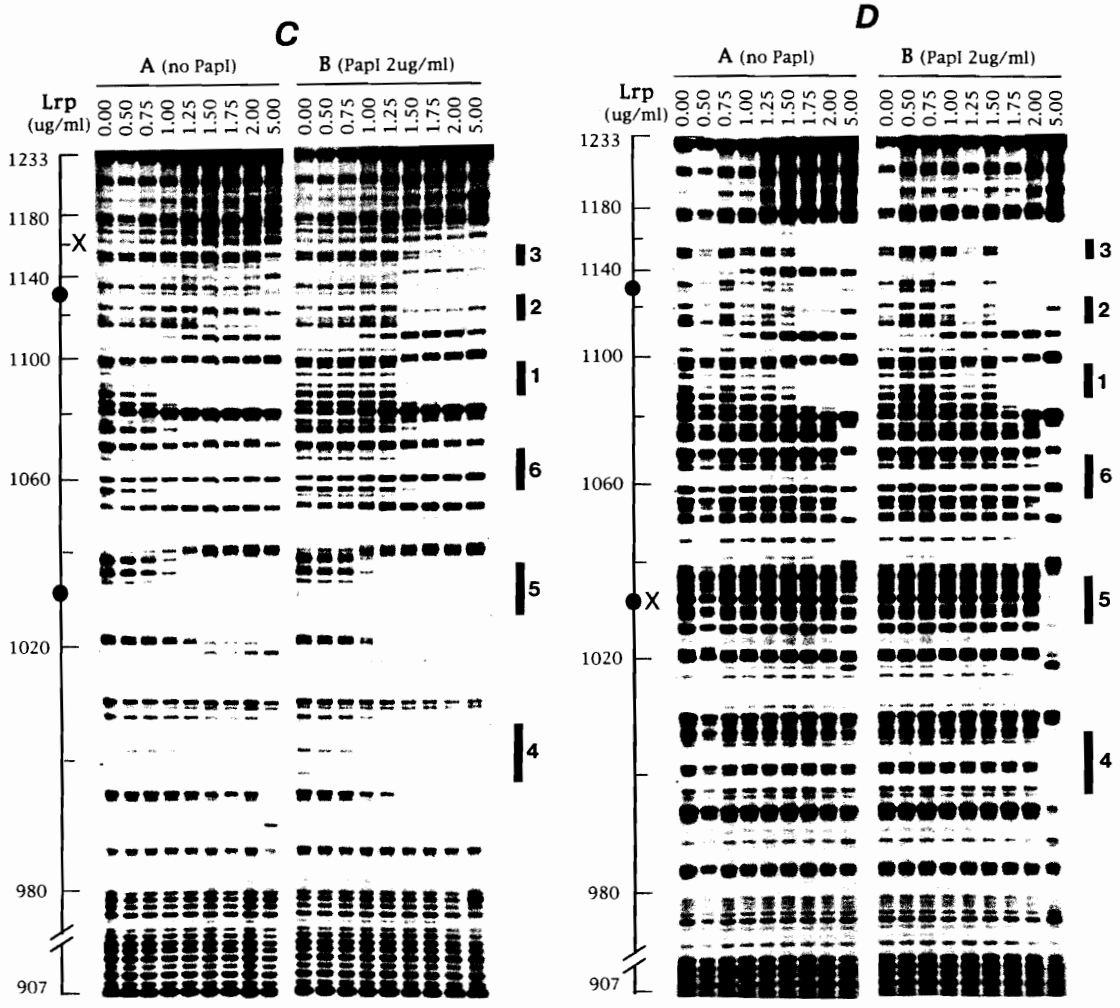


Fig. 3. A quantitative DNase I footprint analysis of non-methylated wild-type and Lrp binding site mutant *pap* DNAs. The left portion of each figure (section A) shows the addition of increasing amounts of Lrp (μg/ml) to *pap* DNAs, whereas the right portion of each figure (section B) shows the addition of increasing amounts of Lrp in the presence of PapI (2 μg/ml). To the left of each figure are shown the *pap* nucleotide positions (corresponding to Figure 1), with the GATC-II site at 1130–1133 bp and the GATC-I site at 1028–1031 bp (both GATC sites are shown as closed circles). To the right, the six Lrp binding sites, identified by a methylation protection analysis, are shown. (**A**) Wild-type *pap* DNA; (**B**) mutant #12 DNA; (**C**) mutant #13 DNA; (**D**) mutant #14 DNA.

The GATC-I site plays an essential role in the modulation of Lrp binding by PapI, as shown by an analysis of mutant #14 which contains a 6 bp substitution of GATC-I. Although the affinity of Lrp for sites [4,5] of mutant #14 DNA was not affected, PapI did not increase the affinity of Lrp for these sites as it did with wild-type *pap* DNA (compare Figure 3A with D). Thus, the stable binding of Lrp/PapI to sites [4,5] requires the GATC-I sequence. The 'G' of GATC-I appears to be in contact with Lrp (Figure 2B); it is essential for transcriptional activation because a G to T transversion (mutation #7) gave a locked OFF phenotype. Although we have not yet analyzed the binding of Lrp to this point mutant, it seems likely that it will reduce the affinity of Lrp/PapI for sites [4,5], as observed with mutant #14. Our previous results indicated that the adenosine of GATC-I must also come into close contact with Lrp because methylation of this 'A' blocks the binding of Lrp/PapI to the GATC-I region (Nou *et al.*, 1993).

The mutation of binding site 3 (mutant #13) resulted in a locked ON transcription phenotype which, in contrast to wild-type cells, was independent of PapI and Dam methylation (Tables II and III). An *in vitro* footprint analysis showed that Lrp bound to sites [5,6,1] of mutant #13 DNA with highest affinity (Figure 4C). An *in vivo* analysis of methylation patterns showed that the GATC-I site was almost fully protected from methylation, consistent with the *in vitro* binding results, indicating that Lrp binds with high affinity near site 5, which encompasses the GATC-I region. Although the binding of Lrp to site 5 and the transcription activation occurred in the absence of PapI, the high affinity binding of Lrp to site 4 (Figure 4C) and the methylation protection of GATC-I (Tables II and IV) both required PapI. These data are consistent with the observation that *pap* transcription was activated in a Dam overproducer background, conditions under which both *pap* GATC sites are fully methylated (Table III). This raises the possibility that the binding of Lrp to mutant #13 DNA site 5, containing GATC-I, is not blocked by methylation of this site or is not critical for binding to sites [6,1]. This could result from the fact that Lrp binds to three sites ([5,6,1]) of mutant #13 DNA with high affinity, which might stabilize binding compared with



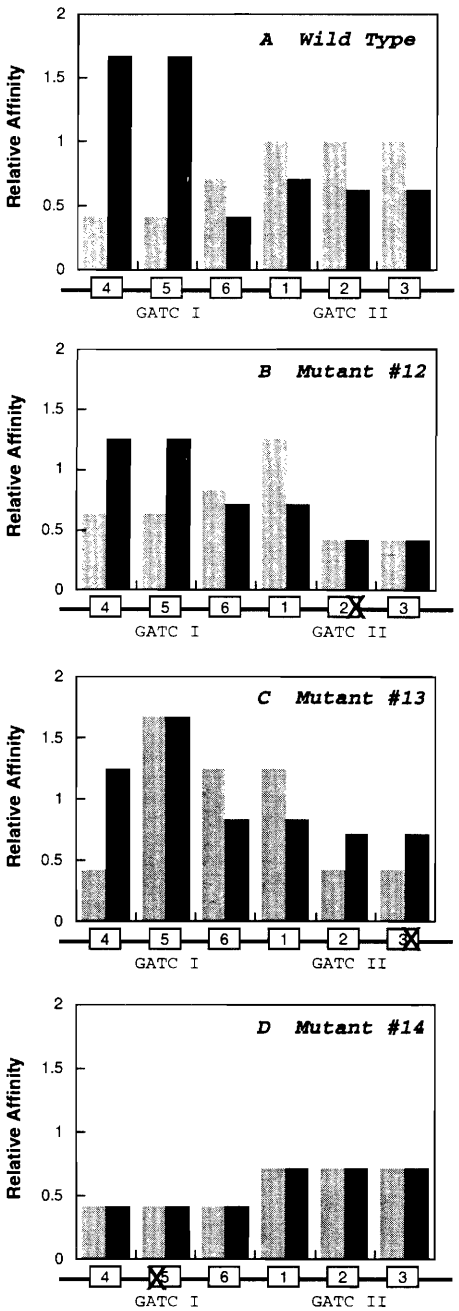


Fig. 4. Summary of the quantitative footprint analyses of Lrp binding to wild-type and mutant *pap* DNAs. In each panel, the relative affinity of Lrp for each *pap* DNA binding site is shown on the left. The affinity of Lrp for sites [1,2,3] of wild-type *pap* DNA has been set at a value of 1.0. Each of the six Lrp binding sites is shown at the bottom of the figure, including the location of the *pap* GATC-I and GATC-II sites. The binding affinity of Lrp is shown by light bars, whereas the affinity of Lrp in the presence of PapI is shown by dark bars. The binding affinities were calculated as the concentration of Lrp at which protection from DNase I cleavage was one-half maximal. Each panel corresponds to the panels shown in Figure 3.

wild-type *pap* in which Lrp/PapI binds with high affinity to only two sites ([4,5]). This hypothesis was supported by our observation that mutation of site 5 does not block transcription when combined with mutation #13 in site 3 (see double mutant #47, Table I). Thus, in contrast to wild-type *pap*, the activation of transcription in mutant #13 does not require site 5 which contains GATC-I.

Our previous results indicated that the methylation of GATC-II by Dam methylase is essential for *pap* transcription (Braaten *et al.*, 1994). We hypothesized that the methylation of this site might inhibit Lrp binding to the GATC-II region, thus preventing the Lrp-mediated repression of transcription (van der Woude *et al.*, 1995). Our finding here, that mutant #13 remained phase locked in the phase ON transcription state in the absence of Dam methylase, is consistent with this hypothesis because the binding of Lrp to the GATC-II region is already greatly reduced by the mutation of site 3, precluding the need for methylation (Tables II and III, and Figure 4). Similarly, mutant #12 lacking the GATC-II site was phase locked ON in the absence of Dam (Table III), in contrast to our previous finding that a GCTC-II point mutation, which also prevents Dam methylation, was locked in the OFF state (Braaten *et al.*, 1994). We hypothesize that the different phenotypes of these mutations in GATC-II are a result of the 6 bp substitution mutation in GATC-II (mutation #12) disrupting the binding of Lrp to sites [2,3] (Figure 4B), whereas the GCTC-II point mutation does not (Braaten *et al.*, 1994). Thus, in the latter case, the high affinity binding of Lrp to sites [1,2,3] cannot be disrupted by Dam, which prevents the phase ON transition.

The mutation of site 6 (mutant #57) had only a small effect on *pap* transcription, indicating that the binding of Lrp to site 6 does not play a major role in Pap phase variation. In contrast, mutation #10, located between sites 6 and 1, resulted in a locked OFF phenotype (Figure 2A and Table I). Preliminary results indicate that the mutation of site 1 also gives a locked OFF phenotype (unpublished data), suggesting that the binding of Lrp at or close to site 1 is essential for *pap* transcription. In contrast, the Lrp binding at sites [2,3] blocks *pap* transcription because the mutation of these sites results in locked ON phenotypes (Figure 2A). Thus, although Lrp appears to bind to sites [1,2,3] in phase OFF cells, the role of site 1 appears to differ from that of sites [2,3]. One possible reason for this difference in Lrp binding site activity is that the translocation of Lrp from sites [1,2,3] to sites [4,5] might require the interaction of PapI with Lrp bound at site 1. If this is so, then the PapI-induced translocation of Lrp from GATC-II to GATC-I should not occur in *pap* site 1 mutants. Experiments are in progress to test this hypothesis.

Previous work by Wang and Calvo (1993) has shown that there are five Lrp binding sites which play roles in the transcription activation of the *ilvIH* operon. However, the arrangement of these sites and the binding by Lrp differ significantly from the *pap* operon. The spacing between *ilvIH* sites varies from 20 to 34 bp, compared with a spacing of 30–33 bp between the Lrp binding sites of *pap* (Figure 2B). Lrp binds with high affinity (K_d ~6 nM) to sites [1,2] of *ilvIH*. Lrp has a lower affinity for sites [3,4,5] of *ilvIH*, with a K_d of ~22 nM. The latter value is similar to our estimate of the binding affinity of

Table I. Phenotypes of *pap* regulatory mutants

Mutation in *pap* DNA[a]	Switch frequency (events/cell/generation)[b]		β-Galactosidase activity (Miller units)			Methylation protection (% non-methylated)[d]	
	ON→OFF	OFF→ON	OFF	ON	(%ON)[c]	GATC-I	GATC-II
None (wild type)	2.35×10^{-2}	5.54×10^{-4}	26 ± 4	1440 ± 115	35	31	63
#13 (site 3)	locked ON	NA[e]	NA	5714 ± 899	100	95	0
#12 (site 2, GATC-II)	3.07×10^{-2}	3.45×10^{-3}	106 ± 42	1292 ± 190	31	28	NA
#24 (site 2)	locked ON	NA	NA	5282 ± 662	100	94	0
#11 (between sites 1 and 2)	1.23×10^{-2}	4.24×10^{-3}	369 ± 15	3272 ± 154	60	61	15
#10 (between sites 1 and 6)	NA	locked OFF	3 ± 1	NA		0	90
#57 (site 6)	2.96×10^{-2}	1.90×10^{-4}	20 ± 1	781 ± 40	9.8	ND[f]	ND
#9 (between sites 5 and 6)	2.28×10^{-2}	3.27×10^{-4}	30 ± 9	1703 ± 277	38	35	54
#14 (site 5, GATC-I)	NA	locked OFF	6 ± 1	NA		NA	80
#56 (site 4)	NA	locked OFF	1.9 ± 0.3	NA		ND	ND
#47 (sites 3 and 5) (mutations #13, #14)	locked ON	NA	NA	5053 ± 427	100	NA	0

[a]The mutation number is followed by the Lrp binding site (and GATC site where applicable) affected by the mutation (see Figure 2).
[b]The switch frequency is calculated as the number of switch events per cell per generation (Blyn *et al.*, 1990).
[c]When β-galactosidase activities were measured, the percentage of cells in the ON phase was also determined, as described previously (Blyn *et al.*, 1990).
[d]Methylation protection was calculated as the percentage of GATC sites that were digested by *Mbo*I (Braaten *et al.*, 1994).
[e]Not applicable.
[f]Not done.

Table II. Effects of Lrp and PapI on the phenotypes of *pap* regulatory mutants

Mutation in *pap* DNA	Switch frequency (events/cell/generation)[a]		β-Galactosidase activity (Miller units)		Methylation protection (% non-methylated)[b]	
	ON→OFF	OFF→ON	OFF	ON	GATC-I	GATC-II
In *Lrp*⁻ background						
Wild type	NA[c]	locked OFF	8 ± 9	NA	0	0
#13 (site 3)	NA	locked OFF	6 ± 1	NA	0	0
#12 (site 2, GATC-II)	NA	locked OFF	9 ± 1	NA	0	0
In *PapI*⁻ background						
Wild type	NA	locked OFF	3 ± 1	NA	0	70
#13 (site 3)	locked ON	NA	NA	4320 ± 420	0	0
#12 (site 2, GATC-II)	NA	locked OFF	41 ± 1	NA	0	NA

[a]Switch frequency is calculated as the number of switch events per cell per generation (Blyn *et al.*, 1990).
[b]Methylation protection was calculated as the percentage of GATC sites that were digested by *Mbo*I (Braaten *et al.*, 1994).
[c]Not applicable.

Table III. Effects of Dam methylase on the phenotypes of *pap* regulatory mutants

Mutation in *pap* DNA	Switch frequency (events/cell/generation)[a]		β-Galactosidase activity (Miller units)		Methylation protection (% non-methylated)[b]	
	ON→OFF	OFF→ON	OFF	ON	GATC-I	GATC-II
In *Dam*⁻ strain						
Wild type	NA	locked OFF	84 ± 6	NA[c]	NT[d]	NT
#13 (site 3)	locked ON	NA	NA	4488 ± 411	NT	NT
#12 (site 2, GATC-II)	locked ON	NA	NA	2965 ± 435	NT	NT
#14 (site 5, GATC-I)	NA	locked OFF	7 ± 1	NA	NT	NT
In *Dam* overproducer strain (4-fold up)						
Wild type	NA	locked OFF	70 ± 34	NA	0	0
#13 (site 3)	locked ON	NA	NA	5779 ± 415	0	0
#12 (site 2, GATC-II)	NA	locked OFF	19 ± 1	NA	0	NA
#14 (site 5, GATC-I)	NA	locked OFF	69 ± 2	NA	NA	20

[a]Switch frequency is calculated as the number of switch events per cell per generation (Blyn *et al.*, 1990).
[b]Methylation protection was calculated as the percentage of GATC sites that were digested by *Mbo*I (Braaten *et al.*, 1994).
[c]Not applicable.
[d]Not tested (previous analysis showed that the two *pap* GATC sites are 100% non-methylated in this *dam* background; unpublished data).

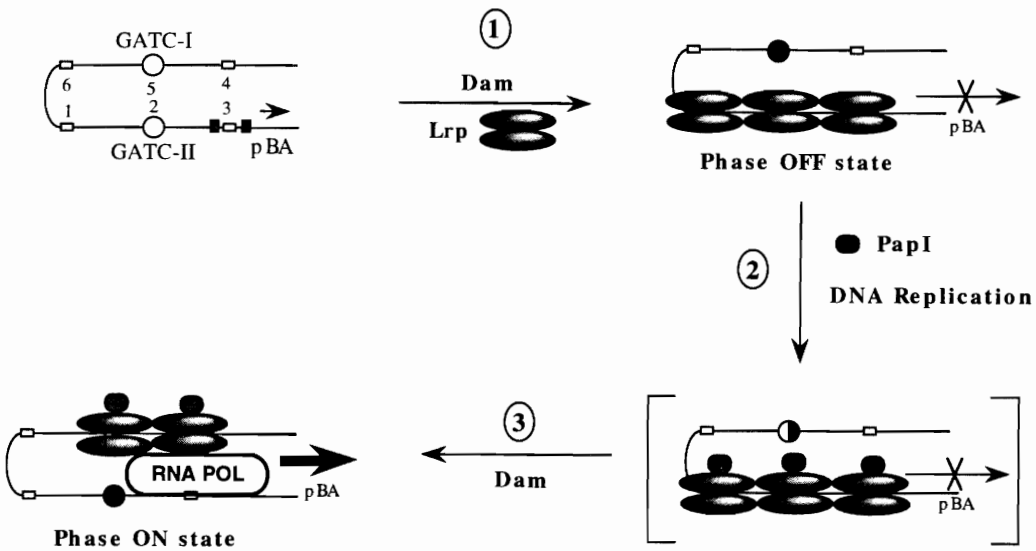


Fig. 5. Pap phase variation model. At the upper left, the six Lrp binding sites identified in this study are shown. Lrp binding sites [1,3,4,6] are shown as boxes, whereas Lrp sites [2,5], which contain the *pap* GATC-II and GATC-I sites, respectively, are shown as circles. Fully methylated GATC sites are shown as closed circles; hemi-methylated GATC sites are shown as half-closed circles. The −35 and −10 RNA polymerase binding sites of the *pap* BA promoter are shown as small shaded boxes flanking site 3. The circled numbers refer to steps in the *pap* transcription phase switching referred to in the text. Step 1: binding of Lrp to sites [1,2,3] blocks basal transcription and generates the phase OFF state. Step 2: PapI binds to Lrp. After DNA replication, a hemi-methylated GATC-II site is generated by DNA replication. Step 3: Lrp-PapI complexes bind to sites [4,5] and the GATC-II site is methylated by Dam, generating the phase ON state.

Lrp for sites [1,2,3] of *pap* (K_d ~27 nM; taken from the data in Figures 3 and 4). It appears that the activation of *ilvIH* transcription involves the occupancy of all five Lrp binding sites, a condition in which Lrp bound to site 5, 70 bp upstream of the transcription start site, might allow a direct contact with RNA polymerase. In contrast, our data for *pap* show that Lrp alternates between binding sites [1,2,3] in the absence of PapI, and sites [4,5] in the presence of PapI. Assuming that this *in vitro* analysis applies *in vivo*, in phase ON cells Lrp bound to *pap* sites [4,5] would be at least 140 bp upstream of the transcript start site, necessitating DNA bending to allow the contact between Lrp and RNA polymerase. It does not seem likely that site 6 is occupied in ON phase cells *in vivo* because PapI, which is required for *pap* transcription, reduces the affinity of Lrp for this site (Figure 4A).

The Lrp consensus binding sequence deduced from this study for *pap* [Gnn(n)TTTt] (Figure 2) differs somewhat from that reported previously for *ilvIH* (AgaATTT-TATtcT), although both contain a 'T' tract. Our mutational analyses indicate that the consensus 'G' in sites [1,3,5] is important for Lrp binding (Figure 2B and Table I). However, sites outside this consensus also appear to be important for Lrp binding (see mutants #10 and #24). These DNA regions could function directly as contact sites for Lrp or indirectly through effects on the DNA structure.

Previous results have shown that the binding of Lrp to *ilvIH* DNA is highly cooperative (Wang and Calvo, 1993). Our data suggest that Lrp binds cooperatively to *pap* DNA sites [2,3]. Footprint analyses showed that Lrp bound to sites [1,2,3] with equal affinity (Figures 3 and 4). The disruption of site 3 inhibited the binding of Lrp to sites [3,2] (mutant #13), as did the disruption of site 2 (mutant #12) (Figure 4). These results indicate that the binding of Lrp to sites [2,3] is cooperative. Based on preliminary data it does not seem likely that site 1 is required for Lrp cooperative interactions because the disruption of this site has only a small effect on Lrp affinity for sites [2,3] (L.Kaltenbach, unpublished data).

A model for Pap phase variation which includes the data presented here is shown in Figure 5. Lrp binds with highest affinity to sites [1,2,3], blocking basal transcription (step 1) (van der Woude *et al.*, 1995). The GATC-I sequence at site 5 is not bound by Lrp, and is methylated by Dam. Under conditions in which PapI levels are induced (Blyn *et al.*, 1990), PapI binds specifically to the Lrp moiety of Lrp–*pap* DNA complexes (step 2) (Kaltenbach *et al.*, 1995). However, Lrp–PapI complexes are locked in position because the methylation of GATC-I prevents the binding of Lrp/PapI to sites [4,5] (Nou *et al.*, 1993; Braaten *et al.*, 1994). We hypothesize that the Lrp/PapI binding transition shown in step 3 occurs following DNA replication because Lrp/PapI binds to the GATC-I region if this GATC site is hemi-methylated (Nou *et al.*, 1993). Following DNA replication, PapI decreases the affinity of Lrp for sites [1,2,3] and increases Lrp affinity for sites [4,5], resulting in the translocation of Lrp from the GATC-II to the GATC-I region (step 3). The methylation of GATC-II is required for *pap* transcription, possibly by inhibiting the binding of Lrp to the GATC-II region (Braaten *et al.*, 1994). The binding of Lrp to sites [4,5] activates *pap* transcription, possibly via the direct contact of Lrp with RNA polymerase (van der Woude *et al.*, 1995). The transition from the phase ON to phase OFF state can occur after DNA replication if Lrp dissociates

Table IV. *Escherichia coli* strains, plasmids and bacteriophage used in this study

E.coli strain, plasmid or bacteriophage	Description	Reference or source
E.coli		
MC4100	*F⁻araD139 Δ(lacIPOZYA-argF)U169 rpsL thi-1*	Casadaban (1976)
DL379	MC4100 λ246 lysogen	Blyn *et al.* (1989)
DL520	MC4100 λ246-16 lysogen	this study
DL524	MC4100 λ246-17 lysogen	this study
DL1504	MC4100 λ354 lysogen	Braaten *et al.* (1994)
DL2207	MC4100 λ354-7 lysogen	this study
DL2078	MC4100 λ354-9 lysogen	this study
DL2079	MC4100 λ354-10 lysogen	this study
DL2080	MC4100 λ354-11 lysogen	this study
DL2120	MC4100 λ354-12 lysogen	this study
DL2121	MC4100 λ354-13 lysogen	this study
DL2122	MC4100 λ354-14 lysogen	this study
DL2124	MC4100 λ354-24 lysogen	this study
DL2193	MC4100 λ354-47 lysogen	this study
DL2496	MC4100 λ354-56 lysogen	this study
DL2497	MC4100 λ354-57 lysogen	this study
DL1611	DL1504 *dam-13*::Tn9 (cam^r)	Braaten *et al.* (1994)
DL2150	DL2120 *dam-13*::Tn9 (cam^r)	this study
DL2151	DL2121 *dam-13*::Tn9 (cam^r)	this study
DL2152	DL2122 *dam-13*::Tn9 (cam^r)	this study
DL1650	DL1504(pPY1025, pTP166)	Braaten *et al.* (1994)
DL2156	DL2120(pPY1025, pTP166)	this study
DL2157	DL2121(pPY1025, pTP166)	this study
DL2158	DL2122(pPY1025, pTP166)	this study
DL2046	DL1504 *mbf-20* (*lrp* null mutation)	this study
DL2161	DL2120 *mbf-20*	this study
DL2162	DL2121 *mbf-20*	this study
DL1741	MC4100 λ354 Δ*papI* lysogen	this study
DL2187	MC4100 λ354-12 Δ*papI* lysogen	this study
DL2188	MC4100 λ354-13 Δ*papI* lysogen	this study
Plasmids		
pTZ18R	*amp* pMB1 replicon	US Biochemical
pDAL262	pTZ18R containing a 271 bp *Taq*I–*Sph*I *papI* DNA fragment	Nou *et al.* (1993)
pRS550	*amp–kan–lacZYA* pMB1 replicon	Simons *et al.* (1987)
pRS551	*amp–kan–lacZYA* pMB1 replicon	Simons *et al.* (1987)
pDAL337	pTZ19U containing a 1.76 kb *papIB* regulatory sequence	Braaten *et al.* (1994)
pDAL354	pRS550 containing a 1.76 kb *papIB* regulatory sequence	Braaten *et al.* (1994)
pDAL354-7	pDAL354 with *pap* bp 1027 G to T transversion	this study
pDAL354-9	pDAL354 with *pap* bp 1048–1053 replaced by 'GCTAGC'	this study
pDAL354-10	pDAL354 with *pap* bp 1070–1075 replaced by 'GCTAGC'	this study
pDAL354-11	pDAL354 with *pap* bp 1101–1106 replaced by 'GCTAGC'	this study
pDAL354-12	pDAL354 with *pap* bp 1048–1053 replaced by 'GCTAGC'	this study
pDAL354-13	pDAL354 with *pap* bp 1161–1166 replaced by 'GCTAGC'	this study
pDAL354-24	pDAL354 with *pap* bp 1117–1121 replaced by 'CATAT'	this study
pDAL354-47	pDAL354 with mutations #13 and #14 in combination	this study
pDAL354-56	pDAL354 with *pap* bp 997–1009 replaced by 'TTGCTAGCTACTA'	this study
pDAL354-57	pDAL354 with *pap* bp 1059–1068 replaced by 'AGCTAGCAAT'	this study
pDAL246	pRS551 containing a 1.6 kb *papIB* DNA sequence	Blyn *et al.* (1990)
pDAL246-16	pDAL246 with *pap* bp 1163 C to A transversion	this study
pDAL246-17	pDAL246 with *pap* bp 1163 C to T transition	this study
pPY1025	pGB2 containing *lacI^q*	P.Youderian
pTP166	pBR322 containing *ptac–dam*	Marinus *et al.* (1984)
Bacteriophage		
λRS45	*amp'-lacZYA imm^{21}*	Simons *et al.* (1987)
λ354	λRS45–pDAL354 recombinant phage	Braaten *et al.* (1994)
λ354-7	λRS45–pDAL354-7 recombinant phage	this study
λ354-9	λRS45–pDAL354-9 recombinant phage	this study
λ354-10	λRS45–pDAL354-10 recombinant phage	this study
λ354-11	λRS45–pDAL354-11 recombinant phage	this study
λ354-12	λRS45–pDAL354-12 recombinant phage	this study
λ354-13	λRS45–pDAL354-13 recombinant phage	this study
λ354-14	λRS45–pDAL354-14 recombinant phage	this study
λ354-24	λRS45–pDAL354-24 recombinant phage	this study
λ354-47	λRS45–pDAL354-47 recombinant phage	this study
λ354-56	λRS45–pDAL354-56 recombinant phage	this study
λ354-57	λRS45–pDAL354-57 recombinant phage	this study
λ246-16	λRS45–pDAL246-16 recombinant phage	this study
λ246-17	λRS45–pDAL246-17 recombinant phage	this study

amp, ampicillin; kan, kanamycin.

from sites [4,5] and rebinds to sites [1,2,3]. This should be favored under low PapI levels (Blyn *et al.*, 1990).

Materials and methods

Bacterial strains, plasmids, bacteriophage and media

The *E.coli* strains, plasmids and bacteriophage used in this study are shown in Table IV. The construction of *E.coli* λ lysogens containing single-copy wild-type and mutant *pap* DNA sequences was carried out as described previously (Braaten *et al.*, 1994). Congenic *dam*$^-$ and *lrp*$^-$ strains were constructed by phage P1 transduction, and Dam overproducing strains were made as described previously (Braaten *et al.*, 1994). Luria–Bertani (LB) broth, LB agar, M9 minimal broth (M9) and M9 agar were prepared according to Miller (1972). The carbon source for M9 medium was glycerol at a final concentration of 0.2%. The chromogenic substrate 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal; Gold Biotechnology) was used at 40 μg/ml of medium. Antibiotics, when used, were at the following final concentrations: ampicillin, 100 μg/ml; kanamycin, 25 μg/ml; spectinomycin, 100 μg/ml; and tetracycline, 15 μg/ml.

Site-specific mutagenesis and construction of double-site and papI deletion mutants

The site-directed mutagenesis protocol of Kunkel (1985) was used to mutate specific *pap* regulatory sequences as described previously, except that T7 DNA polymerase was used for second-strand synthesis (Braaten *et al.*, 1994). Mutagenic primers were made using an Applied Biosystems 380B DNA synthesizer as follows: mutant #7, 5′-TTCATTTAGACTATCTTTTATGC-3′OH; mutant #9, 5′-TTTATGCTGTAAATTGCTAGCGCCATGATGTTTTTA-3′OH; mutant #10, 5′-CCATGATGTTTTTATGCTAGCACCCTCTTGCTATTA-3′OH; mutant #11, 5′-CTATTAGTGTTTTGTTGCTAGCTAATTTTGTTTTGTGG-3′OH; mutant #12, 5′-GTTTTGTGGGTTAAAGCTAGCTTTAAATCAATATTT-3′OH; mutant #13, 5′-ATTTACAACATAAAAGCTAGCATTTAACTTATTGCG-3′OH; mutant #14, 5′-AAATTCATTTAGACGGCTAGCTATGCTGTAAATTCA-3′OH; mutant #56, 5′-GCGTTTTATTTTTCTGCTTGCTAGCTACTAGTAAAAATTCATTTAGACG-3′OH; and mutant #57, 5′-GCTGTAAATTCAATTTGCCATAGCTAGCAATTCTGAGTACCCTCTTTGC-3′OH.

Mutations #16 and #17 were constructed by induction with hydroxylamine as follows. Phage λ246 (Blyn *et al.*, 1990) was treated with hydroxylamine as described previously (Silhavy *et al.*, 1984) and used to lysogenize *E.coli* strain MC4100. Lysogens were plated on M9 minimal medium containing kanamycin and X-gal indicator at 23 and 37°C. Two isolates, DL520 and DL524, that showed locked ON phenotypes at both temperatures, were picked and analyzed. A DNA sequence analysis showed that DL520 contained a C to A transversion at 1163 bp (mutation #16), whereas strain DL524 contained a C to T transition at 1163 bp (mutation #17).

Mutant #24 was constructed by an overlapping PCR using two internal primers (primer A, 5′-TTTGTTTTGTGCATATGGTTAAAAGAT-3′; primer B, 5′-TCTTTTAACCATATGCACAAAACAAA-3′) and two external primers (primer C, 5′-CGCATTTCCTGACCGACT-3′; primer D, 5′-CCTGTTCAGAAATTCCAG-3′). Plasmid pDAL337, containing the entire *pap* regulatory region, was used as a template (Braaten *et al.*, 1994). An overlapping PCR was carried out as described previously (Ausubel *et al.*, 1994).

The double mutant #47, containing mutations #13 and #14, was constructed as follows. Both plasmids pDAL354-13 and pDAL354-14 were digested with *Nsp*V and *Apa*I. The *Nsp*V–*Apa*I DNA fragment was then digested with *Dde*I and the two DNA fragments containing mutations #13 and #14 were religated to the large DNA fragment of vector pDAL354, digested previously with *Nsp*V and *Apa*I (Braaten *et al.*, 1994).

The *papI* deletion derivatives λ354 Δ*papI*, λ354-12 Δ*papI* and λ354-13 Δ*papI* were constructed by replacing the *Sph*I–*Nsp*V DNA fragment containing the entire *papI* coding sequence with a 21mer double-stranded oligonucleotide containing the *Sph*I and *Nsp*V end sites. This double-stranded oligonucleotide was made by annealing the following single-strand DNAs together: 5′-CTCACTGTAACAAAGTTTCTT-3′ and 5′-CGAAGAAACTTTGTTACAGTGAGCATG-3′.

Southern blot, switch frequency and β-galactosidase analyses

A Southern blot analysis was performed as described previously (Blyn *et al.*, 1990), except that chromosomal DNA was isolated using a QIAGEN QIAamp tissue kit following the manufacturer's instructions. A quantitation of radioactivity was carried out using a Bio-Rad model GS-250 imager system. Switch frequencies (number of switch events per cell per generation) were determined as described previously (Blyn *et al.*, 1989), and β-galactosidase activities were determined as described previously (Braaten *et al.*, 1994).

DNA methylation protection, gel retardation and DNase I quantitative footprint analyses

A gel retardation analysis was carried out as described previously with the following modifications. Samples (20 μl) contained 20 000 c.p.m. [^{32}P]dATP-labeled DNA fragment, 3 μg herring sperm DNA and 2 μg acetylated bovine serum albumin in binding buffer (60 mM Tris–HCl, pH 7.5, 40 mM KCl, 100 mM NaCl, 0.1 mM EDTA and 1 mM dithiothreitol). The reaction mixes were incubated 15 min at room temperature and analyzed on high ionic strength polyacrylamide gels, as described previously (Nou *et al.*, 1993). Similar results were obtained on low ionic strength gels, but with a reduced resolution.

DNA footprinting using DMS was performed by treating *pap* DNA–Lrp complexes, formed as described previously (Nou *et al.*, 1993), with 0.2 volumes of 3 mM DMS in reaction buffer (0.3 M sodium cacodylate, pH 7.3, 1 mM EDTA) for 3 min at room temperature. The reaction was stopped by the addition of an equal volume (0.2 volumes of DNA–protein mix) of 2.5 M β-mercaptoethanol. DNA was precipitated by ethanol addition and washed twice with 70% ethanol before resuspension in 7 μl 20 mM sodium acetate. Samples were then heated to 90°C for 30 min, and piperidine was added to 10% (v/v) in a final volume of 70 μl. Samples were heated to 90°C for 30 min and vacuum evaporated. DNA samples were then resuspended in 5 μl formamide loading buffer and analyzed on a 6% polyacrylamide sequencing gel.

Quantitative DNase I footprinting was carried out as described previously for DNase I footprinting (Braaten *et al.*, 1994) with the following modifications. Highly purified Lrp (Braaten *et al.*, 1994) was a gift from Felix Vajdos and Chris Hill (University of Utah, UT). PapI was isolated from a glutathione-S-transferase–PapI fusion protein, purified by affinity chromatography on glutathione agarose. After cleavage with thrombin, PapI containing two additional amino acids at the N-terminus (gly–ser) was isolated. A constant level of PapI (sufficient to supershift 100% of Lrp–DNA complexes) and increasing amounts of Lrp were added to 100 μl of a mixture containing 5 ng ^{32}P end-labeled *pap* DNA probe (1×10^5 c.p.m./μl), 2 μg sonicated herring sperm DNA and 5 μg acetylated bovine serum albumin in DNA binding buffer (40 mM Tris–HCl, pH 7.5, 60 mM KCl, 1 mM dithiothreitol, 0.1 mM EDTA). The binding was allowed to reach equilibrium by incubating for 20 min at room temperature. DNA–protein complexes were treated with 10 μl 1 μg/ml DNase I in 22 mM CaCl and 22 mM MgCl for 1 min. The reaction was stopped by the addition of 50 μl stop buffer containing 50% glycerol and 60 mM EDTA. 10 μl were then loaded on a native high ionic strength polyacrylamide gel to monitor the binding (Ausubel *et al.*, 1994). The remaining sample was precipitated by ethanol addition, resuspended in 10 μl formamide loading buffer and analyzed on a 6% polyacrylamide sequencing gel. The quantitation of radioactivity was carried out using a Bio-Rad model GS-250 imager system.

Acknowledgements

We thank Marjan van der Woude and Barbara Graves for helpful discussions, and the Protein–DNA Core Facility, Cancer Center, University of Utah, UT, for the production of the oligonucleotides used in this study. The facility is supported by National Cancer Institute grant no. 5P30-CA42014-07. We are very grateful to Felix Vajdos and Chris Hill at the University of Utah, UT, for supplying highly purified Lrp, and to Brian Ernsting and Rowena Matthews for supplying the Lrp-overproducing strain used for purification. This work was supported by National Institutes of Health grant No. 2R01 AI23348 to D.A.L.

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1994) *Current Protocols in Molecular Biology*. John Wiley and Sons, New York, NY.

Blyn,L.B., Braaten,B.A., White-Ziegler,C.A., Rolfson,D.H. and Low,D.A. (1989) Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: evidence for transcriptional regulation. *EMBO J.*, **8**, 613–620.

Blyn,L.B., Braaten,B.A. and Low,D.A. (1990) Regulation of *pap* pilin

phase variation by a mechanism involving differential *dam* methylation states. *EMBO J.*, **9**, 4045–4054.

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

Braaten,B.A., Platko,J.V., Woude,M.W.v.d., Simons,B.H., Graaf,F.K.d., 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*. *Proc. Natl Acad. Sci. USA*, **89**, 4250–4254.

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

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

Casadaban,M. (1976) Transposition and fusion of the *lac* genes to selected promoters in *E. coli* using bacteriophage lambda and Mu. *J. Mol. Biol.*, **104**, 541–555.

Hochschild,A. (1991) Detecting cooperative protein–DNA interactions and DNA loop formation by footprinting. *Methods Enzymol.*, **208**, 343–364.

Kaltenbach,L.S., Braaten,B.A. and Low,D.A. (1995) Specific binding of PapI to Lrp-*pap* DNA complexes. *J. Bacteriol.*, in press.

Kunkel,T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.

Marinus,M.G., Poteete,A. and Arraj,J.A. (1984) Correlation of DNA adenine methlyase activity with spontaneous mutability in *Escherichia coli* K-12. *Gene*, **28**, 123–125.

Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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

Platko,J.V. and Calvo,J.M. (1993) Mutations affecting the ability of *Escherichia coli* Lrp to bind DNA, activate transcription, or respond to leucine. *J. Bacteriol.*, **175**, 1110–1117.

Silhavy,T.J., Berman,M.L. and Enquist,L.W. (1984) *Experiments with Gene Fusions*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Simons,R.W., Houman,F. and Kleckner,N. (1987) Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene*, **53**, 85–96.

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

van der Woude,M.W., Kaltenbach,L.S. and Low,D.A. (1995) Leucine-responsive regulatory protein plays dual roles as both an activator and a repressor of the *E. coli pap* operon. *Mol. Microbiol.*, **17**, 303–312.

Wang,Q. and Calvo,J.M. (1993) Lrp, a global regulatory protein of *Escherichia coli*, binds co-operatively to multiple sites and activates transcription of *ilvIH*. *J. Mol. Biol.*, **229**, 306–318.

Wissman,A. and Hillen,W. (1991) DNA contacts probed by modification protection and interference studies. *Methods Enzymol.*, **208**, 365–379.

Received on June 2, 1995; revised on September 7, 1995

CHAPTER V

SUMMARY AND DISCUSSION

Previous studies by Blyn *et al.* (10) have established that Pap pili phase variation is controlled at the transcriptional level. Unlike other well-known phase variation systems in which phase variation is accompanied by DNA rearrangement or alteration (1, 42, 56, 71), Pap pili phase variation does not involve changes in the primary DNA sequence. Instead, differential methylation by deoxyadenosine methylase (Dam) at two GATC sites, located strategically in two inverted repeats in the *pap* regulatory region, plays a very important role in determining the expression states of Pap pili (9). Therefore, Pap pili phase variation is controlled by an epigenetic mechanism that involves DNA methylation. A chromosome encoded protein, Mbf, which recently has been found to be identical to the global regulatory protein Lrp (15), and a *pap* encoded local regulatory protein, PapI, are required for such differential methylation, as well as for the expression of Pap pili (13). Whereas methylation protection at the GATC site proximal to the *papBA* promoter (denoted GATC-II or $GATC_{1130}$) requires Lrp, methylation protection at the distal site (denoted GATC-I or $GATC_{1028}$) requires both Lrp and PapI (13). It has been proposed that binding of Lrp to the GATC-II region inhibits the transcription of *papBA*, leading to the phase OFF phenotype. Binding of Lrp/PapI to the GATC-I region activates the transcription of *papBA*, leading to the phase ON phenotype (8).

Evidence suggests that the binding of Lrp or Lrp/PapI complex to the GATC sites blocks DNA methylation. Conversely, methylation of GATC sites by Dam methylase can also block or significantly reduce the affinity of the binding by Lrp or Lrp/PapI complex to that region. The interplay between the differential DNA methylation and the differential protein binding is a key feature of this complex phase variation process.

The Role of Lrp in Pap Pili Phase Variation

Lrp as Both Negative and Positive Regulator of *papBA*

Lrp was originally identified in this laboratory as Mbf (**M**ethylation **B**locking **F**actor) (13). Mutations in *mbf* (*lrp*) abolished *pap* transcription and Pap pili phase

variation. Subsequent analysis showed that both GATC sites in the *pap* regulatory region were methylated in *lrp* mutants, suggesting that Lrp might regulate *pap* transcription by binding to and affecting the conformation of the *pap* regulatory region (13).

Lrp not only positively or negatively regulates different operons, but it can also either positively or negatively regulate the same operon, *papBA*, depending on the location it binds to *pap* DNA. The *papBA* promoter has a basal transcription activity independent of Lrp. However, this intrinsic activity is repressed in the presence of either H-NS or Lrp (68). It has been proposed that Lrp binds to the promoter region to sterically interfere with the binding of RNA polymerase. This repression results in a phase OFF phenotype. The basal transcription from the *papBA* promoter can also be stimulated eightfold in the presence of Lrp and PapI. This stimulation cannot be achieved in the absence of Lrp (68), though the need for PapI can be circumvented in some special circumstances, indicating that Lrp is critically involved in the transcriptional activation of *papBA* operon (44). Therefore, Lrp plays dual roles in repressing and activating the transcription from the *papBA* promoter. The switch of Lrp functions requires that Lrp binds to different locations in the *pap* regulatory region. When Lrp binds to the promoter region, it represses *papBA* transcription. When Lrp binds to an upstream region, it activates *papBA* transcription (44).

Specificity of Lrp Interaction with *pap* Regulatory DNA

As a global regulatory protein that affects the expression of dozens of operons (19), Lrp could regulate the transcription of *papBA* operon directly or indirectly. Our experiments show that Lrp exerts its regulatory function on *papBA* expression by directly binding to the intercistronic regulatory region of *papBA* operons.

Lrp indeed binds to the regulatory region of *pap* DNA specifically. This has been shown by mobility shift experiments using a fragment of DNA containing the regulatory region of *pap* and crude cell extracts from wild-type *Escherichia coli* K-12 strain and congenic *lrp* mutant strains, as well as using purified Lrp (44, 45). Whereas *pap* DNA fragments from the intercistronic regulatory region are retarded in a dose responsive manner

by cell extract containing Lrp or purified Lrp, but not by cell extract from *lrp* mutants, DNA fragments from outside of the *pap* intercistronic region are not retarded by the presence of Lrp, as evidenced by the absence of mobility shift on native polyacylamide gels.

Lrp can also bind to DNA nonspecifically at high concentration *in vitro,* consistent with the hypothesis that Lrp may have a role in organizing the chromosome (21). However, nonspecific binding by a protein is efficiently eliminated by the presence of excessive amounts of nonspecific DNA, whereas the specific binding is competable only by DNA containing the specific binding sites. The mobility shift resulting from Lrp binding to *pap* DNA can be eliminated only by competing with an excess amount of unlabeled *pap* DNA. A 5000-fold excess of sonicated herring sperm DNA does not compete efficiently with Lrp binding, indicating that the observed binding is specific (45).

The specificity of Lrp-*pap* DNA interactions is also directly demonstrated by DNase I footprinting experiments (44, 45). At high binding stringency or low Lrp concentrations, a region of about 120 bp upstream of the *papBA* promoter region is footprinted. This footprinted region contains GATC-II, but not GATC-I. At low binding stringency and high concentrations of Lrp, an additional 60 bp further upstream is also footprinted, which includes GATC-I. However, the DNA sequences outside of the *pap* regulatory region are not footprinted under any experimental conditions. Therefore, our mobility shifting and footprinting experiments show that Lrp binds specifically to the *pap* regulatory region.

Factors Affecting Lrp Binding to *pap* Regulatory DNA

PapI

The specificity of Lrp binding to the *pap* DNA regulatory region is controlled by PapI (44, 45). In the mobility shift experiments, a supershift is observed when purified PapI or a crude cell extract from a strain producing PapI, but not from a strain containing vector alone, is added along with Lrp to a DNA fragment containing the *pap* regulatory region. However, purified PapI or crude cell extract containing PapI does not shift the DNA fragment containing the *pap* regulatory region. These data indicate that PapI alone does not

interact with the *pap* regulatory region. PapI interacts with the *pap* regulatory region only when Lrp is present. Recent evidence indicates that PapI binds to the Lrp moiety of the Lrp-*pap* DNA complex (35). It is also possible that PapI interacts with *pap* DNA sequences in the complex, though there has been no evidence supporting such a hypothesis.

DNase I footprinting experiments show that PapI affects the specificity of Lrp binding to the *pap* regulatory DNA. As expected, PapI alone does not footprint the *pap* regulatory region, indicating that no direct contact between PapI and the *pap* regulatory region occurs. However, PapI alters the Lrp footprinting pattern in the *pap* regulatory region. In the absence of PapI, Lrp preferentially footprints a region proximal to the *papBA* promoter, which contains the GATC-II site. In the presence of both PapI and Lrp, the region distal to the *papBA* promoter containing the GATC-I site is preferentially footprinted (44). This altered footprinting pattern should be due to the enhanced Lrp binding affinity to the GATC-I region by PapI, rather than to PapI binding directly to the DNA, because the footprinting pattern at the GATC-I region is identical to that observed when a high concentration of Lrp alone is used (44).

DNA methylation

DNA methylation is another factor that affects the specificity of Lrp binding to the *pap* regulatory region. Interestingly, the methylation states of the DNA fragment containing the *pap* regulatory region does not seem to change the pattern of mobility shift caused by Lrp or Lrp/PapI binding. The same shifting pattern is observed when fully methylated, fully nonmethylated, or hemimethylated DNA fragments containing the entire *pap* regulatory region are used (45). However, DNA methylation does reduce the affinity of Lrp for individual sites, as shown by the reduced binding of Lrp to a *pap* DNA fragment containing either the promoter proximal or the promoter distal region (halves of the entire Lrp binding region) rather than the whole regulatory region of *pap* DNA when the DNA is methylated (43).

Lrp still binds to the GATC-II region when the two GATC sites in the *pap* regulatory region are methylated. However, the effect of PapI on Lrp binding to the GATC-I region is not observed under such conditions (45). It seems that the binding by Lrp/PapI to the GATC-I region is more sensitive to methylation than the binding of Lrp alone to the GATC-II region. This is consistent with the finding that overexpression of Dam methylase inhibits *papBA* transcription (9). When the GATC sites are hemimethylated, Lrp/PapI can bind to the GATC-I region, but the affinity is reduced (45). This is consistent with our model that DNA replication, which results in transient hemimethylation at the GATC sites, is required for Pap pili phase variation (9).

Leucine

The activity of Lrp is often regulated by leucine (19). Binding of Lrp to *ilvIH* regulatory DNA is reduced by leucine (54). In contrast, leucine does not appear to affect Lrp binding to *pap* regulatory DNA (18). These results are consistent with the observation that the *ilvIH* operon is leucine responsive whereas *papBA* is not (19).

Lrp Binding Consensus Sequences

A common feature of Lrp binding to DNA is that a long DNA region, generally more than a hundred base pairs, is protected from DNase I cleavage and other footprinting agents (26, 44, 45, 72). Periodic hypersensitive bands corresponding to the DNA helical turns are found in the footprints, reflecting the curvature of the DNA resulting from backbone strain (30, 31). This phenomenon makes the DNase I footprinting data of little significance in determining the binding sites of the protein, because it is impossible to distinguish the bases protected by direct protein-DNA contact from those protected indirectly by the formation of nucleoprotein complexes.

Using DMS (Dimethyl sulfate), which modifies "G" residues in DNA molecules and "A" residues at a lower efficiency, six potential Lrp binding sites were identified in the regulatory region of *pap* DNA, of which two are overlapping with GATC-I and GATC-II

(44). A common feature of these binding sites is that they all contain a sequence of Gnn(n)TTTT (except site 4, which is GGacTTTc). In all cases, the consensus guanine was protected from modification by DMS. This sequence feature positions the "G" residue directly facing an "A" track across the major groove. However, not all sequences with such a feature are protected from DNase I cleavage. Therefore, whether this feature is important for Lrp binding remains to be determined. Another interesting feature is that the potential Lrp binding sites are all evenly spaced, with 32 or 33 bp, or roughly three helical turns, spanning between the centers of the "T" tracks in any two adjacent Lrp binding sites. However, the orientations of the binding sites are not the same. Whereas binding site 1, 5, and 6 have an "A" track on the bottom strand, sites 2, 3, and 4 have an "A" track on the top strand. Therefore, Lrp would bind to different sides of the DNA molecules at these binding sites.

Such sequence features are substantially different from those proposed as Lrp binding consensus sequences by either Wang *et al.* (AgaATTTTATtcT) (74) or by Rex *et al.* (TTTATTCtnaAT) (53), though a "T:A" track is common among all of the proposed consensus sequences. In an attempt to define a common feature for Lrp binding, the footprints in the regulatory regions of *papBA, ilvIH, lysU and ompC* (23, 26, 44, 72) were compared (Figure 5-1). The transcription of *papBA* and *ilvIH* is positively regulated by Lrp and that of *lysU* and *ompC* is negatively regulated (19). The four Lrp bound sequences are lined up by the transcription start point. First, it is evident that Lrp binds to different locations relative to the transcription start point, though in most cases Lrp is found bound to an extended region upstream of the promoter. This is consistent with the diverse function of Lrp (19). Second, no significant consensus nucleotide sequence is found among the footprinted regions. The consensus sequences proposed by Wang *et al.* (74) and Rex *et al.* (53) are not present in all regulatory regions of the four operons. Such sequences, when present, are not always protected in footprinting assays. The regulatory region of *lysU* contains a sequence matching the proposed consensus sequence TTTATTCtnaAT, but

Figue 5-1. Comparison of Lrp footprints in *pap, ilvIH, lysU* and *ompC* regulatory regions. The DNA sequences are lined up by the transcription star point. The -10 and -35 sequences of the promoters are underlined. The footprinted sequences are boxed. Only the top strand is shown.

ompC ATCCGGTTGA AATAGGGGTA AACAGACATT CAGAAATGAA TGACGGTAAT
lysU GGCACAGCAG ACGACAGAGT CGCAAATTAG CGGTATGGCA ATTGCAGCTT
ilvIH TGGATGGAAG AGCAATTAGT CTCAATTTGC AAACGCTAAC TGATTGCAGA
papBA AATAATAAAA ATCATGCTCT CTGTTATCAA CGGAAAGGTA TTTTTATTCT

AAATAAAGTT AATGATGATA GCGGGAGTTA TTCTAGTTGC GAGTGAAGGT TTTGTTTTGA
ACCAGCGATT CTTTTCTCAG ATGGGAGAGT GGACGTTGGC TTGTGTCGCG ATCATCGTCG
ATAGGTCAGA CATGAATGTC TGGTTTATTC TGCATTTTTT ATTGAATGTA GAATTTTATT
CTATGTTTGC TTTATTTGTT CAATTTAGTG AATTTGCTTT TTATTGGATT TATTTGATGT

CATTCAGTGC TGTCAAATAC TTAAGAATAA GTTATTGATT TTAACCTTGA ATTATTATTG
TATTGGCCTT TGCTACCCGT TTTCTGTTTA GCACGCCGAC GAATATGATA CAGGAGAGCA
CTGAATGTGT GGGCTCTCTA TTTTAGGATT AATTAAAAAA ATAGAGAAAT TGCTGTAAGT
GTATCACATT TTGCGTTTTA TTTTTCTGCG AAAAGAAAGT CCGTAAAAAT TCATTTAGAC

CTTGATGTTA GGTGCTTATT TCGCCATTCC GCAATAATCT TAAAAAGTTC CCTTGCATTT
ACGATTAATG TCGTAAAAAC AATTGGTTAT GAATAAATTA ACTTTCAGTT TATAACACTA
TGTGGGATTC AGCGGATTTA TTATCAATTT AATCCTCTGT AATGGAGGAT TTTATCGTTT
GATCTTTTAT GCTGTAAATT CAATTTGCCA TGATGTTTTT ATCTGAGTAC CCTCTTGCTA

ACATTTTGAA ACATCTATAG CGATAAATGA AACATCTTAA AAGTTTTAGT ATCATATTCG
ACAAATCGAG ATTAATAAGT CATTTAAACC ATTTTGATGG TTATTTATTA GTGATATCAA
CTTTTCACCT TTCCTCCTGT TTATTCTTAT TACCCCGTGT TTATGTCTCT GGCTGCCAAT
TTAGTGTTTT GTTCTAGTTT AATTTTGTTT TGTGGGTTAA AAGATCGTTT AAATCAATAT

TGTTGGATTA TTCTGCATTT TTGGGGACAA TGGACTTGCC GACTGATTAA TGAGGGTTAA
CTTGAGGTAA GCGTTAGTTT CGATAAGATA AACTGAGTTA CTAATAGTCG AGGCAGATAA
TGCTTAAGCA AGATCGGACG GTTAATGTGT TTTACACATT TTTTCCGTCA AACAGTGAGG
TTACAACATA AAAAACTAAA TTTAACTTAT TGCGTGAAGA GTATTTCCGG GCCGGAAGCA

-35 **-10** **+1**

TCAGTATGCA GTGGCATAAA AAAGCAAATA AAGGCATATA ACAGAGGGTT AATAACATGA
TACAGTGTAC CGATCTGTCT CTTTATCTAC GCTAAATTGA AAGCTGGATT TAGAGGAACC
CAGGCCATGG AGATGTTGTC TGGAGCCGAG
TATATCCAGG GGCCCGACAG AAGGGGGAAA CATGGCGCAT CATGAAGTCA TCAGTCGGTC

site-directed mutagenesis of this sequence had little effect on either Lrp binding or *lysU* transcription (26).

However, in all cases studied so far, Lrp is found bound to an extended region over 100 bp which is generally AT rich. In most cases, periodic AT base pairs corresponding to the DNA helical turns are preferentially protected, suggesting a possible role of DNA conformation in recognition by Lrp. The current data do not allow a conclusion as to whether Lrp binds to the major or minor groove of the DNA. In the case of *pap*, protection across the major groove is most evident (44), which is consistent with the Dam methylation-sensitive nature of Lrp binding to *pap* DNA (45), as the methyl group on the N^6 position of adenine is exposed in the major groove. The presence of a helix-turn-helix motif in Lrp (19, 51) also suggests that Lrp binds to the major groove, since the HTH motif is involved in the specific interactions with nucleotide chemical groups exposed in the major groove (48). However, the Lrp HTH is less than typical. The recognition helix matches very poorly. In the case of *lysU*, however, it appears that Lrp binds across the minor groove (26).

In conclusion, it is not yet clear what sequences Lrp recognizes based on the comparison of the regulatory regions of four operons to which footprint analyses have been performed. It seems reasonable to suggest that Lrp differs from the classic transcription regulatory proteins which recognize their cognate DNAs by specific nucleotide sequence. Instead, the mode of Lrp binding to DNA shows substantial similarity with that of the so-called type II DNA binding proteins, such as the HU protein of *E. coli* and the TF1 protein of *B. subtilis* phage SPO1, which recognize the targets primarily through the three-dimensional structure of the DNA (52, 63). It is possible that Lrp recognizes not only a "consensus sequence" but also a certain subtle three-dimensional feature of the DNA, such as periodically spaced "AT" tracks, which affect the conformation of the DNA molecules by causing compression in the minor groove (29, 65). Such sequence features exist in the *pap* and *lysU* regulatory regions. To this context, it is noteworthy that in both the *lysU* and *pap* regulatory regions some single base pair deletions affect transcription and Lrp binding more

severely than base pair substitutions (26, 38). This hypothesis is consistent with the finding that DNA binding proteins can recognize the target DNA both by specific interaction between complementing side chains of amino acids and nucleotide bases and by recognizing the overall conformation of the DNA (65). Thus, in some cases, Lrp could recognize a particular sequence of DNA when the specific hydrogen bonding was favored. In some other cases, the overall conformation of the DNA could play a major role in binding of Lrp.

The chromatin-associated protein H-NS (67) may provide an example of a DNA binding protein recognizing DNA in multiple ways. Like Lrp, H-NS also shows an ambiguity regarding the "consensus sequence" of DNA binding. On the one hand, H-NS exhibits sequence-specific binding to DNA fragments with a loosely defined consensus sequence, which is generally AT rich (55). On the other hand, Lrp binds avidly to DNA molecules with sharp static bends introduced by phased A:T tracks (47, 77), though in some cases the curvature of the DNA is not essential for high affinity binding by H-NS (39).

The *Bacillus subtilis* AbrB protein is another example of a DNA binding protein recognizing certain three-dimensional features of the target DNA. AbrB is a pleiotropic regulatory protein that regulates the sporulation and gene expression during the post-exponential growth stage (62). Like Lrp, AbrB binds specifically to the regulatory regions of the operons under its control and protects an extended DNA region from attack by a variety of footprinting agents (63). However, the HTH motif of the protein is dispensable for normal DNA binding. Footprinting analyses of at least 16 target DNA sequences have failed to reveal any candidate sequence as a "consensus sequence" for AbrB binding. It was proposed that AbrB recognizes a subtle three-dimensional feature of the target DNA that can arise from a finite subset of different nucleotide sequences (62, 63). Sequence-directed statically bent DNA is one of the candidates for such three-dimensional features. However, an analysis of six AbrB binding regions revealed that only three of them had detectable intrinsic bending (61). Those observations are consistent with the hypothesis that a DNA protein can recognize its target DNA by different means.

To understand the true feature that Lrp recognizes when it binds to DNA, more mutational analyses are required. It is important to establish whether the spacing between the potential Lrp binding sites is vital to the binding of Lrp. To achieve this, different numbers of bases should be inserted or deleted between the potential Lrp binding sites and the phenotype of the mutants should be assessed both in term of *pap* transcription and Lrp binding patterns to the *pap* regulatory region. However, as CAP is sensitive to the phasing of its binding site to the promoter region, proper compensating deletion or insertion mutations should be constructed accordingly so that the CAP phasing is not changed.

Lrp Binding Patterns

Six potential Lrp binding sites in the *pap* regulatory region, each separated by 32-33 bp or roughly three helical turns, have been identified. Under higher stringency or at lower Lrp levels, the three sites proximal to the *papBA* promoter (sites 1, 2, and 3), including the one overlapping the GATC-II sequence (site 2), are cooperatively bound by Lrp. The three sites distal to the *papBA* promoter (sites 4, 5, and 6), including the one overlapping with GATC-I (site 5), are bound by Lrp at high concentration or low stringency (44). These observations indicate that sites 1, 2, and 3 are the primary binding sites of Lrp, because Lrp has the highest affinity to these sites. However, in the presence of PapI, the two sites distal to the *papBA* promoter (site 4 and 5) are bound with highest affinity, followed by sites proximal to the *papBA* promoter (sites 1, 2, and 3) (44). These data indicate that sites 4 and 5 are the primary binding sites of Lrp in the presence of PapI. The implication of the above observation is that at low Lrp levels, which are more physiologically relevant, Lrp binds to different sets of binding sites depending on the presence or absence of PapI. Therefore, the availability of PapI may determine the transcription states *in vivo*.

Site-directed mutagenesis showed that the *in vitro* Lrp binding pattern is correlated with the *in vivo* transcriptional activity. Interruption of Lrp binding site 2 or 3 results in either locked ON or enhanced phase OFF to phase ON switching rates (44). When the phenotype is phase-locked ON, such as in the mutants with mutations in Lrp binding site 3

(mutants 13, 16 and 17), DNase I footprinting showed that Lrp binds with high affinity to sites 5 and 6 independent of PapI (43, 44). Southern blot analysis also showed that GATC-I is protected from methylation and GATC-II is completely methylated (43). In contrast, mutations in Lrp binding sites 4 and 5 result in locked OFF phenotypes. DNase I footprint analysis showed that Lrp no longer binds to GATC-I (site 5), even in the presence of PapI (44). This is consistent with the Southern blot analysis data showing that Lrp binds to the GATC-II site in the presence of PapI (43).

We do not known why Lrp preferentially binds to the GATC-II region and Lrp/PapI to the GATC-I region, even though the two GATC sites are encompassed by nearly perfect inverted repeats of 27 nucleotides (9). This differential binding should not be caused by the slight difference of the primary sequences in the two inverted repeats, because a single base pair substitution outside the inverted repeats (G1163) allows Lrp to bind preferentially to the GATC-I region in the absence of PapI (43). It is more likely that the preferential binding by Lrp to the GATC-II region results from the cooperativity of binding to sites 1, 2, and 3. Moreover, this differential binding is not affected by Lrp overproduction *in vivo*. An eightfold overproduction of Lrp does not facilitate Lrp binding to the GATC-I region in the absence of PapI *in vivo* (43), though Lrp alone binds to both GATC-I and GATC-II regions *in vitro* when the concentration increases (44). It seems that there is a mechanism which prevents Lrp or Lrp/PapI from binding to both the GATC-I and the GATC-II regions. This mechanism would involve a conformational change of the *pap* regulatory DNA caused by Lrp or Lrp/PapI binding to one GATC region which precludes the binding to the other GATC region.

Stoichiometry and Cooperativity of Lrp Binding to *pap* DNA.

The stoichiometry of Lrp binding to the *pap* regulatory region has not been well studied. In the case of *ilvIH*, Lrp reportedly binds as a dimer to each binding site (19). Though this is also likely to be the case with *pap* DNA, there is currently no evidence to support or argue against this hypothesis. Nevertheless, there seems to be a fixed number of

Lrp molecules binding to the *pap* regulatory region, as the binding reaction is saturated at a relatively low concentration of Lrp, as evidenced by the appearance of a single dominant shifted band in a mobility shift assay. This is in sharp contrast to the DNA binding by the protein H-NS, in which the binding reaction is saturated until the entire DNA molecule is bound by H-NS (67).

Lrp has been shown to bind cooperatively to two binding sites in the *ilvIH* regulatory region at lower concentration and to another three at higher concentration (72). This cooperative binding may play an important role in the formation of nucleoprotein structure. Cooperative binding of Lrp to the *pap* regulatory region has not been specifically tested. However, based on the observations from other experiments, we infer that Lrp binding to the *pap* regulatory region is cooperative.

First, Lrp binding affinity to a truncated *pap* DNA fragment containing Lrp binding sites 2 and 3 or sites 4 and 5 is significantly reduced, probably due to the lack of a cooperative binding site (43). Second, when the *pap* DNA fragment containing sites 1, 2, and 3 is used in a mobility shift experiment, usually only a single shifted species is observed; intermediate bands, which would represent intermediate complexes, are not observed (43, 44). Third, footprinting experiments indicate that even at very low concentration of Lrp, sites 1, 2, and 3 are simultaneously occupied by Lrp (44). Fourth, a mutation interrupting Lrp binding site 3 also abolishes the binding of Lrp to site 2 (44).

Lrp Binding and the Conformation Change of *pap* DNA

DNA bending or curvature is a common theme in regulating prokaryotic gene transcription. Bacterial promoters often contain static bends, and bacterial RNA polymerase is known to further bend DNA at the promoter region (50). Intrinsically bent or curved DNA upstream of the promoter region in the right orientation generally stimulates gene transcription. Binding of transcription regulatory proteins to the cognate DNA regulatory sequences often induce bending.

The best studied transcription regulatory proteins that bend DNA are IHF and CAP. IHF binds to the minor grove and bends DNA as much as 140° (64). IHF positively or negatively regulates the transcription of numerous genes, either directly or in concert with other transcription regulatory proteins . The hypothesis that IHF regulates gene transcription by bending DNA is supported by the finding that DNA sequences causing equivalent static bends can functionally substitute for the authentic IHF binding site (27). Similarly CAP also binds DNA and causes a bending of at least 90° (58). The function of CAP requires that the binding sites be located on the same side of RNA polymerase binding site. In some cases the cAMP-CAP binding site can be functionally substituted by the binding site of other DNA bending proteins or even by sequences directing natural bent (16, 49), indicating DNA bending can be sufficient for transcription activation. In some other cases direct interaction between CAP and RNA polymerase is required for transcription activation (17, 22, 76, 78).

Using different footprinting techniques, it has been shown in the *ilvIH*, *papBA*, *ompC,* and *lysU* systems that the binding of Lrp results in periodic hypersensitive bands dispersed in the footprint in an extended region, indicating that Lrp also bends DNA and forms a nucleoprotein complex upon binding (23, 26, 44, 45, 72). Lrp binding to a single binding site in the *ilvIH* regulatory region has been shown to bend DNA by 55° and by at least 130° when two binding sites are present (73). Though DNA bending has not been shown directly from binding of Lrp to the *pap* regulatory region, the extensive footprinting patterns we observed using DNase I, hydroxyl radical, MPE, and DMS footprinting techniques also suggest that complex nucleoprotein structure is formed upon Lrp or Lrp/PapI binding to the *pap* regulatory region (43, 44, 45). However, it is not known whether the bending of DNA is essential for transcription activation or it is just a consequence of Lrp binding. Because intrinsically bent DNA is known to facilitate gene transcription (50), we believe that DNA bending caused by Lrp or Lrp/PapI binding might play an important role in transcription activation.

There are nine DNA helical turns between the Lrp binding sites encompassing GATC-I and GATC-II, and the two sites are in opposite orientations. Therefore, Lrp binding to the GATC-I region and to the GATC-II region would be on opposite sides of the DNA helix and induce bending to opposite directions. If DNA bending induced by Lrp/PapI binding to the GATC-I region is required for *papBA* transcription activation, this bending must occur in such a way that it either facilitates the assembly of the transcription complex or promotes the interaction of transcription activators with the basal transcription complex. As it is hypothesized that Lrp bends DNA to opposite directions when it binds to GATC-I or GATC-II region, it is possible that Lrp binding to the GATC-II region inhibits *papBA* transcription by repressing the assembly of the transcription complex and/or by repressing the interaction between RNA polymerase and the activator proteins.

Though all of the evidence suggests that a nucleoprotein structure is formed as a consequence of binding of Lrp or Lrp/PapI to *pap* DNA, there is no evidence directly indicating the nature of such complex. Either DNA looping between two protein binding sites or DNA wrapping around protein molecules can result in a higher order nucleoprotein structure (41, 50). However, DNA wrapping around protein molecules would change the supercoiling of DNA in a closed circle or in a constrained region, whereas simple looping between two sites would not lead to such a change. Therefore, testing the supercoiling change of a plasmid containing the *pap* regulatory region as a consequence of Lrp binding may shed light on whether *pap* regulatory DNA wraps around Lrp to form a nucleoprotein structure. Preliminary experimental results suggest that Lrp binding does change the supercoiling of a plasmid carrying the *pap* regulatory region. However, it is not clear whether the change of supercoiling is due to Lrp binding to the *pap* regulatory region or to Lrp nonspecific binding to the whole plasmid DNA (43).

Electron microscopy is a powerful tool to determine the topology of DNA-protein interactions. An electron microscopic examination of the Lrp-*pap* DNA complex should help clarify the nature of the nucleoprotein structure.

Although it is not clear how Lrp binding affects the overall conformation of the *pap* regulatory region, it might be instructive to compare the binding of Lrp and H-NS to the target DNA molecules. Both Lrp and H-NS are pleiotropic global regulatory proteins that positively or negatively regulate a large number of genes; both Lrp and H-NS show sequence specific binding to DNA with a loosely defined consensus sequence whereas the overall conformation of DNA also seems to play a role in the interaction. Both Lrp and H-NS bind to an extended region on the target DNA as judged by the lengthy footprints. H-NS has been shown to constrain supercoils and compact DNA (59, 66), probably by wrapping DNA (67). Therefore, it would not be surprising if Lrp and H-NS affect DNA conformation in a similar way.

Interaction of Lrp with the Transcription Machinery.

Though the binding of Lrp or Lrp/PapI to *pap* regulatory region causes a DNA conformation change that may be essential for transcription, it is not sufficient for transcription activation of *papBA* based on the analysis of *lrp* activation mutants (68). Lrp produced by either the wild-type strain or the *lrp* activation mutant strains binds to the *ilvIH* or the *pap* regulatory region with similar affinity. The *pap* DNA footprinting pattern using the mutant Lrp is similar to that obtained using wild-type Lrp. Nevertheless, transcription activation is not achieved in such mutants. This finding suggests that Lrp interacts directly with another factor in the transcription machinery or another factor involved in transcription activation. It is interesting to note that activation of transcription from the *papBA* promoter requires binding of Lrp or Lrp/PapI to the distal Lrp binding sites (GATC-I region), which is adjacent to the binding site of cAMP-CAP. This observation suggests that there might be a concerted effect of cAMP-CAP and Lrp on RNA polymerase or direct interaction between cAMP-CAP and Lrp.

If the activation mutants are unable to activate *pap* transcription due to the inability to interact properly with the transcription machinery or other factors involved in transcription activation, second site suppressor mutants of the *lrp* activation mutants should prove to be

valuable in understanding how Lrp binding to *pap* DNA activates transcription. If such mutants are found in the genes encoding RNA polymerase subunits or CAP, the possibility of a direct interaction between Lrp and RNA polymerase or CAP should be further investigated.

Conclusions

There are two sets of Lrp binding sites in the *pap* regulatory region. Each set is composed of two or three Lrp binding sites, which are separated by approximately three DNA helical turns. Lrp can specifically and cooperatively bind to either of those two sets of sites, depending on the methylation states of the *pap* regulatory region and the concentration of PapI. Lrp binding to sites proximal to the *papBA* promoter (binding sites 1, 2, and 3) inhibits *pap* transcription, probably by either sterically interfering with the binding of RNA polymerase or bending the DNA so as to inhibit binding of RNA polymerase. Conversely, binding of Lrp to the sites distal to the *papBA* promoter (binding sites 4 and 5) stimulates *pap* transcription. This stimulation involves the formation of complex nucleoprotein structures and the interaction between Lrp and the transcription machinery or other factors involved in transcriptional activation.

The Role of PapI in Pap Pili Phase Variation

PapI Affects *pap* Transcription and Methylation Protection.

PapI is a regulatory protein that is conserved among a number of different pili producing enteric bacteria (70). PapI has been shown to functionally substitute for FaeA in modulating Lrp binding to the *fae* regulatory region and for *fae* transcription (33). Though both PapB and PapI are required for the transcription of the *papBA* operon (3), expression of PapI *in trans* relieves the requirement for PapB (45). Therefore we have proposed that PapI but not PapB is critically involved in the transcription activation of the *papBA* promoter. The transcription from the *papI* promoter, in turn, is dependent on PapB (45).

In contrast to Lrp, which is required for the methylation protection of both GATC-I and GATC-II, PapI is required only for the methylation protection of GATC-I (13). This raises the possibility that PapI might be a DNA binding protein which binds specifically to *pap* DNA in the GATC-I region. Alternatively, PapI may be involved in the methylation protection of GATC-I by affecting the binding of another protein, such as Lrp.

pap DNA-dependent PapI-Lrp Interactions

It has been shown that both Lrp and PapI are required for binding to the GATC-I region (44). The requirement of both Lrp and PapI for binding to the GATC-I region could be due to two possible events. On the one hand, PapI could bind directly to the GATC-I region in the presence of Lrp. On the other hand, the binding specificity of Lrp could be altered by PapI in such a way that the affinity for the GATC-I region is increased and the affinity for the GATC-II region is decreased. There is strong evidence suggesting that Lrp binding specificity is altered in the presence of PapI. First, both the GATC-I and the GATC-II regions are footprinted at high Lrp concentration in the absence of PapI, indicating Lrp alone is able to bind to the GATC-I region. Second, Lrp alone binds to the GATC-I region when the binding sites in the GATC-II region are mutated. These observations indicate that Lrp alone has the potential for binding to the GATC-I region, yet this potential is only maximized in the presence of PapI (44).

The alteration of Lrp binding specificity by PapI could be achieved either by chemical modification of Lrp caused by catalytic reactions of PapI or by a conformational change of Lrp caused by physical binding to PapI. PapI interacts with Lrp very poorly in solution, as evidenced by data from cross-linking and gel filtration chromatography experiments by Kaltenbach and Low (35). Lrp and PapI mixed in solution crosslink very weakly and form separate peaks on gel filtration chromatography. However, in the presence of *pap* regulatory DNA, the efficiency of the interaction is dramatically enhanced, as shown by the strong crosslinking and the unique peak observed following sizing column fractionation. This enhancement is not achieved by adding sonicated herring sperm DNA or

a DNA fragment containing the regulatory region of *ilvIH*, to which Lrp specifically binds. Therefore, this DNA dependent protein-protein interaction between PapI and Lrp is specific to *pap* regulatory DNA (35).

It is not known why the Lrp-PapI interaction is dependent on the presence of DNA of the *pap* regulatory region. We can assume that the PapI interacting domain in Lrp is normally buried and Lrp binding to DNA changes the conformation of Lrp in such a way that the masked PapI-interacting domain is exposed. This assumption is reasonable because DNA binding is known to change protein conformation (46, 75). In some cases, DNA binding may cause the refolding of the entire protein (2). However, for this assumption to be acceptable, we must assume again that the Lrp conformation change caused by binding to *pap* DNA is unique, since Lrp binding to *ilvIH* regulatory DNA does not facilitate Lrp-PapI interaction (35).

Interestingly, the Lrp-PapI interaction requires *pap* DNA containing both the intact GATC-I and GATC-II regions. DNA fragments containing mutated GATC-I or GATC-II are not efficient in stimulating Lrp-PapI interactions, suggesting that the Lrp-PapI interactions may involve both GATC boxes (35). Consistent with this hypothesis, PapI is found in the Lrp-*pap* DNA complex only when both GATC sites are occupied by Lrp (35).

Regardless of the mechanism, *pap* DNA dependent Lrp-PapI interaction is advantageous for the bacterial cells. First, Lrp is an important transcription regulator that participates in a variety of cellular activities. This specific DNA-dependent interaction with another protein would prevent an operon specific regulator from interfering with Lrp regulating other operons. By this mechanism, only Lrp molecules engaged in regulating *papBA* can bind to PapI. Second, this mechanism would also prevent PapI from being diluted out by Lrp.

PapI-independent *pap* Transcription.

PapI is absolutely required in the wild-type background for *pap* transcription activation (3, 45). However, in some of the phase locked ON mutants, such as the ones

with mutated Lrp binding site 3 (mutants 13, 16, and 17) or with a single base change in GATC-I which prevent methylation (mutant 1) , PapI is no longer required for transcription, though Lrp is still required (43, 44). DNase I footprint analysis showed that Lrp alone binds to the GATC-I region with highest affinity in such mutants (44). However, Southern blot analysis showed that both GATC sites in the *pap* regulatory are methylated in the absence of PapI, even in the PapI-independent mutants (43). We cannot explain why GATC-I is methylated whereas the transcription is active in these mutants. We believe that Lrp still binds to the GATC-I region because Lrp is still required for transcription (44). Perhaps Lrp binding to the GATC-I region in the absence of PapI is not strong enough to protect the GATC site from being methylated. This point can be clarified by *in vivo* footprinting experiments. Nevertheless, the PapI independent transcription suggests that PapI affects the function of Lrp in such a way that the modification caused by interacting with PapI can be compensated for by preventing Lrp binding to the GATC-II region. The finding of PapI-independent mutants also suggests that PapI is not directly involved in *pap* transcription activation.

Translocation of Lrp.

As discussed above, Lrp binds with highest affinity to the GATC-II region in the absence of PapI, blocking the transcription from the *papBA* promoter. In the presence of PapI, Lrp binds with highest affinity to the GATC-I region, activating *papBA* transcription (44). Therefore, PapI plays a primary role in the translocation of Lrp from the GATC-II region to the GATC-I region. PapI could act by stabilizing Lrp binding to the GATC-I region and/or destabilizing Lrp binding to the GATC-II region. Currently it is not clear how such translocation is achieved. Stable binding of PapI to Lrp may not be required for such translocation, because PapI is not present in the Lrp-*pap* DNA complex until both GATC regions are occupied by Lrp *in vitro* (35).

Recently, Lrp mutants unable to activate *papBA* transcription but still capable of activating *ilvIH* transcription are isolated and are being characterized in this laboratory (34).

Possibly these mutants are unable to interact with PapI properly. It will be interesting to determine whether they can activate *papBA* transcription in the PapI-independent mutants. It is likely that further characterization of these mutants will shed light on the understanding of the mechanism by which PapI alters Lrp binding specificity.

A current hypothesis is that PapI forms an unstable complex with Lrp bound to *pap* DNA in the GATC-II region. This unstable complex changes the conformation of Lrp in such a way that binding to the GATC-I region is favored. This hypothesis predicts that the free energy for Lrp binding to the GATC-II region is lower than that for binding to the GATC-I region and the conformation change of Lrp caused by the formation of unstable complex with PapI lowers the free energy barrier for binding to the GATC-I region. If this is the case, removal of PapI from the Lrp/PapI complex bound to GATC-I region by adding anti-PapI antibody will result in Lrp binding to the GATC-II region, because Lrp will revert back to the conformation which favors binding to the GATC-II region. If instead there is a chemical modification to Lrp by PapI, removal of PapI would not revert the presumed Lrp conformation change and the altered binding specificity. This can be tested by DNase I footprinting of Lrp/PapI-*pap* DNA complexes prior to and after PapI antibody treatment.

Meanwhile, this hypothesis predicts that Pap pili phase variation involves Lrp present in at least three distinct conformations: unengaged Lrp, Lrp bound to the GATC-II region of the *pap* DNA (Lrp-II), and Lrp bound to the GATC-I region of the *pap* DNA (Lrp-I), with or without PapI being in the complex. Transcriptional activation of *pap* would require the sequential Lrp to Lrp-II to Lrp-I conformation change. If this is true, it should be possible to isolate *lrp* mutants that activate *papBA* transcription independent of PapI (locked on Lrp-I conformation) and that interact with PapI independent of *pap* regulatory DNA (locked on Lrp-II conformation).

Conclusions

PapI does not bind to the *pap* regulatory DNA directly and does not interact with the transcription machinery. PapI exerts its role as a transcription activator of *pap* by interacting

with Lrp in Lrp-*pap* DNA complexes and possibly altering the conformation of Lrp so that Lrp/PapI complexes binds preferentially to the GATC-I region, which is required for *pap* transcription activation.

The Role of DNA Methylation in Pap Pili Phase Variation

DNA Methylation Patterns and *pap* Transcription

DNA methylation by Dam methylase has been shown to affect transcription and the binding of transcription regulatory proteins (5, 11, 60). Generally DNA methylation reduces the binding of transcription regulatory proteins, probably by steric hindrance. Therefore DNA methylase can either stimulate or inhibit gene transcription by inhibiting the binding of repressor or activator proteins.

A fundamental difference between the role of DNA methylation in regulating the expression of *pap* versus other documented systems is the presence of methylation patterns in the case of *pap*. On the one hand, the GATC sites affecting the binding of other regulatory proteins are generally methylated or hemimethylated, corresponding to the DNA replication cycle (40). On the other hand, the two GATC sites in the *pap* regulatory region are always differentially methylated, resulting in distinct methylation patterns corresponding to the expression states of the *papBA* operon (13, 14). The differential DNA methylation pattern is established and maintained by the differential binding of the regulatory protein Lrp. Another aspect of the *pap* methylation pattern is that it also dictates the location of Lrp binding, which determines the expression state of the *papBA* operon. Therefore, the role of DNA methylation in *pap* transcriptional regulation is not simply blocking the binding of a transcriptional regulatory protein, but it also responds to and determines the binding of Lrp.

The active transcription of *papBA* requires that GATC-I be nonmethylated and GATC-II methylated (13). Such a DNA methylation pattern presumably stabilizes Lrp binding to the GATC-I region but not the GATC-II region. DNA fragments with such differentially methylated GATC sites have not been used in our *in vitro* binding studies.

However, we have constructed mutants that closely mimic the differential methylation *in vivo* and *in vitro* by introducing single base pair substitutions at both GATC sites. When GATC-I is mutated so that only GATC-II can be methylated, the transcription of *pap* is locked ON. Conversely, when GATC-II is mutated so that only GATC-I can be methylated, the transcription of *pap* is locked OFF (14). *In vitro* analysis shows that Lrp/PapI is unable to bind to the GATC-I region when this site is methylated. Lrp binding to the GATC-I region can be observed even in the absence of PapI when only GATC-II is methylated (14). These results indicate that DNA methylation patterns indeed regulate *pap* transcription by controlling the binding of Lrp and Lrp/PapI to *pap* DNA.

Balance between Lrp/PapI and Dam

The function of DNA methylation in regulating Pap pili phase variation depends on the delicate balance of the activities of Dam methylase and the methylation blocking factor, Lrp. The transcription of *papBA* is shut-off in a *dam* background, as a result of the absence of methylation at GATC-II. Overexpression of Dam by as little as fourfold also leads to the shut-off of *papBA* transcription, possibly as a result of methylation of both GATC sites (9, 13).

An interesting observation is that the two GATC sites in the regulatory region are always differentially methylated *in vivo*, unless the *dam* gene is mutated or overexpressed. This observation suggests that there is a delicate balance between Dam methylase and Lrp/PapI which prevents both GATC sites being methylated or protected from methylation. Alternatively, the binding of Lrp or Lrp/PapI to one GATC site changes the DNA conformation in such a way that it precludes the binding to the other GATC site, which is then available for methylation.

An eightfold overproduction of Lrp renders the transcription of *papBA* in a wild-type background locked OFF (43). Southern blot analysis shows that methylation protection is virtually 100% at GATC-II and 0% at GATC-I (43). The same level of overproduction of Lrp does not affect either the *papBA* expression state or the methylation protection pattern in

the locked ON mutants (those having a mutation preventing Lrp binding to GATC-II, mutants 13 and 24) (43). These observations support the hypothesis that binding of Lrp or Lrp/PapI to one GATC site precludes the binding at the other. This hypothesis contradicts what is observed *in vitro* by quantitative DNase I footprinting, which shows that at high levels Lrp binds to both GATC sites (44). However, this *in vitro* observation was made using an unphysiologically high concentration of Lrp: hence it might not reflect the events that occur *in vivo*. Alternatively, weak binding of Lrp around GATC-I or GATC-II might not block methylation at these GATC sites.

Lrp binds preferentially to the GATC-II region, and the transcription of *papBA* is predominantly inactive in a wild-type background, presumably because the PapI level is very low. Expressing PapI *in trans* at a low level increases the fraction of cells that have the phase ON DNA methylation pattern and are transcriptionally active. However, overexpression of PapI at high levels reduces *papBA* transcription in the wild-type background. Subsequent Southern blot analysis revealed both GATC sites are protected from methylation, indicating that Lrp/PapI binds to both sites. Similar overexpression of PapI in the locked ON mutants does not affect either the expression state or the methylation pattern (43). This observation is consistent with the idea that PapI affects the binding of Lrp but contradicts the hypothesis that protein binding to one GATC site precludes binding to the other GATC site.

Methylation Pattern Switch

Pap pili phase variation requires a change in the *pap* DNA methylation patterns. Because no DNA demethylase activity is found in *E. coli*, alteration of DNA methylation must involve DNA replication (13). Immediately following the passage of the replication fork, the original methylated GATC site becomes hemimethylated. This would provide a time window when the switch of methylation patterns could occur. Since DNA methylase preferentially targets hemimethylated GATC sites (40), the hemimethylated GATC site usually will quickly become methylated, preventing the switch of the methylation pattern

and preventing phase variation. However, at low frequencies, Lrp or Lrp/PapI binds to the hemimethylated GATC region, leaving the other GATC site available for methylation, resulting in the switch of the methylation patterns.

Conclusions

DNA methylation patterns in the *pap* regulatory region are the result of competition between Dam methylase and Lrp for binding to the GATC-I and the GATC-II regions. The methylation patterns, in turn, determine the binding of Lrp or Lrp/PapI and the activation of transcription. However, when the binding potential of Lrp to one of the two GATC regions is greatly reduced by mutation, DNA methylation no longer affects transcription activation.

The Roles of Other Regulatory Factors in Pap Pili Phase Variation

H-NS

H-NS is a chromatin-associated global regulatory protein that shows low sequence specificity in binding to numerous binding sites on the chromosome. H-NS generally represses transcription, possible by forming a chromatin-like nucleoprotein complex. However, the transcription of some genes may also be positively regulated by H-NS (67). It has been proposed that H-NS plays a central role in the regulation of *pap* transcription by silencing the *papBA* promoter and that PapB and CAP activate transcription by acting as antirepressors which antagonize the silencing effects of H-NS (25). This conclusion is based on the observation that mutation in the *hns* gene relieves the requirement for PapB and CAP for *papBA* transcription. However, this observation was made using *pap* genes cloned on a multicopy plasmid, which has been show to affect transcriptional regulation (37).

Observations from this laboratory by van der Woude *et al.* (68) do not indicate that H-NS plays a central role in Pap pili phase variation. The *papBA* promoter is intrinsically active. They also have shown that it is this intrinsic basal activity of the promoter that is repressed by H-NS. Changes in the supercoiling of plasmid DNA might be involved in the

repression mediated by H-NS, because H-NS is known to affect the supercoiling of DNA (47, 66). Lrp binding to the GATC-II region has the same repressing effects. Removal of H-NS by introducing *hns* mutant alleles does not affect *pap* transcription and Pap pili phase variation in a single copy wild-type background. Therefore, it was concluded that H-NS is not critically involved in the regulation of transcription from the *papBA* promoter, though H-NS does repress the basal activity of the *papBA* promoter in the absence of Lrp.

PapB

PapB is required for the transcription of *pap* operons. It also represses *papBA* transcription when overexpressed. It has been shown to bind at -240 of *papBA* promoter with high affinity and to the *papBA* promoter region with lower affinity (24). Forsman *et al.* (25) have proposed that PapB, together with CAP, acts as antirepressor to antagonize the silencing effect of H-NS (28).

Our experiments, however, show that PapB is not required for transcription from the *papBA* promoter when PapI is expressed *in trans* (45). In addition, transcription from the *papBA* promoter in the PapI-independent locked ON mutants is not affected by a mutation in the start codon of *papB* (12). Therefore, we believe that the function of PapB is to positively regulate the transcription of *papI* and autogeneously regulate the transcription of *papBA*.

cAMP-CAP

The transcription of *papBA* requires the binding of cAMP-CAP to a region about 210 bp upstream of the *papBA* transcription start site and about 120 bp upstream of the *papI* transcription start site. It has been proposed that CAP, together with PapB, acts as antirepressor to neutralize the silencing effects caused by H-NS (25).

Observations from this laboratory suggest that CAP is directly involved in the transcription activation of *papBA* operon and Pap pili phase variation. When either the *crp* gene or the CAP binding site in the *pap* regulatory DNA is mutated, *papBA* transcription is

shut off, resulting in a phase locked OFF phenotype. This is true even with the PapI-independent locked on mutants (mutants 13 and 17, mutation in Lrp binding site 3). Moreover, a CAP activation mutant , which binds to CAP binding sites normally, also failed to activate *papBA* transcription (38). These results indicate that mere binding by CAP is not sufficient for transcription activation.

We conclude that CAP activates *pap* transcription in a way other than acting as an H-NS antagonist. CAP interaction with RNA polymerase or other components of the transcription machinery might be essential for *papBA* transcription activation. Judging by the location of the CAP binding site and the requirement of other regulators (Lrp and PapI) for transcription activation, the *papBA* promoter should be classified as a class III CAP-dependent promoter (17, 22). In this case, the interaction between CAP and RNA polymerase requires a mechanism which bring CAP and RNA polymerase into contact with one another. Lrp binding to and bending of *pap* DNA may provide such a mechanism.

A Working Model for Pap Pili Phase Variation

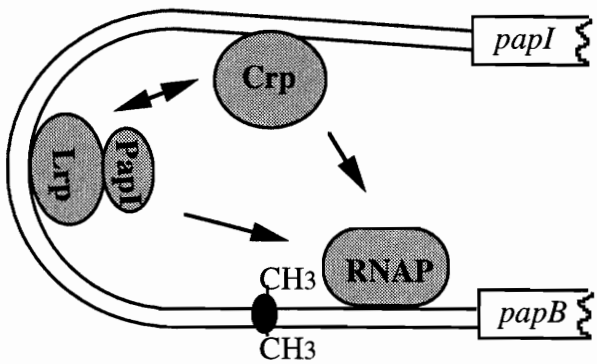
Pap pili phase variation is a complex process involving the coordinated interactions of multiple protein factors with the *pap* regulatory region. The differential methylation of *pap* regulatory DNA is central to this process. The methylation state is determined by the location of binding of Lrp or Lrp/PapI to *pap* DNA, and in turn, it also affects the binding of the regulatory proteins.

Phase OFF State

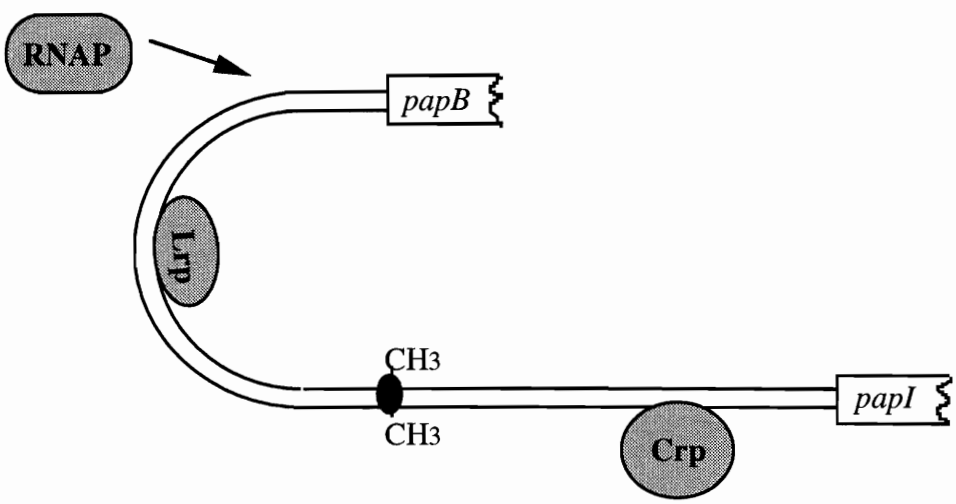
In the phase OFF state, the *pap* regulatory DNA is methylated at GATC-I and Lrp is bound to the GATC-II region, which is partially overlapping the *papBA* promoter region. The binding of Lrp at the promoter region represses the transcription of *papBA* either by spatially interfering with the binding of RNA polymerase or by bending the DNA in the opposite direction that is required for transcription initiation to occur (Figure 5-2 A). The

Figure 5-2. Proposed DNA conformation in phase ON and OFF cells. (A). Phase ON DNA conformation. Binding of Lrp and PapI to GATC-I bends the *pap* DNA in such a way that Lrp and CAP can interact with RNA polymerase and with one another. (B). Phase OFF DNA conformation. Binding of Lrp to GATC-II interferes with the binding of RNA polymerase and Lrp, CAP and RNA polymerase are unable to interact with one another.

A. **Phase ON cell DNA conformation**



B. **Phase OFF cell DNA conformation**



bending in this wrong direction would change the DNA conformation in such a way that the initiation complex cannot be assembled and the required protein-protein interactions of RNA polymerase, Lrp, and CAP are spatially unfavorable.

Phase ON State

In the phase ON state, the *pap* regulatory DNA is methylated at GATC-II, and Lrp is bound to the GATC-I region. It is not yet clear if PapI remains stably bound to the Lrp moiety of the Lrp-*pap* DNA complex. Lrp binding to the GATC-I region also puts Lrp to the vicinity of CAP bound to *pap*. We propose that Lrp or Lrp/PapI binding to the GATC-I region bends the *pap* regulatory region in the direction which enables the proper assembly of the transcription initiation complex at the *papBA* promoter region and enables both Lrp and CAP to interact with RNA polymerase, which may be necessary for the transcription activation (Figure 5-2 B).

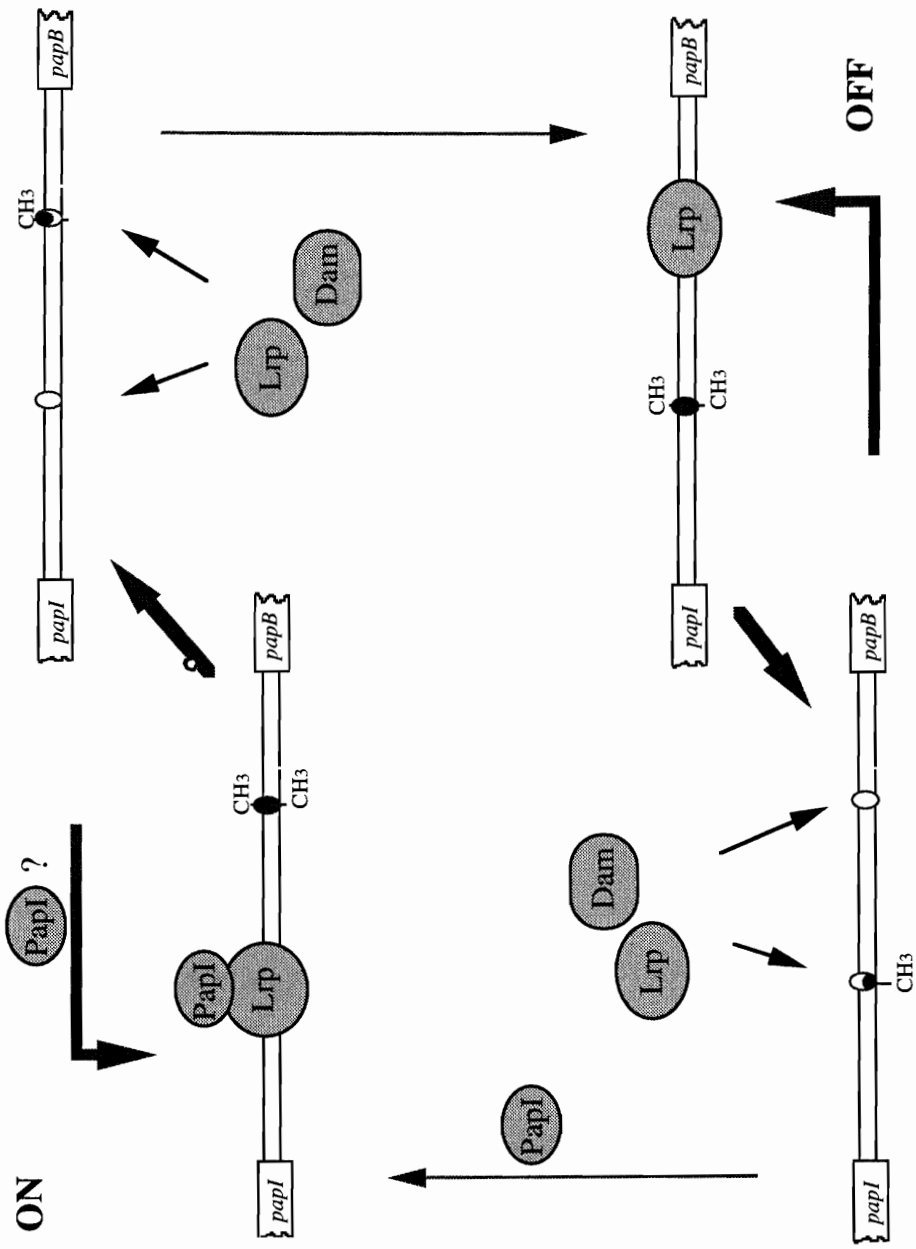
Transition State

For the purpose of discussion, the transition state is defined as a period of time following DNA replication when one of the two GATC sites is hemimethylated and the protein originally bound is displaced by the process of DNA replication. As further discussed below, this is the stage when DNA methylation and protein binding switches between the two GATC sites. PapI is critically involved in this process.

Phase OFF to Phase ON Switch

Figure 5-3 presents a model explaining how the phase of *papBA* expression is epigenetically maintained and how phase variation occurs. After DNA replication, the *pap* regulatory DNA in a phase OFF cell becomes hemimethylated at GATC-I and nonmethylated at GATC-II. As discussed before, since Dam methylase has a higher affinity for hemimethylated GATC sites compaired with nonmethylated sites (40), the hemimethylated GATC-I would be preferentially methylated, rendering it unavailable for Lrp or Lrp/PapI to bind. At the same time, Lrp would bind to the GATC-II region, for

Figure 5-3. Model of Pap pili phase variation.


ON
PapI ?
papI
papB
CH3
CH3
PapI
Lrp
Lrp
Dam
PapI
Dam
Lrp
CH3
CH3
Lrp
papB
papI
OFF
papB
CH3
papI

which Lrp has a higher affinity than the GATC-I region. The methylation of GATC-I and Lrp binding to the GATC-II region restores the original phase OFF conformation. This process provides an epigenetic mechanism that stably maintains the expression phase OFF state.

However, at low frequencies, possibly due to slightly elevated PapI levels or statistical chance, Lrp/PapI binds to the hemimethylated GATC-I region before this site becomes methylated and blocks the access of Dam methylase for further methylation. Binding of Lrp/PapI to the GATC-I region may require that Lrp binds to the GATC-II region first, because the interaction between Lrp and PapI is dependent on the presence of *pap* DNA (35). We propose that PapI interacting with Lrp bound to the *pap* GATC-II region alters the conformation of Lrp either by chemical modification or by physical contact. Lrp conformation change is followed by translocation of Lrp or Lrp/PapI to the GATC-I region by an as yet unknown mechanism, leaving GATC-II accessible to Dam methylase. Alternatively, the possible low frequency unstable binding of Lrp to the GATC-I region is stabilized by a Lrp conformation change triggered by interaction with PapI. The ensuing methylation of GATC-II will prevent Lrp from binding to the GATC-II region. The binding of Lrp or Lrp/PapI to the GATC-I region will change the conformation of the *pap* regulatory region in a way which facilitates the assembly of the transcription machinery and/or the interaction of Lrp, CAP, and RNA polymerase, leading to a phase ON phenotype. After a new round of DNA replication, one of the two daughter cells becomes nonmethylated at GATC-I and hemimethylated at GATC-II, which is opposite to that in a phase OFF cell following DNA replication. Now the preferential binding of Lrp or Lrp/PapI to the GATC-I region and the preferential methylation of GATC-II epigenetically maintains the phase ON expression state.

Phase ON to Phase OFF Switch

For a phase ON cell switching to phase OFF, it requires only that Lrp binds to the GATC-II region, which occurs when the local PapI level is low. A reciprocal process to that discussed above should also apply to the phase ON to OFF switching.

Locked Phase ON and Locked Phase OFF

If one of the two GATC regions is mutated so that the affinity for Lrp binding to this region is dramatically reduced, the mutant become phase locked ON or locked OFF. In this case, PapI becomes irrelevant because the affinity difference for Lrp binding to the intact and the mutated GATC region is so significant that Lrp either can bind to the GATC-I region independent of PapI or is unable to bind there even with the help of PapI.

This model is consistent with the previous finding that Pap pili phase variation is sensitive to the expression level of the *dam* gene (9, 13). Our mutagenesis studies so far also support this model (12, 34, 43, 44). However, there are still many aspects of this model that need to be further refined. Mutational analyses, especially those regarding the three-dimensional structure of the *pap* regulatory DNA and the domains of Lrp and PapI, will significantly advance our understanding of the mechanisms of the Pap pili phase variation.

Implications for Other Pili Phase Variation Systems

A variety of pathogenic *E. coli* and other bacteria produce different types of pili that facilitate the adhesion to different types of epithelial cells in different hosts. Though these types of pili are encoded by genes located on the chromosome or plasmids, the comparison of the gene clusters shows similar organization (36). These results suggest that they might be evolutionally related. F1845 pili, K88 pili, and S pili, which are encoded by *daa, fae,* and *sfa* operons (4, 7, 57), respectively, are particularly interesting. Besides the similarity

of gene cluster organization with the *pap* operons, they also encode regulatory proteins homologous to PapI and PapB. Like *pap*, the intercistronic regulatory region of *daa, fae,* and *sfa* also has conserved inverted repeats containing GATC sites, corresponding to the GATC-I and GATC-II boxes of *pap*. Furthermore, the spacing, rather than the sequence, between the GATC boxes is conserved (70). It has been proposed that F1854, K88, and S pili are regulated by a phase variation mechanism similar to that of Pap pili (69).

Like *pap*, when present in single copy, both *daa* and *sfa* undergo phase variation, which is sensitive to the Dam methylase level *in vivo*. The GATC sites corresponding to GATC-I and GATC-II of *pap* are differentially methylated in cells from phase ON and phase OFF populations. The GATC-I but not GATC-II is fully methylated in cells from a phase OFF population, and the converse pattern is observed with cells from the phase ON population (69).

The transcription of both *daa* and *sfa* is positively regulated by Lrp (6, 69). Moreover, PapI can functionally substitute for DaaF and SfaC in modulating Lrp binding, indicating that the function of the PapI-like protein is conserved. Lrp specifically binds to the regulatory regions of *daa* and *sfa,* as evidenced by the shift in mobility on a native gel. PapI can form an even slower moving complex with Lrp and *daa* or *sfa* regulatory DNA, though by itself it does not form a complex with either *daa* or *sfa* regulatory DNA. As in the case of *pap*, Lrp binding to the *sfa* regulatory region protects a region covering the GATC-II site based on a DNase I footprinting assay. Addition of PapI extends the footprinting to a region covering the GATC-I site, which is not observed when the GATC-I is methylated. Lrp binding to the *daa* regulatory region footprints a region covering both GATC-I and GATC-II. Nevertheless, addition of PapI significantly increases the affinity of Lrp for GATC-I (69).

The expression of K88 pili has been studied using multicopy plasmids (32, 33). Phase variation has not been reported in this system, though differential DNA methylation patterns are observed. There are three GATC sites in the *fae* regulatory region. Besides

those corresponding to GATC-I and GATC-II of *pap*, a third GATC site (GATC-III) is found 31 bp downstream of GATC-II. GATC-I is always methylated and GATC-II nonmethylated. GATC-III is either methylated or nonmethylated. Mutation of GATC-I stimulates *fae* transcription by twofold. Mutation of GATC-III decrease *fae* expression by less than twofold. Unlike *pap*, *fae* expression is negatively regulated by Lrp and the PapI homologue FaeA. The apparent absence of phase variation and the less pronounced effects of DNA methylation on transcription regulation in the *fae* system could be due to IS1 insertions in the regulatory region. Alternatively, this also could be due to the location of *fae* genes on a multicopy plasmid, which abrogates phase variation.

Evidence indicates that Lrp mediated differential DNA methylation is a common mechanism in the transcriptional regulation of numerous pili operons in *E. coli*. The regulation of Pap pili phase variation can serve as a model system of these phase variation systems. Understanding how the interactions of Lrp, PapI, and the *pap* regulatory DNA form the phase ON and phase OFF methylation patterns and how these patterns, in turn, activate or repress *papBA* transcription will contribute significantly to the understanding of the other *E. coli* pili phase variation systems. On the other hand, advances in the studies of the other pili phase variation systems should provide new insights into our understanding of the regulation of Pap pili phase variation.

Concluding Remarks

Although the work concerning the mechanisms of Pap pili phase variation is still very preliminary and our understanding in this regard is superficial, the current study raises several intriguing questions regarding DNA-protein and protein-protein interactions. First, how does Lrp recognize its targets? A consensus sequence was obtained by Lrp binding selection-PCR amplification (20). This sequence seems to be conserved in the Lrp binding sites in the *ilvIH* regulatory region (72, 74). However, such a "consensus sequence" does not seem responsible for Lrp binding to a variety of DNA recognized by Lrp (23, 26, 44).

The possibility that Lrp recognizes both specific sequences and specific subtle three-dimensional structure should be further investigated. Second, how does Lrp function both as transcription activator and repressor? In the case of *ilvIH*, Lrp binding to the sites adjacent to the promoter is required for the transcription activation (72). In the case of *papBA*, Lrp binding to sites adjacent to the promoter represses the transcription, and binding to sites distal to the promoter is required for the transcription activation (14, 44). Though in all cases a complex nucleoprotein structure is implied, the nature of such a complex is unknown. Third, how does Lrp interact with PapI and with other factors involved in *papBA* transcription activation? The *pap* DNA-dependent Lrp-PapI interaction demonstrated by Kaltenbach *et al.* (35) is an intriguing mechanism for the interaction of global-local regulatory proteins and merits study in great detail. Fourth, why is there differential DNA methylation in the *pap* regulatory region? We still do not understand why Lrp binding to the GATC-II region could inhibit binding to the GATC-I region by Lrp/PapI, and vice-versa, *in vivo*. I propose that this involves a three-dimensional structure change in the *pap* regulatory region.

Bacteria, especially pathogenic bacteria, have evolved mechanisms to evade the host immune systems by varying the expression states of important surface components. The common phase variation mechanisms usually involve specific DNA recombination or alteration (56). DNA methylation-mediated Pap pili phase variation is an excellent example of phase variation occurring without alteration of primary DNA sequences. However, this mechanism is not unique to the Pap pili of *E. coli*. Similar mechanisms have been discovered in the phase variation of the S pili and F1845 pili (69) and are likely to be found in a variety of other phase variation systems.

References

1. **Abraham, J. M., C. S. Freitag, J. R. Clements, and B. I. Eisenstein**. 1985. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. **82:** 5724-5727.

2. **Andera, L., C. J. Spangler, A. Galeone, L. Mayol, and E. P. Geiduschek**. 1994. Interrelation of secondary structure stability and DNA-binding affinity in the bacteriophage SPO1-encoded type II DNA binding protein TF1. J. Mol. Biol. **236:** 139-150.

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

4. **Bakker, D., P. T. J. Willemsen, L. H. Simons, F. G. van Zijderveld, and F. K. de Graaf**. 1992. Characterization of the antigenic and adhesive properties of FaeG, the major subunit of K88 fimbriae. Mol. Microbiol. **6:** 247-255.

5. **Barras, F. and M. G. Marinus**. 1989. The great GATC: DNA methylation in *E. coli*. Trends Genet. **5:** 139-143.

6. **Bilge, S. S., J. Apostol, J. M., K. J. Fullner, and S. L. Moseley**. 1993. Transcriptional organization of the F1845 fimbrial adhesin determinant of *Escherichia coli*. Mol. Microbiol. **7:** 993-1006.

7. **Bilge, S. S., C. R. Clausen, W. Lau, and S. L. Moseley**. 1989. Molecular characterization of a fimbrial adhesin, F1845, mediating diffuse adherence of diarrhea-associated *Escherichia coli* to Hep-2 cells. J. Bacteriol. **171:** 4281-4289.

8. **Blyn, L. B.** 1991. The Regulation of Pap Pilin Phase Variation. Ph. D. Thesis. The University of Utah.

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

10. **Blyn, L. B., B. A. Braaten, C. A. White-Ziegler, D. A. Rolfson, and D. A. Low**. 1989. Phase-variation of pyelonephritis-associated pili in *Escherichia coli*: Evidence for transcriptional regulation. EMBO J. **8:** 613-620.

11. **Bolker, M. and R. Kahmann**. 1989. The *Escherichia coli* regulatory protein OxyR discriminates between methylated and unmethylated states of the phage Mu *mom* promoter. EMBO J. **8:** 2403-2410.

12. **Braaten, B.** Unpublished data (University of Utah).

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

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

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

16. **Bracco, L., D. Kotlarz, A. Kolb, S. Diekmann, and H. Buc**. 1989. synthetic curved DNA sequence can act as transcriptional activators in *Escherichia coli*. EMBO J. **8:** 4286-4296.

17. **Busby, S., A. Kolb, and H. Bug.** 1995. The *E. coli* cyclic AMP receptor protein. p. 177-191. *In* F. Eckstein and D. M. J. Lilley (ed.), Nucleic Acids and Molecular Biology. Vol. 9. Springer-Verlag, Berlin.

18. **Calvo, J. M.** Personal communications with D. Low.

19. **Calvo, J. M. and R. G. Matthews**. 1994. The Leucine-responsive regulatory protein, a global regulator of metabolism in *Escherichia coli*. Microbiol. Rev. **58:** 466-490.

20. **Cui, Y., Q. Wang, G. D. Strormo, and J. M. Calvo**. 1995. A consensus sequence for binding of Lrp to DNA. J. Bacteriol. **177:** 2872-4880.

21. **D'Ari, R., R. T. Lin, and E. B. Newman**. 1993. The leucine-responsive regulatory protein: more than a regulator. Trends Biochem. Sci. **18:** 260-263.

22. **Ebright, R. H.** 1993. Transcription activation of Class I CAP-dependent promoters. Mol. Microbiol. **8:** 797-802.

23. **Ferario, M., B. R. Ernsting, D. W. Borst, D. E. Wiese II, R. M. Blumenthal, and R. G. Matthews**. 1995. The leucine-responsive regulatory protein of *Escherichia coli* negatively regulates transcription of *ompC* and *micF* and positively regulates translation of *ompF*. J. Bacteriol. **177:** 103-113.

24. **Forsman, K., M. Gorannson, and B. E. Uhlin**. 1989. Autoregulation and multiple DNA interactions by a transcriptional regulatory protein in *E. coli* pili biogenesis. EMBO J. **8:** 1271-1277.

25. **Forsman, K., B. Sonden, M. Goransson, and B. E. Uhlin**. 1992. Antirepression function in *Escherichia coli* for the cAMP-cAMP receptor protein transcriptional activator. Proc. Natl. Acad. Sci. USA **89:** 9880-9884.

26. **Gazeau, M., F. Delort, M. Fromant, P. Dessen, S. Blanquet, and P. Plateau**. 1994. Structure-function relationship of the Lrp-binding region upstream of lysU in *Escherichia coli*. J. Mol. Biol. **241:** 378-389.

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

28. **Goransson, M., B. Sonden, P. Nilsson, B. Dagberg, K. Forsman, K. Emanuelsson, and B. E. Uhlin**. 1990. Transcriptional silencing and thermoregulation of gene expression in *Escherichia coli*. Nature (London) **344:** 682-685.

29. **Grothers, D. M., T. E. Haran, and J. Nadeau**. 1990. Intrinsically bent DNA. J. Biol. Chem. **265:** 7093-7096.

30. **Hochschild, A.** 1991. Detecting cooperative protein-DNA interactions and DNA loop formation by footprinting, p. 343-364. *In* R. T. Sauer (ed.), *Methods in Enzymology*. Protein-DNA interactions. Academic Press. New York

31. **Hochschild, A. and M. Ptashne**. 1986. Cooperative binding of λ repressor to sites separated by integral turns of the DNA helix. Cell **44:** 681-687.

32. **Huisman, T. T., D. Bakker, P. Klaasen, and F. K. de Graaf**. 1994. Leucine-responsive regulatory protein, IS1 insertions, and the negative regulator FaeJA control the expression of the fae (K88) operon in *Escherichia coli*. Mol. Microbiol. **11:** 525-536.

33. **Huisman, T. T. and F. K. de Graf**. 1995. Negative control of fae (K88) expression by the 'global' regulator Lrp is modulated by the 'local' regulator FaeA and affected by DNA methylation. Mol. Microbiol. **16:** 943-953.

34. **Kaltenbach, L. S.** Unpublished data (University of Utah).

35. **Kaltenbach, L. S., B. A. Braaten, and D. A. Low**. 1995. Specific binding of PapI to Lrp-*pap* DNA complexes. J. Bacteriol. **173:** 1789-1800.

36. **Kusters, J. G. and W. Gaastra.** 1994. Fimbrial operons and evolution, p. 179-196. *In* P. Klemm (ed.), Fimbriae: Adhesion, Genetics, Biogenesis and Vaccines. CRC Press, Boca Raton.

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

38. **Low, D. A.** Unpublished data (University of Utah).

39. **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:** 6678-6686.

40. **Marinus, M. G.** 1987. DNA methylation in *Escherichia coli*. Ann. Rev. Genet. **21:** 113-131.

41. **Matthews, K. S.** 1992. DNA looping. Microbiol. Rev. **56:** 123-136.

42. **Meyer, T. F., C. P. Gibbs, and R. Haas**. 1990. Variation and control of protein expression in *Neisseria*. Ann. Rev. Microbiol. **44:** 451-477.

43. **Nou, X.** Unpublished data (University of Utah).

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

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

46. **O'Neil, K. T., J. D. Shuman, C. Ampe, and W. F. Degrado**. 1991. DNA-induced increase in the alpha-helical content of C/EBP and GCN4. Biochemistry **30:** 9030-9034.

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

48. **Pabo, C. O. and R. T. Sauer**. 1984. Ann. Rev. Biochem. **53:** 293-321.

49. **Perez-Martin, J. and M. Espinosa**. 1993. Protein induced bending as a transcriptional switch. Science **260:** 805-807.

50. **Perez-Martin, J., F. Rojo, and V. Lorenzo**. 1994. Promoters responsive to DNA bending: a common theme in prokaryotic gene expression. Microbiol. Rev. **58:** 268-290.

51. **Platko, J. V. and J. M. Calvo**. 1993. Mutations affecting the ability of the *Escherichia coli* Lrp to bind to DNA, activate transcription, or respond to leucine. J. Bacteriol. **175:** 11110-11117.

52. **Pontiggia, A., A. Negri, M. Beltrame, and M. Bianchi**. 1993. Protein HU binds specifically to kinked DNA. Mol. Microbiol. **7:** 343-350.

53. **Rex, J. H., B. D. Aronson, and R. L. Somerville**. 1991. The *tdh* and *serA* operons of *E. coli*: Mutational analysis of the regulatory elements of leucine-responsive genes. J. Bacteriol. **173:** 5944-5953.

54. **Ricca, E., D. A. Aker, and J. M. Calvo**. 1989. A protein that binds to the regulatory region of the *ilvIH* operon of *Escherichia coli*. J. Bacteriol. **171:** 1658-1664.

55. **Rimsky, S. and A. Spassky**. 1990. Sequence determinants for H1 binding on *Escherichia coli lac* and *gal* promoters. Biochemistry **29:** 3765-3771.

56. **Robertson, B. D. and T. F. Meyer**. 1992. Genetic variation in pathogenic bacteria. Trends Genet. **8:** 422-427.

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

58. **Schultz, S. C., G. C. Shield, and T. A. Steitz**. 1991. Crystal structure of a CAP-DNA complex: The DNA is bent by 90°. Science **253:** 1001-1007.

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

60. **Sternberg, N.** 1985. Evidence that adenine methylation influences DNA-protein interactions in *Escherichia coli*. J. Bacteriol. **164:** 490-493.

61. **Strauch, M. A. and M. Ayazifar**. 1995. Bent DNA is found in some, but not all, regions recognized by the *Bacillus subtilis* AbrB protein. Mol. Gen. Genet. **246:** 756-760.

62. **Strauch, M. A. and J. A. Hoch**. 1993. Transition-state regulators: sentinels of *Bacillus subtilis* post-exponential gene expression. Mol. Microbiol. **7:** 337-342.

63. **Strauch, M. A., G. B. Spiegelman, M. Perego, W. C. Johnson, D. Burbulys, and J. A. Hoch**. 1989. A transition state transcription regulator *abrB* of *Bacillus subtilis* is a DNA binding protein. EMBO J. 1615-1621.

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

65. **Travers, A. A.** 1989. DNA conformation and protein binding. Ann. Rev. Biochem. **58:** 427-452.

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

67. **Ussery, D. W., J. Hinton, B. Jordi, P. Granum, A. Serafi, R. Stephen, A. Tupper, G. Berrige, and C. Sidebotham**. 1994. The chromatin-associated protein H-NS. Biochimie **76:** 968-980.

68. **van der Woude, M., L. S. Kaltenbach, and D. A. Low**. 1995. Leucine-responsive regulatory protein plays dual roles as both an activator and a repressor of the *E. coli pap* fimbrial operon. Mol. Microbiol. **17**: 303-312.

69. **van der Woude, M. and D. A. Low**. 1994. Leucine-responsive regulatory protein and deoxyadenosine methylase control the phase variation and expression of the *E. coli sfa* and *daa* pili operons. Mol. Microbiol. **11:** 605-618.

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

71. **van Ham, S. M., L. van Alphen, F. R. Mooi, and J. P. M. van Putten**. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. Cell **73:** 1187-1196.

72. **Wang, Q. and J. M. Calvo**. 1993. Lrp, a global regulatory protein of *Escherichia coli*, binds co-operatively to multiple sites and activates transcription of *ilvIH*. J. Mol. Biol. **229:** 306-318.

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

74. **Wang, Q., M. Sacco, E. Ricca, C. T. Lago, M. DeFelice, and J. M. Calvo**. 1993. Organization of Lrp-binding sites upstream of *ilvIH* in *Salmonella typhimurium*. Mol. Microbiol. **7:** 883-891.

75. **Weiss, M. A., T. Ellenberger, C. R. Wobbe, J. P. Lee, S. C. Harrison, and K. Struhl**. 1990. Folding transition in the DNA binding domain of GCN4 on specific binding to DNA. Nature (London) **347:** 575-578.

76. **West, D., R. Williams, V. Rhodius, B. A., N. Sharma, C. Zou, N. Fujita, A. Ishihama, and S. Busby**. 1993. Interaction between the *Escherichia coli* cyclic AMP receptor protein and RNA polymerase at Class II promoters. Mol. Microbiol. **10:** 789-797.

77. **Yamada, H., S. Muramatsu, and T. Mizuno**. 1990. An *Escherichia coli* protein that preferentially binds to sharply curved DNA. J. Biochem. **108:** 420-425.

78. **Zhou, Y., T. J. Mercel, and R. H. Ebright**. 1994. Characterization of the activating region of Escherichia coli catabolite gene activator protein (CAP). II. Role at class I and class II CAP-dependent promoters. J. Mol. Biol. **243:** 603-610.