Identification of Bacterial Genes Regulated by the *Escherichia coli* Transposable Phage Ner Protein Homologue, Nlp/Sfs 7, a Histone-Like Protein

Submitted by: **Manuelle Rongy**

Department of Microbiology and Immunology
McGill University
Montreal, Québec, Canada

January, 2002

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science

Abstract

DNA-binding proteins such as IHF play a crucial role in bacterial survival. IHF is involved in many biological processes in *E. coli* and possibly in other bacteria. It was hypothesized that another protein, Nlp, belongs to a family of conserved DNA-binding proteins, and this protein was shown to be non-essential for cell viability, but highly conserved among the *Enterobacteriaceae*. As Nlp appears to play a role in gene regulation, we wished to identify all Nlp-regulated genes. *nlp* was cloned into pBAD18-Kan generating pMM5, where Nlp expression was under the control of arabinose. Strain LF20300 was transformed with pMM5 and then lysogenized with the *lacZ* transcriptional fusion reporter vector Mu dI to create a library of approximately 10,000 clones. This library was screened on plates containing either glucose/X-gal or arabinose/X-gal in order to identify genes whose expression was altered upon production of Nlp. Two clones, 90-6 and 205-15, were identified which displayed increased β-galactosidase expression in the presence of Nlp. Further studies with these two clones have shown the insertion of a single Mu dI phage within clone 90-6 and a double insertion within clone 205-15. Cloning and sequencing of one of the Mu dI insertions of strain 205-15 has identified the *yqhG* gene as being one of the genes whose expression may be altered in the presence of Nlp.

Résumé

Les protéines se liant à l'ADN telles IHF jouent un rôle crucial dans la survie des bactéries. IHF est impliquée dans de nombreux processus biologiques chez *E. coli* et peut-être également chez d'autres bactéries. Nlp, une autre protéine, pourrait également appartenir à une famille de protéines se liant à l'ADN. Cette protéine n'est pas essentielle à la survie bactérienne, bien que fortement conservée chez les *Enterobactériaceae*. Nlp paraissant jouer un rôle dans la régulation des gènes, l'identification des gènes régulés par Nlp a été entreprise. Le gène *nlp* a été cloné dans pBAD18-Kan générant pMM5, dans lequel l'expression de Nlp était sous contrôle de l'arabinose. La souche LF20300 a été transformée avec pMM5 puis lysogénisée par le vecteur Mu dI pour créer une banque de fusion transcriptionnelle avec le gène *lacZ* d'environ 10.000 clones. Cette banque a été criblée sur un milieu contenant glucose/X-gal ou arabinose/X-gal pour identifier les gènes dont l'expression était altéré par la production de Nlp. Deux clones, 90-6 et 205-15, montrant une expression accrue de β-galactosidase en présence de Nlp ont été identifiés. Une étude approfondie de ces deux clones a montré la présence d'une seule insertion de Mu dI dans le clone 90-6 et de deux insertions dans le clone 205-15. De plus, le clonage et le séquençage d'une des deux insertions de Mu dI dans la souche 205-15 ont identifié le gène *yqhG* comme étant un des gènes dont l'expression pourrait être altérée par la présence de Nlp.

Acknowledgments

I would like to thank all the people who helped me during the past three years, during this marathon master... I especially thank Dr. Michael DuBow for giving me the chance to conduct research in his lab, Dr. Malcolm Baines for giving me the opportunity to get a teaching experience, and all the lab members for their support.

A very special thank goes to Kirsty Salmon and Madani Thiam, for their friendship but also for the incredible help they provided me. I really appreciated Madani's support throughout my research progression and his help in the lab when I was pregnant. Kirsty's encouragement, support and helpful suggestions about my thesis were also very much appreciated.

Je voudrais surtout remercier mes parents de m'avoir permis d'étudier à des millers de kilomètres de la maison, malgré la douleur de l'éloignement. Je les remercie pour leur amour, leur confiance, leur aide inconditionnelle et tous les sacrifices qu'ils ont faits pour leurs filles. Je remercie aussi mes sœurs, pour nos fous rires et tous les bons moments passés ensembles. Enfin, mille pensées pour les deux personnes qui me remplissent de bonheur: Omar, choukrane pour tout, mais surtout pour ta patience cette dernière année; Neyla, ton petit air espiègle est un rayon de soleil. Je vous promets de vous consacrer plus de temps...

Je ne pourrais pas nommer toutes les personnes dont l'amitié m'est très chère (amis ou famille); j'aurais trop peur d'oublier un nom! Ils se reconnaîtront. Merci d'avoir été si patients lorsque je n'avais pas le temps de donner des nouvelles.

Financial support was provided by fellowships from the Department of Microbiology and Immunology (F.C. Harrison Fellowship), and Dr. M.S. DuBow.

Preface to the thesis

This thesis was prepared in accordance with the 'Guidelines for Thesis Preparation' obtained from the Faculty of Graduate Studies and Research, McGill University, Montreal (QC).

This thesis consists of a general introduction (Chapter 1), a review of literature (Chapter 2), an experimental chapter (Chapter 3), and a final conclusion and summary (Chapter 4). References for all chapters are listed by citation order at the end of the thesis.

I was responsible for all the research described in Chapter 3 with the exception of the β-galactosidase experiments, for which I got help from Faiz Ahmad.

Table of Contents

List of Figures and Tables

Chapter 1 General Introduction

DNA-binding proteins play a crucial role in bacterial physiology and survival. In this thesis, two DNA-binding proteins were studied. Chapter 2 is a review of the literature concerning IHF (Integration Host Factor), a histone-like protein which has been extensively studied and was shown to be involved in many different biological processes in the *E. coli* cell. Interestingly, this protein has homologues in bacteria more or less related to *E. coli* suggesting that this important protein has been conserved throughout evolution.

Chapter 3 (the experimental portion of this thesis) focuses on another protein, Nlp (Ner-Like Protein), which is hypothesized to belong to a family of conserved DNA-binding proteins. However, no roles are known yet for this novel protein which appears to play a role in gene regulation and is conserved among the *Enterobacteriaceae*. The goal of the work presented in Chapter 3 was to identify and characterize bacterial genes regulated by the *E. coli* Nlp protein. A study of the role of Nlp in *E. coli* cell growth and physiology was therefore undertaken using a gene fusion approach to better understand this putative family of regulatory proteins.

Chapter 2 Histone-Like Proteins Conserved in Bacteria: a Review with an Emphasis on IHF (Integration Host Factor)

1.1 Introduction

1.1.1 Histone-like proteins of bacteria

The genome of *Escherichia coli* is associated with ten to twenty different DNA-binding proteins, forming what is called the nucleoid (1). These histone-like proteins, or nucleoid-associated proteins, share several properties with eukaryotic histones, which themselves are responsible, in higher organisms, for compacting DNA into nucleosomes. They are small, basic, abundant, DNA-binding proteins and their primary structure is highly conserved (2). However, the bacterial histone-like proteins have not been as well characterized as their eukaryotic counterparts. The idea that there could be histone-like proteins in bacteria came from the observation that a bacterial protein, known as HU (heat-unstable nucleoid protein), displayed many similarities with eukaryotic histones. The HU protein is a small, basic, abundant, DNA-binding protein which is capable of wrapping DNA, and whose primary structure is highly conserved among bacterial species (3, 4, 5, 6) (Table 1.1). Several reviews on histone-like proteins have been published, including reviews by Pettijohn (7), Oberto, Drlica, *et al.* (2) and Hayat & Mancarella (8).

Other DNA-binding proteins, such as H-NS (histone-like nucleoid structuring protein; also known as protein H1a) (9, 10, 11), Fis (factor for inversion stimulation) (12, 13), and IHF (Integration Host Factor) (14) were discovered and classified in the same group as HU on the basis of their high abundance, low molecular weight, solubility at low pH, and association with DNA (15) (Table 1.1).

HU is the most abundant DNA-binding protein of *E. coli* and displays an extensive amino acid sequence homology with IHF (16). However, although IHF can replace HU in some processes such as Tn*10* transposition (17), or open complex formation at the origin of replication of the *E. coli* chromosome (18), their binding properties are very different. While several HU monomers can bind to a single DNA fragment in a non-specific manner, only one IHF monomer recognizes a specific sequence (19, 20).

The Fis protein is basic, and although it was identified for its role in site-specific DNA recombination reactions, there is substantial evidence that it also participates in essential cell processes such as rRNA and tRNA transcription and chromosomal DNA replication (13). The neutral H-NS protein has also been classified as a histone-like protein, since H-NS is thought to bind DNA in a rather non-specific manner (although it binds preferentially to curved DNA) and to be able to compact it (9, 21, 22).

IHF, HU, and H-NS have been shown to be required for cell growth, and that elimination of more than one of these proteins results in dramatic effects on bacterial viability (23). Interestingly, unlike IHF, H-NS, and Fis, HU does not display any apparent DNA-binding recognition specificity (2, 24). Another important finding is that IHF, H-NS, and Fis are growth phase regulated, and the abundance of *E. coli* IHF and H-NS increase during stationary phase (10, 25, 26, 27, 28) (Table 1.1).

In the case of *E. coli* Fis, there is a dramatic increase in the protein level when stationary-phase cells are sub-cultured into growth medium, prior to the first cell division, and a rapid shut off of synthesis when cells enter exponential growth (28, 29, 30, 31) (Table 1.1). Thus, it appears that the relative level of the different histone-like proteins is regulated by environmental changes.

The nucleoid-associated protein family also includes other proteins such as CbpA (curved DNA-binding protein A) (32), CbpB or Rob (curved DNA-binding protein B or right arm of the replication origin binding protein) (33, 34, 35), DnaA (DNA-binding protein A) (36, 37, 38), Dps (DNA-binding protein from starved cells) (39, 40, 41), Hfq (host factor for phage Q_β) (42, 43, 44), IciA (inhibitor of chromosome initiation A) (45, 46, 47), Lrp (leucine-responsive regulatory protein) (48), StpA (suppressor of *td* phenotype A) (49), and H protein or ribosomal protein S3 (50). Of these proteins, three (CbpB, DnaA and Lrp) were found to bind to specific DNA sequences while the others (CbpA, Dps, Hfq, IciA and StpA) do not seem to display any marked sequence specificity for DNA-binding (24) (Table 1.1).

The roles of IHF have been extensively studied in *E. coli*, and this histone-like protein has been shown to be involved in many different processes such as site-specific recombination, DNA transposition, DNA replication, gene expression and phage packaging [reviewed in ((51, 52, 53, 54)]. This protein has also been found in other

Table 1.1. Characteristics of some histone-like proteins of *E. coli*.

The complete name of the proteins listed are: HU, heat-unstable nucleoid protein; H-NS, histone-like nucleoid structuring protein; Fis, factor for inversion stimulation; IHF, integration host factor; CbpA, curved DNA-binding protein A; CbpB, curved DNA-binding protein B; DnaA, DNA-binding protein A; Dps, DNA-binding protein from starved cells; Hfq, host factor for phage Q_β; IciA, inhibitor of chromosome initiation A; Lrp, leucine-responsive regulatory protein; StpA, suppressor of *td*⁻ phenotype A; and H protein or ribosomal protein S3.

Abbreviations: Ex, exponential phase; S, stationary phase; ES, early stationary phase; LS, late stationary phase; Un, undetectable; ND, no data.

Protein name (M, in kDa)		*Gene name*	*Molecules/cell*	*Specific sequence DNA binding*	*Functions*	*References*
Major species						
HU	HUα (9.5)	*hupA*	15,000 LS	No	DNA wrapping, transposition, inversion, recombination, and replication	(2, 6, 16, 24, 28)
	HUβ (9.2)	*hupB*	55,000 Ex			
IHF	IHFα (11.2)	*ihfA*	12,000 Ex	Yes	DNA wrapping; λ site-specific recombination; control of gene expression; DNA transposition; and DNA replication	(2, 16, 24, 28, 51, 69, 86)
	IHFβ (10.5)	*ihfB*	55,000 ES			
H-NS (15.6)		*hns*	6,500 LS	No	DNA packaging; general repressor and/or silencer of transcription	(9, 10, 11, 21, 24, 28)
			20,000 Ex			
Fis (11.2)		*fis*	Un S	Yes	DNA inversion, DNA replication; site-specific recombination; global transcription regulator	(12, 13, 24, 28, 30)
			> 50,000 Ex			

Table 1.1 - continued

Protein name (M, in kDa)	*Gene name*	*Molecules/cell*	*Specific sequence DNA binding*	*Functions*	*References*
Other DNA-binding proteins					
CbpA (33.4)	*cbpA*	Un Ex 15,000 LS	No	Analogue of the DnaJ molecular chaperone: role in DNA replication, protein folding and protein translocation	(24, 28, 32)
CbpB (33.0)	*rob*	7,000 LS 10,000 Ex	Yes	Transcription regulation of genes involved in oxidative stress response and multiple antibiotic resistance	(24, 28, 33, 34, 35)
DnaA (53.0)	*dnaA*	900 ES 2,300-2,700 LS-Ex	Yes	Regulation of replication initiation; transcription regulation of some genes	(24, 28, 36, 37, 38)
Dps (19.0)	*dps*	6,000 Ex > 180,000 LS	No	Conversion of nucleoid DNA into stress-resistant state; change of gene expression between Ex to S; defence against H_2O_2	(24, 28, 39, 40, 41)
Hfq (11.2)	*hfq*	18,000 ES 55,000 Ex	No	Replication of phage Q_β RNA; regulation of translation of certain mRNAs	(24, 28, 42, 43, 44)
IciA (33.5)	*iciA*	300 LS 800 Ex	No	Inhibition of initiation of DNA replication; transcriptional activator	(24, 28, 45, 46)
Lrp (19.0)	*lrp*	300 ES 3,000 Ex	Yes	Global transcription factor; determinant of nucleoid structure	(24, 28, 48)
StpA (15.3)	*stpA*	8,000 LS 25,000 Ex	No	Chaperone activity, stimulation of self-splicing by promoting RNA assembly *in vitro*; structurally and functionally homologous to H-NS	(24, 28)
H protein (28.0)	*rpsC*	120,000 Ex 120,000 S	ND	Inhibiton of transcription; DNA replication, DNA topoisomerase activity and DNA-dependent ATP hydrolysis *in vitro*	(50, 55, 56)

bacteria (57, 58, 59 60, 61, 62, 63, 64, 65, 66) (Tables 1.2 and 1.3) and it is of great interest to study the evolutionary conservation of this important protein, especially given the rapid rise in bacterial genome sequences. A study in 1991 showed that genes coding for the IHF protein are conserved in Gram-negative bacteria (67). Amino acid sequences of IHF protein homologues of the Gram-negative bacteria, as well as other bacteria, will be compared to highlight the conserved and non-conserved portions of the protein.

Table 1.2. Bacterial IHFα protein homologues.

The accession number in the Entrez Protein database of the National Center for Biotechnology Information is given for each protein. Homologous proteins and their identity percentage to *E. coli* IHFα were found using BLAST searches at the National Center for Biotechnology Information website (**www.ncbi.nlm.nih.gov/**) using the BLOSUM62 matrix (68). Proteins are listed by identity degree to *E. coli* IHFα.

Protein name	***Accession number***	***Length (amino acids)***	***Identity to E. coli IHFα***	***Reference***
Escherichia coli IHFα	P06984	99	100%	(69, 70)
Erwinia chrysanthemi IHFα	P37982	99	96%	(60)
Salmonella typhimurium IHFα	P15430	99	96%	(57)
Serratia marcescens IHFα	P23302	99	94%	(67)
Yersinia. pseudotuberculosis IHFα	CAB46606	98	88%	(71)
Vibrio cholerae IHFα	AAF94381	98	86%	(72)
Pseudomonas putida IHFα	Q52284	100	85%	(59)
Pseudomonas aeruginosa IHFα	Q51472	100	84%	(65)
Xanthomonas campestris IHFα	S67817	99	78%	Direct submission
Xylella fastidiosa IHFα	AAF83553	99	73%	(73)
Pasteurella multocida IHFα	AAK02712	98	71%	(74)
Pasteurella haemolytica IHFα	P95516	99	69%	(63)
Neisseria meningitidis IHFα	AAF41142	100	69%	(75)
Haemophilus influenzae IHFα	P43723	96	66%	(76)
Buchnera species APS IHFα	BAB12849	102	64%	(77)
Buchnera aphidicola IHFα	P57231	102	62%	(77)
Mesorhizobium loti IHFα	NP_108523	107	50%	(78)
Zymomonas mobilis IHFα	AAD53893	113	49%	Direct submission
Myxococcus xanthus IHFα	CAC01236	129	48%	(79)
Rhodobacter capsulatus IHFα	P30787	100	47%	(80, 58)
Caulobacter crescentus IHFα	AAK23351	100	47%	(81)

Table 1.2 - continued

Neisseria gonorrhoeae IHFα	AAG10095	104	35%	(82)
Rickettsia rickettsii IHFα	CAC33712	95	31%	Direct submission
Rickettsia typhi IHFα	CAC33761	95	31%	Direct submission
Rickettsia montanensis IHFα	CAC33647	95	30%	Direct submission
Rickettsia prowazekii IHFα	G71630	95	29%	(83)
Chlamydophila pneumoniae IHFα	AAD18560	100	26%	(84)
Chlamydia trachomatis IHFα	AAC67860	100	26%	(84, 85)

Table 1.3 Bacterial IHFβ protein homologues.

The accession number in the Entrez Protein database of the National Center for Biotechnology Information is given for each protein. Homologous proteins and their identity percentage to *E. coli* IHFβ were found using BLAST searches at the National Center for Biotechnology Information website (**www.ncbi.nlm.nih.gov/**) using the BLOSUM62 matrix (68). Proteins are listed by identity degree to *E. coli* IHFβ.

Protein name	***Accession number***	***Length (amino acids)***	***Identity to E. coli IHFβ***	***Reference***
Escherichia coli IHFβ	P08756	94	100%	(86)
Serratia marcescens IHFβ	P23303	94	93%	(67)
Erwinia chrysanthemi IHFβ	P37983	94	91%	(60)
Vibrio cholerae IHFβ	AAF95062	92	81%	(72)
Pseudomonas aeruginosa IHFβ	Q51473	94	73%	(65)
Pseudomonas putida IHFβ	Q52285	100	73%	(59)
Buchnera species APS IHFβ	BAB13017	94	68%	(77)
Xylella fastidiosa IHFβ	AAF85236	126	68%	(73)
Pasteurella multocida IHFβ	AAK02884	94	65%	(74)
Buchnera aphidicola IHFβ	Q44654	93	64%	(64)
Pasteurella haemolytica IHFβ	P95519	93	64%	(63)
Haemophilus influenzae IHFβ	P43724	94	57%	(76)
Neisseria meningitidis IHFβ	AAF41677	104	56%	(75)
Caulobacter crescentus IHFβ	AAK25548	96	52%	(81)
Mesorhizobium loti IHFβ	NP_104360	94	50%	(78)
Rhodobacter sphaeroides IHFβ	AAD29258	93	49%	(87)
Rhodobacter capsulatus IHFβ	Q06607	95	47%	(58)
Agrobacterium tumefaciens IHFβ	CAC15177	102	42%	(88, 89)
Neisseria gonorrhoeae IHFβ	AAG10094	96	36%	(66)
Rickettsia rickettsii unknown	AAC05204	90	35%	Direct submission
Aquifex aeolicus IHFβ	G70486	100	35%	(90)
Chlamydia muridarum IHFβ	F81691	100	32%	(91)
Chlamydophila pneumoniae IHFβ	B72080	100	31%	(84, 91)

1.1.2 Discovery and identification of the IHF protein

The IHF protein was first discovered in a mutant strain of *E. coli* which was defective for bacteriophage lambda (λ) recombination (92). Integration of λ occurs by site-specific recombination between a unique segment of the viral genome, the phage attachment site, *attP*, and a unique region in the *E. coli* chromosome, the bacterial attachment site, *attB* (93). This recombination event requires a phage-encoded protein called integrase (Int) which carries out the strand exchange reaction. Williams *et al.* (92) treated a culture of *E. coli* with a mutagen and then plated it on tryptone agar. They found that λ was not able to form stable lysogens of one mutant strain even though it was able to form turbid plaques on this strain. These results indicated that there was an integration defect; the responsible host mutation was therefore named *hid*, for host integration defective. The *hid* mutation was then located to about 37-38 minutes on the *E. coli* genetic map. This mutation did not seem to have any significant effect on cell growth, sensitivity to UV light, burst size following λ infection or plaque morphology with λ. Integrative and excisive recombination abilities were tested in both a *hid* $^{+}$ and a *hid* $^{-}$ strain using a test λ phage. They were both greatly reduced in the *hid* $^{-}$ strain compared to the *hid* $^{+}$ strain. After further studies on integrative recombination *in vitro*, it was shown that the *hid* mutation did not affect *int* gene expression. As the addition of extract from a *hid* $^{+}$ nonlysogen could overcome the defect in *hid* $^{-}$ extracts, it was stated that the *hid* $^{-}$ strains were missing an active host component required for λ integrative recombination. These results confirmed previous studies that showed that the product of the λ *int* gene (the λ-encoded protein integrase) and an activity present in uninfected *E. coli* were required for the integration of bacteriophage λ into the *E. coli* chromosome (94, 95, 96).

Genetic and biochemical studies of *E. coli* mutants defective in λ site-specific recombination showed that these mutants could be classified into two groups depending on the location of the mutation on the *E. coli* chromosome: these two classes of mutations were called *himA* for host integration mediator A (97) and *hip* for host integration protein (98). Extracts from these mutant cells were unable to supplement the

Int protein for λ recombination but extracts from *himA* or *hip* cells could complement each other *in vitro* to allow Int-promoted recombination (98). This *E. coli* protein necessary for λ site-specific recombination was named IHF for Integration Host Factor (14, 99). When trying to purify the bacterial component required for λ site-specific recombination, Nash and Robertson obtained two bands after gel electrophoresis that co-migrated with IHF activity (14). These two polypeptide chains were called α and β. The α subunit of IHF is in fact encoded by the *himA* gene (69, 99) and the β subunit by the *hip* gene (86, 100).

A new nomenclature for the genes encoding IHF was suggested in 1996 (101) to remove the confusion over the name of these genes. The *ihfA* name was substituted for *himA* and *hid* for the gene encoding the IHFα subunit, and *ihfB* for *himD* and *hip* for the gene encoding IHFβ. In this review, only the new names (*ihfA* and *ihfB*) will be used.

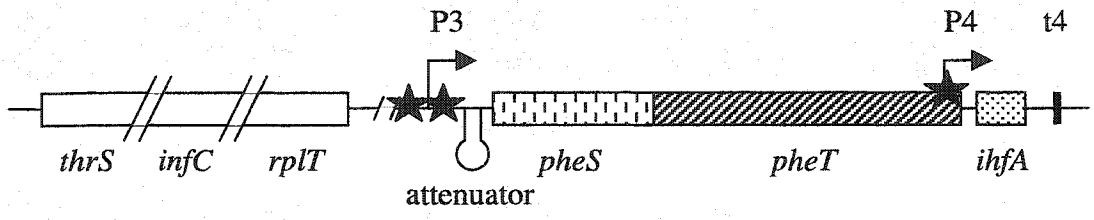
The IHFα chain is a 99 amino acid protein with a molecular weight of about 11.2 kDa and is encoded by the *ihfA* gene (69). This gene was identified in 1985 by Mechulam *et al.* (70) as they were sequencing the *E. coli pheST* operon. This operon codes for the two subunits of phenylalanyl-tRNA synthetase (102). As mentioned above, the *hid* mutation (92) as well as the *himA* mutation (98) were located at about 37-38 min on the *E. coli* chromosome and interestingly, the *pheST* operon was mapped to 38 minutes (70). Indeed, an open reading frame of 297 nucleotides, named *prp*, was found a few nucleotides downstream from the *pheT* gene. A region resembling a rho-independent transcription terminator (103) was found 15 nucleotides downstream from the stop codon of *prp*. Several facts [*pheST* and *ihfA* co-transduction by phage P1 (99), coincidence of the molecular weight of IHFα and the putative *prp* product (70)] suggested that the *ihfA* gene was situated very close to the *pheST* operon. Complementation experiments using an *ihfA* mutant and the *prp* gene showed that *prp* corresponded to the *ihfA* structural gene. (70) (Fig. 1.1).

Concerning the β subunit of the IHF protein, it was suggested in 1981 that it was encoded by the *hip* gene, now known as the *ihfB* gene (100). This gene was isolated and sequenced in 1985 by Flamm and Weisberg using a set of deletion mutants constructed by exonucleolytic digestion. The IHFβ chain was found to be a 93 amino acid protein with a molecular weight of about 10.5 kDa (86). The *ihfB* gene is located near

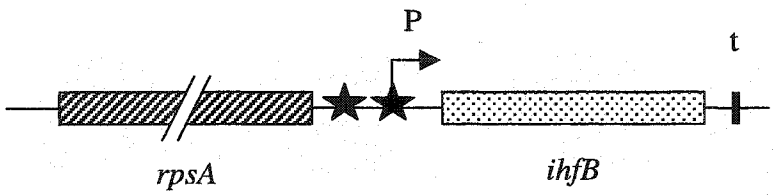
Figure 1.1 Map of the *ihfA* (A) and *ihfB* (B) gene chromosomal regions in *E. coli* (104, 105).

The arrows represent the promoters (P3, P4 and P) and the vertical black bar, the transcription terminators (t4 and t). The attenuator is represented between the P3 promoter and the *pheS* gene in the *ihfA* region. The *thrS* gene encodes the threonyl-tRNA synthetase; *infC*, the initiation factor IF3; *rplT*, the ribosomal protein L20; *pheS* and *pheT*, the two subunits of the phenylalanyl-tRNA synthetase; and *rpsA*, the ribosomal protein S1. (★) are putative IHF binding sites.

A


P3
P4
t4
thrS
infC
rplT
attenuator
pheS
pheT
ihfA

B


P
t
rpsA
ihfB

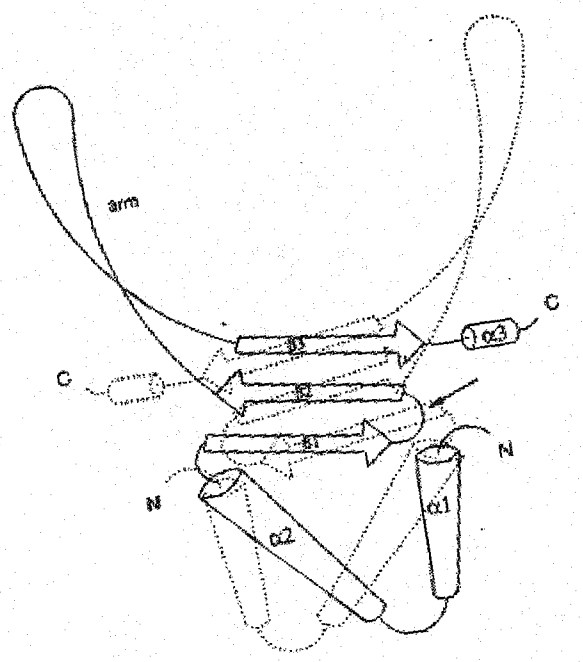
minute 20 on the linkage map (106) and immediately downstream of the *rpsA* gene (86) (Fig. 1.1).

1.1.3 Characteristics of IHF

The two non-identical subunits of IHF, IHFα and IHFβ (14), are both related to a family of nonspecific DNA-binding proteins, the type II or HU family of DNA-binding proteins (16). The IHF heterodimer (Fig. 1.2) can bind and bend DNA at specific sites (107, 108, 109). The IHFα and IHFβ subunits are very similar to each other (30% of their amino acids are identical), but they are both required *in vivo* for integrative recombination and other IHF-dependent processes (97, 106). However, it has been shown that when either IHFα or IHFβ are present in sufficient quantity, they can bind as a homodimer specifically to IHF sites and satisfy the *in vitro* requirement for IHF heterodimer in λ site-specific recombination. Since the stability of the IHFα-DNA and IHFβ-DNA complexes are reduced compared to that of the heterodimer IHF-DNA complexes, it is unlikely that both homodimers bind *ihf* sites *in vivo*, a result in agreement with previous studies that showed that both chains are required for IHF function *in vivo* (110, 111).

As mentioned previously, IHF is homologous to the HU protein, and they share 50-62% identity at the amino acid level. While a defect in either protein is very well tolerated, cell growth is dramatically impaired when both are deficient (112). These observations indicate that both protein can substitute for each other, and that they may function in a similar way (112).

Figure 1.2 Schematic model of the IHF dimer [reprinted from (111), Copyright 1994, with permission from Oxford University Press].
Black lines and dotted lines indicate the two subunits of IHF. Each subunit contains three α-helical domains, α1, α2, and α3; three anti-parallel β-sheets, β1, β2, and β3; and a β-sheeted arm (between β2 and β3) protruding from the body of the protein. The contact point of the IHFα and IHFβ proteins is indicated by a black arrow.


arm
C
C
N
N
α1
α2
α3

1.2 Expression of IHF

The cellular abundance of IHF is growth-phase regulated, but it is not affected by growth conditions such as altered incubation temperature (e.g. 32°C or 42°C instead of 37°C) or growth medium (minimal-glucose salts medium instead of LB broth) (26). The amount of IHF increases from 0.5-1.0 ng of IHF subunits per μg of total protein in exponentially growing cells to 5-6 ng/μg in late-stationary-phase cells. Cells in exponential phase contain approximately 8,500 to 17,000 IHF dimers per cell, a number that is comparable to the levels of ribosomes and HU protein. Another study showed that there were approximately 15,000 IHF subunits in the exponential phase, a number which increased up to 55,000 in early stationary phase (28) (Table 1.1). This large number is surprising for a site-specific DNA-binding protein having only a limited number of specific sites. However, since small decreases in IHF amount have significant effects on several IHF dependent functions, it is possible that the level of IHF in exponential-phase is not in large excess of the minimum required for occupancy of physiologically important IHF-binding sites, and that most of the intracellular IHF is not free. As there are a number of known IHF sites that can only bind a small fraction of all the IHF present in the cell (even if some others were to be found), Ditto *et al.* (1994) suggested that the majority of the IHF proteins could be bound to low-affinity DNA sites with unknown roles. The release of IHF from these low-affinity binding sites could promote some of the modifications that are associated with entry into or transition out of stationary phase (113, 114). However, since in IHF-deficient strains there is no important growth defects nor any problem associated with stationary-phase recovery, growth-phase regulation by IHF, if it exists, may only be subtle, redundant, or both (26). As IHF is a stable protein, changes in its abundance are due to modifications in relative rates of transcription or translation, but not due to protein degradation (26). In fact, the increase in the level of IHF in the entry to stationary phase is in part due to an increase in *ihfA* and *ihfB* transcription and not to increased mRNA stability (105).

1.2.1 Regulation of transcription of the *ihfA* gene

Transcription of the different parts of the *pheST-ihfA* operon was elucidated thanks to the technique of gene fusions (104). Expression of the *ihfA* gene is accomplished in two different ways It can be expressed from a polycistronic transcript initiating at the P3 promoter, located upstream of the *pheS* gene, as well as from a monocistronic transcript initiating at P4, located within *pheT* (Fig. 1.1). Indeed, most *ihfA* transcription originates from the P4 promoter (75%), as approximately 80% of the transcripts originating from P3 terminate at the attenuator (located in the leader mRNA region between the P3 promoter and the *pheS* gene). However, cotranscription between *pheST* and *ihfA* can occur (shown by *in vitro* transcription and S_1 nuclease mapping techniques) since there is only one transcription terminator (t4) located downstream from *ihfA*. The level of transcription at the P3 and P4 promoters is negatively controlled by the *ihfA* product alone or associated with another molecule (104). Interestingly, a putative IHF binding sequence (115), flanked by a 12 base direct repeat, is present between the -35 and -10 regions of P4 and two (one on each DNA strand) overlapping the -35 region of the P3 promoter. These observations seem to indicate that both promoters controlling the transcription of *ihfA* can be negatively controlled by IHF (104, 105) (Fig. 1.1). The P3 and P4 promoters are also under SOS control and previous work indicated that *ihfA* transcription is derepressed in a *lexA*-deficient strain (100). However, no sequence homologous to the LexA consensus binding site was found in the P3 and P4 promoter region, whereas all other genes known to be under SOS control all contain a LexA binding site (104). Moreover, the signal molecule ppGpp also plays a major role in the growth-phase variation of activity at the *ihfA* P4 promoter It affects this promoter indirectly via RpoS (105).

In summary, transcription of the *ihfA* gene appears to be controlled by two different mechanisms. The first one is a classical operator-repressor type mechanism, involving the *ihfA* product, ppGpp, RpoS, and the SOS network, though it is not clear how the SOS network is involved in *ihfA* transcription control. The second mechanism controlling the expression of *ihfA*, which acts independently from the first one, is a mechanism of attenuation control. Thanks to this second mechanism, a link may be

established between the intracellular concentration of IHFα and the functional state of the translational machinery (104, 105).

1.2.2 Regulation of transcription of the *ihfB* gene

Thanks to primer extension and S1 analyses, the transcription start site of *ihfB* was identified (105). The *ihfB* promoter, like the *ihfA* promoter, exhibited a reduced growth-phase response in the absence of RpoS, but the effect was observed to be weaker than at the *ihfA* promoter. The *ihfB* promoter is in fact controlled by both ppGpp and RpoS, acting independently. Transcription from the *ihfB* promoter is negatively regulated by IHF, like the *ihfA* gene. In fact, two putative IHF binding-sites are present in the *ihfB* promoter region: one overlapping the *ihfB* promoter and one directly upstream (105) (Fig. 1.1).

It is of interest to note that within the IHF-binding sites in the *ihfA* and *ihfB* promoter regions, the IHF-recognition sequence differs by one or two nucleotides from the consensus sequence (115). These sites were shown to display a ten-fold lower binding affinity to IHF, which can result from the deviation from the consensus and also, as previously shown (116), from the sequence context of the binding sites. Since these sites display low binding affinity to IHF, it was hypothesized that autoregulation can only take place when sufficient levels of IHF accumulate in the cell (105).

1.3 Structural characteristics of IHF

1.3.1 Interaction of IHF with its specific binding sites

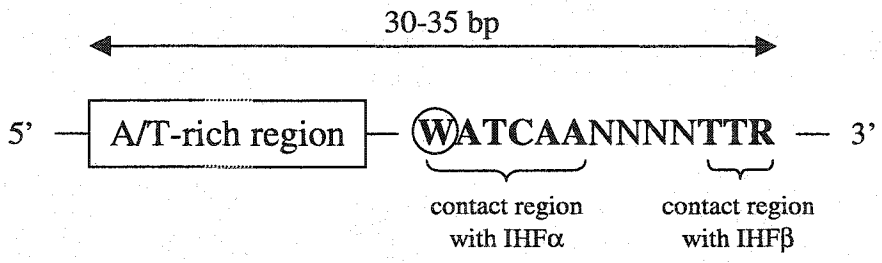
1.3.1.1 IHF consensus sequence

In 1981, Nash and Robertson first demonstrated that purified IHF could bind DNA (14) and in 1984, IHF was shown to be a specific DNA-binding protein which recognized a DNA-binding site of approximately 30 to 35 bp long (115). Binding of IHF to nucleic acids requires neither cofactors nor additional proteins. Craig and Nash (115) initially established the IHF consensus sequence thanks to nucleotide sequence comparison and methylation patterns of different IHF binding sites. The final consensus sequence recognized by IHF has been identified principally by protection assays and elecrophoretic mobility shift assays (EMSAs). Protection from attack by hydroxyl radicals has provided the most precise estimate of the size of an IHF site, typically 25-35bp (20), (117, 118). This consensus site is now widely accepted as WATCAANNNN TTR (W = dA or dT, R = dG or dC and N = any nucleotide) (51) (Fig. 1.3). The four central nucleotides are designated by the letter N since comparison of known IHF sequences have not shown nucleotide preferences within this region (116). Interestingly, the IHF binding site is asymmetric, a property which is thought to be important for IHF-DNA complex formation (119).

IHF binds to unique sites with an affinity of 1-25 nM (120, 121, 122, 123). When IHF binds its target sites, it does so with a specficity ratio of roughly 1,000 to 10,000 for its natural targets vs. nonspecific nucleic acids (122, 123). In addition, in the few cases where it has been examined, cooperativity between IHF sites is weak or undetectable (120, 124).

When IHF binds to DNA, it induces a bend in the DNA molecule (107). Using EMSA techniques such as circular permutation analysis (125) and phasing (126), as well as electron microscopy (127), it has been determined that IHF introduces a very substantial distortion of the double helix (109),. Although there is no definitive basis for

Figure 1.3 Characteristics of the IHF consensus sequence.
The bolded WATCAA and TTR parts of the consensus are believed to contact the IHFα and IHFβ subunits, respectively (128). The circled letter (W) is thought to be the centre of the IHF-induced bend (129).


30-35 bp
5'
A/T-rich region
WATCAANNNNTTR
3'
contact region with IHFα
contact region with IHFβ

quantitating the parameters of the distortion, studies employing most methods agree that, if IHF were to induce a hinge-like bend, the bend angle would be 140° or greater (108, 109, 127, 130, 131). This would convert the shape of the DNA from something approximating a straight line to something resembling a hairpin. Interestingly, IHF has little or no effect on the overall topology of the DNA. This has been deduced from the very similar affinity of IHF for sites on supercoiled and nonsupercoiled templates (122). In addition, IHF does not seem to rely on hydrogen bond donors and acceptors of the base pairs to ensure DNA-binding specificity, as is normally the case (132), but instead appears to rely on indirect readout (133).

1.3.1.2 Importance of the 5' A/T-rich region

The consensus sequence is not the only element involved in IHF binding to DNA: the sequences adjacent to the consensus site also play a role (134). Their importance was demonstrated by mutational analyses (107, 135, 136) and by protection assays showing that a large region of DNA was protected from nuclease attack upon IHF binding (115, 137). Interestingly, IHF binding sites are usually found in A/T-rich regions (137) and a computer analysis of 27 IHF binding sites found the presence of an expanded consensus sequence which includes an A/T-rich region upstream of the core consensus element (116) (Fig. 1.3). However, if the core consensus and the 5'-A/T-rich region are important for IHF binding, these two elements are not always simultaneously present, as shown by the study of the H' and H1 sites in the *attP* region. The core consensus sequence alone is sufficient for IHF binding to the H1 site, but both elements are required for binding of IHF to the H' site (138). In fact, IHF binding is enhanced when a run of adenines or an A-tract is positioned in the 5' region of the consensus sequence. Moreover, IHF appears to bend more DNA at A/T-rich sites than when associated with sites lacking it. Thus, the 5' domain of the IHF binding site appears to be very important for the binding and bending of DNA by this protein and the DNA may be bent differently when an A/T-rich element is present at the 5' end of the consensus sequence (139). Indeed, it was recently shown that the 5' region contributes significantly to the structure of IHF-DNA nucleoprotein complexes but little to their affinity or their

function (133). The idea that the regions flanking the consensus sequence are important in the binding of IHF is supported by the fact that some binding sites, which show a perfect match with the consensus, are bound only weakly by IHF while some sites showing a weak core consensus can be very good IHF sites (107, 117, 136, 137). It has been suggested that the A/T-rich region might increase the flexibility or the bending ability of the DNA in the IHF binding site region, thus providing a local conformation more favorable to IHF recognition, or that it might also be important to position IHF in a unique orientation (135, 138). Sun *et al.* (131) demonstrated that IHF contacts the adenines on the 3'-side of the AT-rich region and thereby induces DNA bending into the minor groove in proximity. This bending may, in turn, stabilize the interaction. Upon IHF binding, structural perturbations, such as DNA unwinding, occur in the vicinity of the IHF contact region but there is no strand separation. (131). The DNA bend induced by IHF binding is more or less symmetrical and the center of the bend is located near the W residue of the consensus sequence (129) (Fig. 1.3).

In summary, IHF binding sites are approximately 25 to 35 bp long and can be divided into at least two domains: a degenerate 5' domain which is typically A/T-rich, and a conserved 3' domain containing the consensus sequence WATCAANNNNTTR (Fig. 1.3). Furthermore, local DNA conformation influences the affinity of IHF binding and bending, thereby providing one mechanism to regulate IHF activity. This modulation may be very important since the consensus sequence of the IHF site is degenerate (140).

1.3.2 Deformation of DNA upon IHF binding

Goodman and Nash (141), working with bacteriophage lambda, suggested that the role of IHF is only to bend DNA into a functional conformation. They postulated that binding of IHF to its site helps bring different parts of the lambda DNA recombination site *attP* closer together, and helps the interaction of other components of the "intasome" (the complex formed by Int, IHF and the DNA) (141). Other studies also led to the hypothesis that IHF only acts as an architectural element. For example, when

IHF binding sites are replaced with stably bent kinetoplast DNA or binding sites for other proteins which bend DNA, site-specific recombination occurs independently of IHF (141, 142, 143, 144). Furthermore, the nonspecific DNA-binding protein HU can substitute both *in vitro* and *in vivo* for IHF to perform excisive recombination of a plasmid containing the prophage *att* sites (HU is only moderately effective at replacing IHF), but is not able to perform integration between the *attP* and *attB* sites (143). This result confirms the hypothesis that IHF functions primarily as an architectural element and that deformation of DNA is a crucial function of IHF (143).

IHF is able to bend DNA even when it binds in a non-specific manner. This property was shown when mutations were created in the IHF site of *attL* (one of the two attachment sites involved in λ excisive recombination). The cooperativity between Int and IHF was retained and stable "intasomes" were created on the mutated *attL*. In addition, IHF mutants defective in specific binding are still able to bend DNA (145). However, when flexible sequences (single-stranded DNA) replace the IHF-induced DNA bending, bacteriophage λ site-specific recombination is enhanced in the absence of IHF but the replacement is not perfect. This result suggests that IHF may have additional functions other than simply deforming DNA (146).

1.3.3 Model of the IHF-DNA complex: the Yang and Nash model

Yang and Nash have established that only one IHF molecule binds to one IHF binding site, although this small protein protects a large region of DNA (20). They also suggested that IHF contacts and recognizes DNA via the minor groove, which is unusual for sequence-specific DNA-binding proteins. A member of the HMG-I family of chromosomal nonhistone proteins, the α protein, and the octanucleotide factor of the specific protein-DNA complex that governs immediate early gene expression of herpes simplex virus, are two possible eukaryotic relatives of IHF since they also appear to contact DNA in the minor groove (147, 148). However, they do not appear to be structurally related to the HU/IHF family as shown by amino acid sequence comparisons (20). A model for the interaction of IHF and its DNA target was described after

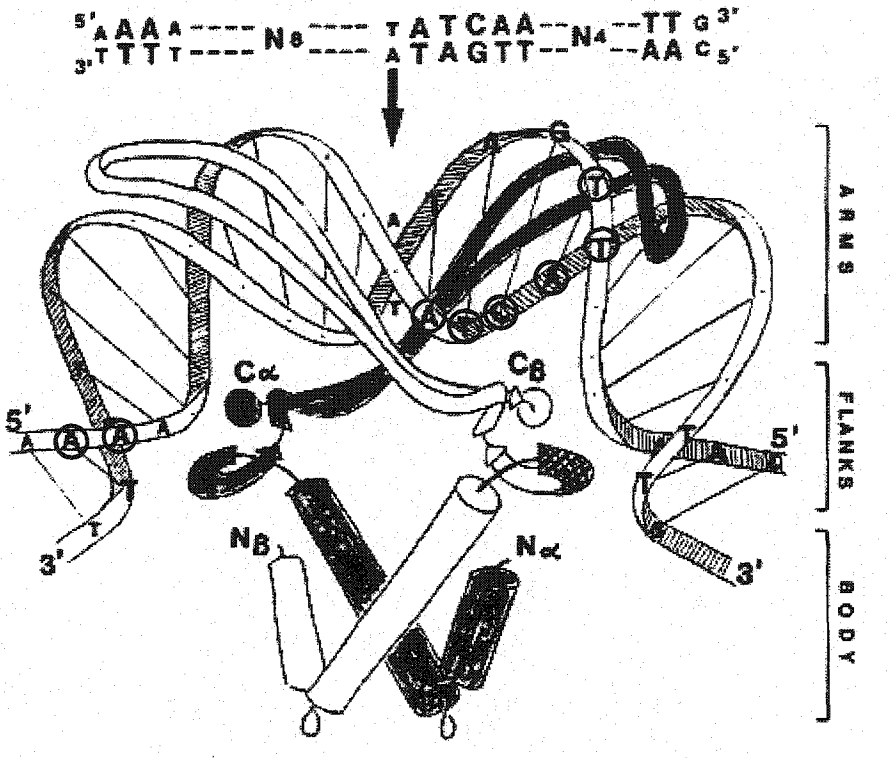
extensive footprint analysis of the three IHF-DNA complexes that are formed at the attachment site of λ and crystal structure analysis of the HU protein from *Bacillus stearothermophilus* determined by Tanaka *et al.* (149). It was suggested that the HU protein structure could be taken as a reasonable approximation for the IHF structure since there is important similarity in amino acid sequence between the two proteins (20). In this model, the two anti-parallel β-sheet arms of IHF encircle the DNA by binding in the minor groove at two sites. A DNA bend is then introduced through interaction of the body of the IHF molecule with sequences flanking the DNA occupied by the arms (20) (Fig. 1.4). Photocrosslinking experiments of IHF-DNA complexes supported the Yang and Nash model and also indicated that both IHF subunits contact the DNA target (150).

The presence of close interactions between the IHFα subunit and the 3' end of the WATCAA portion of the consensus IHF binding site, and between the IHFβ subunit and the TTR part was suggested by Lee *et al.* (128) (Fig. 1.4). Their studies, as well as other genetic studies (121, 151) are consistent with the structural model of Yang and Nash (20). Furthermore, it has been suggested that the strongest interactions between IHF and the DNA may lie within the WATCAR and TTR elements of the consensus sequence where IHF interacts closely with the minor groove of the DNA (123).

The binding specificity of IHF seems to be due, at least in part, to the C-terminal α3 helix of the IHFα subunit, since deletion of this region suppresses all IHF specificity (119). The α3 helix of IHF may not directly interact with the DNA but it may act indirectly by stabilizing the structure of the DNA-binding domain and thus ensure the specificity of the binding reaction (119). Zulianello *et al.* proposed that the IHFα α3 helix could interact directly with the IHFβ arm and thus stabilize its structure, or interact with the turn between the α2 helix and the β1 sheet of IHFβ, thus stabilizing the IHFα arm. At first, the C-terminal parts of IHFα and IHFβ were thought to be responsible for the binding specificity of IHF since they are more extended than those of the two identical subunits of HU. However, deletions of these regions showed that the modified IHF protein still bound DNA in a very specific way. On the other hand, the C-terminal α3 helix of the IHFβ subunit does not affect the specificity of the binding but rather its

Figure 1.4 Schematic model of an IHF-DNA complex at the H' binding site of the *attP* site of bacteriophage λ [reprinteded from (129), Copyright 1995, with permission from Academic Press].

The model was established on the basis of different studies (13, 149, 151). The arms of the two subunits extend from β-sheets into the minor groove, the flanks consist of the first β-sheets and the third C-terminal α-helices, and bodies comprise the first two N-terminal α-helices. The arm of the IHFα subunit (shaded) and the flank of the IHFβ subunit (blank) contact closely the 5' ATCAA part and the TT 3' part of the consensus IHF binding site, respectively (circled letters on right). The flank of the IHFα subunit is believed to enhance binding affinity by contacting a distal A/T-rich region (circled letters on left) (129).


5' A A A A ---- N8 ---- T A T C A A -- N4 -- T T G 3'
3' T T T T ---- N8 ---- A T A G T T -- N4 -- A A C 5'
ARMS
FLANKS
BODY
Cα
Cβ
NB
Nα
5'
3'

affinity, suggesting a possible role for that last α helix of IHFβ in the stability of the IHF-DNA complex. It was also shown that the turn between β-sheets 1 and 2 of IHFβ, particularly the Arg-46 residue which is conserved in all IHFβ proteins known so far except for *Neisseria gonorrhoeae*, may interact with the TTR segment of the consensus sequence and might contribute to the bending of the DNA. In fact, this arginine residue may even be involved in the recognition of the TTR element, since arginine side chains are known, for different proteins, to interact with the minor groove of specific DNA sequences. However, even if the TTR element is important, specific IHF-DNA complexes can still form in the absence of this element, but with decreased affinity. Surprisingly, the four N nucleotides located between the WATCAA and the TTR elements of the consensus, which are not conserved between the different IHF sites, seem to be important for the IHF-DNA complex formation as deletion of these nucleotides prevents the formation of specific complexes. They may indeed be involved in interactions with the IHF protein, mediated by the phosphates since there are no nucleotide preferences in this region (119), or perhaps are only required for spacing purposes.

The co-crystal structure of IHF bound to one of its natural binding sites (H' site of phage λ) was established by Rice *et al.* in 1996 (130) and the stoichiometry of the complex and the orientation of the IHF protein on the DNA agree with solution studies previously performed (20, 128, 150). The study by Rice *et al.* (130) demonstrated that the DNA is bent by 180° into a virtual U-turn and confirmed that the arms encircle the DNA and only interact with the minor groove. In order to bend DNA, IHF positions its positively charged body on the inside of the bend to counteract the symmetric repulsion of the phosphate backbone charges and inserts a hydrophobic residue between base pairs into each of the two kinks it creates (130, 152). The crystal structure was obtained by Rice *et al.* with a DNA molecule containing a single nick, positioned in one of the two kinks that IHF generates upon binding, rather than a completely intact duplex. The structure obtained with the nicked DNA could therefore differ from the global solution structure with intact DNA. However, it was shown using fluorescence resonance energy transfer (FRET) that the global structure was hardly changed, although some subtle changes could be distinguished (153).

1.4 Roles of IHF

1.4.1 Site-specific recombination

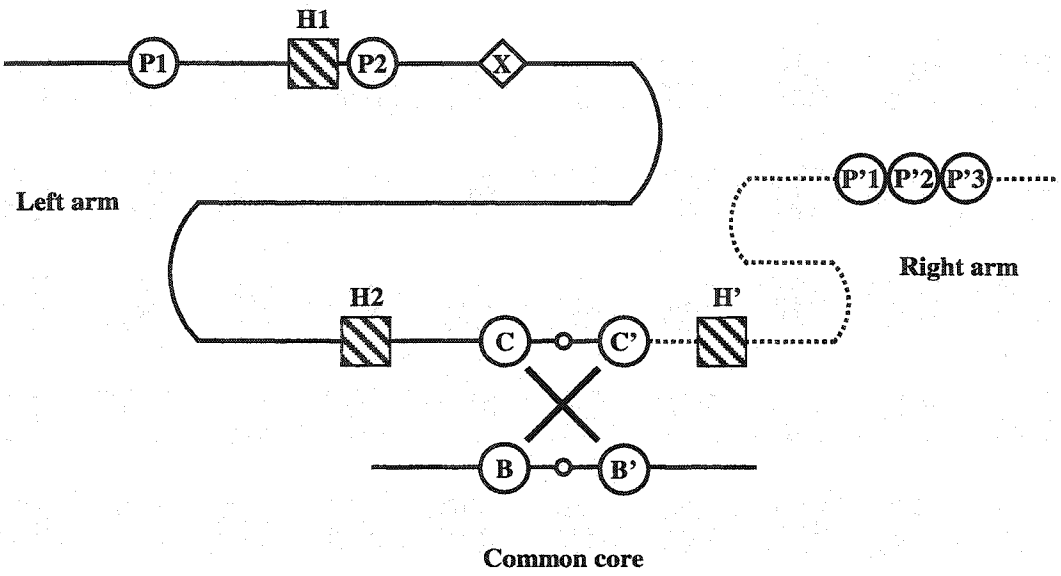
As previously discussed, the IHF protein was originally discovered for its essential role in site-specific λ DNA recombination into the bacterial chromosome both *in vivo* (92, 94, 97, 98) and *in vitro* (95). For reviews on integration and excision of bacteriophage λ, see the articles written by Nash (154) and Weisberg and Landy (155).

Integrative recombination by phage λ occurs between a 250-bp site on the phage chromosome (*attP*) and a 25-bp site on the bacterial chromosome (*attB*) (Fig. 1.5), and leads to the formation of a prophage delimited by two attachment sites, *attL* and *attR*. The *attP* and *attB* sites share a 15-bp homologous common core where the crossover event occurs (Fig. 1.5). In both sites, this common core is flanked by two arm sequences which are not identical. At *attP*, three IHF binding sites, called H1, H2, and H', are located in the arm sequences, while *attB* does not contain any (115) (Fig. 1.5). Interaction of IHF with all three binding sites in *attP* is required for integrative recombination since disruption of any of them decreases the integrative recombination event (120). The common core is also found in the two attachment sites created from the integrative recombination, *attL* and *attR*. Both the phage-encoded integrase protein, Int, as well as the bacterial protein IHF are required in this process. Excision of the prophage is also possible by recombination between these two sites to regenerate *attP* and *attB*, using Int, IHF, and another phage-encoded protein, Xis (154, 155, 156).

However, while Int is responsible for performing the strand exchange by nicking and ligating the DNA (96), IHF seems to function only as an accessory protein. This was suggested by experiments showing that, *in vitro*, a mutant Int protein alone is able to promote integrative recombination, although at a much reduced efficiency (157, 158). However, studies also showed that IHF can promote the specific binding of the Int protein to the core region of *attP* and thereby promote recombination (115, 159).

Concerning λ excisive recombination, IHF is required. However, *in vitro*, high

Figure 1.5 Distribution of IHF binding sites in the *attP* and *attB* regions [adapted from (51), Copyright 1988, with permission from Elsevier Science].
Squares represent the IHF sites: two, H1 and H2, are in the left arm of the phage genome and the third one, H', is in the right arm. The Int sites are represented in circles: P1 and P2 in the left arm; P'1, P'2, and P'3 in the right arm; B, B', C, and C' in the common core of *attB* and *attP*, respectively. There is also a Xis site (X) on the left arm. The region of crossover is indicated by a large cross (51).


H1
P1
P2
X
Left arm
P'1
P'2
P'3
Right arm
H2
H'
C
C'
B
B'
Common core

IHF concentrations inhibit this process. On the other hand, higher concentrations of IHF are required for integration than for excision. As the concentration of IHF in the *E. coli* cell increases during entry into stationary phase, it is possible that IHF is important in the control of the phage life cycle: the level of IHF would prevent a premature excision at a time when conditions are not favorable for a successful phage burst (25).

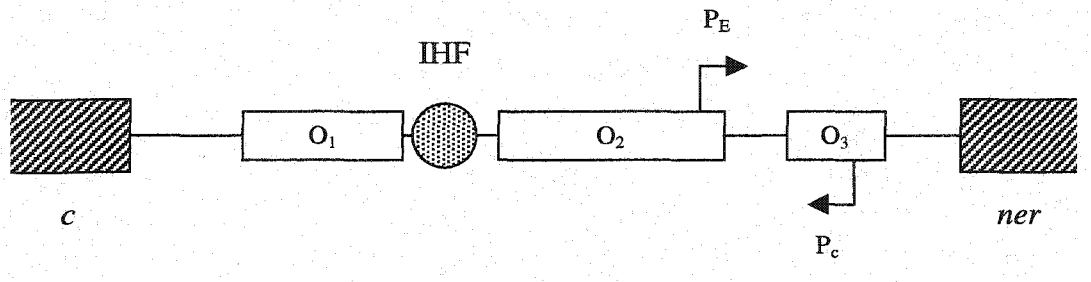
Interestingly, IHF has also been thought to be involved in site-specific recombination of coliphage ϕ80 and *Salmonella typhimurium* phage P22 (160). Indeed, similarities were found in the location and orientation of the IHF binding sites in the attachment sites of coliphages λ and ϕ80 and P22 suggesting a similar role for IHF in the mechanism of site-specific recombination in all three phages (160). However, a recent *in vivo* study showed that IHF was not required for site-specific recombination of bacteriophage P22 (161).

1.4.2 Control of transposition

IHF is known to be required for coliphage Mu transposition *in vitro*: it reduces the negative supercoiling of the Mu donor molecule to the minimal level required for transposition to occur (162, 163). An IHF binding site is present in the Mu left end intergenic region, located between the repressor binding sites O1 and O2. This site stimulates the rate of strand transfer 100-fold and therefore functions as an enhancer-like element (16, 163, 164) (Fig. 1.6). IHF is believed to act at this transpositional enhancer and stimulate the Mu strand-transfer reaction by inducing a sharp bend at this site (162, 163, 165).

IHF also acts as an architectural catalyst in the modulation of Tn*10*/IS*10* transposition (166, 167). The Tn*10* transposon is composed of two IS*10* insertion elements flanking a DNA segment which consists of a gene involved in tetracycline resistance. An IHF binding site is present near the outside end of Tn*10*. It was recently shown that IHF first promotes transpososome assembly (a complex formed by the binding of transposase to both transposon ends) and is then ejected from it. In a second

Figure 1.6 Map of the operator region of coliphage Mu (16).
O1, O2, and O3 are the repressor binding sites. Transcription of the early genes (the first of which is *ner*) of Mu initiates at the early promoter P_E. P_c promotes transcription of the repressor gene, *c*. An IHF binding site is located in the intergenic region between sites O1 and O2.


P_E
IHF
O1
O2
O3
c
ner
P_c

reaction, IHF rebinds Tn*10* and alters transpososome conformation to allow the subsequent steps of transposition to occur (166, 167).

IHF was also thought to be involved in insertion element IS*1* transposition since it had been shown to bind specifically to a site on each end of this transposable element (107, 137). However, according to Shiga *et al.* (1999), a recent unpublished study suggests that IHF is not required for this process to occur; instead, another histone-like protein, H-NS, appears to be responsible (168). IHF is also thought to be involved in the excision process of the genetic element e14, a non-essential component of 14.4kb of the *E. coli* K12 chromosome (169), however further studies are required.

1.4.3 Control of phage gene expression

IHF indirectly regulates the synthesis of two phage-encoded proteins involved in the lysogenic pathway of bacteriophage λ: Int, necessary for λ site-specific recombination, and the cI repressor which shuts off expression of lytic functions (170, 171). *In vivo*, this regulation seems to be accomplished via the translational control of another phage protein, λcII, probably by binding to a duplex region of RNA (171, 172). However, *in vitro* studies suggested that IHF stimulates the expression of the *c*II gene at the transcriptional level, possibly by acting as an anti-terminator (173, 174). IHF is also known to be directly involved in stimulation of the phage lambda P_L promoter by enhancing promoter recognition by the RNA polymerase protein (175, 176). This λ P_L promoter is responsible for the expression of at least 14 genes, including the *N* gene required for the λ lytic cycle and the *c*III gene involved in the lysogenic cycle (175, 177). IHF recognizes two sites upstream from the P_L promoter, induces DNA bending upon binding and appears to contact the α C-terminal domain of RNA polymerase (αCTD) (175, 176, 178). Moreover, the DNA bend induced by IHF brings into proximity the αCTD and the UP element, a distal element located upstream from the promoter, thereby allowing the establishment of close contacts between them and increasing the affinity of RNA polymerase to the P_L promoter region (179).

IHF is also involved in the transcriptional activation of some coliphage Mu genes. This protein modulates transcription differentially from promoters P_E and P_c: while stimulating transcription from P_E, it weakens the P_c promoter thereby favoring the lytic cycle (164, 180). Moreover, like the λ P_L promoter, activation of P_E by IHF is dependent on the presence of the αCTD of the α subunit of RNA polymerase (181). Binding of IHF to the IHF binding site located between operator sites O1 and O2 was shown to be responsible both *in vitro* and *in vivo* for the activation of early transcription in the absence of the c repressor (182). In fact, in the IHF binding region, two IHF consensus sequences named *ihfa* and *ihfb* are present. These two sequences are partially overlapping and in opposite orientation (136, 183). However, IHF shows a strong preference for *ihfb*, primarily because of the bent structure of the flanking DNA in the Mu regulatory region and indeed, *ihfb* is the only site active in stimulation of transcription from both promoters (136). The differential regulation of P_E and P_c by IHF bound to the same site might be accomplished by the different nature of the IHF-DNA complexes it can produce. Activation of transcription at the P_E promoter only requires a weak IHF-DNA complex and may not be mediated by DNA bending but, presumably, by protein-protein interactions between IHF and RNA polymerase (136). On the other hand, activation of transcription at the P_c promoter requires a strong IHF-DNA complex and in this case, conformational changes induced by IHF seems to be crucial (136). It was further shown that by binding to its binding site in the Mu operator region, IHF plays a dual role in the early transcription of coliphage Mu. In the presence of the phage-encoded c repressor, IHF enhances repression of early transcription *in vivo* by stimulating repressor binding to the O1 and O2 operators, due to the formation of a stable nucleoprotein complex with the early operator P_E (184, 185). However, when repressor is absent, IHF activates early transcription leading to the lytic cycle (182). Furthermore, IHFs role appears to be primarily architectural by building a stable repression complex in cooperation with the Mu repressor c: IHF stabilizes the interactions between Mu repressor and the early operator (182). Additionally, IHF is able to counteract the H-NS mediated repression of the P_E promoter both *in vivo* and *in vitro* by interfering with the formation of a stable H-NS/DNA complex (181). The exact mechanism allowing this alleviation of H-NS repression by IHF is not well understood

(181), but seems to be crucial for the lytic development of bacteriophage Mu. The phage cannot become lytic in the presence of H-NS and the absence of IHF (180, 181).

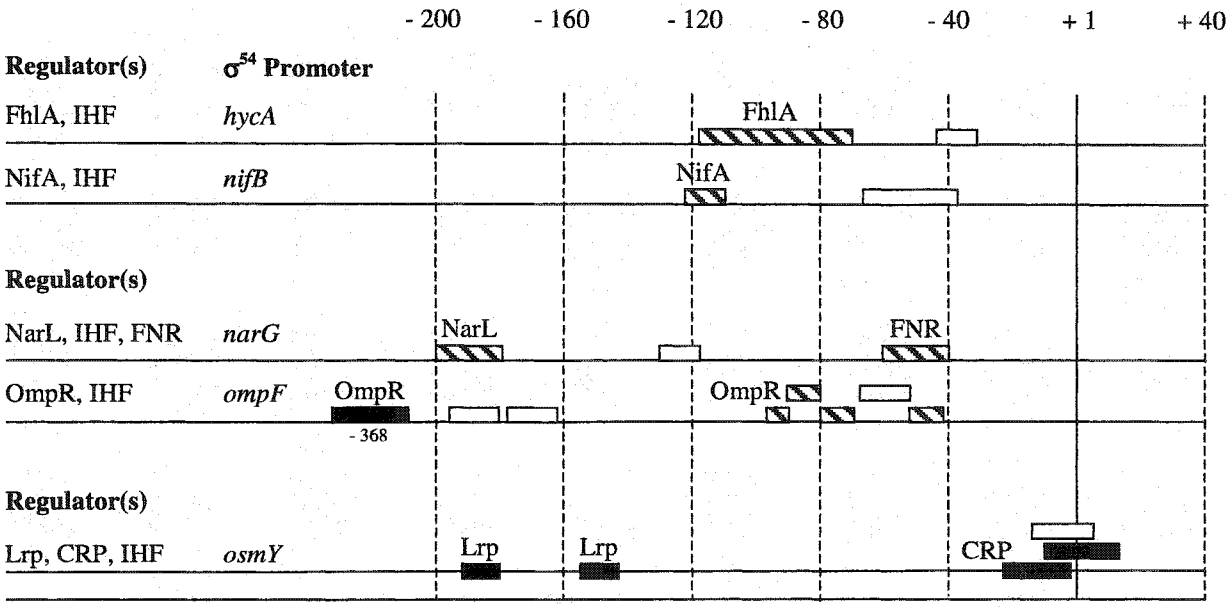
1.4.4 Control of bacterial gene expression

Statistical analyses and two-dimensional gel electrophoresis suggested that the *E. coli* chromosome contains approximately 80 to 100 putative IHF binding sites (116, 186) suggesting a role for this protein in gene regulation in *E. coli*. Indeed, several studies suggested that IHF interacted with many sites on the bacterial chromosome and was directly involved in the regulation of expression of many genes in *E. coli* [reviewed in (51, 52, 53, 54); see also (187)].

1.4.4.1 Coactivation of σ^{54} promoters

IHF consensus sequences are located around -40 to -50 bp from the start of transcription in about half of the known nitrogen-regulated σ^{54} promoters, but does not directly stimulate transcription from these promoters (188). Instead, IHF appears to bring into proximity to the promoter region an activator, bound to the upstream activator site (UAS) (Fig. 1.7). IHF binding sites are normally located between the UAS and the RNA polymerase recognition sequence (Fig. 1.7) and can therefore coactivate transcription by bringing the activator and the σ^{54}-RNA polymerase together by the formation of a loop. The activator is therefore brought into close proximity to the RNA polymerase and interactions can occur between them (53, 188, 189, 190). IHF can also enhance the binding of the activator to the DNA (191), the binding of the σ^{54}-RNA polymerase to the promoter (190), and it may stimulate the interaction between the activator and the RNA polymerase (53). Furthermore, IHF can coactivate transcription thanks to direct contacts it can form with the RNA polymerase or with the activator itself (53). Further experiments have shown that transcription initiation requires a promoter with high affinity for σ^{54}-RNA polymerase, as well as a bend in the DNA, intrinsic or induced by IHF, located between the enhancer and the promoter in the case of linear

Figure 1.7 Examples of σ^{54}, σ^{70}, and σ^{S} promoter organization (188, 192).
The organization of promoters *hycA*, *nifB*, *narG*, and *ompF* was adapted from (188).
Symbols : ▭ IHF binding site; ▨ activator binding site (upstream activator site; UAS); ▬ repressor binding site.


- 200
- 160
- 120
- 80
- 40
+ 1
+ 40
Regulator(s)
σ54 Promoter
FhlA, IHF
hycA
FhlA
NifA, IHF
nifB
NifA
Regulator(s)
NarL, IHF, FNR
narG
NarL
FNR
OmpR, IHF
ompF
OmpR
- 368
OmpR
Regulator(s)
Lrp, CRP, IHF
osmY
Lrp
Lrp
CRP

DNA. However, when the DNA is supercoiled, transcription initiation only requires one of these characteristics, but not both (144, 189, 193).

Examples of σ^{54} promoters coactivated by IHF include *glnAp2* and *glnHp2* in *E. coli* and *Salmonella typhimurium*, and the *nif* promoters (*nifB*, *E*, *LA*, *H*, *J* and *U*) in *Klebsiella pneumoniae* and in many nitrogen-fixing organisms classified in the "purple bacteria" phylum (53, 144, 194, 195, 196, 197) (Fig. 1.7). Other examples of promoters coactivated by IHF include the *hycA* and *hypA* promoters of the formate hydrogenase system of *E. coli* (198) (Fig. 1.7), the *pspA* promoter (phage shock protein operon of *E. coli*) (191, 199, 200), and the Pu promoter of the upper operon (involved in the oxidative transfomation of toluene/xylenes to the corresponding benzoate/toluates) of the *Pseudomonas putida* TOL plasmid (201).

Interestingly, the IHF protein may not only coactivate transcription from σ^{54} promoters by helping in the formation of an active complex but it may also inhibit non specific activation by heterologuous regulators due to a distortion of the DNA structure (200, 202).

1.4.4.2 Regulation of transcription at σ^{70} promoters

As previously seen, IHF is a negative repressor of its own synthesis by acting at the *ihfA* and *ihfB* σ^{70} promoters (104, 105) (Fig. 1.1). However, from DNA microarray experiments, IHF appears to be involved (either directly or indirectly) in the expression of numerous genes expressed from σ^{70} promoters (187).

Interestingly, it was shown that where IHF is an activator of transcription at σ^{70} promoters, it does not interact directly with RNA polymerase to activate initiation. On the contrary, IHF has only an architectural role consisting of bending the DNA to allow the interaction between a region of DNA upstream from the IHF binding site and RNA polymerase (203). The IHF protein has been shown to be involved in positive regulation of expression of the isoleucine and valine (Ilv) enzymes from the *ilvBN* and *ilvGMEDA* operons, possibly by decreasing transcriptional pausing and termination (204, 205, 206). IHF also positively regulates the *pst* operon (genes coding for the inducible inorganic phosphate transport system) by binding weakly to a site located immediately

downstream from the promoter, and as a consequence, indirectly controls the expression of *phoA* (alkaline phosphatase gene) since they both belong to the PHO regulon (207). IHF also stimulates expression of the *xyl* genes (205), the group 2 capsule gene clusters (K antigens) (208), the nitrate reductase *narGHJI* operon (209, 210) (Fig. 1.7), the biodegradative threonine deaminase *tdc* operon (211), the L-cysteine biosynthesis *cysJIH* operon (212), the conjugal transfer operon (*tra*) (213), and the *pifCAB* operon (involved in inhibition of phage T7 growth) (51, 214) of the plasmid F, to name a few.

However, IHF can also exert an inhibitory effect on σ^{70} promoters, and repressor effects have been observed in σ^{70} promoters with IHF-binding sites, although the mechanisms are still unknown. Still, many IHF effects appear to be direct with no other regulatory protein involved (214), (215). For example, IHF was found to inhibit transcription of the *ompB* operon, coding for the OmpR and EnvZ regulatory proteins of *ompF* and *ompC*, by directly binding in the -10 and -35 region of the major promoter (216). Expression of the two outer membrane porin proteins OmpF and OmpC are also inhibited by IHF as a consequence of its negative effect on *ompB* expression. However, the *ompF* gene is also repressed indirectly by modulation of the activity of the regulator protein OmpR by IHF (217) (Fig. 1.7). Expression of *ompC* is also repressed through conformational changes in the DNA induced by IHF (215). In *E. coli*, IHF was shown to repress a *Chlamydomonas* chloroplast promoter, P_A, presumably by binding to a site overlapping the promoter along with the binding of other proteins upstream from the promoter (218).

1.4.4.3 Regulation of σ^S promoters

IHF was shown to participate in stationary-phase-induced expression of the *osmY* gene, coding for a periplasmic protein, and the *dps* gene, coding for a non-specific DNA-binding protein from starved cells important for defence against hydrogen peroxide (219, 220). In the case of the *osmY* gene, IHF seems to help in establishing σ^S (σ^{38} or stationary phase σ factor) selectivity at the *osmY* promoter. In fact, IHF inhibits transcription initiation at the *osmY* promoter along with regulatory proteins CRP and Lrp, but the repression is much stronger with σ^{70}-RNA polymerase than with σ^S-RNA

polymerase (192, 219) (Fig. 1.7). As for *dps* gene regulation, expression from this gene in stationary phase is activated by IHF which binds upstream of the *dps* promoter along with the transcriptional activator OxyR (220).

1.4.5 Control of DNA replication

IHF was shown to be involved in DNA replication in 1986 with the study of plasmid pSC101 replication *in vitro* (221). An IHF binding site was found in the essential replication region of the plasmid and located between the replication initiator proteins DnaA and RepA binding sites. IHFs role in DNA replication was hypothesized to be structural by allowing the formation of a replication complex (221, 222). IHF also enhances the binding of DnaA to *ori*γ of plasmid R6K *in vitro* (223) and the IHF-induced bend was shown to be crucial in initiation of replication of pKL1, a small cryptic plasmid of *E. coli* (224). For replication of these three plasmids, pSC101, R6K, and pKL1, IHF enhances the binding of DnaA and a plasmid-encoded initiator protein (RepA for pSC101 and pKL1, and π protein for R6K) at distant sites cooperatively (223, 224).

The same role for IHF was also suggested in the case of *E. coli oriC*, where it is also thought to be a component of the multiprotein complex which forms at the origin of replication and seems to be important in the stabilization of the open complex at the replication origin (225, 226). In the case of *E. coli* chromosome replication, IHF seems to be a crucial part in the cell cycle initiation timing mechanism (227): IHF-induced bending enhanced a redistribution of the DnaA protein at *oriC* from stronger to weaker sites and thereby reduced the DnaA level required to unwind *oriC* and trigger initiation of DNA replication (227).

1.4.6 Role of IHF in other processes

IHF has also been shown to be involved in a variety of other processes in the cell. For example, IHF is involved in phage λ and lambdoid phage 21 packaging (127, 228): IHF is thought to facilitate binding of the terminase protein to phage 21 DNA (228) and enhance terminase binding and *cos* site cutting via modification of λ DNA structure (127). IHF is also known to be directly involved in plasmid P1 partitioning since it forms the partition complex, along with the ParB protein, at the centromere-like *parS* site (229, 230, 231). Moreover, IHF may be involved in F and R100 conjugal plasmids transfer since their transfer is reduced in *ihfA* and *ihfB* mutants (213, 232). Three putative IHF binding sites were also found near *oriT* of R100, where the nick initiating the transfer of single-stranded DNA appears (51, 233). Furthermore, IHF may also be involved in phase variation expression of flagellin (234) and type-1 fimbriae (235) through site-specific DNA inversion processes.

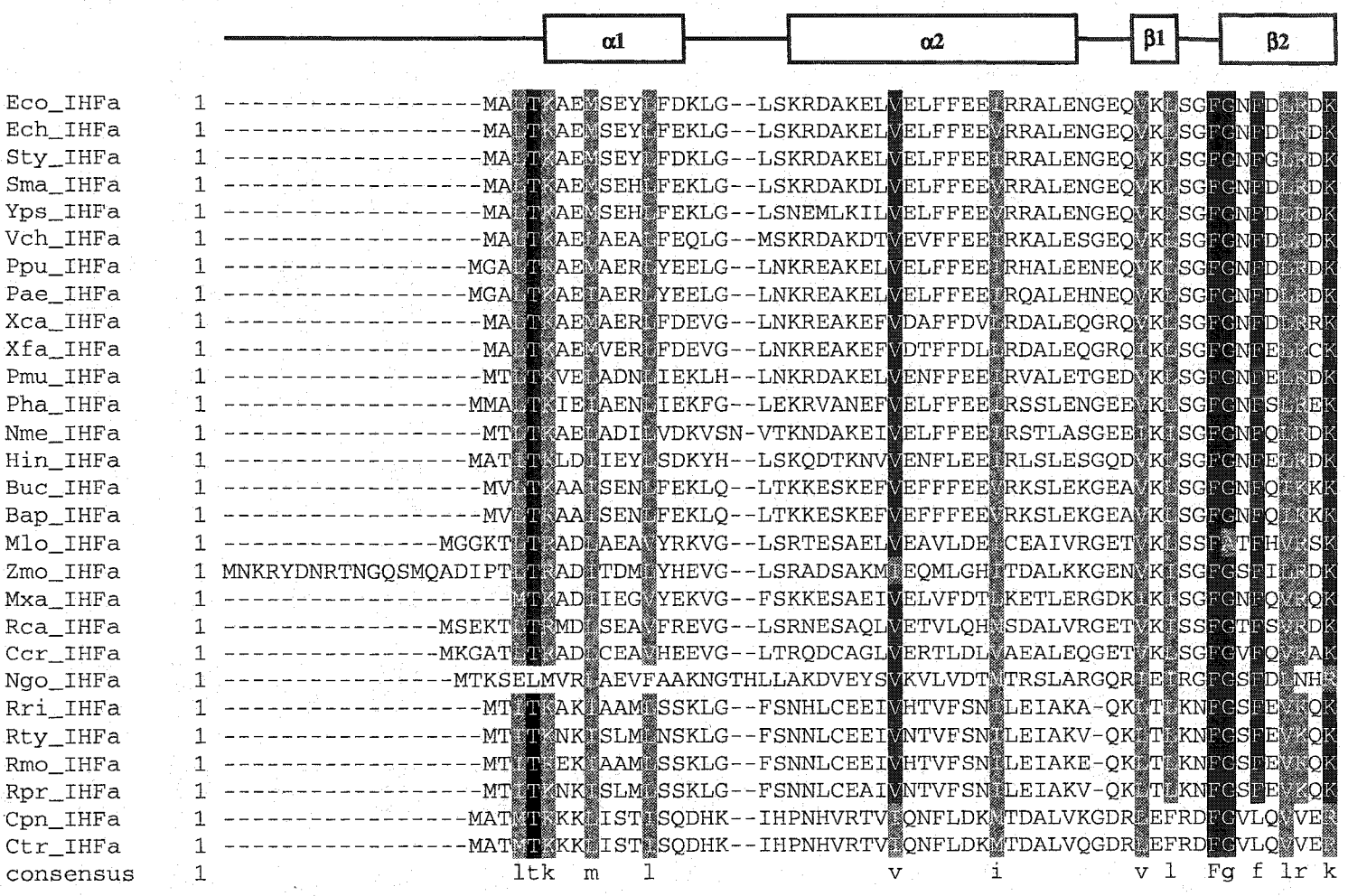
Finally, IHF was found to bind to a site located between two inverted conserved repetitive extragenic palindromic sequences (REP) in *E. coli* (236) forming a RIP (repetitive IHF-binding palindromic) element (237), also labelled RIB (reiterative *ihf* bacterial interspersed mosaic elements) (238). These REP sequences, also called PU (palindromic units), are dispersed throughout the genome of *E. coli* and could account for about 1% of the chromosome (236, 239). There are often associated with other repetitive elements forming about 500 clusters in the *E. coli* chromosome termed BIME (bacterial interspersed mosaic elements) (240). The RIP elements are a subclass of BIME, called BIME-1 (241). Approximately 70 to 100 RIP elements of about 100 bp long were found in the *E. coli* chromosome (237, 238). The role of these RIP elements is still unknown. It has been suggested that IHF may stabilize the formation of a DNA gyrase-REP complex (242) or that it may enhance cooperative binding of DNA gyrase to two adjacent REP sequences (237). The higher-order DNA structure created by IHF and DNA gyrase binding may also help maintain negative supercoiling and influence transcription of adjacent genes (237, 242).

After reviewing its known roles in *E. coli*, it is clear that IHF is a pleiotropic protein with involvement in many different systems and mechanisms in the cell. Its participation is also important in the cycle of "genetic free agents" (51) like phages, transposons and plasmids, since it is involved in the processes of recombination, replication, partitioning and transfer (51). However, IHF is always an accessory factor which modulates many processes in the cell more or less markedly, as its effects may be more subtle like in promoter activation or more pronounced like in λ site-specific recombination. The IHF protein has been extensively studied in *E. coli*, but homologous proteins have also been found in related bacteria and have also been named IHF. Comparing their amino acid sequence could be a method to understand how this protein was conserved through evolution and to give insight into the importance of this protein in bacteria.

1.5 Evolutionary conservation of IHF

Many studies have suggested that the IHF protein is present in bacteria other than *E. coli* and that it is relatively conserved. In 1991, Haluzi *et al.* compared the *ihfA* and *ihfB* genes of *Serratia marcescens* to the IHF genes of *E. coli* (67). They showed that both genes had a very similar sequence and that they were located downstream from the same genes as in *E. coli* (*pheT* for *ihfA* and *rpsA* for *ihfB*). However, the sequences following both genes were highly divergent. The protein sequences of these genes were also highly conserved: the IHFα and IHFβ subunits of *S. marcescens* only had five and six amino acids different, respectively, from the *E. coli* subunits and eight of these changes were conservative substitutions (Fig. 1.8 and 1.9). They also showed that the *ihfA* and *ihfB* genes (sequences unknown) of *Aeromonas proteolytica*, like for *S. marcescens*, could substitute for the *E. coli* genes in various assays. They hypothesized that the IHF genes were probably evolutionary conserved in Gram-negative bacteria (67). Other *ihf* genes were found in different bacteria with the same chromosomal localization as in *E. coli*. The *ihfA* gene of *Salmonella typhimurium* (57), *Erwinia chrysanthemi* (60), *Pasteurella haemolytica* (63), and *Buchnera aphidicola* (77) were all found downstream of the *pheT* gene or a gene with a similar sequence to the *E. coli pheT* gene, with the exception of *Rhodobacter capsulatus ihfA* gene (80). As well, the *ihfB* gene of *N. gonorrhoeae* (66), *P. haemolytica* (63), *E. chrysanthemi* (60), and *B. aphidicola* (64) were found downstream of the *rpsA* gene. Moreover, other IHF proteins displayed similar characteristics as *E. coli* IHF. For example, *N. gonorrhoeae* IHF displays a similar primary structure, similar binding characteristics and seems to be functionally similar to *E. coli* IHF (66). However, both *ihfA* and *ihfB* mRNA levels decline when cells enter stationary phase unlike in *E. coli* where they increase, and the regulatory mechanism for transcription of the *N. gonorrhoeae ihf* genes is different from that in *E. coli* (66). Interestingly, transcription from *P. haemolytica ihfA* is also different from the one in *E. coli*, since the potential promoter region was not localized to the same location (63). Therefore, it appears that, as suggested by Hill *et al.* (66), regulation of IHF expression may not be very well conserved in Gram-negative bacteria.

Figure 1.8 Comparison of IHFα amino acid sequences from different bacteria. Amino acid sequences were obtained from the Entrez Protein database of the National Center for Biotechnology Information with the accession numbers listed in Table 1.2. Multiple alignments were generated by ClustalW 1.75 (243) and elaborated with Boxshade 3.21 (244). Amino acid sequences are aligned underneath the secondary structure prediction established by Rice *et al.* (130). IHFα amino acid sequences from the following organisms were analyzed: *E. coli* (Eco), *E. chrysanthemi* (Ech), *S. typhimurium* (Sty), *S. marcescens* (Sma), *Y. pseudotuberculosis* (Yps), *V. cholerae* (Vch), *P. putida* (Ppu), *P. aeruginosa* (Pae), *X. campestris* (Xca), *X. fastidiosa* (Xfa), *P. multocida* (Pmu), *P. haemolytica* (Pha), *N. meningitidis* (Nme), *H. influenzae* (Hin), *Buchnera* species strain APS (Buc), *B. aphidicola* (Bap), *M. loti* (Mlo), *Z. mobilis* (Zmo), *M. xanthus* (Mxa), *R. capsulatus* (Rca), *C. crescentus* (Ccr), *N. gonorrhoeae* (Ngo), *R. rickettsii* (Rri), *R. typhi* (Rty), *R. montanensis* (Rmo), *R. prowazekii* (Rpr), *C. pneumoniae* (Cpn), *C. trachomatis* (Ctr); for references, see Table 1.2. To improve the homology score, gaps (-) were added in the sequence by the ClustalW program. A black font indicates that the amino acid residues are identical for 90% of the sequences aligned, while a grey font indicates conservative substitutions.


α1
α2
β1
β2
Eco_IHFa 1
Ech_IHFa 1
Sty_IHFa 1
Sma_IHFa 1
Yps_IHFa 1
Vch_IHFa 1
Ppu_IHFa 1
Pae_IHFa 1
Xca_IHFa 1
Xfa_IHFa 1
Pmu_IHFa 1
Pha_IHFa 1
Nme_IHFa 1
Hin_IHFa 1
Buc_IHFa 1
Bap_IHFa 1
Mlo_IHFa 1
Zmo_IHFa 1
Mxa_IHFa 1
Rca_IHFa 1
Ccr_IHFa 1
Ngo_IHFa 1
Rri_IHFa 1
Rty_IHFa 1
Rmo_IHFa 1
Rpr_IHFa 1
Cpn_IHFa 1
Ctr_IHFa 1
consensus 1

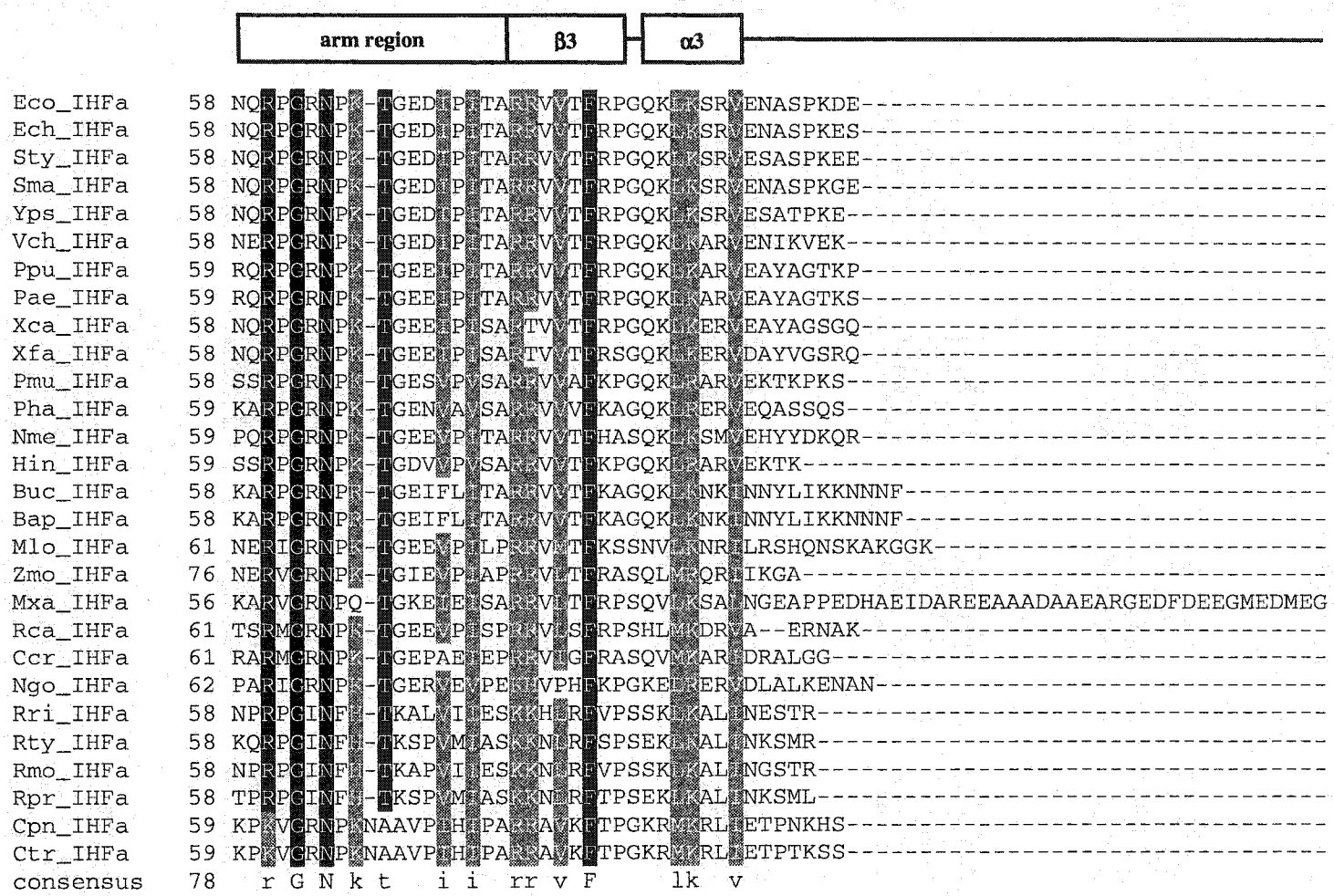
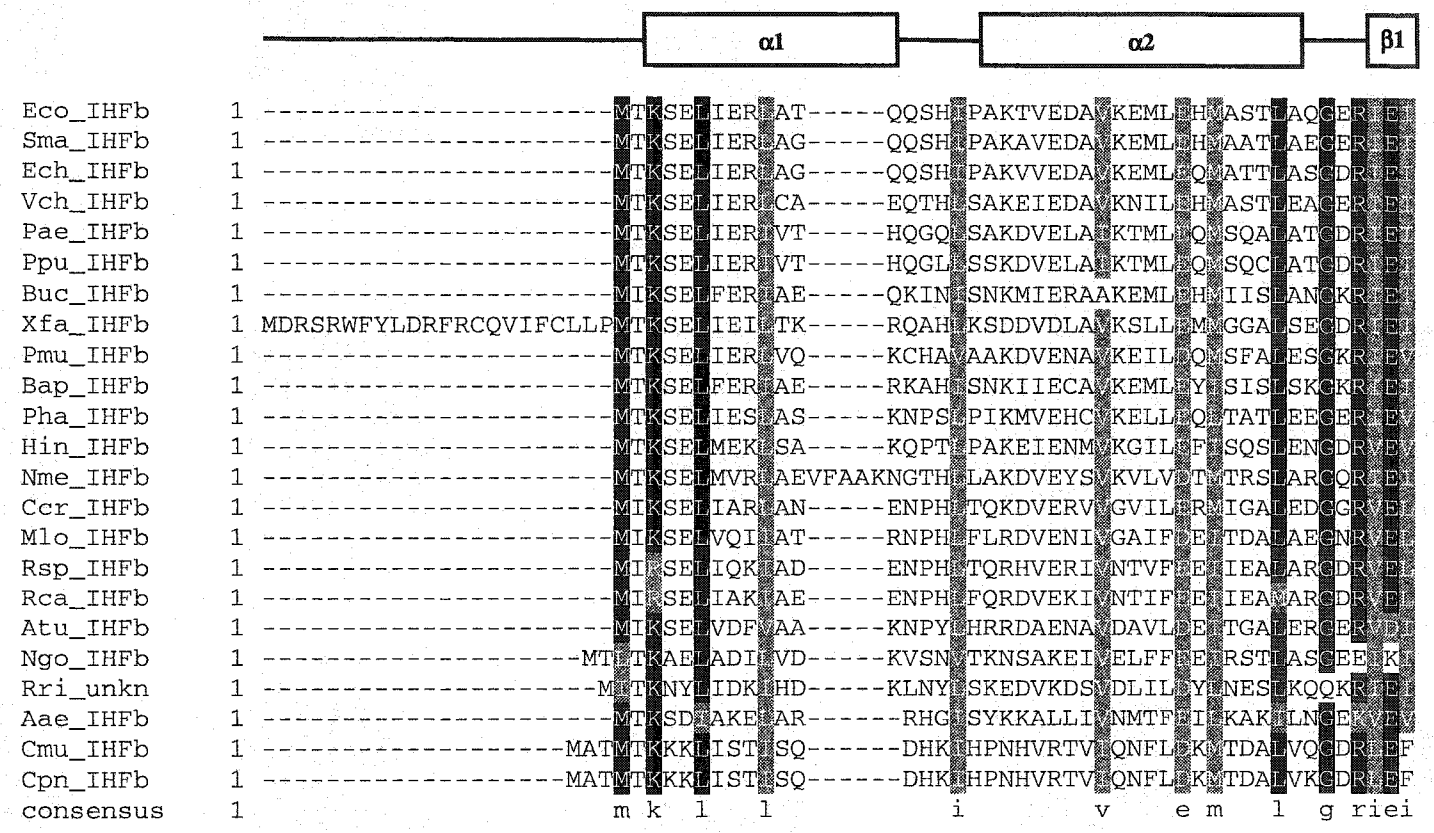
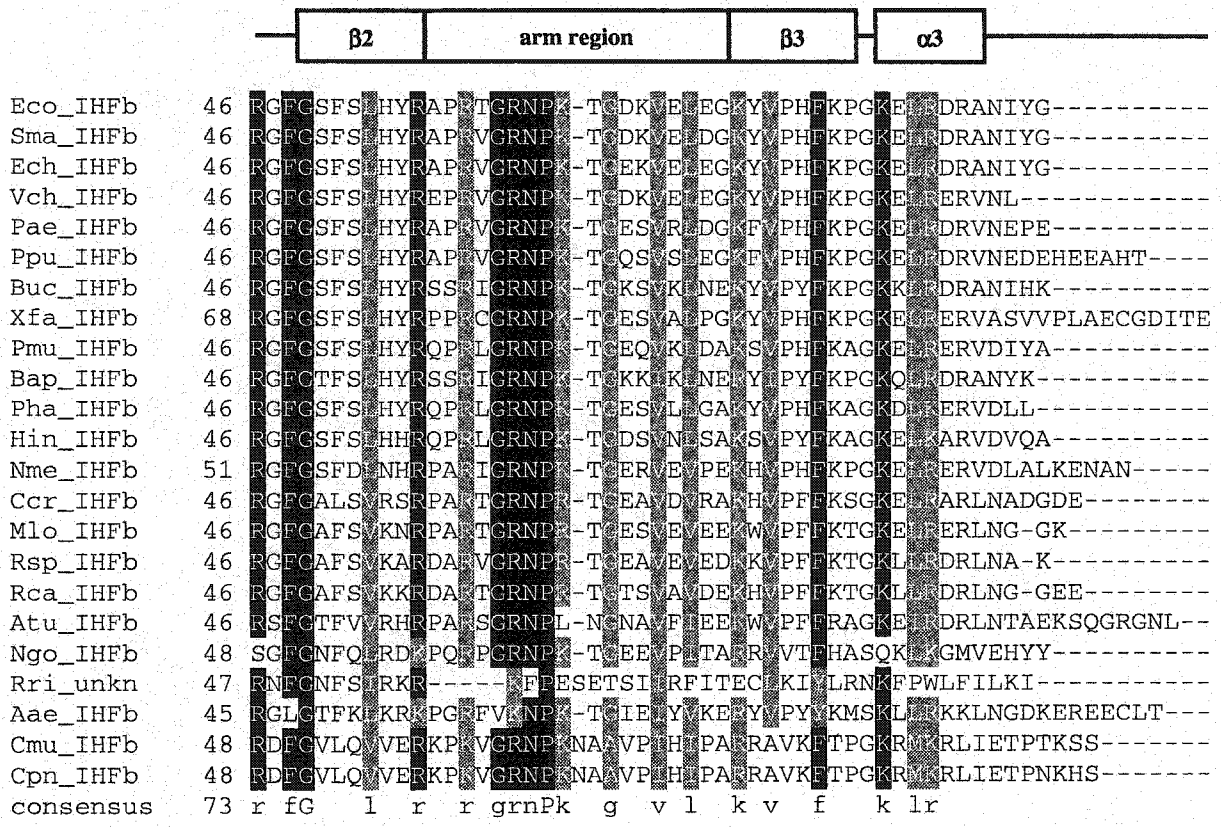

arm region
β3
α3
Eco_IHFa 58 NQRPGRNPK-TGEDIPITARRVVTFRPGQKLKSRVENASPKDE
Ech_IHFa 58 NQRPGRNPK-TGEDIPITARRVVTFRPGQKLKSRVENASPKES
Sty_IHFa 58 NQRPGRNPK-TGEDIPITARRVVTFRPGQKLKSRVESASPKEE
Sma_IHFa 58 NQRPGRNPK-TGEDIPITARRVVTFRPGQKLKSRVENASPKGE
Yps_IHFa 58 NQRPGRNPK-TGEDIPITARRVVTFRPGQKLKSRVESATPKE
Vch_IHFa 58 NERPGRNPK-TGEDIPITARRVVTFRPGQKLKARVENIKVEK
Ppu_IHFa 59 RQRPGRNPK-TGEEIPITARRVVTFRPGQKLKARVEAYAGTKP
Pae_IHFa 59 RQRPGRNPK-TGEEIPITARRVVTFRPGQKLKARVEAYAGTKS
Xca_IHFa 58 NQRPGRNPK-TGEEIPISARTVVTFRPGQKLKERVEAYAGSGQ
Xfa_IHFa 58 NQRPGRNPK-TGEEIPISARTVVTFRSGQKLKERVDAYVGSRQ
Pmu_IHFa 58 SSRPGRNPK-TGESVPVSARRVVAFKPGQKLKARVEKTKPKS
Pha_IHFa 59 KARPGRNPK-TGENVAVSARRVVVFKAGQKLKERVEQASSQS
Nme_IHFa 59 PQRPGRNPK-TGEEVPITARRVVTFHASQKLKSMVEHYYDKQR
Hin_IHFa 59 SSRPGRNPK-TGDVVPVSARRVVTFKPGQKLKARVEKTK
Buc_IHFa 58 KARPGRNPR-TGEIFLITARRVVTFKAGQKLKNKINNYLIKKNNNF
Bap_IHFa 58 KARPGRNPR-TGEIFLITARRVVTFKAGQKLKNKINNYLIKKNNNF
Mlo_IHFa 61 NERICRNPK-TGEEVPILPRRVVTFKSSNVLKNRILRSHQNSKAKGGK
Zmo_IHFa 76 NERVCRNPK-TGIEVPLAPRRVVTFRASQLMKQRIIKGA
Mxa_IHFa 56 KARVCRNPQ-TGKEIESARRVVTFRPSQVMKSALNGEAPPEDHAEIDAREEAAADAAEARGEDFDEEGMEDMEG
Rca_IHFa 61 TSRMCRNPK-TGEEVPISPRRVLSFRPSHLMKDRVA--ERNAK
Ccr_IHFa 61 RARMCRNPK-TGEPAEIEPRRVLGFRASQVMKARIDRALGG
Ngo_IHFa 62 PARICRNPK-TGERVEVPERRVPHFKPGKELRERVDLALKENAN
Rri_IHFa 58 NPRPGINFI-TKALVIIESKKHIRFVPSSKLKALINESTR
Rty_IHFa 58 KQRPGINFI-TKSPVMIASKKNIRFSPSEKLKALINKSMR
Rmo_IHFa 58 NPRPGINFI-TKAPVIIESKKNIRFVPSSKLKALINGSTR
Rpr_IHFa 58 TPRPGINFI-TKSPVMIASKKNIRFTPSEKLKALINKSML
Cpn_IHFa 59 KPKVGRNPKNAAVPIHIPARRAVKFTPGKRMKRLIETPNKHS
Ctr_IHFa 59 KPKVGRNPKNAAVPIHIPARRAVKFTPGKRMKRLIETPTKSS
consensus 78 r G N k t i i rr v F lk v

Figure 1.9 Comparison of IHFβ amino acid sequences from different bacteria. Amino acid sequences were obtained from the Entrez Protein database of the National Center for Biotechnology Information with the accession numbers listed in Table 1.3. Multiple alignments were generated by ClustalW 1.75 (243) and elaborated with Boxshade 3.21 (244). Amino acid sequences are aligned underneath the secondary structure prediction established by Rice *et al.* (130). IHFβ amino acid sequences from the following organisms were analyzed: *E. coli* (Eco), *S. marcescens* (Sma), *E. chrysanthemi* (Ech), *V. cholerae* (Vch), *P. aeruginosa* (Pae), *P. putida* (Ppu), *Buchnera* species strain APS (Buc), *X. fastidiosa* (Xfa), *P. multocida* (Pmu), *B. aphidicola* (Bap), *P. haemolytica* (Pha), *H. influenzae* (Hin), *N. meningitidis* (Nme), *C. crescentus* (Ccr), *M. loti* (Mlo), *R. spaeroides* (Rsp), *R. capsulatus* (Rca), *A. tumefaciens* (Atu), *N. gonorrhoeae* (Ngo), *R. rickettsii* (Rri), *A. aeolicus* (Aae), *C. muridarum* (Cmu), *C. pneumoniae* (Cpn); for references, see Table 1.3. To improve the homology score, gaps (-) were added in the sequence by the ClustalW program. A black font indicates that the amino acid residues are identical for 90% of the sequences aligned, while a grey font indicates conservative substitutions.


α1
α2
β1
Eco_IHFb 1
Sma_IHFb 1
Ech_IHFb 1
Vch_IHFb 1
Pae_IHFb 1
Ppu_IHFb 1
Buc_IHFb 1
Xfa_IHFb 1
Pmu_IHFb 1
Bap_IHFb 1
Pha_IHFb 1
Hin_IHFb 1
Nme_IHFb 1
Ccr_IHFb 1
Mlo_IHFb 1
Rsp_IHFb 1
Rca_IHFb 1
Atu_IHFb 1
Ngo_IHFb 1
Rri_unkn 1
Aae_IHFb 1
Cmu_IHFb 1
Cpn_IHFb 1
consensus 1


β2
arm region
β3
α3
Eco_IHFb 46 RGFGSFSLHYRAPRTGRNPK-TGDKVELEGKYVPHFKPGKELRDRANIYG----------
Sma_IHFb 46 RGFGSFSLHYRAPRVGRNPK-TGDKVELDGKYVPHFKPGKELRDRANIYG----------
Ech_IHFb 46 RGFGSFSLHYRAPRVGRNPK-TGEKVELEGKYVPHFKPGKELRDRANIYG----------
Vch_IHFb 46 RGFGSFSLHYREPRVGRNPK-TGDKVELEGKYVPHFKPGKELRERVNL-----------
Pae_IHFb 46 RGFGSFSLHYRAPRVGRNPK-TGESVRLDGKFVPHFKPGKELRDRVNEPE----------
Ppu_IHFb 46 RGFGSFSLHYRAPRVGRNPK-TGQSVSLEGKFVPHFKPGKELRDRVNEDEHEEAHT----
Buc_IHFb 46 RGFGSFSLHYRSSRIGRNPK-TGKSVKLNEKYVPYFKPGKKLRDRANIHK----------
Xfa_IHFb 68 RGFGSFSLHYRPPRCGRNPK-TGESVALPGKYVPHFKPGKELRERVASVVPLAECGDITE
Pmu_IHFb 46 RGFGSFSLHYRQPRLGRNPK-TGEQVKLDASVPHFKAGKELRERVDIYA----------
Bap_IHFb 46 RGFGTFSLHYRSSRIGRNPK-TGKKLKLNEKYVPYFKPGKQLRDRANYK----------
Pha_IHFb 46 RGFGSFSLHYRQPRLGRNPK-TGESVLLGAKYVPHFKAGKDLRERVDLL----------
Hin_IHFb 46 RGFGSFSLHHRQPRLGRNPK-TGDSVNLSAKSVPYFKAGKELRARVDVQA----------
Nme_IHFb 51 RGFGSFDLNHRPARIGRNPK-TGERVEVPEKHVPHFKPGKELRERVDLALKENAN-----
Ccr_IHFb 46 RGFGALSVRSRPARTGRNPR-TGEAVDVRAKHVPFFKSGKELRARLNADGDE--------
Mlo_IHFb 46 RGFGAFSVKNRPARTGRNPR-TGESVEVEEKWVPFFKTGKELRERLNG-GK---------
Rsp_IHFb 46 RGFGAFSVKARDARVGRNPR-TGEAVEVEDKKVPFFKTGKLLRDRLNA-K----------
Rca_IHFb 46 RGFGAFSVKKRDARTGRNPK-TGTSVAVDEKHVPFFKTGKLLRDRLNG-GEE--------
Atu_IHFb 46 RSFGTFVVRHRPARSGRNPL-NGNAVFIEEKWVPFFRAGKELRDRLNTAEKSQGRGNL--
Ngo_IHFb 48 SGFGNFQLRDKPQRPGRNPK-TGEEVPLTARRVTFHASQKLKGMVEHYY----------
Rri_unkn 47 RNFGNFSIRKR-----RFPESETSIRRFITECKIYLRNKFPWLFILKI-----------
Aae_IHFb 45 RGLGTFKLKRKPGRFVKNPK-TGIEVYKEKYVPYYKMSKLLRKKLNGDKEREECLT---
Cmu_IHFb 48 RDFGVLQVVERKPKVGRNPKNAAVPIHIPARRAVKFTPGKRMKRLIETPTKSS-------
Cpn_IHFb 48 RDFGVLQVVERKPKVGRNPKNAAVPLHLPARRAVKFTPGKRMKRLIETPNKHS-------
consensus 73 r fG l r r grnPk g v l k v f k lr

However, the most striking fact is that IHF roles in other bacteria appear to be similar to that observed in *E. coli*. This was suggested by the localization of its binding sites and the cellular processes (site-specific recombination, transposition, gene expression) in which it was shown to play a role (62, 80, 190, 194, 201, 212, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261).

The amino acid sequences of the different IHF proteins analysed thus far were compared, and they appear to be relatively well conserved (Tables 1.2 and 1.3; Figs. 1.8 and 1.9). Of the 27 *E. coli* IHFα homologues, 15 are more than 62% identical to *E. coli* IHFα at the amino acid level (Table 1.2). These 15 IHFα polypeptides are found in bacteria belonging to the gamma-subdivision of Proteobacteria, like *E. coli*, except for *N. meningitidis* which is in the beta-subdivision. The remaining IHFα homologues are found in bacteria of the alpha-, beta- and delta-subdivisions of the Proteobacteria order, and of the Chlamydiales order with the least identical protein being 26% identical at the amino acid level to *E. coli* IHFα. Of the 22 *E. coli* IHFβ homologues, 11 are more than 57% identical to *E. coli* IHFα at the amino acid level (Table 1.3) and they are all found in bacteria belonging to the gamma-subdivision of Proteobacteria. The other IHFβ homologues are found in bacteria of the alpha- and beta-subdivisions of the Proteobacteria order, and of the Aquificales and Chlamydiales orders with the least identical protein being 31% identical at the amino acid level to *E. coli* IHFβ. It appears from these two tables that the majority of IHF polypeptides analysed thus far are found in the Proteobacteria, except for the IHFβ protein of *A. aeolicus*, which belongs to the Aquificales, a lineage of thermophilic bacteria which may have been one of the earliest divergences in the eubacterial tree (262), and for the IHFα proteins of *C. pneumoniae* and *C. trachomatis* and the IHFβ proteins of *C. pneumoniae* and *C. muridarum*, which are all obligate intracellular eubacteria belonging to the Chlamydiales order (84).

Multiple alignments of the available sequences for all IHF proteins identified were generated by the ClustalW 1.75 program (243) and elaborated with Boxshade 3.21 (244) (Figs. 1.8 and 1.9). This comparison of IHF protein sequences in different bacteria shows that 17 amino acids for IHFα and 16 for IHFβ are conserved in all proteins known so far. Most of the conserved amino acid residues are not identical in all proteins but have similar physico-chemical properties and appear to be conservative

changes. Indeed, only four amino acids are identical in all IHFα proteins and two in IHFβ proteins. The conserved amino acids may indicate positions where an important evolutionary constraint was applied in order to maintain IHF structure and functions. The arm regions of *E. coli* IHFα and IHFβ are believed to be involved in the binding of IHF in the minor groove of the DNA (20) and interestingly, four residues are conserved in the arm region of all IHFα proteins known and seven out of the 18 residues constituting the arm region are conserved in 90% of the proteins (Fig. 1.8). The arm region of *E coli* IHFα is also thought to contact the 5'-ATCAA element of the IHF consensus sequence (129). It can be hypothesized that the IHF binding domains in bacterial species studied may have a high degree of similarity to the *E. coli* IHF consensus sequence since this was shown for the *N. gonorrhoeae pilE* promoter (257) and since *E. coli* IHF was shown to substitute for the *C. crescentus*, *R. spaeroides*, and *S. typhimurium* IHF proteins in different studies (246, 252, 261). In the case of IHFβ, three residues are conserved in all IHFβ proteins and half of the residues of the arm are conserved in 90% of the proteins (Fig. 1.9). The IHFα α3 helix, which is thought to be important for the binding specificity of IHF (119), also contains three conserved amino acids in all IHFα proteins studied (Fig. 1.8). This α3 helix and the IHFα β1 sheet, which is very well conserved (Fig. 1.8), are thought to contact the distal A/T-rich element present upstream of the IHF consensus sequence (129). In the IHFβ α3 helix, which has been proposed to be important for the stability of the IHF-DNA complex and for the binding affinity of IHF (119), three amino acid residues are conserved in all proteins studied except *N. gonorrhoeae* IHFβ for Q84β [glutamine (Q) amino acid residue number 84 of the β subunit] and *R. rickettsii* IHFβ for P86β and W87β (Fig. 1.9). This α3 helix and the IHFβ β1 sheet, which is very well conserved (Fig. 1.9), are thought to contact the TT-3' element of the *E. coli* IHF consensus sequence (119, 129). The *E. coli* R46β residue is also thought to be involved in the TT-3' element recognition (119) and it is present in all IHFβ subunit studied with the exception of *N. gonorrhoeae* IHFβ protein (Fig. 1.9). The N-termini of α1 and α3 helices of IHFβ are very well conserved in almost all IHFβ proteins known so far (Fig. 1.9). Indeed, they have been proposed to interact with DNA via the formation of hydrogen bonds, and mutations in

these regions interfered with DNA-binding (130). However, these regions are not as well conserved in IHFα proteins, though the same importance was suggested for them in DNA-binding (130) (Fig. 1.8). Moreover, Rice *et al.* (130) showed that two prolines are intercalated between DNA base pairs upon binding of IHF to the DNA, P65α and P64β (130). While P65α is conserved in almost all IHFα proteins (except for the IHFα proteins of the Rickettsia group), P64β is unchanged in all IHFβ subunits (Figs. 1.8 and 1.9). Several amino acid residues (S47α, R63α, R60α, K66α, P61α, I73α, I71α, R42β, E44β, R46β) were suggested to be directly involved in DNA-binding thanks to hydrogen bonding or hydrophobic contacts (128, 130). These amino acid residues tend to be conserved in all proteins studied, suggesting their important role in IHF binding and function (Figs. 1.8 and 1.9). It is interesting to note that the C-terminal ends of both IHFα and IHFβ, which were shown to be unessential for the recognition of the *ihf* site and for the bending of DNA (119), are not well conserved in the different IHF proteins studied (Figs. 1.8 and 1.9).

IHF has also been implicated in cellular processes in numerous other bacteria however, their *ihf* genes have not yet been sequenced. *E. coli* IHF was shown to play a role in gene regulation [for example via *in vitro* transcription-translation experiments(194)] in *Azotobacter vinelandii* and *Rhodospirillum rubrum* (194), *Rhizobium meliloti* and *R. leguminosarum* (61, 263), *Shigella flexneri* (264), *Bradyrhizobium japonicum* (194, 263, 265), *Thiobacillus ferrooxidans* (194, 263), *Herbaspirillum seropedicae* (266, 267), and *Bacillus megaterium* (268). There are also many bacteria for which only the *ihfA* or *ihfB* gene is known. While sequencing of the entire genome may provide the sequence of the other gene, it is also possible that some bacteria have homodimer IHF proteins (α2 or β2). Of the bacteria studied known to encode for either IHFα or IHFβ, but not both (Tables 1.2 and 1.3), *R. prowazekii*, *C. trachomatis*, *R. sphaeroides*, *A. aeolicus*, and *C. muridarum* genomes have been completely sequenced and they only contain one gene homologous to either of the two subunits of IHF. It can therefore be hypothesized that these bacteria may have homodimer IHF proteins. It is also possible that the second subunit is less similar to the corresponding subunits in other bacteria and was not found by homology searches.

Interestingly, if IHFα and IHFβ are related to each other, they are actually more related to the HU proteins at the amino acid level (Table 1.4). It has been suggested by Drlica and Rouviere-Yaniv (16) that IHFα, IHFβ, and a primitive HU protein may have descended from a common ancestor, and that this primitive HU may have diverged giving rise to HUα and HUβ, which are more identical to one another (70%) than to IHFα (36% and 32%, respectively) and IHFβ (34% and 35%, respectively) (Table 1.4). The structure of the IHF and HU proteins are also well conserved, such that the structure of the HU protein of *B. stearothermophilus* (149) was used by Yang and Nash to establish the first predicted model of IHF structure (20) which appeared to be correct following crystal structure studies (130, 152, 153). Both *E. coli* IHF subunits also display a high degree of homology with histone-like proteins (HLP) and DNA-binding protein type II (DBPII) from other bacteria as exemplified in Table 1.4 with *Sinorhizobium meliloti* HLP and *Clostridium pasteurianum* DBPII. Moreover, some DNA-binding proteins able to bend DNA upon binding were found which are neither strict IHF nor HU analogues; examples of these include transcription factor 1 (TF1) from *Bacillus subtilis* phage SPO1 (16, 269) and the Hbb protein of *Borrelia burgdorferi* (270) (Table 1.4). When the amino acid sequence of these proteins are compared (Fig. 1.10), it appears that there are many conserved amino acids and that these proteins belong to the same family of DNA-binding proteins. Moreover, most of the conserved amino acids are located in structurally important regions of the proteins when the IHFα structural motifs are taken as an approximation of the structure of the other proteins (Fig. 1.10).

The histone-like IHF protein was shown to be involved in many different biological processes in the *E. coli* cell. The presence of similar proteins in other bacteria, more or less related to *E. coli*, suggests that this important protein was conserved throughout evolution and conserved many structural characteristics. Even if the roles of IHF in other bacteria are not yet known, evidence suggests that they may be very similar to the ones played in *E. coli*. The important similarity between the site-specific DNA-binding IHF protein, the nonspecific DNA-binding HU protein, and many other DNA-binding proteins like Hbb and SPO1 TF1, as well as the idea that IHF and

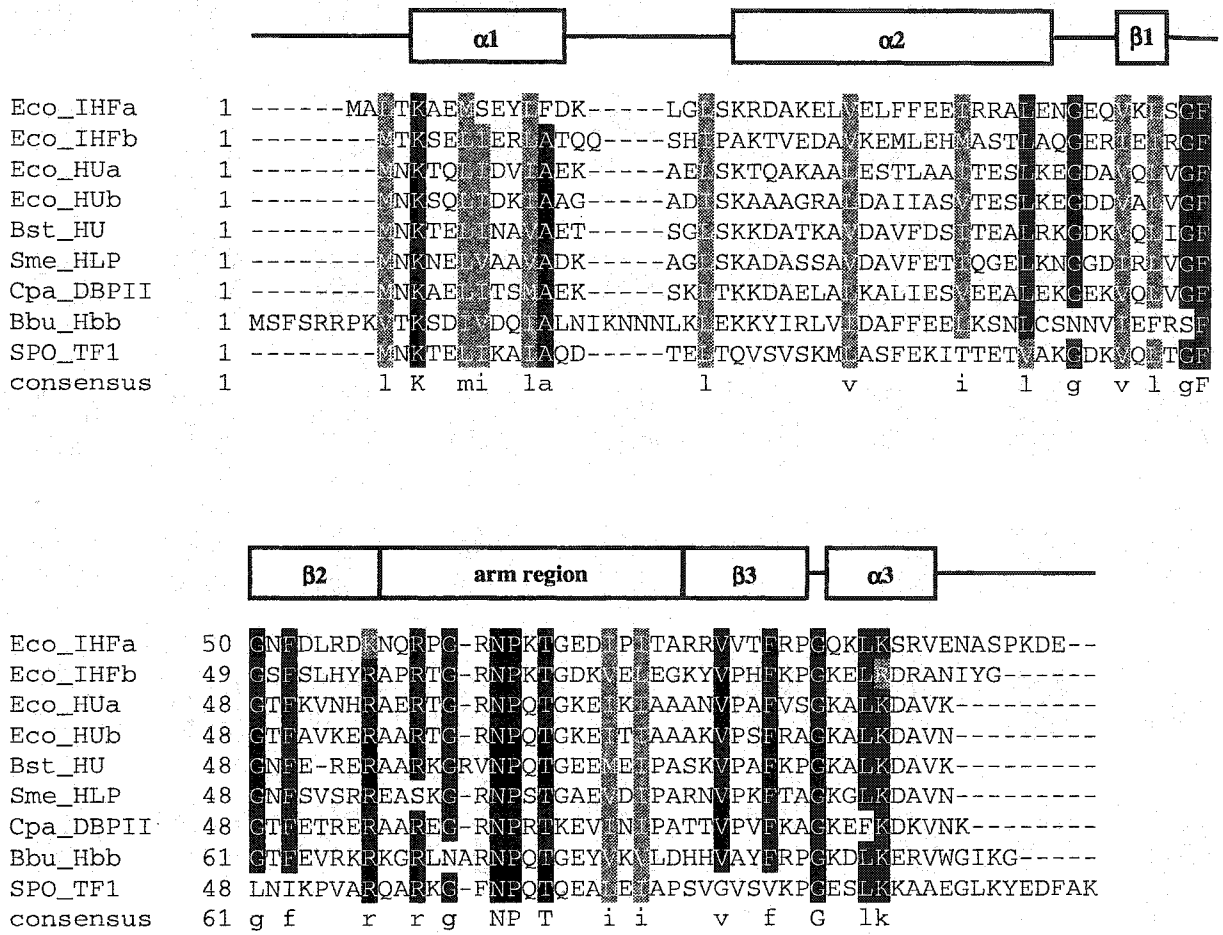
Table 1.4 Examples of homologous histone-like proteins in bacteria.
The accession number in the Entrez Protein database of the National Center for Biotechnology Information is given for each protein. Homologous proteins and their identity percentage to *E. coli* IHFα and IHFβ were found using BLAST searches at the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/) using the BLOSUM62 matrix (68).
HLP: histone-like protein; DBPII: DNA-binding protein II; TF1: transcription factor 1; Hbb: Hu and IHF homologue.

Protein name	***Accession number***	***Length (amino acids)***	***Identity to E. coli***		***Reference***
			IHFα	***IHFβ***	
Escherichia coli IHFα	P06984	99	100%	32%	(69, 70)
Escherichia coli IHFβ	P08756	94	32%	100%	(86)
Escherichia coli HUα	AAC76974	90	36%	34%	(271)
Escherichia coli HUβ	AAC73543	90	32%	35%	(271)
Bacillus stearothermophilus HU	2106156A	90	42%	39%	(272)
Sinorhizobium meliloti HLP	AAF05301	90	42%	35%	(273)
Clostridium pasteurianum DBPII	1011218A	91	40%	39%	(274)
Borrelia burgdorferi Hbb	AAB41461	108	35%	30%	(275)
Bacillus subtilis phage SPO1 TF1	P04445	99	25%	27%	(276)

HU subunits may have arisen from a single ancestor protein (16), support the idea that there is an IHF/HU family of DNA-binding proteins with a crucial role in gene regulation and many other processes in the bacterial cell and perhaps, like in eukaryotes, in compacting DNA into a higher order structure.

Figure 1.10 Comparison of the amino acid sequence of some histone-like proteins of bacteria.

Amino acid sequences were obtained from the Entrez Protein database of the National Center for Biotechnology Information with the accession numbers listed in Table 1.4. Multiple alignments were generated by ClustalW 1.75 (243) and elaborated with Boxshade 3.21 (244). Amino acid sequences are aligned underneath the secondary structure prediction established by Rice *et al.* (130) for IHFα. Amino acid sequences from the following proteins were analyzed: *E. coli* (Eco) IHFα, IHFβ, HUα and HUβ, *B. stearothermophilus* (Bst) HU, *B. subtilis* phage SPO1 (SPO) TF1, *B. burgdorferi* (Bbu) Hbb, *S. meliloti* (Sme) histone-like protein, and *C. pasteurianum* (Cpa) DNA-binding protein type II; for references, see Table 1.4. To improve the homology score, gaps (-) were added in the sequence by the ClustalW program. A black font indicates that the amino acid residues are identical for 90% of the sequences aligned, while a grey font indicates conservative substitutions.


α1
α2
β1
β2
arm region
β3
α3
Eco_IHFa
Eco_IHFb
Eco_HUa
Eco_HUb
Bst_HU
Sme_HLP
Cpa_DBPII
Bbu_Hbb
SPO_TF1
consensus

The IHF protein, first shown to be crucial for λ site-specific recombination (92), has also been shown to be involved in numerous fundamental processes in bacteria such as gene expression and DNA replication [reviewed in (51, 52, 53, 54)] and is now known to be present in many bacteria with very conserved structure and functions as explained in the preceding chapter. Moreover, IHF belongs to a family of proteins sharing similar characteristics and participating in essential cell processes, the histone-like proteins. Many families of regulatory proteins are crucial for bacterial growth and have been more or less well-charaterized. The Ner/Nlp/TMF family is a novel family of evolutionary conserved DNA-binding proteins comprising the Ner proteins of bacteriophages Mu and D108 (277, 278, 279), the novel Ner-like protein (Nlp) of *E. coli* (280) and the HIV-1 TATA element modulatory factor (TMF) of humans (281). No roles are known yet for the Nlp protein which appears to play a role in gene regulation and identification of bacterial genes regulated by *E. coli* Nlp was the goal of the work presented in this thesis.

Chapter 3 Identification of Bacterial Genes Regulated by the *Escherichia coli* Transposable Phage Ner Protein Homologue, Nlp/Sfs 7

2.1 Introduction

The *nlp* gene was discovered while Choi *et al.* were attempting to isolate a putative guanylate cyclase gene (*cyg*) from the *Escherichia coli* chromosome using complementation of a *crp**1/*cya*$^-$ mutant (strain MK2001) to stimulate maltose fermentation (280). The *crp*1* gene encodes for an altered cAMP receptor protein (CRP*) (282). This CRP* protein is still functional in the expression of most sugar fermentation genes, such as *lac*, *ara* and *man*, in the absence of cAMP (282). However, it does not fully activate the expression of the *mal* gene in the absence of cAMP. As it was shown that cGMP could be substituted for cAMP as an effector of CRP*, the idea arose that this *crp**1/*cya*$^-$ strain could be used to isolate the putative guanylate cyclase gene (*cyg*). Indeed, the *cyg* gene was not found. However several other genes were found to be able to stimulate maltose fermentation. One of these novel genes mapped to 69.3 minutes on the *E. coli* chromosome and was named *nlp* (Ner-like protein) (280), due to the high level of identity of the protein it encoded to the Ner proteins of bacteriophages Mu and D108 (respectively 65% and 59% identity) (Table 2.1). This gene was also named *sfs*7 (for sugar fermentation stimulation gene number 7) because of its ability to stimulate fermentation of maltose (283). The Nlp protein is composed of 92 amino acid residues, and has a predicted molecular weight of 10.4 kDa. Furthermore, studies on the amino acid sequence of this protein have suggested the presence of a putative DNA-binding domain located between amino acids 51 and 69 (280). Thus, the strong homology between Nlp, the Ner proteins of transposable coliphages Mu and D108, as well as TMF (TATA element modulatory factor) in humans, has further supported the hypothesis of a Ner/Nlp/TMF family of evolutionary conserved DNA-binding proteins (281, 284) (Table 2.1; Fig. 2.1). More recently, other proteins homologous to Nlp have been found in *Photorhabdus luminescens*, Pln, *Neisseria meningitidis*, Nlp, *Pasteurella multocida*, Ner, pathogenic *E. coli* strain O157:H7, probable DNA-binding protein, and *Haemophilus influenzae*, Nlp (Table 2.1). Multiple alignment of the available sequences for all Nlp homologous proteins identified shows that nine amino acids are conserved in all proteins (Fig. 2.1). Most of the conserved

Table 2.1 Nlp homologues.

The accession number in the Entrez Protein database of the National Center for Biotechnology Information is given for each protein. Homologous proteins and their identity percentage to *E. coli* Nlp were found using BLAST searches at the National Center for Biotechnology Information website (www.ncbi.nlm.nih.gov/) using the BLOSUM62 matrix (68).

Protein name	***Accession number***	***Length (amino acids)***	***Identity to E. coli Nlp***	***References***
Escherichia coli Nlp	P18837	92	100%	(280)
Photorhabdus luminescens Pln	AAF22961	70	68%	Direct submission
Mu Ner	AAF01082	75	65%	(277)
D108 Ner	P06903	73	59%	(278, 279)
Neisseria meningitidis Nlp	CAB85106	87	55%	(285)
Pasteurella multocida Ner	AAK02502	69	46%	(74)
Haemophilus influenzae Nlp	AAC23124	89	41%	(76)
E. coli [O157:H7] probable DBP	BAB13017	82	42%	(286)
Human TMF; aa 764-848	A47212	85	16%	(281)

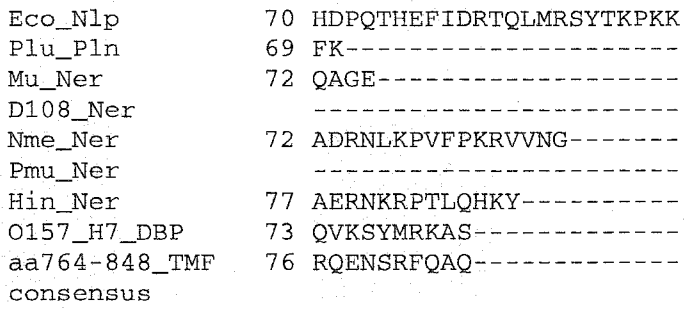
amino acid residues are not identical in all proteins but have similar physico-chemical properties and appear to be conservative changes. Indeed, only one amino acid is identical in all proteins. However, 33 amino acids are conserved in 80% of the proteins and it can be suggested that these amino acids may be important for this family of DNA-binding proteins function. The putative DNA-binding domain is also very well conserved in all Nlp homologues except for the TMF protein (Fig. 2.1). This important amino acid conservation could indicate the importance of this family of proteins for regulation of gene expression.

It has been shown that *nlp* is highly expressed as a monocistronic transcript of about 300 nucleotides when cells exit stationary phase (287), but is non-essential for cell viability under laboratory conditions (284). The role of Nlp in the cell, therefore, is not understood, however, this protein is nonetheless highly conserved among the *Enterobacteriaceae*, as shown by Southern blot analysis (284). This fact strongly suggests that the *nlp* gene may encode a protein which is functionally important for regulation of gene expression.

The goal of the work presented in this thesis was to identify and characterize bacterial genes regulated by the *E. coli* Nlp protein. A study of the role of Nlp in *E. coli* cell growth and physiology was therefore undertaken to better understand this putative family of regulatory proteins. As the phage and human homologues of Nlp appear to play a role in gene regulation, a search for *E. coli* genes which may be positively or negatively regulated by Nlp was undertaken, using a gene fusion approach.

Figure 2.1 Comparison of the amino acid sequence of Nlp homologues.
Amino acid sequences were obtained from the Entrez Protein database of the National Center for Biotechnology Information with the accession numbers listed in Table 2.1. Multiple alignments were generated by ClustalW 1.75 (243) and elaborated with Boxshade 3.21 (244). Amino acid sequences of the following proteins were analyzed: *E. coli* (Eco) Nlp, *P. luminescens* (Plu) Pln, Mu Ner, D108 Ner, *N. meningitidis* (Nme) Nlp, *P. multocida* (Pmu) Ner, *H. influenzae* (Hin) Ner, *E. coli* [O157:H7] probable DNA-binding protein (DBP), amino acids number 764 to 848 of human TMF protein; for reference, see Table V. To improve the homology score, gaps (-) were added in the sequence by the ClustalW program. A black font indicates that the amino acid residues are identical for 80% of the sequences aligned, while a grey font indicates conservative substitutions.


Eco_Nlp 1 -----MESN-F--IDWHPADIAAGLRKKGTSAAESRRNGLSSSTLANALSRPVPKGEMIIAKAIGTDPWVIWPSRY
Plu_Pln 1 ----MIERH-----DWHKADIAAALKKRGTSLAAVSREAGLASSTLSNVLHRPVPKGERIIATIINCEPSEIWPSRY
Mu_Ner 1 -----MCSN-EKARDWHRADVIAGLKKRKLSLSALSRQFGYAPTTLANALERHVPKGEQIIANAIETKPEVIWPSRY
D108_Ner 1 ---MHMNKR-TNRQDWHRADIVAELRKRNMSLAELGRSNHLSSSTLKNALDKRVPKAEKIIADAIGMTPQDIWPSRY
Nme_Ner 1 -----MQKN-ATPKNWHRADIVAALKKKGWSLRALSIEAGLSPNTLRSALAAPVLKGERIIAAAIGVEPEEIWPERY
Pmu_Ner 1 -----MKK---SKKNMHRAYIAAAIKEKGSSLAQLSVQAGLHPRTLNNALDKKVPKGEKIIADFIGVPVQEIWPERY
Hin_Ner 1 MSVLEKPKK-TAEQDWHRADIAAELKKNGWSLRSLAKEGQVSYNTLKTVLDKSVPKMERIVANAIGVPPEVIWAGRF
O157_H7_DBP 1 -----MSRREVSDHDWPPEILKARLMAGLSLRSLSLKAGYSRDSLKSVLRTPCRPYMQIIADAIGVSPEEIWPSRY
aa764-848_TMF 1 --VSSTTRPLLRQIENLQATLGSQTSSWEKLEKNLSDRLGESQTLLAAAVERERAATEELIANKIQMSSMESQNSLL
consensus 1 m r dwh a iia lr sl ls g tL l r w k e iiA l p iwp ry


Eco_Nlp 70 HDPQTHEFIDRTQLMRSYTKPKK
Plu_Pln 69 FK----------
Mu_Ner 72 QAGE----------
D108_Ner ----------
Nme_Ner 72 ADRNLKPVFPKRVVNG-------
Pmu_Ner ----------
Hin_Ner 77 AERNKRPTLQHKY----------
O157_H7_DBP 73 QVKSYMRKAS----------
aa764-848_TMF 76 RQENSRFQAQ----------
consensus

2.2 Materials and Methods

2.2.1 Bacterial strains and phage

All bacteria and phages used in this study are listed in Table 2.2.

2.2.2 DNA manipulations

All restriction endonucleases were purchased from Pharmacia Canada Inc., New England Biolabs, Amersham International or Gibco-Bethesda Research Laboratories Inc. (Gibco BRL). Restriction endonuclease digestions of DNA were performed at 37°C (unless otherwise recommended) in the digestion buffer provided by the manufacturer for 1-4 hours using 3 units of enzyme per μg of DNA. To remove the 5'-phosphate following a restriction digestion, where necessary, calf intestine alkaline phosphatase (Gibco BRL) was used prior to ligation.

For isolation, DNA fragments larger than 1 kbp for cloning were subjected to electrophoresis through 0.75% agarose gels in 1X TAE buffer (40 mM Trizma base; 20 mM Acetic acid; 2 mM EDTA; pH 8.1) and then eluted from the agarose using the GeneClean II Kit (Bio101) according to the manufacturer's instructions. DNA fragments smaller than 1 kbp were isolated from 5% polyacrylamide gels in 1X TBE buffer (89 mM Trizma base; 89 mM Boric acid; 2 mM EDTA; pH 8.3) (288) using the "crush and soak" method of Maxam and Gilbert (289).

Ligations were performed using T4 DNA ligase (Gibco BRL) with an insert to vector DNA ratio of 3:1. Ligation mixtures were incubated overnight at 15°C followed by heat-inactivation for 20 minutes at 65°C. Ligated products were transformed directly into cells using the RbCl procedure (290) or dialyzed on Milipore disks before being used for electroporation of competent cells (see below).

Amplification of the *nlp* gene by the Polymerase Chain Reaction (PCR) was accomplished using the synthetic oligonucleotide primers MM3 (5'-CCCGAATTCG

Table 2.2 List of *Escherichia coli* strains and plasmids used in this study.

Name	*Characteristics*	*References*
***E. coli* strains**		
40	F' Δ*proAB-lac rpsL trp8am thi*	(291)
DH5α	*supE44 hsdR17 recA1 endA1 gyrA1 thi-1 relA1*	(290)
JM109	*recA1 supE44 endA1 hsdR17 gyrA96 relA1 thi* Δ(*lac-proAB*)	(292)
LF20300	*E. coli* 40 *nlp::luxAB*, tet^R	(293)
Mal103	Δ(*gpt-proAB-argF-lac*)*XIII rpsL* [Mu dI(*lac, amp*)] (Mucts62)	(294)
90-6	LF20300/Mu dI containing a putative Nlp responsive gene	This study
205-15	LF20300/Mu dI containing a putative Nlp responsive gene	This study
Plasmids		
pNLP1.7	Amp^R *E. coli nlp* (1.7 kb *Pst*I-*Hin*dIII), (P_{tac}) P/O containing vector	(284)
pCR®2.1	Kan^R, Amp^R, (P_{lac}) P/O TA cloning vector	Invitrogen Inc.
pUC120	Amp^R (P_{lac}) P/O cloning vector	(295)
pBAD18-Kan	Kan^R cloning vector, P_{BAD} promoter, *araC*	(296)
pBAD18-Cm	Cm^R cloning vector, P_{BAD} promoter, *araC*	(296)
pMM4	pCR®2.1/*nlp*	This study
pMM5	pBAD18-Kan/*nlp*	This study
pMM6	pUC120/*nlp*	This study

ACTAACTTAAGGAGTGAGG-3') and MM4 (5'-CCCTGCAGCGAATAATCGTCT GAGAGCTGGC-3') purchased from Gibco BRL. Amplification of *nlp* was carried out using 80 pg of pNLP1.7, 1X PCR buffer (50 mM Tris-HCl pH 8.0; 25mM KCl; 1 μg/mL BSA), 2 mM $MgCl_2$, 400 μM of each dNTP, 0.5 pmol of each primer and 2.5 units of *Taq* polymerase (Promega). The PCR mixture was overlaid with mineral oil and cycled in an MJ Research PTC100 temperature cycler (MJ Research Inc.) programmed to carry out a first denaturation step of 3 min at 95°C followed by 30 cycles of 1.5 min at 94°C, 1.5 min at 55°C and 45 sec at 72°C followed by a final elongation step of 10 min at 72°C.

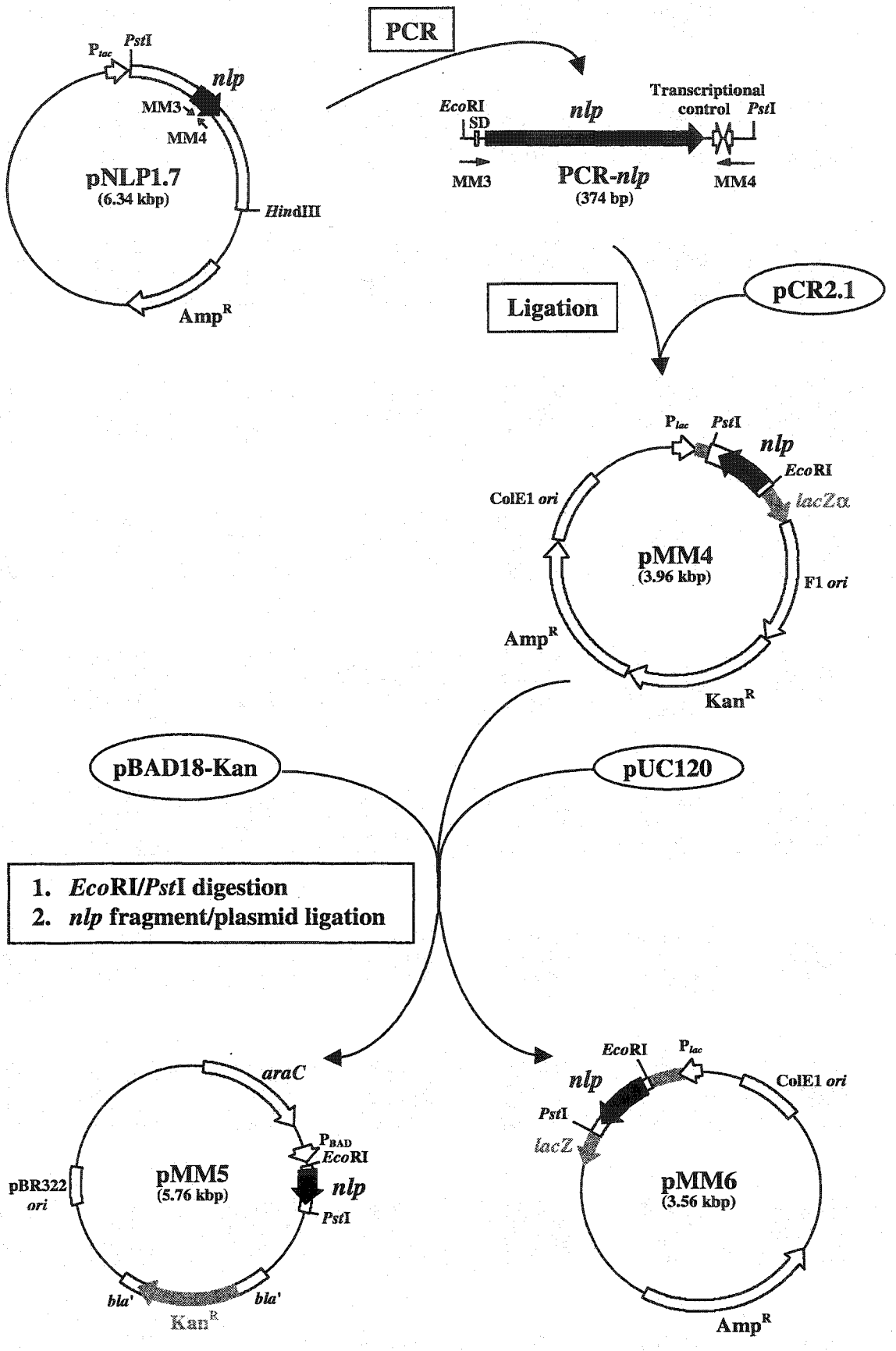
Double-stranded DNA sequencing of the PCR-amplified *nlp* gene in pMM4 was performed using the Sequenase version 2.0 kit (United States Biochemical Corp.) and universal primers as recommended by the manufacturer.

2.2.3 Plasmid construction

All plasmids used and constructed in this study are listed in Table 2.2.

Cloning of the *nlp* gene, from the transcription start to the end of the first transcriptional terminator loop, was accomplished by first specifically amplifying the gene sequence from pNLP1.7 (284) using PCR (Fig. 2.2). The primers used, MM3 and MM4, created unique *Eco*RI and *Pst*I sites at the 5' and 3' ends of the amplified fragment. The *Eco*RI-*Pst*I amplified fragment was then ligated into the cloning vector pCR®2.1 (Invitrogen, Inc) generating the plasmid pMM4. Sequencing of the *nlp* gene was performed as described above in order to ensure there were no errors due to the amplification procedure. The *nlp* gene was then excised from pMM4 using an *Eco*RI-*Pst*I double digestion and subcloned into pBAD18-Kan (296) and pUC120 (295) creating pMM5 and pMM6, respectively (Fig. 2.2).

Figure 2.2 Construction of the pBAD18-Kan-based plasmid pMM5 and the pUC-based plasmid pMM6 containing the PCR-amplified *nlp* gene.
The *nlp* gene was first PCR-amplified from plasmid pNLP1.7 (284) using *Taq* DNA polymerase (Promega) and primers MM3 and MM4. The resulting fragment was then cloned into the T-vector pCR®2.1 (Invitrogen Inc.) generating pMM4. The *nlp* gene from pMM4 was subcloned into the plasmid vector pBAD18-Kan, generating pMM5. This plasmid places the *nlp* gene under the control of the *ara* P_{BAD} promoter. The *nlp* gene was also subcloned into the plasmid vector pUC120 (295), under the control of the P_{lac} promoter, generating plasmid pMM6.


PCR
P lac
PstI
nlp
MM3
MM4
pNLP1.7
(6.34 kbp)
HindIII
Amp R
EcoRI
SD
nlp
Transcriptional
control
PstI
MM3
PCR-nlp
(374 bp)
MM4
Ligation
pCR2.1
P lac
PstI
nlp
EcoRI
ColE1 ori
lacZα
pMM4
(3.96 kbp)
F1 ori
Amp R
Kan R
pBAD18-Kan
pUC120
1. EcoRI/PstI digestion
2. nlp fragment/plasmid ligation
araC
P BAD
EcoRI
nlp
PstI
pBR322
ori
pMM5
(5.76 kbp)
bla'
bla'
Kan R
EcoRI
P lac
nlp
ColE1 ori
PstI
lacZ
pMM6
(3.56 kbp)
Amp R

2.2.4 Culture Media, Growth Conditions and Antibiotic Selection

Bacteria were grown at 37°C, or 32°C where indicated. Bacteria were routinely grown in Luria-Bertani (LB) broth (297) and on LB plates containing 1.5% (w/v) agar.

Superbroth (298) and TCMG [1% (w/v) trypticase peptone, pancreatic digest of casein; 0.25% (w/v) NaCl; 10 mM $MgSO_4$-$7H_2O$; (299)] plates containing 0.85% (w/v) agar were also used. D-glucose (glc) and L-arabinose (ara) were used at final concentrations of 0.2% (w/v), X-gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside) at 40 μg/mL and IPTG (isopropyl β-D-thiogalactopyranoside) at 2 mM.

The following antibiotics were used at the indicated final concentrations: ampicillin (amp), 50 μg/mL (in liquid culture) or 100 μg/mL (in agar plates); kanamycin (kan), 50 μg/mL; streptomycin (strep), 100 μg/mL; tetracycline (tet), 10 μg/mL (in liquid culture) or 20 μg/mL (in agar plates); chloramphenicol (cm), 50 μg/mL unless otherwise indicated.

2.2.5 Transformation and transduction

Transformations of *E. coli* were carried out using the rubidium chloride procedure (290) or the calcium chloride method (288) or by electro-transformation (300) using the Bio-Rad Pulse Controller and Gene Pulser apparatus (Bio-Rad Laboratories). Cells were then incubated in 1 mL of LB at 32°C or 37°C for 1 hour with shaking to improve the recovery of transformants before being plated onto LB plates containing the appropriate antibiotics and incubated 16-18 hours at 37°C.

Transduction with bacteriophage Mu dI (*amp lac*) to construct *lacZ* fusions to chromosomal genes was done as described by Casadaban and Cohen (294).

2.2.6 Chloramphenicol Release Assay

Expression of Nlp from the PCR-amplified *nlp* gene from pMM6 was visualized using the chloramphenicol release procedure of Neidhardt *et al.* (301).

2.2.7 Assay of β-galactosidase activity

β-galactosidase assays were performed as described by Miller (297) using the chloroform-sodium dodecyl sulfate cell lysis procedure. Cells were grown for 36 hours on TCMG plates containing ampicillin, tetracycline, and kanamycin or chloramphenicol and contained either glucose, arabinose or no sugar. Cells were scraped from the plates, suspended in 3 mL of liquid TCMG and grown to an A_{600} of 0.4 to 0.6. The assay was then performed as described by Miller.

2.2.8 DNA isolation

Small scale plasmid DNA preparations were performed using the alkaline lysis procedure of Sambrook *et al.* (302).

Large scale plasmid DNA preparations were performed from 1L LB cultures. Plasmids were amplified at a cell density of $4X10^8$ cells/mL by treatment with 75 μg/mL Cm for 16 hours at 32°C. Extraction of DNA was done using the cleared lysate technique of Clewell and Helinski (303) followed by ultracentrifugation in cesium chloride/ethidium bromide gradients (288).

Isolation of chromosomal DNA from Nlp-responsive clones was done as follows: 10 mL overnight cultures were subjected to centrifugation for ten minutes at 5,000 rpm, 4°C. Pelleted cells were resuspended in 1.4 mL 10X TE (100 mM Tris-HCl, pH 7.5, 10 mM EDTA, pH 8.0). Sodium dodecyl sulfate (20% w/v) and RNaseA (1 mg/mL in 10 mM Tris-HCl, pH 7.6) were added to final concentrations of 0.53% (w/v) and 0.21 mg/mL respectively. The mixture was then incubated for two hours at 37°C. After the

addition of pronase (20 mg/mL in 10 mM Tris-HCl, pH 7.6) to a final concentration of 1.9 mg/mL, the incubation at 37°C was resumed for two hours. The DNA was purified by two phenol extractions (one volume of phenol added; sample mixed by inversion and centrifugation for ten minutes at 5,000 rpm, 25°C) followed by two ether extractions. The DNA was then precipitated in two volumes of 100% ethanol and slowly mixed until the precipitated DNA became visible. The DNA was immediately isolated with a micropipette, air dried at 65°C and resuspended in 200 µL 1X TE.

2.3 Results

2.3.1 Cloning of *nlp* into pUC120 and expression of the Nlp protein

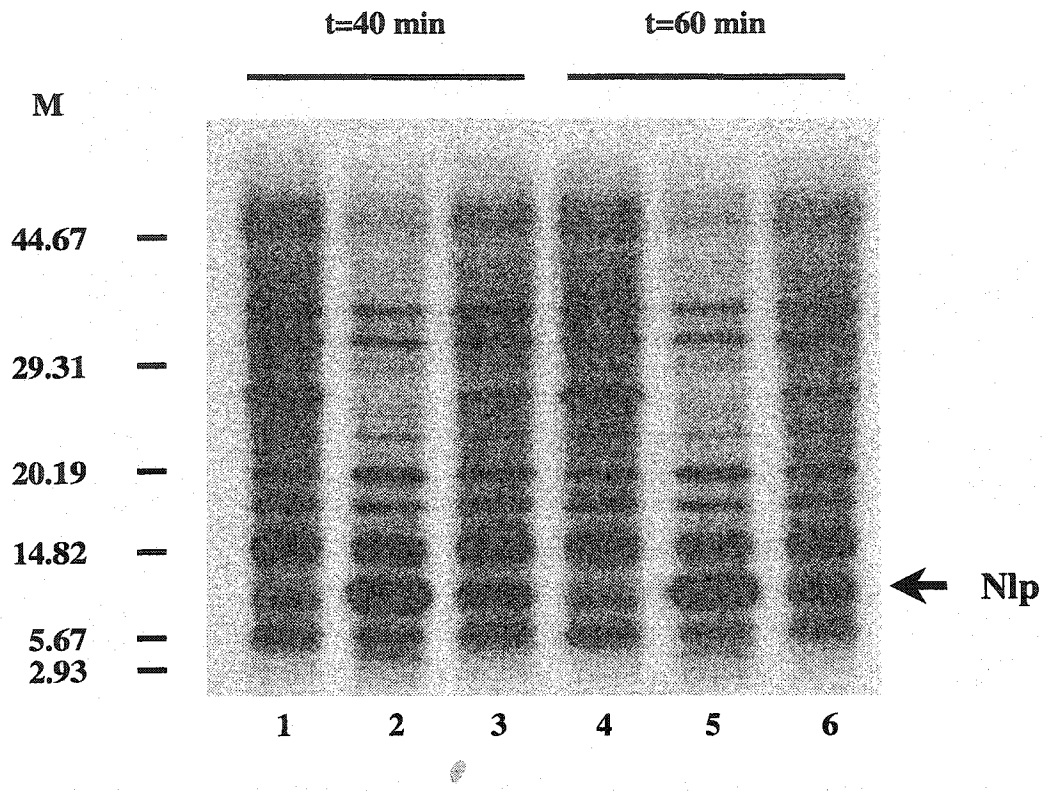
Plasmid pNLP1.7 contains the *nlp* gene in a 1.7 kb *Pst*I-*Hin*dIII chromosomal region of *E. coli* cloned into pKK223-3 (284) (Fig. 2.2). Portions of other genes are also present in pNLP1.7. The end of the *ispB* gene, encoding an enzyme involved in the synthesis of isoprenoid quinones (these quinones are essential components of the respiratory chain of *E. coli*) (304, 305), and the end of *murA* (also called *murZ*), a gene coding for an enzyme which catalyzes the first committed step of peptidoglycan biosynthesis (306). These two regions may encode polypeptides which could possibly confuse this study. Therefore, the *nlp* gene was amplified from the transcription start to the end of the first loop involved in transcriptional control, by the polymerase chain reaction method (PCR). The two oligonucleotide primers used for the PCR, MM3 and MM4, were designed to create an *Eco*RI site at the 5'-end and a *Pst*I site at the 3'-end of the amplified fragment (Fig. 2.2).

Following PCR, the amplified fragment was ligated into the cloning vector, pCR®2.1 (Invitrogen Inc), generating pMM4 (Fig. 2.2). The cloned *nlp* gene in pMM4 was sequenced to verify that no mutation had occurred during the PCR amplification. Using an *Eco*RI/*Pst*I double digestion, the *nlp* gene was isolated from pMM4 and then subcloned into pUC120 digested with the same restriction enzymes generating plasmid pMM6. This directional cloning allowed for the *nlp* gene to be placed under the control of the P_{lac} promoter in pUC120.

In order to verify that the Nlp protein could effectively be produced from the PCR-amplified *nlp* gene, the chloramphenicol release assay of Neidhardt *et al.* (301) was performed (Fig. 2.3). Protein synthesis was induced using 2 mM IPTG and labelled with [^{35}S]-methionine (51 μCi/mL; Amersham) for 40 minutes (lanes 1, 2 and 3) or 60 minutes (lanes 4, 5 and 6). Figure 2.3 shows that in addition to polypeptides encoded by strain JM109 containing the pUC120 plasmid (lanes 1 and 4), the Nlp protein, with an

Figure 2.3 Expression of the Nlp protein.

SDS-polyacrylamide gel electrophoretic analysis of the gene products produced by a chloramphenicol release assay (see Materials and Methods) of a JM109 strain containing pUC120 (lanes 1 and 4), pNLP1.7 (lanes 2 and 5) or pMM6 (lanes 3 and 6) induced with 2mM IPTG and labelled with [^{35}S]-methionine (Amersham) for 40 (lanes 1, 2 and 3) or 60 (lanes 4, 5 and 6) min. M: marker in kDa [molecular weight (range 3.0-43.0 Da) protein standards from Gibco BRL].


t=40 min
t=60 min
M
44.67
29.31
20.19
14.82
5.67
2.93
Nlp
1
2
3
4
5
6

apparent molecular weight of 10.4 kDa, was over-expressed only when the strain contains pNLP1.7 (lanes 2 and 5) or pMM6 (lanes 3 and 6) at both 40 and 60 minutes after post-induction. The Nlp protein was therefore expressed from pMM6.

2.3.2 Cloning of *nlp* into the pBAD expression vector

In order to study the role of *nlp* in the *E. coli* cell, the *nlp* gene was cloned into an *ara* pBAD expression vector allowing for the tight regulation of transcription of *nlp* (294). Using an *Eco*RI/*Pst*I double digestion, the *nlp* gene was isolated from pMM4 and subcloned into pBAD18-Kan digested with the same restriction enzymes generating plasmid pMM5 (Fig. 2.2). This directional cloning allowed the *nlp* gene to be placed under the control of the P_{BAD} promoter.

2.3.3 Construction of a library of random, chromosomal, promotorless, *lacZ* transcriptional fusions

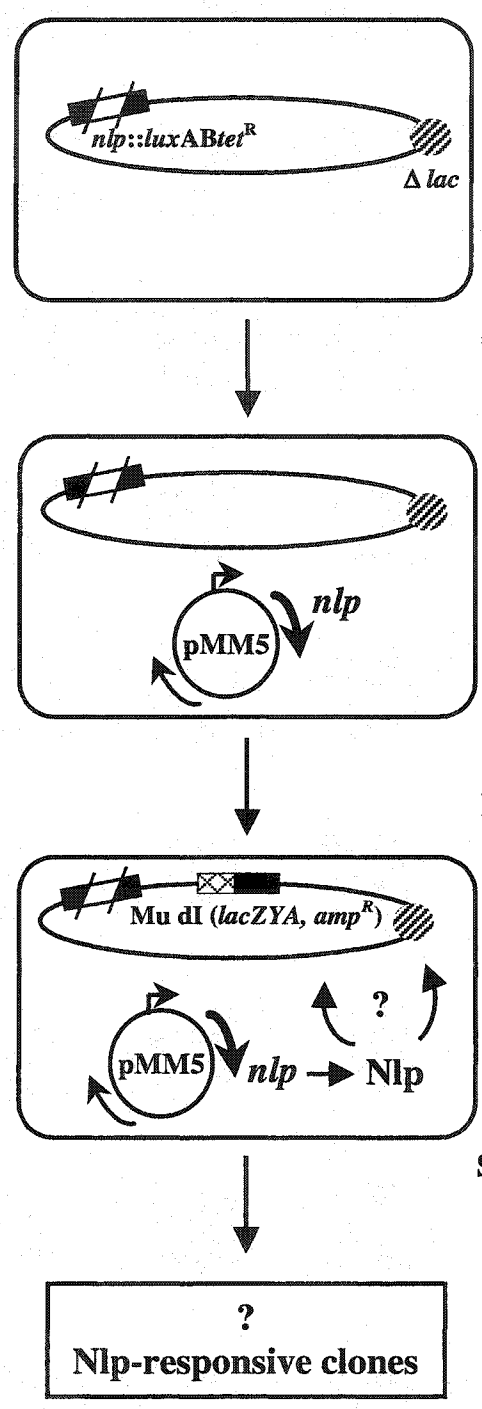
The expression vector pMM5 was transformed into the Δ*pro-lacIZYA E. coli* 40 strain LF20300, in which the *nlp* gene had been disrupted by a *luxAB::tet*^{R} insertion (293) (Fig. 2.4). The resulting strain LF20300/ pMM5 was then lysogenized using a Mu dI lysate obtained from *E. coli* strain Mal103 (294). As the Mu dI bacteriophage contained a promoterless *lacZ* reporter gene, the activity of any exogenous promoter after which Mu dI was inserted could be monitored by levels of β-galactosidase expressed from the *lacZ* gene (Fig. 2.4). A library of 10,000 clones was screened.

2.3.4 Screening of the LF20300/pMM5/Mu dI (*amp lac*) library

The library was screened on minimal medium (TCMG) containing either 0.2% glucose (no Nlp protein produced) or 0.2% arabinose (Nlp produced) as well as X-gal. Due to the X-gal indicator, clones producing the β-galactosidase enzyme were blue,

Figure 2.4 Construction of the *lacZ* fusion library.

A *lacZ* fusion library was constructed by transforming *E. coli* strain LF20300 (*nlp*::*luxAB*, tet^R) (291) with the pBAD18-Kan vector (294) containing the *nlp* gene (pMM5). The resultant strain was lysogenized with Mu dI (*lacZYA*, amp^R) (292), and lysogens were screened for β-galactosidase expression on media containing either 0.2% arabinose or 0.2% glucose.

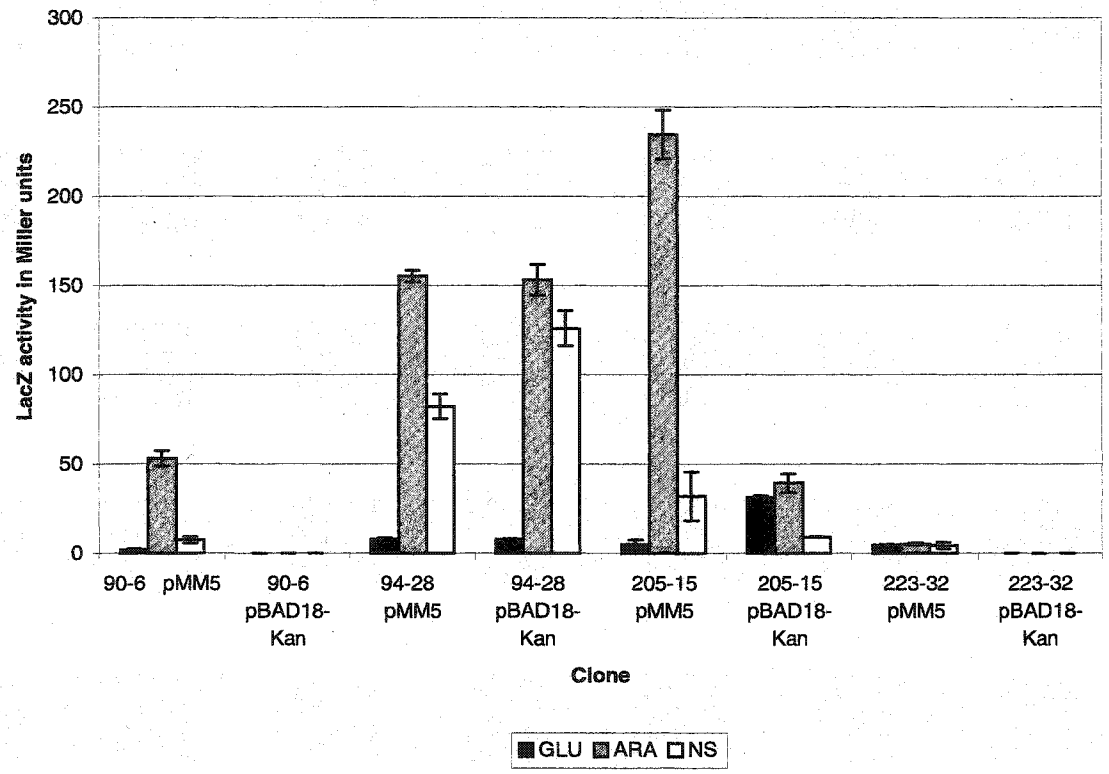

nlp::luxABtet^R
Δ lac
LF20300
(E. coli 40, nlp::luxABtet^R)
Transformation with pMM5
(pBAD18-Kan/nlp)
pMM5
nlp
Lysogenization with
Mu dI (lacZYA , amp^R)
Mu dI (lacZYA, amp^R)
?
pMM5
nlp → Nlp
Screening on 0.2% glc
or 0.2% ara,
plus X-gal
?
Nlp-responsive clones

while clones producing no β-galactosidase were white. Therefore, clones were screened for a change in colour between the two media, as this colour change indicated a change in the expression of the *lacZ* gene from the exogenous promoter, depending on the expression of Nlp. After an initial screening, 95% of the clones were eliminated. Of the remaining 500 clones, 450 appeared to have gene expression repressed in the presence of arabinose, while 50 were induced (i.e. changed from white to blue in the presence of arabinose). After three further subcloning and rescreening procedures, only 17 clones were kept: for nine of them, the β-galactosidase expression was higher in the presence of arabinose than with glucose, and for the other eight, it was the opposite. In order to eliminate the arabinose- and glucose-responsive clones, these 17 clones were then transformed with pBAD18-Cm which is a plasmid of the same incompatibility group (307) as pMM5 (they both have the same origin of replication, pBR322 *ori*), using the calcium chloride transformation procedure as previously described. The transformed bacteria were then plated on plates containing chloramphenicol, but not kanamycin, to select clones that retained pBAD18-Cm and lost the *nlp*-containing plasmid pMM5. The 17 positive clones, now containing pBAD18-Cm and not pMM5, were re-screened on plates containing X-gal and either 0.2% (w/v) glucose or 0.2% (w/v) arabinose. Sugar-responsive clones presenting a different color on the two media were eliminated. Only four clones were retained: clones 90-6 and 205-15, induced in the presence of arabinose (changed from white on glucose to blue on arabinoase), and clones 94-28 and 223-32, repressed in the presence of arabinose (changed from blue on glucose to white on arabinose). In these four clones, Mu dI is inserted downstream form a promoter regulated by Nlp which is produced from pMM5 only in the presence of arabinose.

β-galactosidase activity assays were performed as previously described with these four clones (containing pMM5) to confirm the effect of glucose and arabinose on the level of β-galactosidase expression from the promotorless reporter *lacZ* gene (Fig. 2.5). No significant difference in expression of *lacZ* was noticeable for clones 94-28 and 223-32 with pMM5 or pBAD18-Kan and with glucose, arabinose or no sugar. The β-galactosidase activity from clone 94-28 was different depending on the sugar in the medium, but these differences were the same when this clone contained pMM5 or pBAD18-Kan. Clone 94-28 is therefore sugar-responsive and should have been

Figure 2.5 Results of the liquid β-galactosidase activity assay performed on the four positive clones (90-6, 94-28, 205-15, and 223-32) obtained from the library screening.

The assay was done on clones containing either the pBAD vector alone (pBAD18-Kan), obtained by plasmid replacement from the corresponding pBAD18-Cm strains, or the pBAD vector with the *nlp* gene (pMM5). The clones were grown on glucose (Glu), arabinose (Ara) or with no additional sugar (NS).


300
250
200
150
100
50
0
LacZ activity in Miller units
90-6 pMM5
90-6 pBAD18-Kan
94-28 pMM5
94-28 pBAD18-Kan
205-15 pMM5
205-15 pBAD18-Kan
223-32 pMM5
223-32 pBAD18-Kan
Clone
GLU ARA NS

eliminated when re-screended on glucose or arabinose-containing medium after plasmid replacement with pBAD18-Cm. In the case of clone 223-32, the expression from the *lacZ* gene was very low with pMM5 but identical in the three different conditions and undetectable with pBAD18-Kan. This clone containing pMM5 should not have been retained on the four screenings performed on glucose and arabinose. This clone may not have been isolated correctly and was possibly a mixture of several clones, explaining the change in color seen depending on the sugar present. Moreover, the difference in colour between clones on different conditions were evaluated with human eyes and errors of judgement could have occurred. In these two clones, 94-28 and 223-32, the gene in which the Mu dI phage was inserted was not regulated by Nlp since the activity of the promoter was independent from the expression of Nlp. For both 90-6 and 205-15 clones, the β- galactosidase activity was high in the presence of arabinose, but weak in the presence of glucose or with no sugar added to the medium with plasmid pMM5. However, *lacZ* expression was not detectable in clone 90-6 with pBAD18-Kan, in the three different conditions or low in clone 205-15 with pBAD18-Kan (more than 7 fold decrease). These results confirm that in clones 90-6 and 205-15, the expression of a gene was induced by Nlp. Therefore, the genes in which the Mu dI phage was inserted in both 90-6 and 205-15 clones need to be sequenced and studied further to characterize their role in the *E. coli* cell and the physiological importance of the regulation by Nlp.

2.4 Discussion

The *nlp* gene, discovered by Choi *et al.* when attempting to isolate a putative guanylate cyclase gene (*cyg*) using complementation of a *crp**1/*cya*$^-$ mutant (strain MK2001) to stimulate maltose fermentation (280), was shown to code for a protein, Nlp (Ner-like protein), which is homologous the Ner proteins of bacteriophages Mu and D108 (respectively 65% and 59% identity) and hypothesized to belong to a family of conserved DNA-binding proteins. This gene was also named *sfs*7 (for sugar fermentation stimulation gene number 7) because of its ability to stimulate fermentation of maltose (283). However, no role other than stimulation of maltose fermentation is known yet for this novel protein. The presence of a putative DNA-binding domain located between amino acids 51 and 69 (280) suggests a role for Nlp in gene regulation. Moreover, the strong homology between Nlp, the Ner proteins of transposable coliphages Mu and D108, as well as TMF (TATA element modulatory factor) in humans, has supported the hypothesis of a Ner/Nlp/TMF family of evolutionary conserved DNA-binding proteins (281, 284). More recently, other proteins homologous to Nlp have been found in *Photorhabdus luminescens*, Pln, *Neisseria meningitidis*, Nlp, *Pasteurella multocida*, Ner, pathogenic *E. coli* strain O157:H7, probable DNA-binding protein, and *Haemophilus influenzae*, Ner. The putative DNA-binding domain is also very well conserved in all Nlp homologues except for the TMF protein. This important amino acid conservation could indicate the importance of this family of proteins for regulation of gene expression.

To better understand this putative family of regulatory proteins, and more specifically, to identify and characterize bacterial genes regulated by the *E. coli* Nlp protein, a study of the role of Nlp in *E. coli* cell growth and physiology was therefore initiated. As the phage and human homologues of Nlp appear to play a role in gene regulation, a search for *E. coli* genes which may be positively or negatively regulated by Nlp was undertaken, using a gene fusion approach. The *nlp* gene was first amplified by the PCR reaction, cloned into plasmid pUC120 resulting in plasmid pMM6, and expression of the Nlp protein from this construction was confirmed by the

chloramphenicol release assay of Neidhardt *et al.* (301). *nlp* was then cloned into a pBAD expression vector allowing for the tight regulation of transcription of *nlp* from the P_{BAD} promoter in vector pMM5 (294). The LF20300/pMM5/Mu dI (*amp lac*) library of random, chromosomal, promotorless, *lacZ* transcriptional fusions was then constructed in *E. coli* 40 strain LF20300, in which the *nlp* gene had been disrupted by a *luxAB::tet*^{R} insertion (293). As the Mu dI bacteriophage contained a promoterless *lacZ* reporter gene (294), the activity of any exogenous promoter after which Mu dI was inserted could be monitored by levels of β-galactosidase expressed from the *lacZ* gene library. The library was screened on minimal medium containing either 0.2% glucose (no Nlp protein produced) or 0.2% arabinose (Nlp produced) as well as X-gal. Due to the presence of the X-gal indicator, clones producing the β-galactosidase enzyme were blue while clones producing no β-galactosidase were white. Therefore, clones were screened for a change in colour between the two media: this colour change indicated a change in the expression of the *lacZ* gene from the exogenous promoter, depending on the expression of Nlp. Out of 17 clones found to have a different β-galactosidase expression depending on the presence or absence of Nlp, 13 were found to respond to the change of sugar and not to Nlp. These sugar responsive clones were eliminated and β-galactosidase activity assays were performed on the four remaining clones to confirm the effect of glucose and arabinose on the level of β-galactosidase expression from the promotorless reporter *lacZ* gene. Finally, only two clones were retained, clones 90-6 and 205-15, in which expression of an *E. coli* gene was clearly induced in the presence of Nlp.

The present study has enabled us to identify two genes which are activated by Nlp, however, their identity has not yet been determined. not allowed to determine the role of Nlp. Therefore, further studies are needed in order to identify and characterize these two genes induced by Nlp. However, it can already be suggested that Nlp does not act on as broad a range of genes as IHF does since only two clones where found to be responding to a change in Nlp expression. Following this study, a Southern Blot analysis using a radiolabeled *lacZ* probe was performed using genomic DNA isolated from the two Nlp-responsive clones in order to determine the number of Mu dI insertions in each strain (308). The results of this study showed that Mu dI had inserted

once in the chromosome of clone 90-6, but twice in clone 205-15. The genomic DNA spanning the Mu dI right-end in one of the Mu dI fusions in clone 201-15 was cloned and sequenced and the results suggested that Mu dI may have inserted within the *yqhG* gene (308). Expression from the gene *yqhG* of *E. coli* may therefore be induced by Nlp. As of now, little is known about the gene *yqhG* of *E. coli* and no function has been determined yet for the encoded protein. However, since two Mu dI vectors lysogenized clone 205-15, it is not clear whether this gene is regulated by Nlp or if it is the other one in which Mu dI inserted or if it is both. One way to determine this would be to first clone and sequence the second Mu dI insertion site and then the effect of Nlp should be studied on both genes separately, perhaps by creating individual *lacZ* fusions in a genetically clean background strain. In addition, the gene whose expression seems to be regulated by Nlp in clone 90-6 still needs to be determined.

Although the role of Nlp has not yet been elucidated, the potential regulation by Nlp on the promoter for *yqhG* and the other two putative genes may give insight into how Nlp functions. If it is confirmed that Nlp induces the expression of *yqhG* (and the two other genes which still need to be sequenced), the mechanism by which Nlp regulates expression from these genes can be studied by various protein-DNA interaction studies including electrophoretic mobility shift assays and footprint analyses (e.g. DnaseI, hydroxy-radical and/or copper phenanthroline). Another approach to study the difference of protein expression in the presence and in the absence of Nlp is the use of bidimensional electrophoresis of protein extracts from *E. coli* and other *Enterobacteriaceae* as well as through the use of microarray technology.

Chapter 4 Summary and Conclusion

While the histone-like IHF protein is known to be involved in many different biological processes in the *E. coli* cell, no role is yet known for *E. coli* Nlp, a novel protein which appears to play a role in gene regulation. However, both proteins share several characteristics. For example, both proteins have known (IHF) or predicted (Nlp) helix-turn-helix domains, and both proteins also have homologues in other bacteria more or less related to *E. coli* or its phages.

The goal of the work presented in this thesis was to identify and characterize bacterial genes regulated by the *E. coli* Nlp protein and therein understand better this putative Ner/Nlp/TMF family of evolutionary conserved DNA-binding proteins. As the phage and human homologues of Nlp appear to play a role in gene regulation, a search for *E. coli* genes which may be positively or negatively regulated by Nlp was undertaken, using a gene fusion approach.

During this study, two clones were identified which displayed differential expression in the presence and in the absence of Nlp. Recent work, still in progress, recently determined, using Southern Blot analyses, that one of the clones (90-6) contained a single gene potentially regulated by Nlp, and that the second clone (205-15) had two putative Mu dI insertions leading to the prediction that two different promoters could be potentially Nlp-regulated.

Following cloning and sequencing, one of the two genes disrupted by the Mu dI insertion in clone 205-15 was identified as *yqhG*, for which nothing is yet known. Further studies are now required to confirm and elucidate a possible regulation of the expression of *yqhG*, as well as that of the other two putative Nlp-regulated genes. This will involve the cloning and sequencing of the remaining two genes identified in this thesis. Further studies will also be needed to characterize the binding pattern of Nlp and to determine if a consensus sequence is recognized , which amino acid residues contact the DNA, and what kind of DNA conformational changes the binding of Nlp creates.

If Nlp is confirmed to regulate the expression of one or all three of these genes, the function of these genes and the mechanism of their regulation by Nlp may help us understand the role of Nlp in *E. coli* cell growth and physiology. In addition, information concerning the putative Ner/Nlp/TMF family may also be gained and help us to understand better the members of this unique family of DNA-binding proteins.

Reference List

1. **Pettijohn, D. E.** 1996. The nucleoid, p. 158-166. *In* F. C. Neidhardt and et al. (eds.), *Escherichia coli* and *Salmonella*: cellular and molecular biology. ASM Press, Washington, D.C.

2. **Oberto, J., K. Drlica, and J. Rouviere-Yaniv**. 1994. Histones, HMG, HU, IHF: Meme combat. Biochimie **76**:901-908.

3. **Rouviere-Yaniv, J. and F. Gros**. 1975. Characterization of a novel, low-molecular-weight DNA-binding protein from *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A **72**:3428-3432.

4. **Rouviere-Yaniv, J.** 1978. Localization of the HU protein on the *Escherichia coli* nucleoid. Cold Spring Harb. Symp. Quant. Biol. **42 Pt 1**:439-447.

5. **Rouviere-Yaniv, J., M. Yaniv, and J. E. Germond**. 1979. *E. coli* DNA binding protein HU forms nucleosomelike structure with circular double-stranded DNA. Cell **17**:265-274.

6. **Rouviere-Yaniv, J. and N. O. Kjeldgaard**. 1979. Native *Escherichia coli* HU protein is a heterotypic dimer. FEBS Lett. **106**:297-300.

7. **Pettijohn, D. E.** 1988. Histone-like proteins and bacterial chromosome structure. J.Biol.Chem. **263**:12793-12796.

8. **Hayat, M. A. and D. A. Mancarella**. 1995. Nucleoid proteins. Micron. **26**:461-480.

9. **Cukier-Kahn, R., M. Jacquet, and F. Gros**. 1972. Two heat-resistant, low molecular weight proteins from *Escherichia coli* that stimulate DNA-directed RNA synthesis. Proc. Natl. Acad. Sci. U.S.A **69**:3643-3647.

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

11. **McGovern, V., N. P. Higgins, R. S. Chiz, and A. Jaworski**. 1994. H-NS over-expression induces an artificial stationary phase by silencing global transcription. Biochimie **76**:1019-1029.

12. **Koch, C., J. Vandekerckhove, and R. Kahmann**. 1988. *Escherichia coli* host factor for site-specific DNA inversion: cloning and characterization of the *fis* gene. Proc. Natl. Acad. Sci. U.S.A **85**:4237-4241.

13. **Finkel, S. E. and R. C. Johnson**. 1992. The Fis protein: it's not just for DNA inversion anymore [published erratum appears in Mol Microbiol 1993 Mar;7(2):1023]. Mol. Microbiol. **6**:3257-3265.

14. **Nash, H. A. and C. A. Robertson**. 1981. Purification and properties of the *Escherichia coli* protein factor required for lambda integrative recombination. J. Biol. Chem. **256**:9246-9253.

15. **Schmid, M. B. and R. C. Johnson**. 1991. Southern revival-news of bacterial chromatin. Prokaryotic chromosomes: structure and function in genome design. Fifth Annual University of Alabama at Birmingham Biochemistry symposium, Panama City, FL, USA, May 8-12, 1991. New Biol. **3**:945-950.

16. **Drlica, K. and J. Rouviere-Yaniv**. 1987. Histone-like proteins of bacteria. Microbiol. Rev. **51**:301-319.

17. **Morisato, D. and N. Kleckner**. 1987. Tn10 transposition and circle formation *in vitro*. Cell **51**:101-111.

18. **Bramhill, D. and A. Kornberg**. 1988. Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. Cell **52**:743-755.

19. **Bonnefoy, E. and J. Rouviere-Yaniv**. 1991. HU and IHF, two homologous histone-like proteins of *Escherichia coli*, form different protein-DNA complexes with short DNA fragments. EMBO J. **10**:687-696.

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

21. **Higgins, C. F., J. C. Hinton, C. S. Hulton, T. Owen-Hughes, G. D. Pavitt, and A. Seirafi**. 1990. Protein H1: a role for chromatin structure in the regulation of bacterial gene expression and virulence? Mol. Microbiol. **4**:2007-2012.

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

23. **Yasuzawa, K., N. Hayashi, N. Goshima, K. Kohno, F. Imamoto, and Y. Kano**. 1992. Histone-like proteins are required for cell growth and constraint of supercoils in DNA. Gene **122**:9-15.

24. **Talukder, A. A. and A. Ishihama**. 1999. Twelve species of the nucleoid-associated protein from *Escherichia coli*. Sequence recognition specificity and DNA binding affinity. J. Biol. Chem. **274**:33105-33113.

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

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

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

28. **Talukder, A. A., A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama**. 1999. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. J. Bacteriol. **181**:6361-6370.

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

30. **Ball, C. A., R. Osuna, K. C. Ferguson, and R. C. Johnson**. 1992. Dramatic changes in Fis levels upon nutrient upshift in *Escherichia coli*. J. Bacteriol. **174**:8043-8056.

31. **Ninnemann, O., C. Koch, and R. Kahmann**. 1992. The *E.coli fis* promoter is subject to stringent control and autoregulation. EMBO J. **11**:1075-1083.

32. **Ueguchi, C., M. Kakeda, H. Yamada, and T. Mizuno**. 1994. An analogue of the DnaJ molecular chaperone in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A **91**:1054-1058.

33. **Kakeda, M., C. Ueguchi, H. Yamada, and T. Mizuno**. 1995. An *Escherichia coli* curved DNA-binding protein whose expression is affected by the stationary phase-specific sigma factor sigma S. Mol. Gen. Genet. **248**:629-634.

34. **Jair, K. W., X. Yu, K. Skarstad, B. Thony, N. Fujita, A. Ishihama, and R. E. Wolf**. 1996. Transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the *Escherichia coli* origin of chromosomal replication. J. Bacteriol. **178**:2507-2513.

35. **Ariza, R. R., Z. Li, N. Ringstad, and B. Demple**. 1995. Activation of multiple antibiotic resistance and binding of stress- inducible promoters by *Escherichia coli* Rob protein. J. Bacteriol. **177**:1655-1661.

36. **Fuller, R. S., B. E. Funnell, and A. Kornberg**. 1984. The dnaA protein complex with the *E. coli* chromosomal replication origin (*oriC*) and other DNA sites. Cell **38**:889-900.

37. **Atlung, T., E. S. Clausen, and F. G. Hansen**. 1985. Autoregulation of the *dnaA* gene of *Escherichia coli* K12. Mol. Gen. Genet. **200**:442-450.

38. **Messer, W. and C. Weigel**. 1997. DnaA initiator-also a transcription factor. Mol. Microbiol. **24**:1-6.

39. **Almiron, M., A. J. Link, D. Furlong, and R. Kolter**. 1992. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. Genes Dev. **6**:2646-2654.

40. **Martinez, A. and R. Kolter**. 1997. Protection of DNA during oxidative stress by the nonspecific DNA- binding protein Dps. J. Bacteriol. **179**:5188-5194.

41. **Ishihama, A.** 1999. Modulation of the nucleoid, the transcription apparatus, and the translation machinery in bacteria for stationary phase survival. Genes Cells **4**:135-143.

42. **Kajitani, M., A. Kato, A. Wada, Y. Inokuchi, and A. Ishihama**. 1994. Regulation of the *Escherichia coli hfq* gene encoding the host factor for phage Q beta. J. Bacteriol. **176**:531-534.

43. **Kajitani, M. and A. Ishihama**. 1991. Identification and sequence determination of the host factor gene for bacteriophage Q beta. Nucleic Acids Res. **19**:1063-1066.

44. **Tsui, H. C., H. C. Leung, and M. E. Winkler**. 1994. Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12. Mol. Microbiol. **13**:35-49.

45. **Hwang, D. S. and A. Kornberg**. 1990. A novel protein binds a key origin sequence to block replication of an *E. coli* minichromosome. Cell **63** :325-331.

46. **Lee, Y. S., H. Kim, and D. S. Hwang**. 1996. Transcriptional activation of the *dnaA* gene encoding the initiator for *oriC* replication by IciA protein, an inhibitor of *in vitro oriC* replication in *Escherichia coli*. Mol. Microbiol. **19**:389-396.

47. **Wei, T. and R. Bernander**. 1996. Interaction of the IciA protein with AT-rich regions in plasmid replication origins. Nucleic Acids Res. **24**:1865-1872.

48. **Newman, E. B. and R. Lin**. 1995. Leucine-responsive regulatory protein: a global regulator of gene expression in *E. coli*. Annu. Rev. Microbiol. **49**:747-775.

49. **Cusick, M. E. and M. Belfort**. 1998. Domain structure and RNA annealing activity of the *Escherichia coli* regulatory protein StpA. Mol. Microbiol. **28**:847-857.

50. **Lutz, H., K. Von Meyenburg, and U. Hubscher**. 1985. Quantitation with monoclonal antibodies of *Escherichia coli* H protein suggests histone function. J. Bacteriol. **162**:1005-1007.

51. **Friedman, D. I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545-554.

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

53. **Collado-Vides, J., B. Magasanik, and J. D. Gralla**. 1991. Control site location and transcriptional regulation in *Escherichia coli*. Microbiol. Rev. **55**:371-394.

54. **Goosen, N. and P. P. van de**. 1995. The regulation of transcription initiation by integration host factor. Mol. Microbiol. **16**:1-7.

55. **Bruckner, R. C. and M. M. Cox**. 1989. The histone-like H protein of *Escherichia coli* is ribosomal protein S3. Nucleic Acids Res. **17**:3145-3161.

56. **Hubscher, U., H. Lutz, and A. Kornberg**. 1980. Novel histone H2A-like protein of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A **77**:5097-5101.

57. **Li, Z. J., D. Hillyard, and P. Higgins**. 1989. Nucleotide sequence of the *Salmonella typhimurium himA* gene. Nucleic Acids Res. **17**:8880.

58. **Toussaint, B., I. Delic-Attree, D. 'A. De Sury, L. David, M. Vincon, and P. M. Vignais**. 1993. Purification of the integration host factor homolog of *Rhodobacter capsulatus*: cloning and sequencing of the *hip* gene, which encodes the beta subunit. J. Bacteriol. **175**:6499-6504.

59. **Calb, R., A. Davidovitch, S. Koby, H. Giladi, D. Goldenberg, H. Margalit, A. Holtel, K. Timmis, J. M. Sanchez-Romero, L. de, V, and A. B. Oppenheim**. 1996. Structure and function of the *Pseudomonas putida* integration host factor. J. Bacteriol. **178**:6319-6326.

60. **Douillie, A., A. Toussaint, and M. Faelen**. 1994. Identification of the integration host factor genes of *Erwinia chrysanthemi* 3937. Biochimie **76**:1055-1062.

61. **Sojda, J., III, B. Gu, J. Lee, T. R. Hoover, and B. T. Nixon**. 1999. A rhizobial homolog of IHF stimulates transcription of *dctA* in *Rhizobium leguminosarum* but not in *Sinorhizobium meliloti*. Gene **238**:489-500.

62. **Palacios, S. and J. C. Escalante-Semerena**. 2000. *prpR*, *ntrA*, and *ihf* functions are required for expression of the *prpBCDE* operon, encoding enzymes that catabolize propionate in *Salmonella enterica* serovar *typhimurium* LT2. J. Bacteriol. **182**:905-910.

63. **Highlander, S. K., O. Garza, B. J. Brown, S. Koby, and A. B. Oppenheim**. 1997. Isolation and characterization of the integration host factor genes of *Pasteurella haemolytica*. FEMS Microbiol. Lett. **146**:181-188.

64. **Thao, M. L. and P. Baumann**. 1997. Nucleotide sequence of a DNA fragment from *Buchnera aphidicola* (Aphid endosymbiont) containing the genes *aspS-trxB-serS-serC-aroA-rpsA-himD-tpiA*. Curr. Microbiol. **35**:68-69.

65. **Delic-Attree, I., B. Toussaint, and P. M. Vignais**. 1995. Cloning and sequence analyses of the genes coding for the integration host factor (IHF) and HU proteins of *Pseudomonas aeruginosa*. Gene **154**:61-64.

66. **Hill, S. A., R. J. Belland, and J. Wilson**. 1998. The *ihf* mRNA levels decline as *Neisseria gonorrhoeae* enters the stationary growth phase. Gene **215**:303-310.

67. **Haluzi, H., D. Goitein, S. Koby, I. Mendelson, D. Teff, G. Mengeritsky, H. Giladi, and A. B. Oppenheim**. 1991. Genes coding for integration host factor are conserved in gram-negative bacteria. J. Bacteriol. **173**:6297-6299.

68. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman**. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389-3402.

69. **Miller, H. I.** 1984. Primary structure of the *himA* gene of *Escherichia coli*: homology with DNA-binding protein HU and association with the phenylalanyl-tRNA synthetase operon. Cold Spring Harb. Symp. Quant. Biol. **49**:691-698.

70. **Mechulam, Y., G. Fayat, and S. Blanquet**. 1985. Sequence of the *Escherichia coli pheST* operon and identification of the *himA* gene. J. Bacteriol. **163**:787-791.

71. **Devalckenaere, A., M. Odaert, P. Trieu-Cuot, and M. Simonet**. 1999. Characterization of IS1541-like elements in *Yersinia enterocolitica* and *Yersinia pseudotuberculosis*. FEMS Microbiol. Lett. **176** :229-233.

72. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, and O. White**. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406**:477-483.

73. **Simpson, A. J., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. Briones, M. R. Bueno, A. A. Camargo, L. E. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El Dorry, A. P. Facincani, A. J. Ferreira, V. C. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. Leite, E. G. Lemos, M. V. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. Madeira, H. M. Madeira, and C. L. Marino**. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. Nature **406** :151-157.

74. **May, B. J., Q. Zhang, L. L. Li, M. L. Paustian, T. S. Whittam, and V. Kapur**. 2001. Complete genomic sequence of *Pasteurella multocida*,Pm70. Proc. Natl. Acad. Sci. U.S.A **98**:3460-3465.

75. **Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, J. F. Peden, R. J. Dodson, W. C. Nelson, M. L. Gwinn, R. DeBoy, J. D. Peterson, E. K. Hickey, D. H. Haft, S. L. Salzberg, O. White, R. D. Fleischmann, B. A. Dougherty, T. Mason, A. Ciecko, D. S. Parksey, E. Blair, H. Cittone, E. B. Clark, M. D. Cotton, T. R. Utterback, H. Khouri, H. Qin, J. Vamathevan, J. Gill, V. Scarlato, V. Masignani, M. Pizza, G. Grandi, L. Sun, H. O. Smith, C. M. Fraser, E. R. Moxon, R. Rappuoli, and J. C. Venter**. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287**:1809-1815.

76. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick**. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269**:496-512.

77. **Shigenobu, S., H. Watanabe, M. Hattori, Y. Sakaki, and H. Ishikawa**. 2000. Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS. Nature **407**:81-86.

78. **Kaneko, T., Y. Nakamura, S. Sato, E. Asamizu, T. Kato, S. Sasamoto, A. Watanabe, K. Idesawa, A. Ishikawa, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, Y. Mochizuki, S. Nakayama, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, and S. Tabata**. 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Res. **7**:331-338.

79. **Moreno, A. J., M. Fontes, and F. J. Murillo**. 2001. *ihfA* gene of the bacterium *Myxococcus xanthus* and its role in activation of carotenoid genes by blue light. J. Bacteriol. **183**:557-569.

80. **Toussaint, B., C. Bosc, P. Richaud, A. Colbeau, and P. M. Vignais**. 1991. A mutation in a *Rhodobacter capsulatus* gene encoding an integration host factor-like protein impairs *in vivo* hydrogenase expression. Proc. Natl. Acad. Sci. U.S.A **88**:10749-10753.

81. **Nierman, W. C., T. V. Feldblyum, M. T. Laub, I. T. Paulsen, K. E. Nelson, J. Eisen, J. F. Heidelberg, M. R. Alley, N. Ohta, J. R. Maddock, I. Potocka, W. C. Nelson, A. Newton, C. Stephens, N. D. Phadke, B. Ely, R. T. DeBoy, R. J. Dodson, A. S. Durkin, M. L. Gwinn, D. H. Haft, J. F. Kolonay, J. Smit, M. B. Craven, H. Khouri, J. Shetty, K. Berry, T. Utterback, K. Tran, A. Wolf, J. Vamathevan, M. Ermolaeva, O. White, S. L. Salzberg, J. C. Venter, L. Shapiro, and C. M. Fraser**. 2001. Complete genome sequence of *Caulobacter crescentus*. Proc. Natl. Acad. Sci. U.S.A **98**:4136-4141.

82. **Echols, H.** 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. Science **233**:1050-1056.

83. **Andersson, S. G., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler, and C. G. Kurland**. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **396**:133-140.

84. **Kalman, S., W. Mitchell, R. Marathe, C. Lammel, J. Fan, R. W. Hyman, L. Olinger, J. Grimwood, R. W. Davis, and R. S. Stephens**. 1999. Comparative genomes of *Chlamydia pneumoniae* and *C. trachomatis*. Nat. Genet. **21**:385-389.

85. **Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis**. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science **282**:754-759.

86. **Flamm, E. L. and R. A. Weisberg**. 1985. Primary structure of the *hip* gene of *Escherichia coli* and of its product, the beta subunit of integration host factor. J. Mol. Biol. **183**:117-128.

87. **Mackenzie, C., A. E. Simmons, and S. Kaplan**. 1999. Multiple chromosomes in bacteria. The yin and yang of *trp* gene localization in *Rhodobacter sphaeroides* 2.4.1. Genetics **153**:525-538.

88. **Thompson, D. V., L. S. Melchers, K. B. Idler, R. A. Schilperoort, and P. J. Hooykaas**. 1988. Analysis of the complete nucleotide sequence of the *Agrobacterium tumefaciens virB* operon. Nucleic Acids Res. **16**:4621-4636.

89. **Schrammeijer, B., A. Beijersbergen, K. B. Idler, L. S. Melchers, D. V. Thompson, and P. J. Hooykaas.** 2000. Sequence analysis of the *vir*-region from *Agrobacterium tumefaciens* octopine Ti plasmid pTi15955. J. Exp. Bot. **51**:1167-1169.

90. **Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, and R. V. Swanson.** 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392**:353-358.

91. **Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser.** 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. **28**:1397-1406.

92. **Williams, J. G. K., D. L. Wulff, and H. A. Nash.** 1977. A mutant of *Escherichia coli* deficient in a host function required for phage lambda integration and excision, p. 357-361. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (eds.), DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

93. **Landy, A.** 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. Annu. Rev. Biochem. **58**:913-949.

94. **Miller, H. I. and D. I. Friedman.** 1977. Isolation of *Escherichia coli* mutants unable to support lambda integrative recombination, p. 349-356. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (eds.), DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

95. **Kikuchi, Y. and H. A. Nash.** 1978. The bacteriophage lambda *int* gene product. A filter assay for genetic recombination, purification of int, and specific binding to DNA. J. Biol. Chem. **253**:7149-7157.

96. **Kikuchi, Y. and H. Nash.** 1979. Integrative recombination of bacteriophage lambda: requirement for supertwisted DNA *in vivo* and characterization of int. Cold Spring Harb. Symp. Quant. Biol. **43 Pt 2**:1099-1109.

97. **Miller, H. I. and D. I. Friedman.** 1980. An *E. coli* gene product required for lambda site-specific recombination. Cell **20**:711-719.

98. **Miller, H. I., A. Kikuchi, H. A. Nash, R. A. Weisberg, and D. I. Friedman.** 1979. Site-specific recombination of bacteriophage lambda: the role of host gene products. Cold Spring Harb. Symp. Quant. Biol. **43 Pt 2**:1121-1126.

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

100. **Miller, H. I., M. Kirk, and H. Echols.** 1981. SOS induction and autoregulation of the *himA* gene for site-specific recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A **78**:6754-6758.

101. **Weisberg, R. A., M. Freundlich, D. Friedman, J. Gardner, N. Goosen, H. Nash, A. Oppenheim, and J. Rouviere-Yaniv.** 1996. Nomenclature of the genes encoding IHF [letter]. Mol. Microbiol. **19**:642.

102. **Fayat, G., S. Blanquet, P. Dessen, G. Batelier, and J. P. Waller**. 1974. The molecular weight and subunit composition of phenylalanyl-tRNA synthetase from *Escherichia coli* K-12. Biochimie **56**:35-41.

103. **Rosenberg, M. and D. Court**. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319-353.

104. **Mechulam, Y., S. Blanquet, and G. Fayat**. 1987. Dual level control of the *Escherichia coli pheST-himA* operon expression. tRNA(Phe)-dependent attenuation and transcriptional operator-repressor control by *himA* and the SOS network. J. Mol. Biol. **197**:453-470.

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

106. **Kikuchi, A., E. Flamm, and R. A. Weisberg**. 1985. An *Escherichia coli* mutant unable to support site-specific recombination of bacteriophage lambda. J. Mol. Biol. **183**:129-140.

107. **Prentki, P., M. Chandler, and D. J. Galas**. 1987. *Escherichia coli* integration host factor bends the DNA at the ends of IS1 and in an insertion hotspot with multiple IHF binding sites. EMBO J. **6**:2479-2487.

108. **Robertson, C. A. and H. A. Nash**. 1988. Bending of the bacteriophage lambda attachment site by *Escherichia coli* integration host factor. J. Biol. Chem. **263**:3554-3557.

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

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

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

112. **Kano, Y. and F. Imamoto**. 1990. Requirement of integration host factor (IHF) for growth of *Escherichia coli* deficient in HU protein. Gene **89**:133-137.

113. **Siegele, D. A. and R. Kolter**. 1992. Life after log. J. Bacteriol. **174**:345-348.

114. **Kolter, R., D. A. Siegele, and A. Tormo**. 1993. The stationary phase of the bacterial life cycle. Annu. Rev. Microbiol. **47**:855-874.

115. **Craig, N. L. and H. A. Nash**. 1984. *E. coli* integration host factor binds to specific sites in DNA. Cell **39**:707-716.

116. **Goodrich, J. A., M. L. Schwartz, and W. R. McClure**. 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18**:4993-5000.

117. **Kur, J., N. Hasan, and W. Szybalski**. 1989. Physical and biological consequences of interactions between integration host factor (IHF) and coliphage lambda late p'R promoter and its mutants. Gene **81**:1-15.

118. **Winkelman, J. W. and G. W. Hatfield.** 1990. Characterization of the integration host factor binding site in the *ilv*PG1 promoter region of the *ilvGMEDA* operon of *Escherichia coli.* J. Biol. Chem. **265**:10055-10060.

119. **Zulianello, L., P. van Ulsen, P. P. van de, and N. Goosen.** 1995. Participation of the flank regions of the integration host factor protein in the specificity and stability of DNA binding. J. Biol. Chem. **270**:17902-17907.

120. **Gardner, J. F. and H. A. Nash** . 1986. Role of *Escherichia coli* IHF protein in lambda site-specific recombination. A mutational analysis of binding sites. J. Mol. Biol. **191**:181-189.

121. **Mengeritsky, G., D. Goldenberg, I. Mendelson, H. Giladi, and A. B. Oppenheim.** 1993. Genetic and biochemical analysis of the integration host factor of *Escherichia coli.* J. Mol. Biol. **231**:646-657.

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

123. **Wang, S., R. Cosstick, J. F. Gardner, and R. I. Gumport.** 1995. The specific binding of *Escherichia coli* integration host factor involves both major and minor grooves of DNA. Biochemistry **34**:13082-13090.

124. **Morse, B. K., R. Michalczyk, and L. D. Kosturko.** 1994. Multiple molecules of integration host factor (IHF) at a single DNA binding site, the bacteriophage lambda *cos* I1 site. Biochimie **76**:1005-1017.

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

126. **Zinkel, S. S. and D. M. Crothers.** 1990. Comparative gel electrophoresis measurement of the DNA bend angle induced by the catabolite activator protein. Biopolymers **29**:29-38.

127. **Kosturko, L. D., E. Daub, and H. Murialdo.** 1989. The interaction of *E. coli* integration host factor and lambda *cos* DNA: multiple complex formation and protein-induced bending. Nucleic Acids Res. **17**:317-334.

128. **Lee, E. C., L. M. Hales, R. I. Gumport, and J. F. Gardner.** 1992. The isolation and characterization of mutants of the integration host factor (IHF) of *Escherichia coli* with altered, expanded DNA-binding specificities. EMBO J. **11**:305-313.

129. **Nunes-Duby, S. E., L. I. Smith-Mungo, and A. Landy.** 1995. Single base-pair precision and structural rigidity in a small IHF- induced DNA loop. J. Mol. Biol. **253**:228-242.

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

131. **Sun, D., L. H. Hurley, and R. M. Harshey.** 1996. Structural distortions induced by integration host factor (IHF) at the H' site of phage lambda probed by (+)-CC-1065, pluramycin, and KMnO4 and by DNA cyclization studies. Biochemistry **35**:10815-10827.

132. **von Hippel, P. H.** 1994. Protein-DNA recognition: new perspectives and underlying themes [comment]. Science **263**:769-770.

133. **Goodman, S. D., N. J. Velten, Q. Gao, S. Robinson, and A. M. Segall.** 1999. *In vitro* selection of integration host factor binding sites. J. Bacteriol. **181**:3246-3255.

134. **Kurumizaka, H., F. Kanke, U. Matsumoto, and H. Shindo**. 1992. Specific and nonspecific interactions of integration host factor with oligo DNAs as revealed by circular dichroism spectroscopy and filter binding assay. Arch. Biochem. Biophys. **295**:297-301.

135. **Lee, E. C., M. P. MacWilliams, R. I. Gumport, and J. F. Gardner**. 1991. Genetic analysis of *Escherichia coli* integration host factor interactions with its bacteriophage lambda H' recognition site. J. Bacteriol. **173**:609-617.

136. **van Rijn, P. A., P. P. van de, and N. Goosen**. 1991. Analysis of the IHF binding site in the regulatory region of bacteriophage Mu. Nucleic Acids Res. **19**:2825-2834.

137. **Gamas, P., M. G. Chandler, P. Prentki, and D. J. Galas**. 1987. *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS1 and to its major insertion hot-spot in pBR322. J. Mol. Biol. **195**:261-272.

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

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

140. **Teter, B., S. D. Goodman, and D. J. Galas**. 2000. DNA bending and twisting properties of integration host factor determined by DNA cyclization. Plasmid **43**:73-84.

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

142. **Giese, K., J. Cox, and R. Grosschedl**. 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. Cell **69**:185-195.

143. **Goodman, S. D., S. C. Nicholson, and H. A. Nash**. 1992. Deformation of DNA during site-specific recombination of bacteriophage lambda: replacement of IHF protein by HU protein or sequence-directed bends. Proc. Natl. Acad. Sci. U.S.A **89**:11910-11914.

144. **Molina-Lopez, J. A., F. Govantes, and E. Santero**. 1994. Geometry of the process of transcription activation at the sigma 54- dependent *nifH* promoter of *Klebsiella pneumoniae*. J. Biol. Chem. **269**:25419-25425.

145. **Segall, A. M., S. D. Goodman, and H. A. Nash**. 1994. Architectural elements in nucleoprotein complexes: interchangeability of specific and non-specific DNA binding proteins. EMBO J. **13**:4536-4548.

146. **Goodman, S. D. and O. Kay**. 1999. Replacement of integration host factor protein-induced DNA bending by flexible regions of DNA. J. Biol. Chem. **274**:37004-37011.

147. **Solomon, M. J., F. Strauss, and A. Varshavsky**. 1986. A mammalian high mobility group protein recognizes any stretch of six A.T base pairs in duplex DNA. Proc. Natl. Acad. Sci. U.S.A **83**:1276-1280.

148. **Pruijn, G. J., R. T. van Miltenburg, J. A. Claessens, and P. C. van der Vliet**. 1988. Interaction between the octamer-binding protein nuclear factor III and the adenovirus origin of DNA replication. J. Virol. **62**:3092-3102.

149. **Tanaka, I., K. Appelt, J. Dijk, S. W. White, and K. S. Wilson.** 1984. 3-A resolution structure of a protein with histone-like properties in prokaryotes. Nature **310**:376-381.

150. **Yang, S. W. and H. A. Nash.** 1994. Specific photocrosslinking of DNA-protein complexes: identification of contacts between integration host factor and its target DNA. Proc. Natl. Acad. Sci. U.S.A **91**:12183-12187.

151. **Granston, A. E. and H. A. Nash.** 1993. Characterization of a set of integration host factor mutants deficient for DNA binding. J. Mol. Biol. **234**:45-59.

152. **Rice, P. A.** 1997. Making DNA do a U-turn: IHF and related proteins. Curr. Opin. Struct. Biol. **7**:86-93.

153. **Lorenz, M., A. Hillisch, S. D. Goodman, and S. Diekmann.** 1999. Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. Nucleic Acids Res. **27**:4619-4625.

154. **Nash, H. A.** 1981. Integration and excision of bacteriophage lambda: the mechanism of conservation site specific recombination. Annu. Rev. Genet. **15**:143-167.

155. **Weisberg, R. A. and A. Landy.** 1983. Site-specific recombination in phage lambda, p. 211-250. *In* R. W. Hendrix, roberts.J.W., F. W. stahl, and R. A. Weisberg (eds.), Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

156. **Mizuuchi, K., R. Weisberg, L. Enquist, M. Mizuuchi, M. Buraczynska, C. Foeller, P. L. Hsu, W. Ross, and A. Landy.** 1981. Structure and function of the phage lambda *att* site: size, int-binding sites, and location of the crossover point. Cold Spring Harb. Symp. Quant. Biol. **45 Pt 1**:429-437.

157. **Miller, H. I., M. A. Mozola, and D. I. Friedman.** 1980. int-h: An int mutation of phage lambda that enhances site-specific recombination. Cell **20**:721-729.

158. **Lange-Gustafson, B. J. and H. A. Nash.** 1984. Purification and properties of Int-h, a variant protein involved in site-specific recombination of bacteriophage lambda. J. Biol. Chem. **259**:12724-12732.

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

160. **Leong, J. M., S. Nunes-Duby, C. F. Lesser, P. Youderian, M. M. Susskind, and A. Landy.** 1985. The phi 80 and P22 attachment sites. Primary structure and interaction with *Escherichia coli* integration host factor. J. Biol. Chem. **260**:4468-4477.

161. **Cho, E. H., C. E. Nam, R. Alcaraz, and J. F. Gardner.** 1999. Site-specific recombination of bacteriophage P22 does not require integration host factor. J. Bacteriol. **181**:4245-4249.

162. **Surette, M. G. and G. Chaconas.** 1989. A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is *Escherichia coli* integration host factor. J. Biol. Chem. **264**:3028-3034.

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

164. **Krause, H. M. and N. P. Higgins**. 1986. Positive and negative regulation of the Mu operator by Mu repressor and *Escherichia coli* integration host factor. J. Biol. Chem. **261**:3744-3752.

165. **Higgins, N. P., D. A. Collier, M. W. Kilpatrick, and H. M. Krause**. 1989. Supercoiling and integration host factor change the DNA conformation and alter the flow of convergent transcription in phage Mu. J. Biol. Chem. **264**:3035-3042.

166. **Chalmers, R., A. Guhathakurta, H. Benjamin, and N. Kleckner**. 1998. IHF modulation of Tn*10* transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. Cell **93**:897-908.

167. **Sakai, J. S., N. Kleckner, X. Yang, and A. Guhathakurta**. 2000. Tn*10* transpososome assembly involves a folded intermediate that must be unfolded for target capture and strand transfer. EMBO J. **19**:776-785.

168. **Shiga, Y., Y. Sekine, and E. Ohtsubo**. 1999. Transposition of IS1 circles. Genes Cells **4**:551-561.

169. **Brody, H. and C. W. Hill**. 1988. Attachment site of the genetic element e14. J. Bacteriol. **170**:2040-2044.

170. **Miller, H. I.** 1981. Multilevel regulation of bacteriophage lambda lysogeny by the *E. coli himA* gene. Cell **25**:269-276.

171. **Hoyt, M. A., D. M. Knight, A. Das, H. I. Miller, and H. Echols**. 1982. Control of phage lambda development by stability and synthesis of cII protein: role of the viral *cIII* and host *hflA*, *himA* and *himD* genes. Cell **31**:565-573.

172. **Mahajna, J., A. B. Oppenheim, A. Rattray, and M. Gottesman**. 1986. Translation initiation of bacteriophage lambda gene *cII* requires integration host factor. J. Bacteriol. **165**:167-174.

173. **Peacock, S., H. Weissbach, and H. A. Nash**. 1984. *In vitro* regulation of phage lambda *cII* gene expression by *Escherichia coli* integration host factor. Proc. Natl. Acad. Sci. U.S.A **81**:6009-6013.

174. **Peacock, S. and H. Weissbach**. 1985. IHF stimulation of lambda *cII* gene expression is inhibited by the *E. coli* NusA protein. Biochem. Biophys. Res. Commun. **127**:1026-1031.

175. **Giladi, H., M. Gottesman, and A. B. Oppenheim**. 1990. Integration host factor stimulates the phage lambda pL promoter. J. Mol. Biol. **213**:109-121.

176. **Giladi, H., S. Koby, M. E. Gottesman, and A. B. Oppenheim**. 1992. Supercoiling, integration host factor, and a dual promoter system, participate in the control of the bacteriophage lambda pL promoter. J. Mol. Biol. **224**:937-948.

177. **Court, D. and A. B. Oppenheim** . 1983. p. 251-277. *In* R. W. Hendrix, roberts.J.W., F. W. stahl, and R. A. Weisberg (eds.), Lambda II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

178. **Giladi, H., K. Igarashi, A. Ishihama, and A. B. Oppenheim**. 1992. Stimulation of the phage lambda pL promoter by integration host factor requires the carboxy terminus of the alpha-subunit of RNA polymerase. J. Mol. Biol. **227**:985-990.

179. **Giladi, H., S. Koby, G. Prag, M. Engelhorn, J. Geiselmann, and A. B. Oppenheim**. 1998. Participation of IHF and a distant UP element in the stimulation of the phage lambda P_L promoter. Mol. Microbiol. **30**:443-451.

180. **Goosen, N. and P. P. van de**. 1984. Regulation of Mu transposition. I. Localization of the presumed recognition sites for HimD and Ner functions controlling bacteriophage Mu transcription. Gene **30**:41-46.

181. **van Ulsen, P., M. Hillebrand, L. Zulianello, P. P. van de, and N. Goosen**. 1996. Integration host factor alleviates the H-NS-mediated repression of the early promoter of bacteriophage Mu. Mol. Microbiol. **21**:567-578.

182. **Betermier, M., P. Rousseau, R. Alazard, and M. Chandler**. 1995. Mutual stabilisation of bacteriophage Mu repressor and histone-like proteins in a nucleoprotein structure. J. Mol. Biol. **249**:332-341.

183. **Goosen, N., M. van Heuvel, G. F. Moolenaar, and P. P. van de**. 1984. Regulation of Mu transposition. II. The *Escherichia coli* HimD protein positively controls two repressor promoters and the early promoter of bacteriophage Mu. Gene **32**:419-426.

184. **Alazard, R., M. Betermier, and M. Chandler**. 1992. *Escherichia coli* integration host factor stabilizes bacteriophage Mu repressor interactions with operator DNA *in vitro*. Mol. Microbiol. **6**:1707-1714.

185. **Gama, M. J., A. Toussaint, and N. P. Higgins**. 1992. Stabilization of bacteriophage Mu repressor-operator complexes by the *Escherichia coli* integration host factor protein. Mol. Microbiol. **6**:1715-1722.

186. **Boffini, A. and P. Prentki**. 1991. Identification of protein binding sites in genomic DNA by two-dimensional gel electrophoresis. Nucleic Acids Res. **19**:1369-1374.

187. **Arfin, S. M., A. D. Long, E. T. Ito, L. Tolleri, M. M. Riehle, E. S. Paegle, and G. W. Hatfield**. 2000. Global gene expression profiling in *Escherichia coli* K12. The effects of integration host factor. J. Biol. Chem. **275** :29672-29684.

188. **Gralla, J. D. and J. Collado-Vides**. 1996. Organization and function of transcription regulatory elements, p. 1232-1245. *In* F. C. Neidhardt and et al. (eds.), *Escherichia coli* and *Salmonella* : cellular and molecular biology. ASM Press, Washington, D.C.

189. **Carmona, M., F. Claverie-Martin, and B. Magasanik**. 1997. DNA bending and the initiation of transcription at sigma54-dependent bacterial promoters. Proc. Natl. Acad. Sci. U.S.A **94**:9568-9572.

190. **Bertoni, G., N. Fujita, A. Ishihama, and L. de, V**. 1998. Active recruitment of sigma54-RNA polymerase to the Pu promoter of *Pseudomonas putida*: role of IHF and alphaCTD. EMBO J. **17**:5120-5128.

191. **Jovanovic, G. and P. Model**. 1997. PspF and IHF bind co-operatively in the *psp* promoter-regulatory region of *Escherichia coli*. Mol. Microbiol. **25**:473-481.

192. **Lange, R., M. Barth, and R. Hengge-Aronis**. 1993. Complex transcriptional control of the sigma s-dependent stationary- phase-induced and osmotically regulated *osmY* (*csi-5*) gene suggests novel roles for Lrp, cyclic AMP (cAMP) receptor protein-cAMP complex, and integration host factor in the stationary-phase response of *Escherichia coli*. J. Bacteriol. **175**:7910-7917.

193. **Carmona, M. and B. Magasanik**. 1996. Activation of transcription at sigma 54-dependent promoters on linear templates requires intrinsic or induced bending of the DNA. J. Mol. Biol. **261**:348-356.

194. **Hoover, T. R., E. Santero, S. Porter, and S. Kustu.** 1990. The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. Cell **63**:11-22.

195. **Cannon, W. V., R. Kreutzer, H. M. Kent, E. Morett, and M. Buck.** 1990. Activation of the *Klebsiella pneumoniae nifU* promoter: identification of multiple and overlapping upstream NifA binding sites. Nucleic Acids Res. **18**:1693-1701.

196. **Claverie-Martin, F. and B. Magasanik.** 1991. Role of integration host factor in the regulation of the *glnHp2* promoter of *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A **88**:1631-1635.

197. **Charlton, W., W. Cannon, and M. Buck.** 1993. The *Klebsiella pneumoniae nifJ* promoter: analysis of promoter elements regulating activation by the NifA promoter. Mol. Microbiol. **7**:1007-1021.

198. **Hopper, S., M. Babst, V. Schlensog, H. M. Fischer, H. Hennecke, and A. Bock.** 1994. Regulated expression *in vitro* of genes coding for formate hydrogenlyase components of *Escherichia coli*. J. Biol. Chem. **269**:19597-19604.

199. **Weiner, L., J. L. Brissette, N. Ramani, and P. Model.** 1995. Analysis of the proteins and cis-acting elements regulating the stress- induced phage shock protein operon. Nucleic Acids Res. **23**:2030-2036.

200. **Dworkin, J., A. J. Ninfa, and P. Model.** 1998. A protein-induced DNA bend increases the specificity of a prokaryotic enhancer-binding protein. Genes Dev. **12**:894-900.

201. **de, L., V, M. Herrero, M. Metzke, and K. N. Timmis.** 1991. An upstream XylR- and IHF-induced nucleoprotein complex regulates the sigma 54-dependent Pu promoter of TOL plasmid. EMBO J. **10**:1159-1167.

202. **Perez-Martin, J. and L. de, V** . 1995. Integration host factor suppresses promiscuous activation of the sigma 54-dependent promoter Pu of *Pseudomonas putida*. Proc. Natl. Acad. Sci. U.S.A **92**:7277-7281.

203. **Engelhorn, M. and J. Geiselmann.** 1998. Maximal transcriptional activation by the IHF protein of *Escherichia coli* depends on optimal DNA bending by the activator. Mol. Microbiol. **30**:431-441.

204. **Friden, P., K. Voelkel, R. Sternglanz, and M. Freundlich.** 1984. Reduced expression of the isoleucine and valine enzymes in integration host factor mutants of *Escherichia coli*. J. Mol. Biol. **172**:573-579.

205. **Friedman, D. I., E. J. Olson, D. Carver, and M. Gellert.** 1984. Synergistic effect of *himA* and *gyrB* mutations: evidence that Him functions control expression of *ilv* and *xyl* genes. J. Bacteriol. **157**:484-489.

206. **Tsui, P. and M. Freundlich.** 1990. Integration host factor bends the DNA in the *Escherichia coli ilvBN* promoter region. Mol. Gen. Genet. **223** :349-352.

207. **Spira, B. and E. Yagil.** 1999. The integration host factor (IHF) affects the expression of the phosphate-binding protein and of alkaline phosphatase in *Escherichia coli*. Curr. Microbiol. **38**:80-85.

208. **Rowe, S., N. Hodson, G. Griffiths, and I. S. Roberts.** 2000. Regulation of the *Escherichia coli* K5 capsule gene cluster: evidence for the roles of H-NS, BipA, and integration host factor in regulation of group 2 capsule gene clusters in pathogenic *E. coli*. J. Bacteriol. **182**:2741-2745.

209. **Rabin, R. S., L. A. Collins, and V. Stewart**. 1992. *In vivo* requirement of integration host factor for *nar* (nitrate reductase) operon expression in *Escherichia coli* K-12. Proc. Natl. Acad. Sci. U.S.A **89**:8701-8705.

210. **Schroder, I., S. Darie, and R. P. Gunsalus**. 1993. Activation of the *Escherichia coli* nitrate reductase (*narGHJI*) operon by NarL and Fnr requires integration host factor. J. Biol. Chem. **268**:771-774.

211. **Wu, Y. F. and P. Datta**. 1992. Integration host factor is required for positive regulation of the *tdc* operon of *Escherichia coli*. J. Bacteriol. **174**:233-240.

212. **Sirko, A., A. Weglenska, and D. Hulanicka**. 1998. Integration host factor positively regulates *cycJIH* transcription. Mol. Gen. Genet. **258**:174-177.

213. **Gamas, P., L. Caro, D. Galas, and M. Chandler**. 1987. Expression of F transfer functions depends on the *Escherichia coli* integration host factor. Mol. Gen. Genet. **207**:302-305.

214. **Kennedy, M., M. Chandler, and D. Lane**. 1988. Mapping and regulation of the *pifC* promoter of the F plasmid. Biochim. Biophys. Acta **950**:75-80.

215. **Huang, L., P. Tsui, and M. Freundlich**. 1990. Integration host factor is a negative effector of *in vivo* and *in vitro* expression of *ompC* in *Escherichia coli*. J. Bacteriol. **172**:5293-5298.

216. **Tsui, P., L. Huang, and M. Freundlich**. 1991. Integration host factor binds specifically to multiple sites in the *ompB* promoter of *Escherichia coli* and inhibits transcription. J. Bacteriol. **173**:5800-5807.

217. **Ramani, N., L. Huang, and M. Freundlich**. 1992. *In vitro* interactions of integration host factor with the *ompF* promoter- regulatory region of *Escherichia coli*. Mol. Gen. Genet. **231**:248-255.

218. **Thompson, R. J. and G. Mosig**. 1988. Integration host factor (IHF) represses a *Chlamydomonas* chloroplast promoter in *E. coli*. Nucleic Acids Res. **16**:3313-3326.

219. **Colland, F., M. Barth, R. Hengge-Aronis, and A. Kolb**. 2000. Sigma factor selectivity of *Escherichia coli* RNA polymerase: role for CRP, IHF and lrp transcription factors. EMBO J. **19**:3028-3037.

220. **Altuvia, S., M. Almiron, G. Huisman, R. Kolter, and G. Storz**. 1994. The *dps* promoter is activated by OxyR during growth and by IHF and sigma S in stationary phase. Mol. Microbiol. **13**:265-272.

221. **Gamas, P., A. C. Burger, G. Churchward, L. Caro, D. Galas, and M. Chandler**. 1986. Replication of pSC101: effects of mutations in the *E. coli* DNA binding protein IHF. Mol. Gen. Genet. **204**:85-89.

222. **Stenzel, T. T., P. Patel, and D. Bastia**. 1987. The integration host factor of *Escherichia coli* binds to bent DNA at the origin of replication of the plasmid pSC101. Cell **49**:709-717.

223. **Lu, Y. B., H. J. Datta, and D. Bastia**. 1998. Mechanistic studies of initiator-initiator interaction and replication initiation. EMBO J. **17**:5192-5200.

224. **Burian, J., S. Stuchlik, and W. W. Kay**. 1999. Replication control of a small cryptic plasmid of *Escherichia coli*. J. Mol. Biol. **294**:49-65.

225. **Hwang, D. S. and A. Kornberg**. 1992. Opening of the replication origin of *Escherichia coli* by DnaA protein with protein HU or IHF. J. Biol. Chem. **267**:23083-23086.

226. **Polaczek, P., K. Kwan, and J. L. Campbell**. 1998. Unwinding of the *Escherichia coli* origin of replication (*oriC*) can occur in the absence of initiation proteins but is stabilized by DnaA and histone-like proteins IHF or HU. Plasmid **39**:77-83.

227. **Grimwade, J. E., V. T. Ryan, and A. C. Leonard**. 2000. IHF redistributes bound initiator protein, DnaA, on supercoiled *oriC* of *Escherichia coli*. Mol. Microbiol. **35**:835-844.

228. **Feiss, M., S. Frackman, and J. Sippy**. 1985. Essential interaction between lambdoid phage 21 terminase and the *Escherichia coli* integrative host factor. J. Mol. Biol. **183**:239-246.

229. **Funnell, B. E.** 1988. Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. Proc. Natl. Acad. Sci. U.S.A **85**:6657-6661.

230. **Hayes, F. and S. Austin**. 1994. Topological scanning of the P1 plasmid partition site. J. Mol. Biol. **243**:190-198.

231. **Bouet, J. Y., J. A. Surtees, and B. E. Funnell**. 2000. Stoichiometry of P1 plasmid partition complexes. J. Biol. Chem. **275**:8213-8219.

232. **Dempsey, W. B.** 1987. Integration host factor and conjugative transfer of the antibiotic resistance plasmid R100. J. Bacteriol. **169**:4391-4392.

233. **McIntire, S. A. and W. B. Dempsey**. 1987. *oriT* sequence of the antibiotic resistance plasmid R100. J. Bacteriol. **169**:3829-3832.

234. **Szekely, E., M. Silverman, and M. Simon**. 1983. Recombinational regulation of flagellin gene expression, p. 117-129. *In* K. F. Chater, C. A. Cullis, D. A. Hopwood, A. A. W. B. Johnston, and H. W. Woolhouse (eds.), Genetic rearrangement - Proceedings of the fifth John Innes symposium "Biological consequences of DNA structure and genome arrangement"-. Sinauer Associates, Inc., Sunderland, Massachusetts.

235. **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.

236. **Gilson, E., J. M. Clement, D. Brutlag, and M. Hofnung**. 1984. A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. **3**:1417-1421.

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

238. **Boccard, F. and P. Prentki**. 1993. Specific interaction of IHF with RIBs, a class of bacterial repetitive DNA elements located at the 3' end of transcription units. EMBO J. **12**:5019-5027.

239. **Stern, M. J., G. F. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins**. 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37**:1015-1026.

240. **Gilson, E., W. Saurin, D. Perrin, S. Bachellier, and M. Hofnung**. 1991. Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. **19**:1375-1383.

241. **Bachellier, S., W. Saurin, D. Perrin, M. Hofnung, and E. Gilson.** 1994. Structural and functional diversity among bacterial interspersed mosaic elements (BIMEs). Mol. Microbiol. **12**:61-70.

242. **Jovanovic, G. and P. Model.** 1997. The RIB element in the *goaG-pspF* intergenic region of *Escherichia coli*. J. Bacteriol. **179**:3095-3102.

243. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22** :4673-4680.

244. **Hofmann, K. and M. Baron.** Boxshade 3.21. Unpublished Work.

245. **Hwang, E. S. and J. J. Scocca** . 1990. Interaction of integration host factor from *Escherichia coli* with the integration region of the *Haemophilus influenzae* bacteriophage HP1. J. Bacteriol. **172** :4852-4860.

246. **Gober, J. W. and L. Shapiro.** 1990. Integration host factor is required for the activation of developmentally regulated genes in *Caulobacter*. Genes Dev. **4**:1494-1504.

247. **De Meirsman, C., C. Van Soom, C. Verreth, A. Van Gool, and J. Vanderleyden.** 1990. Nucleotide sequence analysis of IS427 and its target sites in *Agrobacterium tumefaciens* T37. Plasmid **24**:227-234.

248. **Autexier, C., S. Wragg-Legare, and M. S. DuBow.** 1991. Characterization of the *Pseudomonas aeruginosa* transposable phage D3112 left-end regulatory region. Biochim. Biophys. Acta **1088**:147-150.

249. **Suwanto, A. and S. Kaplan.** 1992. A self-transmissible, narrow-host-range endogenous plasmid of *Rhodobacter sphaeroides* 2.4.1: physical structure, incompatibility determinants, origin of replication, and transfer functions. J. Bacteriol. **174**:1124-1134.

250. **Gober, J. W. and L. Shapiro.** 1992. A developmentally regulated *Caulobacter* flagellar promoter is activated by 3' enhancer and IHF binding elements. Mol. Biol. Cell **3**:913-926.

251. **Mohr, C. D. and V. Deretic.** 1992. *In vitro* interactions of the histone-like protein IHF with the *algD* promoter, a critical site for control of mucoidy in *Pseudomonas aeruginosa*. Biochem. Biophys. Res. Commun. **189**:837-844.

252. **Lee, J. K., S. Wang, J. M. Eraso, J. Gardner, and S. Kaplan.** 1993. Transcriptional regulation of *puc* operon expression in *Rhodobacter sphaeroides*. Involvement of an integration host factor-binding sequence. J. Biol. Chem. **268**:24491-24497.

253. **Highlander, S. K. and G. M. Weinstock.** 1994. Static DNA bending and protein interactions within the *Pasteurella haemolytica* leukotoxin promoter region: development of an activation model for leukotoxin transcriptional control. DNA Cell Biol. **13**:171-181.

254. **Marques, M. V. and J. W. Gober.** 1995. Activation of a temporally regulated *Caulobacter* promoter by upstream and downstream sequence elements. Mol. Microbiol. **16**:279-289.

255. **Delic-Attree, I., B. Toussaint, A. Froger, J. C. Willison, and P. M. Vignais.** 1996. Isolation of an IHF-deficient mutant of a *Pseudomonas aeruginosa* mucoid isolate and evaluation of the role of IHF in *algD* gene expression. Microbiology **142 (Pt 10)**:2785-2793.

256. **Rondon, M. R. and J. C. Escalante-Semerena.** 1997. Integration host factor is required for 1,2-propanediol-dependent transcription of the *cob/pdu* regulon in *Salmonella typhimurium* LT2. J. Bacteriol. **179**:3797-3800.

257. **Hill, S. A., D. S. Samuels, J. H. Carlson, J. Wilson, D. Hogan, L. Lubke, and R. J. Belland.** 1997. Integration host factor is a transcriptional cofactor of *pilE* in *Neisseria gonorrhoeae*. Mol. Microbiol. **23**:649-656.

258. **Fyfe, J. A. and J. K. Davies.** 1998. An AT-rich tract containing an integration host factor-binding domain and two UP-like elements enhances transcription from the *pilEp1* promoter of *Neisseria gonorrhoeae*. J. Bacteriol. **180**:2152-2159.

259. **Kirndorfer, M., A. Jager, and G. Klug.** 1998. Integration host factor affects the oxygen-regulated expression of photosynthesis genes in *Rhodobacter capsulatus*. Mol. Gen. Genet. **258**:297-305.

260. **Horak, R. and M. Kivisaar.** 1998. Expression of the transposase gene *tnpA* of Tn4652 is positively affected by integration host factor. J. Bacteriol. **180**:2822-2829.

261. **Marshall, D. G., B. J. Sheehan, and C. J. Dorman.** 1999. A role for the leucine-responsive regulatory protein and integration host factor in the regulation of the *Salmonella* plasmid virulence (*spv*) locus in *Salmonella typhimurium*. Mol. Microbiol. **34**:134-145.

262. **Pitulle, C., Y. Yang, M. Marchiani, E. R. Moore, J. L. Siefert, M. Aragno, P. Jurtshuk, Jr., and G. E. Fox.** 1994. Phylogenetic position of the genus *Hydrogenobacter*. Int. J. Syst. Bacteriol. **44**:620-626.

263. **Santero, E., T. R. Hoover, A. K. North, D. K. Berger, S. C. Porter, and S. Kustu.** 1992. Role of integration host factor in stimulating transcription from the sigma 54-dependent *nifH* promoter. J. Mol. Biol. **227**:602-620.

264. **Porter, M. E. and C. J. Dorman.** 1997. Positive regulation of *Shigella flexneri* virulence genes by integration host factor. J. Bacteriol. **179**:6537-6550.

265. **Black, L. K. and R. J. Maier.** 1995. IHF- and RpoN-dependent regulation of hydrogenase expression in *Bradyrhizobium japonicum*. Mol. Microbiol. **16**:405-413.

266. **Machado, I. M., M. G. Yates, H. B. Machado, E. M. Souza, and F. O. Pedrosa.** 1996. Cloning and sequencing of the nitrogenase structural genes *nifHDK* of *Herbaspirillum seropedicae*. Braz. J. Med. Biol. Res. **29**:1599-1602.

267. **Wassem, R., E. M. De Souza, M. G. Yates, F. D. Pedrosa, and M. Buck.** 2000. Two roles for integration host factor at an enhancer-dependent *nifA* promoter. Mol. Microbiol. **35**:756-764.

268. **Stevenson, D. M., M. Kunnimalaiyaan, K. Muller, and P. S. Vary.** 1998. Characterization of a theta plasmid replicon with homology to all four large plasmids of *Bacillus megaterium* QM B1551. Plasmid **40**:175-189.

269. **White, S. W., K. Appelt, K. S. Wilson, and I. Tanaka.** 1989. A protein structural motif that bends DNA. Proteins **5**:281-288.

270. **Kobryn, K., D. Z. Naigamwalla, and G. Chaconas.** 2000. Site-specific DNA binding and bending by the *Borrelia burgdorferi* Hbb protein. Mol. Microbiol. **37**:145-155.

271. **Blattner, F. R., G. Plunkett, III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick,**

M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277**:1453-1474.

272. **Kawamura, S., H. Kajiyama, N. Yamasaki, and M. Kimura.** 1995. Cloning of the gene encoding DNA binding protein HU from *Bacillus stearothermophilus* and its expression in *Escherichia coli*. Biosci. Biotechnol. Biochem. **59**:126-129.

273. **Summers, M. L., L. M. Botero, S. C. Busse, and T. R. McDermott.** 2000. The *Sinorhizobium meliloti* lon protease is involved in regulating exopolysaccharide synthesis and is required for nodulation of alfalfa. J. Bacteriol. **182**:2551-2558.

274. **Kimura, M., J. Kimura, and R. Zierer.** 1984. The primary structure of the DNA-binding protein II from *Clostridium pasteurianum*. FEBS Lett. **175**:208-212.

275. **Tilly, K. and J. Campbell.** 1993. A *Borrelia burgdorferi* homolog of the *Escherichia coli rho* gene. Nucleic Acids Res. **21**:1040.

276. **Greene, J. R., S. M. Brennan, D. J. Andrew, C. C. Thompson, S. H. Richards, R. L. Heinrikson, and E. P. Geiduschek.** 1984. Sequence of the bacteriophage SP01 gene coding for transcription factor 1, a viral homologue of the bacterial type II DNA-binding proteins. Proc. Natl. Acad. Sci. U.S.A **81**:7031-7035.

277. **Grimaud, R.** 1996. Bacteriophage Mu head assembly. Virology **217**:200-210.

278. **Tolias, P. P. and M. S. DuBow** . 1985. The cloning and characterization of the bacteriophage D108 regulatory DNA-binding protein ner. EMBO J. **4**:3031-3037.

279. **Mizuuchi, M., R. A. Weisberg, and K. Mizuuchi.** 1986. DNA sequence of the control region of phage D108: the N-terminal amino acid sequences of repressor and transposase are similar both in phage D108 and in its relative, phage Mu. Nucleic Acids Res. **14**:3813-3825.

280. **Choi, Y. L., T. Nishida, M. Kawamukai, R. Utsumi, H. Sakai, and T. Komano.** 1989. Cloning and sequencing of an *Escherichia coli* gene, *nlp*, highly homologous to the *ner* genes of bacteriophages Mu and D108. J. Bacteriol. **171**:5222-5225.

281. **Garcia, J. A., S. H. Ou, F. Wu, A. J. Lusis, R. S. Sparkes, and R. B. Gaynor.** 1992. Cloning and chromosomal mapping of a human immunodeficiency virus 1 "TATA" element modulatory factor. Proc. Natl. Acad. Sci. U.S.A **89**:9372-9376.

282. **Aiba, H., T. Nakamura, H. Mitani, and H. Mori.** 1985. Mutations that alter the allosteric nature of cAMP receptor protein of *Escherichia coli*. EMBO J. **4**:3329-3332.

283. **Kawamukai, M., R. Utsumi, K. Takeda, A. Higashi, H. Matsuda, Y. L. Choi, and T. Komano.** 1991. Nucleotide sequence and characterization of the *sfs1* gene: *sfs1* is involved in CRP*-dependent mal gene expression in *Escherichia coli*. J. Bacteriol. **173**:2644-2648.

284. **Autexier, C. and M. S. DuBow.** 1992. The *Escherichia coli* Mu/D108 phage ner homologue gene (*nlp*) is transcribed and evolutionarily conserved among the *Enterobacteriaceae*. Gene **114**:13-18.

285. **Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, D. Brown, T. Chillingworth, R. M. Davies, P. Davis, K. Devlin, T. Feltwell, N. Hamlin, S. Holroyd, K. Jagels, S. Leather, S. Moule, K. Mungall, M. A. Quail, M. A. Rajandream, K. M. Rutherford, M. Simmonds, J. Skelton, S. Whitehead, B. G. Spratt, and B. G. Barrell.** 2000. Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature **404**:502-506.

286. **Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa.** 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res. **8**:11-22.

287. **Fortin, D. L.** 1997. Characterization of D108 Ner and its *Escherichia coli* Homolog Nlp.

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

289. **Maxam, A. M. and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-560.

290. **Hanahan, D.** 1983. Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**:557-580.

291. **Bukhari, A. I. and M. Metlay.** 1973. Genetic mapping of prophage Mu. Virology **54**:109-116.

292. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103-119.

293. **Seguin, B.** 1995. Unpublished work.

294. **Casadaban, M. J. and S. N. Cohen.** 1979. Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: *in vivo* probe for transcriptional control sequences. Proc. Natl. Acad. Sci. U.S.A **76**:4530-4533.

295. **Norrander, J. M., J. Vieira, I. Rubinstein, and J. Messing.** 1985. Manipulation and expression of the maize zein storage proteins in *E. coli*. Journal of Biotechnology **2**:157-175.

296. **Guzman, L. M., D. Belin, M. J. Carson, and J. Beckwith.** 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. J. Bacteriol. **177**:4121-4130.

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

298. **Bukhari, A. I. and E. Ljungquist.** 1977. Bacteriophage Mu: methods for cultivation and use, p. 749-756. *In* A. I. Bukhari, J. A. Shapiro, and S. L. Adhya (eds.), DNA Insertion Elements, Plasmids and Episomes. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

299. **Schumm, J. W., D. D. Moore, F. R. Blattner, and M. M. Howe.** 1980. Correlation of the genetic and physical maps in the central region of the bacteriophage Mu genome. Virology **105**:185-195.

300. **Dower, W. J., J. F. Miller, and C. W. Ragsdale.** 1988. High efficiency transformation of *E. coli* by high voltage electroporation. Nucleic Acids Res. **16**:6127-6145.

301. **Neidhardt, F. C., R. Wirth, M. W. Smith, and R. Van Bogelen.** 1980. Selective synthesis of plasmid-coded proteins by *Escherichia coli* during recovery from chloramphenicol treatment. J. Bacteriol. **143**:535-537.

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

303. **Clewell, D. B. and D. R. Helinski.** 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an opern circular DNA form. Proc. Natl. Acad. Sci. U.S.A **62**:1159-1166.

304. **Asai, K., S. Fujisaki, Y. Nishimura, T. Nishino, K. Okada, T. Nakagawa, M. Kawamukai, and H. Matsuda.** 1994. The identification of Escherichia coli *ispB* (*cel*) gene encoding the octaprenyl diphosphate synthase. Biochem. Biophys. Res. Commun. **202**:340-345.

305. **Okada, K., M. Minehira, X. Zhu, K. Suzuki, T. Nakagawa, H. Matsuda, and M. Kawamukai.** 1997. The *ispB* gene encoding octaprenyl diphosphate synthase is essential for growth of *Escherichia coli*. J. Bacteriol. **179**:3058-3060.

306. **Marquardt, J. L., D. A. Siegele, R. Kolter, and C. T. Walsh.** 1992. Cloning and sequencing of *Escherichia coli murZ* and purification of its product, a UDP-N-acetylglucosamine enolpyruvyl transferase. J. Bacteriol. **174**:5748-5752.

307. **Datta, N.** 1979. Plasmid classification: incompatibility grouping, p. 3. *In* K. N. Timmis and A. Pühler (eds.), Plasmids of medical, environmental, and commercial importance. Elsevier, Amsterdam.

308. **Roy, D. and DuBow, M. S.** 2001. Unpublished Work.