STUDIES ON GENE EXPRESSION IN A PATHOGENIC BACTERIUM

BY

MD. SHAHIDUL ISLAM

A THESIS

SUBMITTED TO THE UNIVERISTY OF BIRMINGHAM

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

SCHOOL OF BIOSCIENCES

THE UNIVERSITY OF BIRMINGHAM

EDGBASTON

BIRMINGHAM

B15 2TT

UK

JUNE 2011

Synopsis

Enterohaemorrhagic *Escherichia coli* (EHEC) O157:H7 Sakai is an emerging human pathogen. The genes responsible for EHEC virulence are contained in a pathogenicity island named the locus of enterocyte effacement (LEE). The LEE consists of five major polycistronic operons (*LEE1-5*), which co-ordinately encode a type three secretion apparatus and effector molecules that are associated with the attaching and effacing lesions in the intestine.

Expression of the genes in LEE is primarily coordinated by expression of the *LEE1* operon. GrlA is a LEE-encoded transcription regulator that has been proposed to be involved in the regulation of expression of the *LEE1* operon. To study GrlA-dependent effects at the *LEE1* operon regulatory region further, a simple plasmid-based system is described in this work. The work reveals that GrlA can activate transcription initiation at the *LEE1* P1 promoter by binding to a target located within the 18 base pair spacer between the promoter -10 and -35 elements, which were defined by mutational analysis. Shortening this spacer to 17 base pairs increases P1 promoter activity and short-circuits GrlA-dependent activation. Hence, at the P1 promoter, the action of GrlA resembles that of many MerR family transcription activators at their target promoters.

A cryptic promoter, designated P1A, was found that can initiate transcription within the *LEE1* operon regulatory region. Mutational- and biochemical analyses revealed that the P1A promoter overlaps the principal P1 promoter and transcription from the P1A starts at a site located 10 base pairs upstream of the P1 promoter. P1A is likely to compete with the P1 since it can only be active when the P1 promoter is mutated. A single base substitution in the P1 consensus -35 element unmasks P1A promoter activity. In contrast, P1A activity is much less when P1 is inactivated by a mutation in its -10 hexamer element. This suggests that, even when P1 is inactive, a consensus -35 element can sequester RNA polymerase and prevent its access to the P1A promoter.

The *LEE1* operon regulatory region has ~170 bases-long leader sequence. Deletion and mutational analyses revealed that the *LEE1* operon leader sequence contains a short translated

open reading frame, which has a strong ribosome binding site in front of two adjacent alternative translation start sites, followed by a lysine codon and a nonsense codon. Inactivation of this mini-gene significantly reduced the expression of the downstream *ler* gene, suggesting that optimal expression of the *ler* gene needs prior translation of the upstream mini-gene.

Deletion and subsequent reporter gene assays revealed that overexpressed Ler represses expression from the *LEE1* promoter regulatory region i.e., autoregulates its own transcription. In contrast, it activates expression from the *LEE2* promoter by negating H-NS-mediated repression. This suggests that both the GrlA and the mini-gene play positive roles in the expression of the *ler* gene whilst overexpressed Ler exhibits a negative role on its own expression and activates the expression from the *LEE2* promoter.

To my parents for their blessings and support throughout my life

Acknowledgements

This thesis is the result of three years of work whereby I have been accompanied and supported by many people. I have performed all the works except where mentioned. It is a pleasant aspect that I have now the opportunity to express my gratitude to all who supported me.

First and foremost, I offer my deepest gratitude to my supervisor, Professor Steve Busby, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the project. It would have been next to impossible to write this thesis without his support. His overly enthusiastic and integral view on research has made a deep impression on me. I am really glad to have been a student with him. I also owe my sincerest gratitude to Professor Mark Pallen for his help and advice during the period of this study. I am grateful to Professor Jeff Cole and Professor Ian Henderson for reviewing the introduction of the thesis.

I am indebted to Dr Lewis Bingle, Dr David Lee, Dr Kerry Hollands, Dr Mohamed Samir Elrobh and Dr Robert Shaw for their help and guidance with their scientific knowledge. I would like to thank Dr Jennie Mitchell for her patience in reading the thesis and making grammatical corrections in the text. I would like to show my gratitude to Rita, Chris and Lesley for their help and humor.

In my daily lab work, I have been blessed with a friendly and cheerful environment shared by a group of fellow students from both Busby and Cole Groups. I would like to thank all of them, especially Jack, Ian, Chismon, Claire and Amanda. I offer my regards and blessings to all of the others who supported me in any respect during the completion of the project.

I am grateful to the authority of the Commonwealth Scholarship Commission, UK for providing scholastic support.

Lastly, I would like to show my gratitude to my wife, Bentul Mawa for her great patience, encouragement and understanding me throughout the period of this study.

Contents

List of Figures

List of Tables

List of abbreviations

A: Adenosine

AE: Attaching and Effacing

Amp^R: Amplicillin-resistance

APS: Ammonium persulfate

ATP: Adenosine triphosphate

bp: Base pair

BSA: Bovine serum albumin

C: Cytosine

cAMP: Cyclic adenosine monophosphate

CAP: Calf intestinal alkaline phosphate

Cm^R: Chloramphenicol-resistance

CRP: Cyclic AMP receptor protein

CTD: C-terminal domain

CTP: Cytidine triphosphate

DNA: Deoxyribonucleic acid

DNase: Deoxyribonuclease

dNTP: Deoxynucleoside triphosphate (N= A, C, G or T)

DTT: Dithiothreitol

EDTA: Ethylenediaminetetraacetic acid

G: Guanine

GTP: Guanosine-5'-triphosphate

H-NS: Histone like-nucleiod structuring protein

HU: Heat unstable protein

HEPES: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

IHF: Integration host factor

kDa: Kilo Dalton

kb: Kilo base pair

LB: Luria broth

NTD: N-terminal domain

OD: Optical density

ONPG: Ortho-nitrophenyl-b-D-galactopyranoside

PCR: Polymerase chain reaction

RNA: Ribonucleic acid

RNase: Ribonuclease

RNAP: RNA polymerase

rRNA: Ribosomal RNA

SDS: Sodium dodecyl sulphate

T: Thymidine

TBE: Tris/Borate/EDTA buffer

Tet^R: Tetracycline-resistance

TEMED: N,N,N',N'-Tetramethylethylenediamine

tRNA: Transfer RNA

Tris: 2-Amino-2-(hydroxymethyl)-1,3-propanediol

UTP: Uridine triphosphate

v/v: Volume per volum

w/v: Weight per volume

Amino acids are denoted by their standard single or three letter codes

Chapter 1

Introduction

1.1 *Escherichia coli*

Escherichia coli belongs to the large family of enteric bacteria, the *Enterobacteriaceae*. These are facultatively-anaerobic Gram-negative rods, which are found in different environments including gastrointestinal tracts of humans and other warm-blooded animals. *E. coli* can be differentiated from other members of the *Enterobacteriaceae* family by either a lactose-positive or an indole-positive test (Nataro and Kaper, 1998; Murray *et al*., 2009).

E. coli cells are about 2 μm long, 0.5 μm in diameter with a cell volume of 0.6-0.7 μm^3 (Murray *et al*., 2009). An *E. coli* cell consists of cell envelope and cytoplasm. Like other bacteria, the *E. coli* cytoplasm is encased by the cell envelope and is the place where the functions of metabolism, protein production and DNA replication occur (Murray *et al*., 2009). Many essential macromolecular assemblies are distributed in the cytoplasm. For example, the *E. coli* chromosome is a single, double-stranded circle of DNA that is located in a discrete area of the cytoplasm known as the nucleoid. Unlike in eukaryotic cells, bacterial chromosomes are not confined in a separate membrane-enclosed nuclear compartment.

For *E. coli*, the cell envelope consists of an interior cytoplasmic membrane, a periplasmic space containing peptidoglycan, an outer membrane and sometimes a capsule. The cytoplasmic membrane is a phospholipid bilayer that shares many of the properties and functions of the eukaryotic cytoplasmic membrane. Located immediately outside of the cytoplasmic membrane is the periplasmic space. This can comprise up to 40% of the total cell volume and can contain at least 20% of the total cell water (Stock *et al*., 1977; Murray *et al*., 2009). The periplasmic space helps in osmoregulation and thus protects the cytoplasm. It houses numerous proteins including those necessary for nutrient binding, chemosensing, peptidoglycan synthesis and electron transport. It also contains degradative proteins including proteases, nucleases or phosphatases necessary for degradation of large nucleic acid or protein molecules to small transportable sizes. Moreover, it contains a detoxifying enzyme, beta-lactamase, which degrades incoming antibiotics before being penetrated into the cytoplasm or target sites. The peptidoglycan polysaccharide layer located within this space provides the rigidity to the cell and determines the shape to the cell. Peptidoglycan is absent in eukaryotic cells but is essential for bacterial survival. In *E. coli*, the peptydoglycan layer is thin compared with that of Gram-positive bacteria and thus, the *E. coli* cannot retain the crystal violet dye

when washed in a decolorizing solution in the Gram staining protocol and hence it is referred to as Gram-negative. Linked to the peptidoglycan layer is the outer membrane. *E. coli* cells, like all other Gram-negative organisms, have unique outer membranes that are essential for survival; these structures are lacking in Gram-positive bacteria. This outer membrane contains lipopolysaccharides (consisting of lipid A, core polysaccharides and 'O' antigen) and integral membrane proteins, as well as proteins located on the periphery; both on the surface of the organisms and within the periplasmic space (Schilling *et al*., 2001). For Gram-negative bacteria, the outer membrane provides the interface with the environment and acts as a molecular sieve, excluding noxious compounds while allowing the uptake of essential nutrients. The capsule is usually made up of complex polysaccharides that protect the cell from desiccation and phagocytosis (Burns and Hull, 1999). It can also promote adherence to other bacteria and to the host cell surfaces and acts as a major virulence determinant in some disease producing *E. coli* (Hérias *et al*., 1997; Bahrani-Mougeot *et al*., 2002).

Many *E. coli* possess flagella that protrude from the cell envelope and help in bacterial motility and initial colonization during pathogenesis (Darnton *et al*., 2007; Mahajan, *et al*., 2009). Flagella and the other surface structures mentioned above form the basis for serotyping of *E. coli* such that they are differentiated based on their O (lipopolysaccharide), H (flagellar), and K (capsular) surface antigen profiles (Edwards and Ewing, 1972). Fimbriae or pili are other bacterial surface appendages. They extend out from the outside of the cell surface and play an important role in attachment of bacterial cells to surfaces to start biofilm formation, and to host cells during pathogenesis (Hicks *et al*., 1998; Pawar *et al*., 2005). Specialized pili help in the bacterial conjugation process during which plasmid DNA is passed from one bacterium to another (Ou and Anderson, 1970).

Non-pathogenic *E. coli*, particularly *E. coli* K-12 strains, are often considered model organisms for studying the molecular bases of life, and it has been said that "what's true for *E. coli* is true for an elephant." This widely quoted phrase was expressed in 1954 by Nobel Laureate Jacques Monod who won the prize for his operon model, which was derived from research on *E. coli* K-12 (Friedmann, 2004; Hobman *et al*., 2007). The great advantage of using the *E. coli* K-12 strain lies in its rapid growth rate and simple nutritional requirements in laboratory conditions. In 1922, the original laboratory *E. coli* K-12 strain was isolated from

human faeces at Stanford University. During many years of use in laboratory conditions, the strain lost its 'O' surface antigens and this increased its acceptance as it was harmless to the researchers. Hence *E. coli* K-12 is currently the best characterized organism at the molecular level, and it has played pioneering roles in studies of various key aspects of life including DNA replication, transcription, translation, gene regulation, the operon concept, restriction enzymes and horizontal gene transfer (Hobman *et al*., 2007). Also the molecular study of higher organisms largely depends on the use of *E. coli* as a vector. It is thus, true that "all cell biologists have two cells of interest, the one they are studying and *Escherichia coli*" (Neidhardt and Curtiss, 1996) and *E. coli* K-12 is still unchallenged as a model organism playing unparalleled roles in different fields of study including genetic engineering and modern biotechnology, synthetic biology, and systems biology (Baba *et al*., 2006).

Sequence analysis of the *E. coli* K-12 genome revealed a total of 4,639,221 base pairs in which its genes are embedded (Blattner *et al*., 1997). The total G+C content is 50.8%, and genes that encode proteins account for >87 % of the genome. The genome also contains many horizontally acquired insertion sequence elements and phage sequences. Comparison of sequence data from different *E. coli* strains revealed that *E. coli* is a diverse bacterial species where ca. 50% of a genome is common to all strains but where this 50% represents a significantly smaller portion (ca. 20%) of the total gene repertoire identified among more than 100 *E. coli* strains to date (Lukjancenko *et al*., 2010). This diversification is due to acquisition and loss of pathogenic islands and other mobile genetic elements; pathogenic bacteria often have genomes ca. 1Mbp larger than their non-pathogenic counterparts (Croxen and Finlay, 2010).

1.2 Pathogenic *E. coli*

In 1885, Theodor Escherich, a German paediatrician and bacteriologist, dicovered *E. coli* in the faeces of healthy infants (Friedmann, 2006; Shulman *et al*, 2007). The species name is 'coli' as it can be found universally in the large intestine or colon. Human intestines are usually colonized by *E. coli* a few hours after birth and thereafter, *E. coli* remain as harmless commensals, and even provide benefits to the host. For instance, *E. coli* can benefit its host by providing important metabolic contributions, including vitamin K_2 synthesis, as well as preventing colonization by harmful bacteria (Reid *et al*., 2001; Canny and McCormick, 2008).

However, a large number of *E. coli* strains can cause a wide range of diseases in humans as well as in mammals and birds worldwide (Kaper *et al*., 2004). Pathogenic *E. coli* are broadly classified into two groups. One group causes intestinal pathologies, whilst the second group cause extra-intestinal disease. Intestinal pathologies mostly consist of more or less severe diarrhoea and are caused by several pathovars, hence the group is also known as diarrhoeagenic *E. coli*. Six well studied pathovars in this group are: enteropathogenic *E. coli* (EPEC), enterohaemorrhagic *E. coli* (EHEC), enterotoxigenic *E. coli* (ETEC), enteroaggregative *E. coli* (EAEC), enteroinvasive *E. coli* (EIEC) and diffusely adherent *E. coli* (DAEC) (Kaper *et al*., 2004; Croxen and Finlay, 2010). The most common extra-intestinal infections (due to ExPEC; Russo and Johnson, 2003) are urinary tract infections caused by uropathogenic *E. coli* (UPEC) and meningitis caused by neonatal meningitis *E. coli* (NMEC) (Croxen and Finlay, 2010). Recent genome comparisons have demonstrated that the four *Shigella* spp. are in fact *E. coli* and are most similar to the EIEC. *Shigella* spp. causes bacillary dysentery and bloody diarrhoea (Levine *et al*., 1973). *S. dysenteriae* arose independently of the other three *Shigella* strains and is closely related to EHEC (Chaudhuri *et al*., 2010).

Different pathovars can be identified on the basis of their clinical manifestations and repertoire of virulence factors (Nataro and Kaper, 1998). The EAEC, EPEC or DAEC pathovars can be differentiated phenotypically by performing a HEp-2 adherence assay (Vial *et al*., 1990). Furthermore, numerous immunological assays can be used to detect antigens e.g. Shiga-like exotoxin, which is specific to certain pathovars (Sowers *et al*., 1996). Nucleic acid probes can be used to detect enterotoxins or target genes (Moseley *et al*., 1982). Additionally, polymerase chain reaction (PCR) using oligonucleotide primers specific for target genes can be used to differentiate pathovars (Stacy-Phipps *et al*., 1995).

In addition to pathogenic *E. coli*, there are a number of clinically significant bacteria in the *Enterobacteriaceae* family that also cause intestinal infections including *Salmonella enterica* and *Citrobacter rodentium*. *S. enterica* causes gastroenteritis associated with diarrhoera though *S. enterica* serovar Typhi causes a severe desseminated invasive infection called typhoid (Zhang *et al*., 2003). *C. rodentium* is a natural mouse pathogen that colonizes the intestine and causes colitis and transmissible colonic hyperplasia (Schauer *et al*., 1993; Deng

et al., 2001). *C. rodentium* infection in the mouse is also used as a model for EPEC mediated disease in humans due to the similarity of the infectious process.

1.3 *E. coli* pathovars EPEC and EHEC

Among different pathovars of *E. coli*, EPEC is a leading cause of infantile diarrhoea worldwide, particularly in developing countries (reviewed by Nataro and Kaper, 1998; Schmidt, 2010). EPEC was first reported to be a cause of epidemic and sporadic diarrhoea in the 1940's (Frankel *et al*., 1998). The disease can be associated with high rates of mortality (Chen and Frankel, 2005). To promote an infectious cycle, a bacterium needs to be able to exploit its host cell for colonization, multiplication and dissemination by using numerous virulence factors. The 2nd International Symposium on EPEC (Kaper, 1996) defined typical EPEC as those *E. coli* strains possessing specific virulence factors namely bundle-forming pili encoded on the EPEC adherence factor plasmid (EAF) and carry the locus of enterocyte effacement (LEE) pathogenicity island encoding a type III secretion system (T3SS). Bundle-forming pili promote initial adherence, primarily in the small intestine, and interaction with other EPEC bacteria to form microcolonies, a characteristic of localized adherence. The LEE-encoded T3SS promotes the formation of attaching and effacing (A/E) lesions, which is a defining feature of EPEC. Atypical EPEC (aEPEC) that lack the EAF plasmid but possess the LEE pathogenicity island (Nataro, 2006), show diffuse adherence instead of localized adherence (Schmidt, 2010) and can cause prolonged diarrhoea in children (Spano *et al*., 2008).

Like atypical EPEC, EHEC strains lack the EAF plasmid and contain the LEE pathogenicity island. However, in contrast to both typical and atypical EPEC strains, EHEC possess enterotoxins and a ca. 60 MDa plasmid pO157. The toxins are called Shiga-like toxins (Stxs) because of their close resemblance to the Shiga toxins of *S. dysenteriae* (O'Brien *et al*., 1982). The *E. coli* version of Stxs are also called Verotoxins (VTs) since they cause irreversible pathotypic effects on Vero cell cultures (kidney cells of African green monkey) (Konowalchuk *et al*., 1977). It is believed that the EHEC strains have evolved from the prototype EPEC via acquisition of bacteriophage encoding VTs (Whittam *et al*., 1993; Wick *et al*., 2005). The pO157 plasmids encode several virulence factors including cytotoxin haemolysin. The virulence properties of EHEC strains cause A/E lesions on epithelial cells in

the large bowel, and cause haemorrhagic colitis and haemolytic uremic syndrome (HUS) in humans (Karmali, 1989; Bettelheim, 2003; Kaper *et al*., 2004; Orth *et al*., 2006; Spears *et al*., 2006). The *E. coli* O157:H7 is the predominant and most virulent serotype in EHEC. However, other serotypes including O26 and O111, which have the same clinical, epidemiological and pathogenic properties associated with the O157:H7 are also referred to EHEC (Levine, 1987; Karmali, 1989; Sharma, 2002).

Generally, all *E. coli* strains that produce VTs are collectively called verotoxigenic *E. coli* (VTEC). In this respect, EHEC strains are referred to as VTEC strains. There are over 100 VTEC serotypes. Many of them have been associated with human illness (Johnson *et al*., 1996; Bettelheim, 2003). However, it appears that the VTs alone are not enough to cause virulence. To exert virulent effects, VTEC strains need to have other virulence determinants such as the LEE pathogenicity island and pO157 virulence plasmids (Stephan *et al*., 2000; Morabito *et al*., 2001). Therefore, the VTEC is a general term used for all *E. coli* strains that produce VTs, whereas the EHEC is a subset of the VTEC that also contain other virulence factors including the LEE pathogenicity island.

The topic of this thesis is the regulation of the LEE virulence determinant of the EHEC strain O157:H7. Therefore, different aspects of the O157:H7 pathogenicity such as epidemiology, clinical features and virulence factors are described in next few sections.

1.3.1 Epidemiology and outbreak of *E. coli* O157:H7 serotype

Ruminants, predominantly cattle, are the main natural reservoirs of EHEC and are usually considered as healthy carriers (Naylor *et al*., 2005; Spears *et al*., 2006; Karmali *et al*., 2010). EHEC strains are rarely harboured by pigs, dogs, cats, birds, amphibians, and fish (Beutin *et al*., 1993; Ferens and Hovde, 2011). The presence of EHEC in cattle faeces provides the potential for the pathogens to enter the food chain through contamination of milk, contamination of meat with intestinal contents during slaughter, or contamination of fruits and vegetables by contact with faeces or contaminated manure. Consumption of such contaminated foods can trigger EHEC infection and can lead to severe outbreaks (Naylor *et al*., 2005). EHEC O157:H7 serotype was first recognised as the causative agent of outbreaks of HUS in the early 1980's (Riley *et al*., 1983). Since then, it has been emerged as an

important cause of diarrhoea in industrialized countries. Outbreaks are mainly associated with hamburgers, unpasteurized milk or fruit juices and fresh vegetables. A number of major outbreaks including a massive-outbreak in the Japanese city of Sakai that affected approximately 10000 people including nearly 5000 schoolchildren are listed in Table 1.1.

1.3.2 Clinical features of *E. coli* O157:H7 pathogenesis

Studies of EHEC O157:H7 outbreaks provide evidence that EHEC virulence is associated with two distinctive clinical features: haemorrhagic colitis (Riley, 1987) and HUS (Gasser *et al*., 1955; Karmali *et al*., 1983). Haemorrhagic colitis is an illness characterized by sudden onset of severe abdominal cramps followed by diarrhoea, which is initially watery but becomes grossly bloody. The incubation period is usually three to four days. In some cases, it can be prolonged and can be fatal, especially in very young or very old patients. The HUS, on the other hand, is a serious complication that may be preceded by haemorrhagic colitis or diarrhoea, and is defined by the triad: acute renal failure (uremia), thrombocytopenia (low platelet count) and microangiopathic haemolytic anemia. The disease predominantly affects children and may cause 5-10% mortality and recovered patient may face chronic kidney disease and may rely on renal replacement therapy (Corrigan and Boineau, 2001).

1.3.3 Major pathogenic determinants of *E. coli* O157:H7 serotype: toxins

As mentioned earlier, the EHEC chromosome contains phage encoded Stx1 and/or Stx2, and these are the major virulence determinant and the defining characteristic of EHEC. Usually, lambdoid phage-mediated lysis of bacterial cells in response to DNA damage and the SOS response causes release of Stx (Toshima *et al*., 2007). Both Stx1 and Stx2 are ~70 kDa holotoxins consisting of a single A subunit of 32 kDa and five 7.7 kDa B subunits (AB5) (Middlebrook and Dorland, 1984; O'Brien and Holmes, 1987). The A subunit is composed of a 28 kDa peptide (A_1) and a 4 kDa peptide (A_2), which remain attached by a disulfide bond (Takao *et al*., 1988). The A_1 peptide is enzymatically active whereas the A_2 peptide helps to bind the A subunit to B subunits. The B subunits help in binding the holotoxin to a specific glycolipid receptor known as globotriaosylceramides (Gb_3), which are found on the surface of Paneth cells in the human intestinal mucosa and surface of the kidney epithelial cells (Nataro and Kaper, 1998; Schüller *et al*., 2007). The interaction of Stx B with Gb3 induces membrane invaginations that help the holotoxin to be internalised. The internalised holotoxin is then transported to the Golgi apparatus and then to the endoplasmic reticulum. Stx A is

Table 1.1: Major worldwide outbreaks of EHEC 0157: H7 infection

Location	No. of cases	No. of HUS	No. of deaths	Setting	Source	Date	Reference
Ontario	73	12	19	Nursing home	Sandwich meal	Sept 1985	Carter *et al*., 1987
Alberta	15		2	Nursing home	Ground beef	June 1987	Hockin and Lior, 1987
Saitama, Japan	174	14	2	Nursery School	Tap water	1990	Akashi *et al*., 1994
Western USA	700	45	4	Community	Hamburgers	Nov 1992 -Feb 1993	Bell *et al*., 1994; Tuttle *et al*., 1999
Oregon, Canada	14			Community	Dairy milk	Dec 1992- Apr 1993	Keene *et al*., 1997
Western United States and British Colombia	70	14	1		Apple juice	Fall 1996	Cody *et al*., 1999
New haven county, USA	14	2		Connecticut residents	Apple cider	Oct 1996	Hilborn *et al*., 2000
United Kingdom	345	34	16	Central Scotland	Meat	1996	Dundas *et al.*, 2001
Sakai, Japan	9451	122	12	Mostly schoolchildren	White radish sprouts	May-Dec 1996	Higami *et al*., 1998; Michino *et al*., 1999
Montgomery, Pennsylvania	51	8		Dairy farm visitors	Calves	Sept 2000	Crump *et al*., 2002
26 States, USA	199	31	3		Fresh spinach	Sept 2006	Wendel *et al*., 2009

translocated to the cytoplasm, where it acts on the 60S ribosomal subunit. Specifically the A_1 peptide is an *N*-glycosidase that cleaves 28S rRNA of host cells, thereby causing protein synthesis to be ceased. The resulting disruption of protein synthesis leads to the cell death (Melton-Celsa and O'Brien, 1998; Johannes and Römer, 2010).

1.3.4 Major pathogenic determinants of *E. coli* O157:H7 serotype: T3SS

Bacterial pathogens exploit host cell for their own benefit such as seeking out essential nutrients for survival. However, protective responses of host cell cause hostile environment for bacterial pathogens. Thus, bacterial pathogens use a large number of virulence proteins to manipulate host cell responses in the course of an infection. In order to release proteins to the host cytosol or to the external milieu, bacterial pathogens have evolved a number of different complex secretion systems. These are classified as Type I, Type II etc (Pugsley, 1993; Hueck, 1998; Aizawa, 2001). In addition to transport virulence proteins to the host cell cytoplasm, bacteria use protein secretion systems to export proteins necessary for biogenesis of external appendages such as pili or flagella. To date, six different secretion systems (T1SS to T6SS) have been described in many Gram-negative bacteria (Bleves *et al*., 2010). The T2SS and T5SS-mediated protein exportation occurs in two steps. Type I and type V exoproteins are exported to periplasm by the Sec or Tat system first. Then they are exported to the external environment by their dedicated systems. In contrast, T1SS, T3SS, T4SS and T6SS use one step mechanisms in such that the type I, type III, type IV and type VI exoproteins are delivered directly to extracellular medium or host cell cytoplasm by their cognate secretion apparatus (Figure 1.1).

In EHEC and related bacterial pathogens, the T3S apparatus play crucial roles in exportation of bacterial effector proteins to the host cytosol. These nanomachines are bacterial surface appendages shaped like microneedles. Remarkably, the T3SS export apparatus is believed to have a common origin with the bacterial flagellum (Pallen *et al*., 2005b; Troisfontaines and Cornelis, 2005; Pallen and Gophna, 2007). The bacterial flagellum possesses a built-in secretion apparatus like the T3S apparatus and there is a substantial level of similarity both in sequence and physico-chemical properties in core components of these two nanomachines (Aizawa, 2001; Blocker *et al*., 2003, Pallen *et al*., 2005b). Moreover, in both cases, there are many functional similarities including specific chaperones necessary in both systems (Parsot

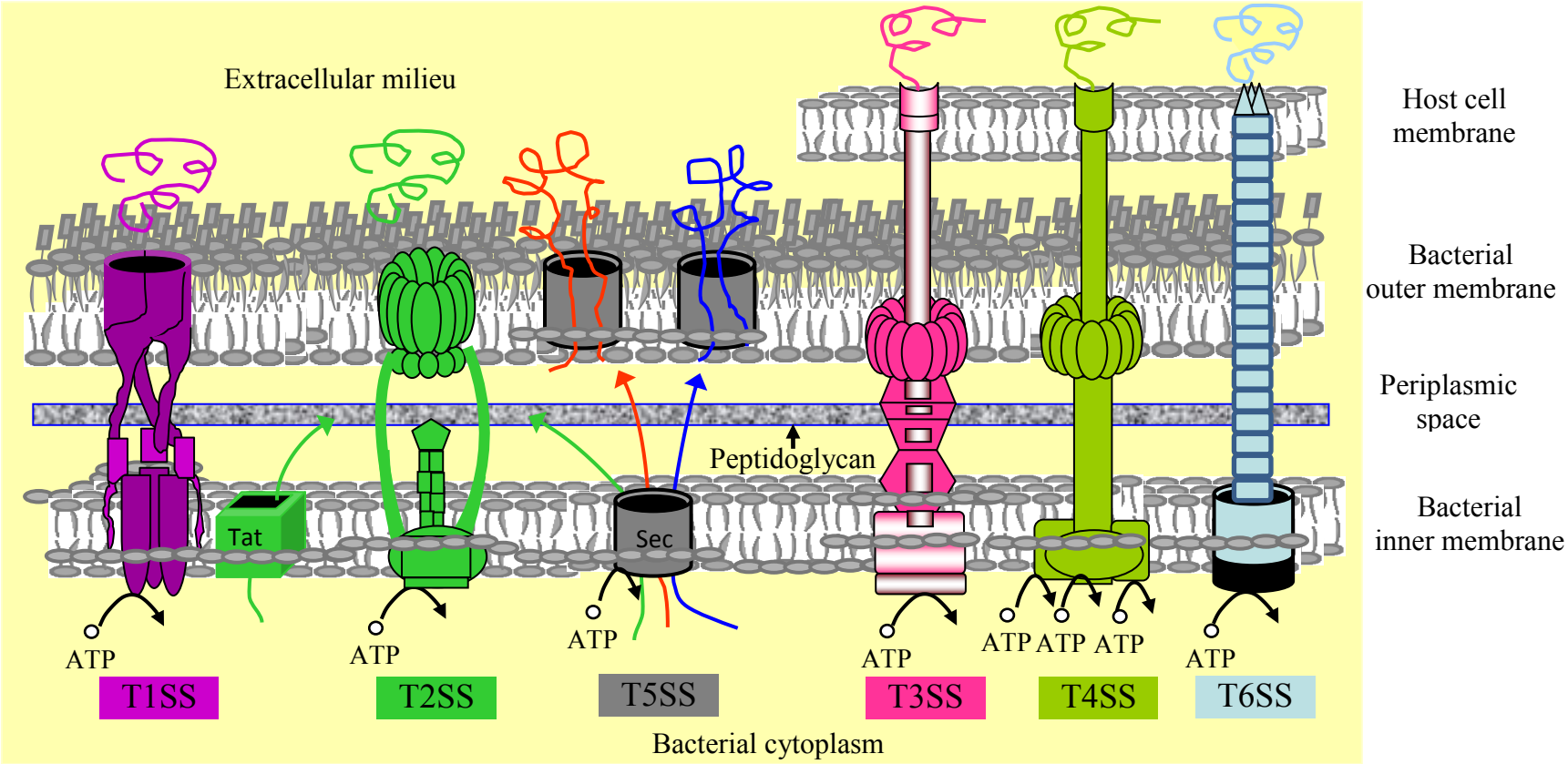


Figure 1.1. Bacterial secretion systems.

The figure shows the schematic diagram of six distinct secretion systems in Gram-negative bacteria. In case of Type II and Type V secretion systems, exoproteins are first exported from bacterial cytoplasm through the inner membrane to the periplasmic space by the Sec or the two arginine (Tat) pathways. After that, proteins are transported through the outer membrane from periplasm to external medium by the dedicated systems. On the other hand, Type 1, Type 3, Type 4 and Type 6 secretion systems directly mediate protein export through inner and outer membranes to the extracellular medium or to the host cytoplasm. The figure adapted from Beleves *et al*. (2010) and KEGG http://www.genome.jp/kegg/pathway/map/map03070.html

et al., 2003), molecular systems used to control the size of the apparatus (Minamino and Pugsley, 2005), exporting distal components (needle subunits) before substrates to be secreted (Blocker *et al*. 2003). Additionally, many reports show that the bacterial flagellum can secret non-flagellar proteins, supporting that it can act as a pure secretory apparatus (Young and Young, 2002; Konkel *et al*., 2004; Lee and Galan, 2004). Taking all into consideration, T3S systems are broadly grouped into two: flagellar- and non-flagellar T3SSs.

Non-flagellar T3S systems are present in EHEC and related bacteria as well as in many other Gram-negative bacteria (reviewed by Pallen *et al*., 2005b; Troisfontaines and Cornelis, 2005). Examples are the plasmid-encoded Ysc T3S from *Yersinia* spp. (Cornelis, 2002), the Psc system of *Pseudomonas aeruginosa*, a pathogen for nematodes, plants, insects, and mammals (Bodey *et al*., 1983), the Asc system of *Aeromonas salmonicida*, a pathogen of fish causing fatal disease in Salmonids (Burr *et al*., 2003) and the Vsc system of *Vibrio* spp., found in association with plankton and juvenile shrimps, prawns, lobsters, and fish causing serous human gastroenteritis (Park *et al*., 2004). T3S systems from intracellular pathogens *S. enterica* and *Shigella* spp., which are known as Inv-Spa and Inv-Mxi T3S systems respectively make the Inv-Mxi-Spa family of T3S system. Members of this T3S family help in invasion of bacterial cells to host cells and trigger actin polymerization and the family includes T3S systems of *Chromobacterium violaceum* and *Burkholderia* spp. that are associated with human diseases (Duran *et al*., 2001; Stevens *et al*., 2002).

In EHEC and related bacteria, the LEE encodes a T3SS that consists of a T3S apparatus, regulators, chaperones, translocators and effectors that co-ordinately cause A/E lesions in host epithelial cells. EHEC and related bacteria have acquired this pathogenic determinant via horizontally gene transfer (Hacker and Kaper, 2000; Deng *et al*., 2001; Pallen and Wren, 2007). Sequence analysis of the entire LEE region from different strains reveals that the EPEC prototype strain E2348/69 contains the core LEE region, which is approximately 35.4 kb in size. The average G+C content of nearly 38% is far below that of the *E. coli* chromosome (50.8% in *E. coli* chromosome; Blattner *et al*., 1997). Like many other pathogenicity islands, the LEE is found to be inserted near tRNA genes, at targets known as recombination hot spots (McDaniel *et al*. 1995; Perna *et al*., 1998; Deng *et al*., 2001; Schmidt, 2010). The size of the LEE differs due to flanking sequence variation in different isolates. In

EHEC and *C. rodentium* isolates, the region is flanked by bacteriophage and insertion sequences, respectively (Perna *et al*., 1998; Müller *et al*., 2009).

The LEE region contains 41 predicted genes encoding proteins with more than 50 amino acids (Elliott *et al*., 1998; Perna *et al*., 1998; Deng *et al*., 2001). Reverse transcriptase-polymerase chain reaction (RT-PCR) analysis of RNAs revealed that the LEE genes are organized into five major polycistronic operons: *LEE1*, *LEE2*, *LEE3*, *LEE4*, *LEE5* (or *TIR*) and several smaller transcriptional units (Mellies *et al*., 1999; Sanchez-SanMartin *et al*., 2001; Deng *et al*., 2004; Lodato and Kaper, 2009; Bhatt *et al*., 2011). Genes from *LEE1-3* encode proteins that are necessary for the formation of the T3S apparatus. The *LEE4* operon encodes proteins necessary for the formation of a translocon for delivering effector molecules to the host cell and disruption of the host cytoskeleton (Nougayrede and Donnenberg, 2004; Dean *et al*., 2010). Finally, the *LEE5* operon encodes proteins necessary for intimate attachment of the bacterial cell to the host cell (Jerse *et al*., 1990; Elliott *et al*., 1999).

The nomenclature of the LEE genes is now well-established (Elliott *et al*., 1998; Pallen *et al*., 2005a). Generally, the names of LEE genes have been adopted following the homology with the most intensively studied T3S system from *Yersinia* spp. The genes that encode *E. coli* secretion proteins were given the generic name *esp* (*E. coli* secreted proteins) (McDaniel *et al*., 1995). Those homologous to the genes of the *Yersinia* T3S (*ysc*) were named as *esc* (*E. coli* secretion) genes with the same suffix as *Yersinia* homologues. The genes that are not homologous to *ysc* but related to T3S were refered to *sep* (secretion of *E. coli* proteins). The genes that function as chaperones for secretion of other proteins were refered to *ces* (chaperone of *E. coli* secretion) (Wainwright and Kaper, 1998). The gene that encode intimin was named *eae* (*E. coli* attaching and effacing) and the intimin receptor gene was named *tir* (translocated intimin receptor) (Jerse *et al*., 1990; Kenny *et al*.,1997b). Apart from these, some genes including *ler*, *grlR*, and *grlA* were named on the basis of their regulatory roles in the expression of other genes (Mellies *et al*., 1999; Deng *et al*., 2004). In Figure 1.2, the organisation of LEE genes is shown and different genes are marked according to their functional properties.

The assembly of a typical T3S apparatus is shown in Figure 1.3 using *E. coli* nomenclature

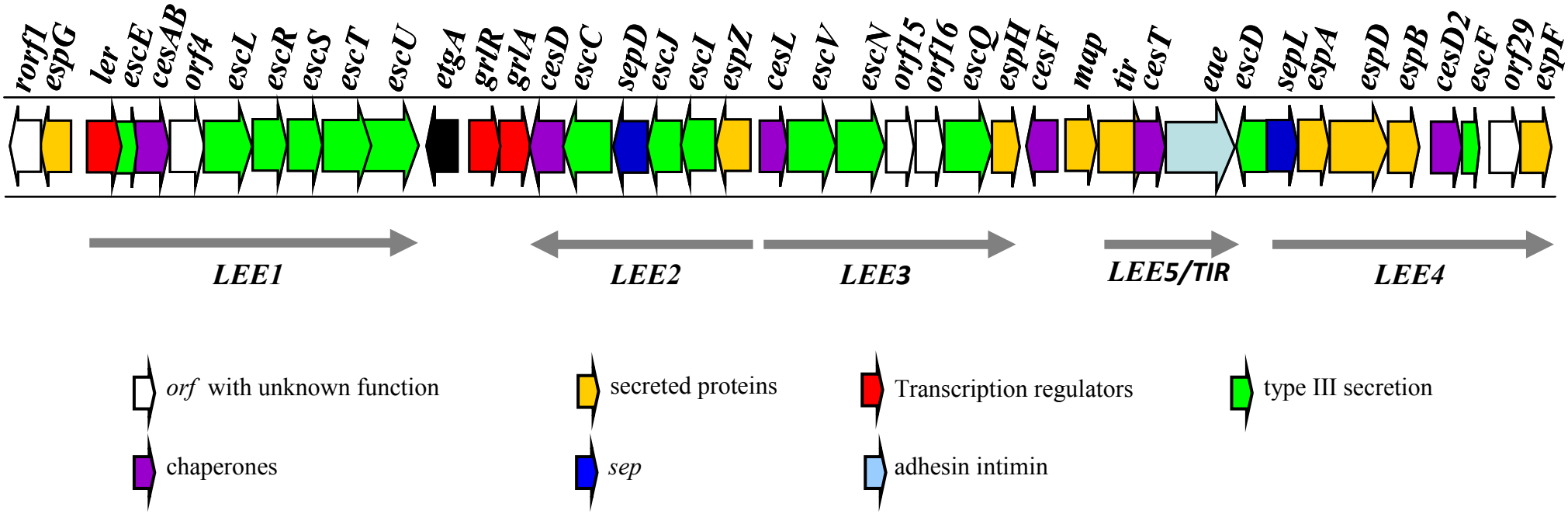


Figure 1.2. Organization of genes in the locus of enterocyte effacement.

The figure shows the arrangement of 41 genes in the LEE region. Genes shown are organised in five major polycistronic operons: *LEE1*, 9 genes; *LEE2*, 6 genes; *LEE3*, 7 genes; *LEE4*, 8 genes and *LEE5* or *TIR*, 3 genes. *orf*, open reading frames with unknown functions*; ler*, encodes LEE-encoded regulator; *esc*, encodes *E. coli* secretion component (homologous to *Yersinia* type III secretion, *ysc*); *grlR*, encodes global regulator of LEE repressor; *grlA*, encodes global regulator of LEE activator; *ces*, encodes chaperone of *E. coli* secretion; *sep*, encode secretion of *E. coli* proteins (not *ysc* homolog but involved in type III secretion); *esp*, encode *E. coli* secreted proteins; *map*, encodes mitochondrial associated protein; *tir*, encodes translocated intimin receptor protein; *eae* (*E. coli* attaching and effacing) encodes intimin; *etgA*, encodes a lytic transglycosylase. This figure has been adapted from Castillo *et al*. (2005). Nomenclature of the genes has been given mainly following Pallen *et al*. (2005a).

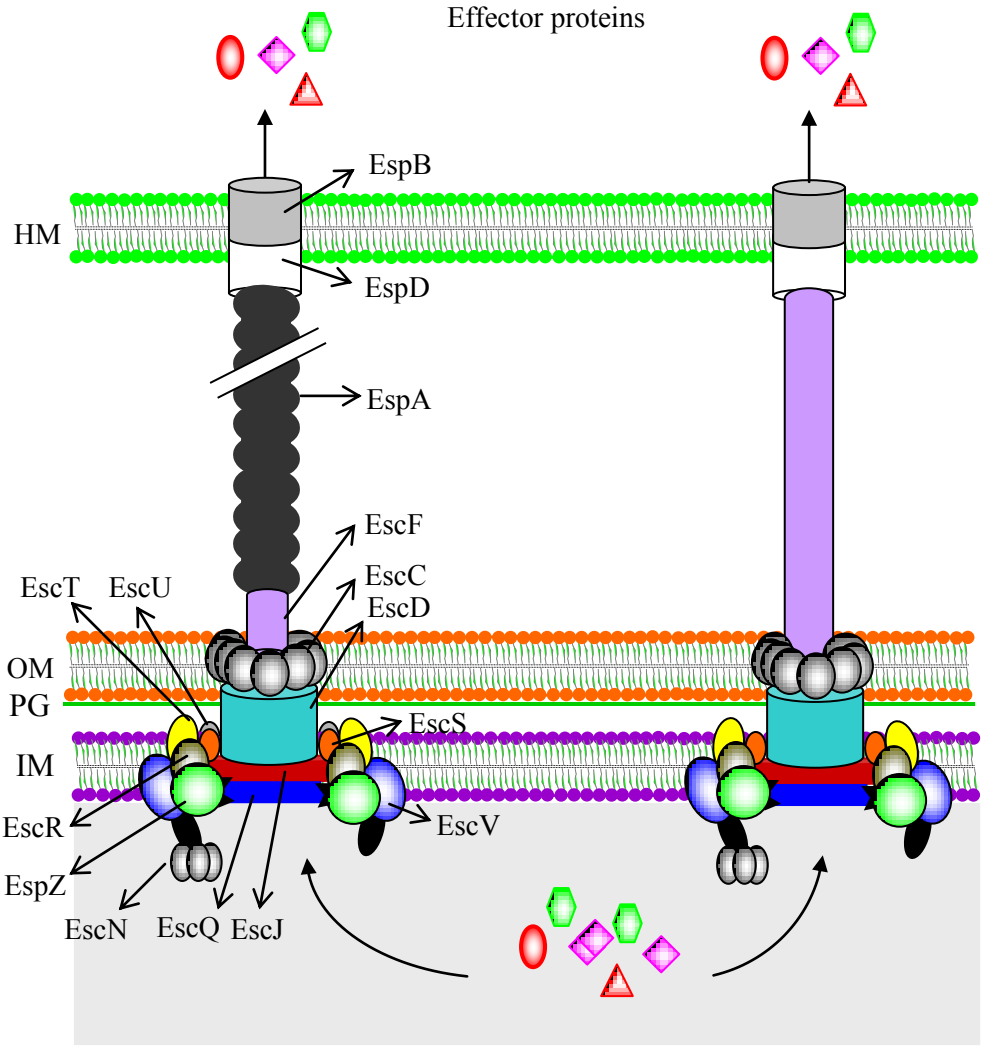


Figure 1.3. Organisation of a type three secretion system.

The left side of the figure illustrates the structure of a type three secretion apparatus encoded by the LEE genes (details in the text). The right side of the figure shows a non-LEE-encoded T3S apparatus that usually lacks a filamentous EspA pilus and uses the short needle complex in protein delivery. HM, host cell membrane; OM, outer membrane; IM, inner membrane; PG, peptidoglycan layer. This figure has been adapted from Pallen *et al*. (2005a,b).

for the different proteins. The basement of the T3S apparatus in the bacteial inner membraneis made up of a number of proteins including EscD, EscQ, EscR, EscS, EscT, EscU, EscV that are mainly encoded by the genes from *LEE1-3* operons (Ogino *et al*., 2006). Moreover, the putative ATPase component, EscN, that supplies energy for transport, is involved in the inner membrane basal structure (Gauthier *et al*., 2003). EscC on the other hand, is associated with the formation of the outer-membrane ring of the structure (Gauthier *et al*., 2003; Ogino *et al*., 2006). EscJ forms a large 24-subunit 'ring' that might connect membrane-bound rings of the structure (Yip *et al*., 2005). EscF form a needle-like structure onto which EspA molecules polymerize to form a filament or translocon, also known as the EspA pilus (Knutton *et al*.,1998; Daniell *et al*., 2003). This EspA filamentous pilus is the characteristic feature of the LEE-encoded T3S system (Daniell *et al*., 2001). The pilus that extends to the host cell, is usually absent in other T3S systems that have only a short needle complex for delivering proteins (Figure 1.3). The EspA pilus is capped by the EspD and EspB molecules, which form a translocation pore in the host cell plasma membrane through which the effector proteins are delivered (Daniell *et al*., 2001).

The LEE encodes several effector proteins including Tir, Map, EspF, EspH, EspZ and EspG (Schmidt, 2010). These proteins are delivered from bacterial cells to the host cell through the T3S apparatus and manipulate host cellular functions (reviewed by Spears *et al*., 2006). EspG encoded in the *espG-rorf1* bicistronic operon modulates the host cytoskeleton and promote colonisation during pathogenicity (Shaw *et al*., 2005). EspH encoded in the *LEE3* causes cytoskeleton disruption (Tu *et al*., 2003). EspF encoded in the *LEE4* operon is injected into the host cell and targeted to the nucleolus and mitochondria playing a role in the cell death pathway (Nougayrede and Donnenberg, 2004; Dean *et al*., 2010; Holmes *et al*., 2010). EspF also disrupts transepithelial cell resistance causing disruption of tight junctions, apoptosis (McNamara *et al*., 2001). Tir encoded in the *LEE5* operon helps in attachment of bacterial cells to the host cells, actin polymerization, and A/E lesion formation (Jerse *et al*., 1990; Elliott *et al*., 1999). Map encoded by the *map* gene causes disruption of tight junctions, mitochondrial membrane potential and filopodium formation (Kenny and Jepson, 2000; Kenny *et al*., 2002).

Apart from the LEE-encoded effector molecules, a number of non-LEE-encoded effectors are

translocated by the T3S apparatus. Some of these are associated with pathogenicity. These effectors are mostly encoded from the genes located at the lambdoid-like prophages that are scattered throughout the chromosome (Garmendia *et al*., 2005; Spears *et al*., 2006; Tobe *et al*., 2006). Effector molecules TccP/$EspF_u$ and EspJ are encoded by a bicistronic operon located on prophage CP-933U. TccP/$EspF_u$ is required for actin polymerization at the cytoplasmic domain of Tir in EHEC, (Campellone *et al*., 2004; Garmendia *et al*., 2004) whilst EspJ is believed to have 'antivirulence' properties (Dahan *et al*., 2005). NleA/EspI effector molecule is encoded by a cistron located on prophage Sp9 and is required for *C. rodentium* to colonise the mouse colon (Deng *et al*., 2004; Mundy *et al*., 2004) and this protein is conserved in *E. coli* pathogens containing the LEE genes (Gruenheid *et al*., 2004; Roe *et al*., 2007).

The cellular target of many bacterial pathogens is the host cytoskeleton. The cytoskeleton of eukaryotic cells consists of actin filaments, microtubules and intermediate filaments, which direct numerous cellular functions from cell shape and structure to programmed cell death. Therefore, alteration of the cytoskeleton is crucial for bacterial pathogenesis. Bacterial pathogens often subvert and control the polymerization of host actin filaments in order to modulate cellular processes, and they do this through the action of delivered effector proteins. During infection, the Tir protein translocates into the host cell and inserts itself into the host cell plasma membrane, a process that requires a cognate chaperone, CesT, for maximum efficiency (Kenny *et al*., 1997b; Abe *et al*., 1999; Sal-Man *et al*., 2009). The receptor protein, Tir interacts with intimin on the bacterial surface, thus firmly attaching the bacterium to the host cell. Then the T3SS serves as a molecular syringe for delivery of effector Esp molecules to the host cell (Figure 1.3) to facilitate the disruption of the cytoskeleton in the formation of AE lesions and subvert the host immune response. During the EPEC infection period, Tir becomes phosphorylated by host protein kinases resulting in binding directly to host adaptor protein, Nck, which activates neural Wiskott-Aldrich syndrome protein (N-WASP) and the actin related protein 2/3 (ARP 2/3) complex to direct actin rearrangements and pedestal formation. In EHEC, Tir is not phosphorylated and actin rearrangement is Nck independent. However, in this case (Figure 1.4), actin rearrangement is mediated by Tir cytoskeleton -coupling protein (TccP or $EspF_u$), which is encoded by prophage CP-933U and delivered by T3S to the host cell (Campellone *et al*., 2004; Garmendia *et al*., 2004). This TccP is linked with the Tir to the host protein insulin receptor tyrosine kinase substrate (IRTKS) and

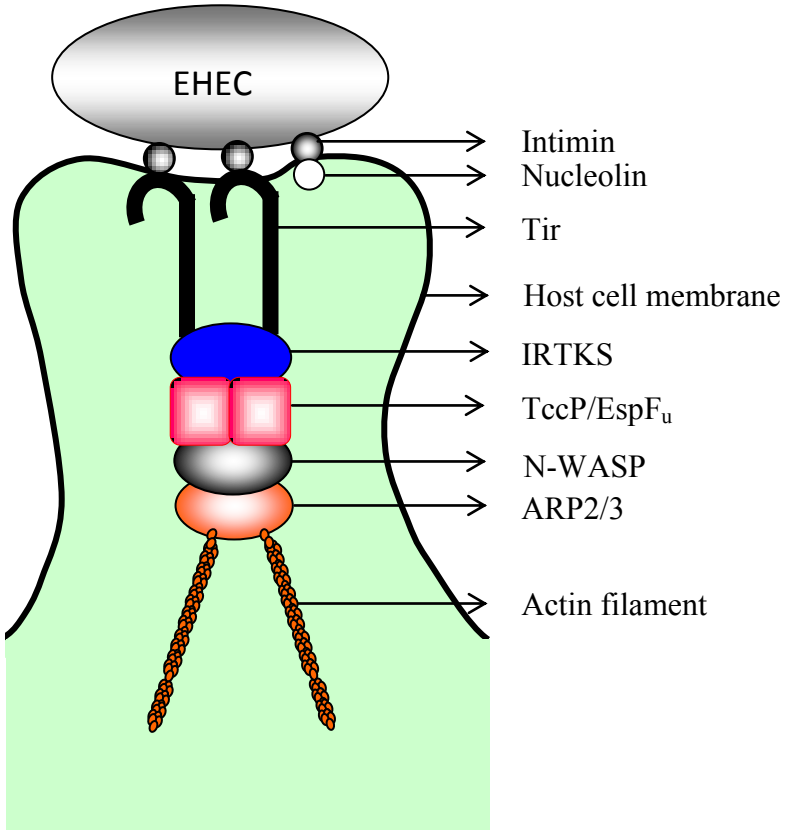


Figure 1.4. Generation of pedestals by an enterohaemorrhagic *E. coli*.

The figure illustrates the mechanism by which A/E bacteria rearrange host cell cytoskeleton resulting in formation of a pedestal like structure beneath the attached bacteria. Tir, translocated intimin receptor protein; IRTKS, host protein insulin receptor tyrosine kinase substrate; TccP, Tir cytoskeleton-coupling protein also known as EspF$_u$; N-WASP, neural Wiskott-Aldrich syndrome protein; ARP2/3, actin related protein 2/3. The figure has been adapted from Croxen and Finlay (2010).

interacts with N-WASP to activate ARP 2/3 complex. This leads to the polymerization of actin filaments beneath the adherent bacterium and this produces pedestal-like structure on the host cell surface (reviewed by Bhavsar *et al*., 2007; Croxen and Finlay, 2010; Schmidt, 2010). Moreover, in response to stress or SOS induction of bacterial prophages, a subset of bacteria lyses and the released Stxs on epithelial cells triggers redistribution of nucleolin to the cell surface. This nucleolin can bind to the bacterial surface protein intimin and can strengthen the attachment (Figure 1.4) (Toshima *et al*., 2007; Tree *et al*., 2009; Croxen and Finlay, 2010).

1.3.5 Coordination of LEE gene activity

The LEE-encoded T3S system must be assembled in a coordinated manner such that the T3S apparatus is first formed and then effector molecules are exported through the apparatus into the host cell (reviewed by Tree *et al*., 2009; Deane *et al*., 2010). Most of the basal apparatus proteins are encoded by the *LEE1-3* operons. These proteins are related to the flagellar proteins. These proteins are exported in a Sec-dependent manner and form basal apparatus, which has a close association with inner and outer membranes through the peptydoglycan layer. Once the basal apparatus has formed, proteins associated with the formation of the needle and filamentous apparatus are assembled. In EPEC and EHEC, basal apparatus protein EscU, which is a functional homologue to *Salmonella* flagella, is likely to be involved in the switch to needle complex assembly (Edqvist, *et al*., 2003). With the completion of needle formation, translocon components EspA, EspD and EspB are assembled and form filamentous translocon. The opening of this filamentous conduit in the host cell leads to changes such as lowering calcium levels that switch to effector protein secretions (Deng *et al*., 2005). It is likely that sequential activities of effector proteins are important for the Esc T3S system. Based on the translocation efficiency, the system has a secretion hierarchy of Tir>EspZ>EspF>EspH>EspG>Map (Mills *et al*., 2008).

From the above discussion, it is clear that, to execute the function of T3S system leading to pathogenesis, the expression of the associated LEE genes need to be orchestrated appropriately in time and space. This is likely to account for the control of the LEE gene expression at transcriptional, post-transcriptional and post-translational levels (Tree *et al*., 2009; Bhatt *et al*., 2011). Since a major part of this thesis concerns the regulation of the LEE genes at the transcriptional level, an overview of transcription initiation and regulation in *E*.

coli K-12 system is discussed in the next section, together with a discussion of relevant aspects of post-transcriptional regulation.

1.4 Regulation of bacterial transcription: an overview

1.4.1 The importance of regulation

Bacteria are highly versatile and responsive organisms. They can rapidly sense environmental changes and adapt themselves to changing environments by controlling metabolic patterns with a wide variety of mechanisms. Their genomes provide the template for the production of proteins necessary for the correct functioning of the cell. Under specific environmental conditions, each gene product is necessary at a certain concentration. Therefore, bacteria have to modulate the level of gene expression in response to environmental changes for survival. The expression of genes can be regulated at many steps, from the initiation of transcription, through to the stability of transcripts, translation initiation and the stability of protein. However, from the point of view of economy, the key step of regulation of a gene is the initiation of RNA-transcript formation. The components of transcriptional regulation include RNA polymerase, σ factors, promoter DNA sequences, small ligands, components of the folded chromosome and transcription factors (Browning and Busby, 2004).

1.4.2 RNA polymerase

RNA polymerase (RNAP) is a multi subunit enzyme that catalyzes DNA-dependent RNA synthesis. RNAP plays a major role in controlling gene expression. In a single *E. coli* cell, transcription of over 4000 genes are catalyzed by a limiting number of RNAP molecules, the majority of which are occupied with transcription of stable RNA genes such as rRNA and tRNA genes. Therefore, only a very little amount of RNAP is available for transcription of the remaining genes (Ishihama, 2000). Since RNAP is in such short supply, it is crucial that transcription initiation is regulated, such that genes are expressed at the required time and under the appropriate environmental conditions. Prudent distribution of RNAP is mediated by the combination and sequence of promoter elements, the availability and activity of alternative sigma factors, and transcription factors (Browning and Busby, 2004; Browning *et al*., 2010).

The form of RNAP, which transcribes genes, is known as the core enzyme that has a subunit

composition of $\alpha_2\beta\beta'\omega$ with a molecular mass of ~400 kDa, where α, β, β′ and ω denote different subunits. Each of the subunits has different size: β' subunit, ~155 kDa; β subunit, ~151 kDa; α subunit, ~37 kDa and ω subunit, ~6 kDa. These core RNAP subunits share considerable sequence, and even more structural homology, with their archaeal and eukaryotic counterparts (Ebright, 2000). Structural studies revealed that core RNAP adopts a structure reminiscent of a crab claw (Zhang *et al*., 1999; Murakami *et al*., 2002b; Browning and Busby, 2004) (Figure 1.5). The two 'pincers' of the 'claw' are made up of the two large subunits β' and β and define a cleft. The cleft has active centre at the base of the two pincers that can bind two Mg^{2+} ions and can accommodate a double-stranded nucleic acid. Thus, the cleft serves as the active center cleft and carries at least part of the binding site for double-stranded nucleic acid. The two identical α subunits are located distal to the cleft. One α subunit interacts with the β; this subunit is designated αI. The other α subunit interacts with β'; this α subunit is designated αII. Each α subunit consists of two domains: an N-terminal domain (αNTD) responsible for interaction with β' and β, and a C-terminal domain (αCTD) responsible for sequence-specific protein-DNA interactions with upstream-bound activators and repressors (Busby and Ebright, 1994; Browning and Busby, 2004). αCTD is connected to αNTD, and thus to the remainder of RNAP, through a long, flexible linker. The linker allows αCTD to occupy different positions relative to the remainder of RNAP in different promoter contexts (Blatter *et al.*, 1994; Busby and Ebright, 1994; Ebright and Busby, 1995). The ω subunit also is located distal to the cleft, near the base of pincer consisting of β' and exclusively interacts with β'. The small ω subunit has no direct role in transcription but seems to function as a chaperone by which it helps to maintain β' in a correct conformation and recruit it to the $\alpha_2\beta$ subassembly to form a functional core enzyme ($\alpha_2\beta\beta'\omega$) (Ghosh *et al*., 2001, 2003; Mathew and Chattergi, 2006).

The core form of RNAP enzyme is capable of carrying out transcript elongation. Specific initiation of transcription, however, requires the holoenzyme which, in addition to the core subunits contains the transcription initiation specific sigma subunit or specificity factor (σ: 70 kDa) (Murakami *et al*., 2002a,b). The σ subunit has multiple functions: to ensure the recognition of specific promoter sequences; to position the RNAP holoenzyme at a target promoter and to facilitate unwinding of the DNA double helix to form a transcription bubble near the transcript start site.

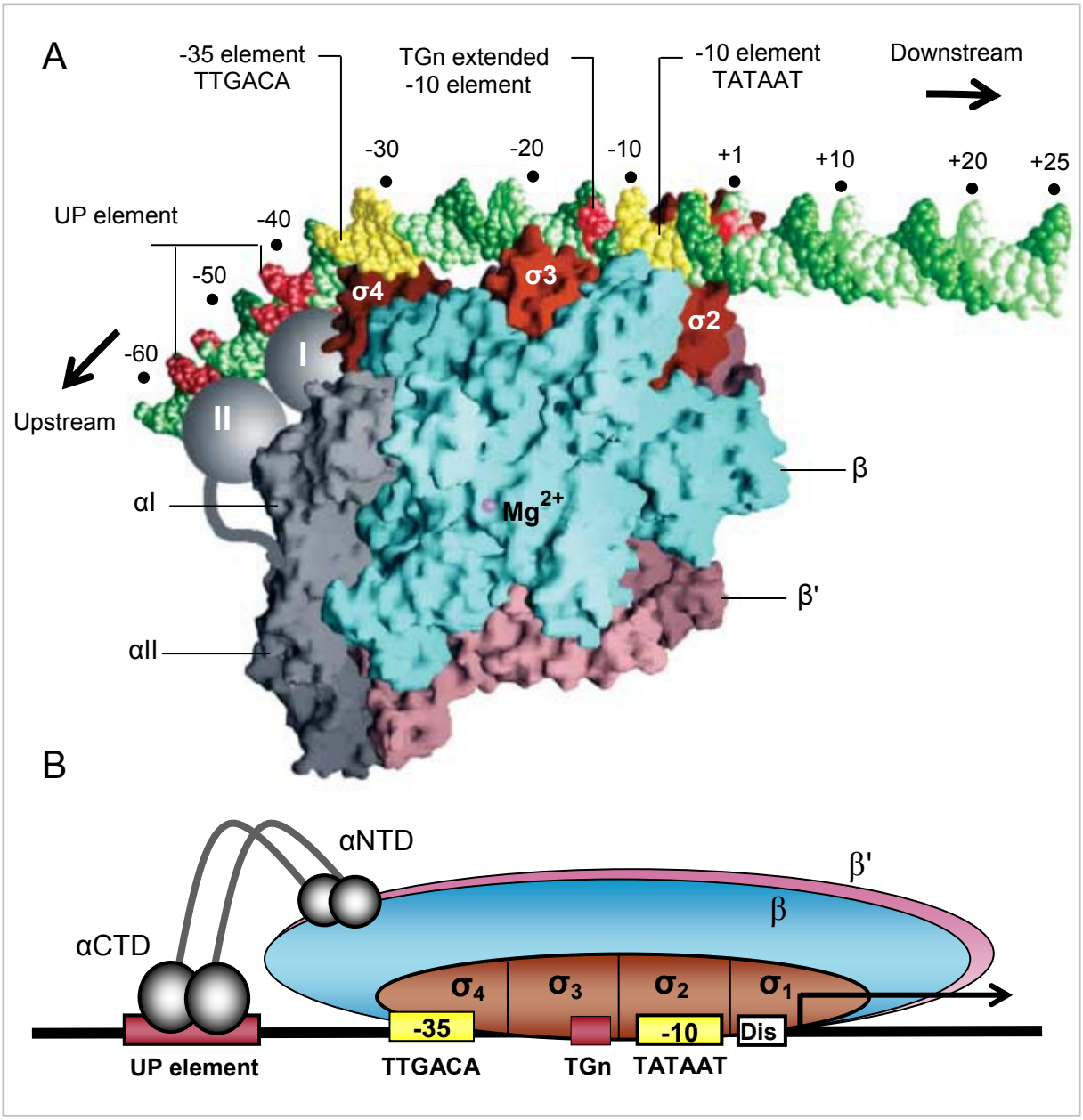


Figure 1.5. RNA polymerase interactions at a promoter.

A. The DNA strands are shown in green, with the –10 and –35 elements highlighted in yellow and the TGn extended –10 and the UP elements highlighted in red. RNA polymerase is shown with the β and β′ subunits coloured light blue and pink, respectively, αNTDs are coloured grey and the different domains of σ are coloured red. Grey spheres labelled I and II, represent the domains of αCTD that bind to the promoter. The RNA polymerase active site is denoted by the Mg^{2+} ion (magenta).

B. Schematic diagram of the different interactions between RNA polymerase holoenzyme and promoter elements shown in part A: αCTDs and the UP element; σ^{70} region 4.2 and the -35 hexamer (consensus TTGACA); σ^{70} region 3.0 and the extended -10 (TGn) element; σ^{70} region 2.4 and the -10 hexamer (consensus TATAAT) and σ^{70} region 1.2 and discriminator (Dis) sequence (not resolved in crystal structure) (adapted from Browning and Busby, 2004).

1.4.3 Promoters

RNAP initiates transcription at segments of DNA known as promoters. A standard promoter has two or more core elements. The two principal conserved core elements are the -10 hexamer (consensus 5′-^{12}TATAAT^{7}-3′) and -35 hexamer (consensus 5′-^{35}TTGACA^{30}-3′) sequences that are located 10 and 35 base pairs upstream from the transcript start site, respectively (Rosenberg and Court, 1979; McClure 1985; Lisser and Margalit, 1993). These two core elements are separated from each other by a spacer having a characteristic length of 17±1 base pairs. Sequence changes in the spacer region have little influence on the promoter activity, while differences in spacer length drastically affect the performance of a promoter (Wagner, 1984; deHaseth and Helmann, 1995). The sequences outside of the core elements can contribute much to the activity of a given promoter. For instance, an extended -10 element that is a 3-4 base pair motif, located one base upstream of the -10 hexamer can increase promoter activity significantly. The major determinant of the extended -10 element is 5′-TG-3′, which is located at positions –15/–14 with respect to the transcription start point (Barne *et al*., 1997; Mitchell *et al*., 2003). At many but not all promoters, upstream or UP elements, which are ~ 20 base pairs sequence located upstream of the promoter -35 element can increase promoter efficiency by as much as 30-fold (Ross *et al*., 1993). In addition, the downstream sequence region, ranging from positions +1 to +20 can also contribute to promoter activity by providing stability to the open complex and influencing the facility with which RNAP escapes from the open complex as the nascent RNA chain is elongated (Wagner, 1984).

Transcription initiation requires recognition of promoter elements by RNAP holoenzyme. The four core promoter elements, i.e., the UP, -35, -10, and extended -10 elements, specify the initial binding of RNAP to a promoter (Figure 1.5). Promoters with core elements closer to the consensus are generally stronger than those that match less well (McClure, 1985; Miroslavova and Busby, 2006). The strength of a promoter can be defined by the number of transcripts it can initiate in a given time. This measure is influenced by how well the promoter binds polymerase initially, how efficiently it supports isomerization and how rapidly the polymerase can then escape. It is important to mention that there is no known naturally occurring promoter in which all the core elements are consensus (Browning and Busby, 2004; Mitchell *et al*., 2003). An artificial promoter with perfect match of all core components could bind RNAP too tightly resulting in reduced level of activity (Miroslavova and Busby, 2006).

Many promoters particularly tRNA and rRNA promoters contain ~20 base pairs A/T rich, factor-independent activation sequence, located upstream of the promoter - 35 element. These are known as UP elements (Ross *et al*., 1993; Estrem *et al*., 1998). Genetic and biochemical analyses revealed that the α subunit C-terminal domain (αCTD) of RNAP holoenzyme is responsible for recognition of UP elements (Ross *et al*., 1993; Busby and Ebright, 1999). UP element sequences consist of two subsites, each of which binds one αCTD (Estrem *et al.,* 1999). Transcription from promoters containing only a consensus proximal UP element subsite require only one αCTD, and this subsite is preferred by αCTD in UP elements containing good matches to consensus in both subsites. Each UP element subsite interacts with a single αCTD and as a result, UP elements containing two good subsites require both αCTDs for maximal stimulation of transcription. The two αCTDs of RNAP are functionally interchangeable with respect to recognition of an UP element consisting of only a consensus proximal subsite. Promoters containing only a consensus distal subsite require both αCTDs for efficient transcription, perhaps because of sequence-non-specific interactions between the proximal subsite region and the second αCTD or because the second αCTD in some other way affects the overall stability of the complex (Gourse *et al*., 2000).

Apart from interaction between promoter UP elements and the RNAP α subunit, other core elements of a promoter are recognized by the RNAP specificity factor or σ factor. Sequence analysis of different σ factors together with structural analysis revealed that members of the σ^{70} family typically share four conserved domains named σ_1, σ_2, σ_3 and σ_4 (Helmann and Chamberlin, 1988; Browning and Busby, 2004) (Figure 1.5). These domains are linked to each other by flexible linkers and consist of conserved region 1.1-1.2, 2.1-2.4, 3.0-3.2, and 4.1-4.2, respectively (Lonetto *et al.*, 1992; Campbell *et al.*, 2002; Murakami *et al*., 2002b). Region 2 is the most highly conserved region playing critical functions including core RNAP binding, promoter -10 element (consensus sequence TATAAT) recognition, melting and interaction with the -10 region nontemplate strand DNA to stabilise the melted state (Siegele *et al*., 1989, Fenton *et al*., 2000; Feklistov and Darst, 2009). Region 3.0 interacts with the 'extended -10' promoter motif (Barne *et al*., 1997; Campbell *et al*., 2002). Region 4 is the most conserved domain of σ^{70} after Region 2. Residues in region 4.2 determine sequence-specific interactions with the -35 element (consensus sequence TTGACA) (Siegele *et al*., 1989; Feklistov and Darst, 2009). Genetic studies revealed that σ region 4 is

dispensable at promoters containing an 'extended -10' element (TGnTATAAT) (Ponnambalam *et al*., 1986; Keilty and Rosenberg, 1987; Barne *et al*., 1997; Mitchell *et al*., 2003). This is because interactions between residues in region 3.0 and the upstream 'TG' motif can substitute for region 4 interactions with the -35 element.

Domain 1.1 is present mainly in the primary σ factors, including σ^{70} (Lonetto *et al.*, 1992). Structural analysis cannot resolve this protease sensitive domain and the function of this domain is not clear. One possibility is that this domain is involved in inhibition of promoter recognition by free σ (Dombroski *et al.*, 1993) and the inhibition is presumed to be relieved upon σ binding to core RNAP. Moreover, it can accelerate open complex formation at some promoters (Vuthoori *et al*., 2001). Additionally, many promoters share an element located immediately downstream of the -10 element, within the discriminator region. The discriminator region includes G+C rich sequence motif that binds to σ region 1.2 (Figure 1.5). High G+C content interferes with strand separation, leading to promoter regulation. Genetic analysis revealed that a base substitution C to G, two nucleotides downstream from the -10 hexamer in the rRNA promoter *rrnB* P1 eliminated its regulation implying that the discriminator region is related to the stability of transcription initiation complex (Haugen *et al*., 2006).

1.4.4 Transcription initiation

Transcription initiation takes place in a number of steps (Murakami and Darst, 2003) (Figure 1.6). Specific recognition of promoter DNA by holo RNAP results in closed complex formation (RP_C) in which double stranded -35 and -10 elements interact with σ_4 and σ_2, respectively (Borukhov and Nudler, 2008). In this form, the DNA remains double stranded, hence the name 'closed complex' and is partially protected by RNAP from about -50 to +5 position (Li and McClure, 1998). In the second step of initiation, the RP_C undergoes a transition to the open complex (RP_O) in which the DNA strands around the transcription start site unwind and melt and the template strand moves into the RNAP active site to form the transcription bubble. During open complex formation, conserved aromatic residues within $\sigma_{2.3}$ function to trigger the destabilization of the DNA duplex at the conserved -11 position (A/T bp) (deHaseth and Helmann, 1995; Murakami *et al*., 2002b). The bending of DNA resulting from duplex unwinding at positions -11 to -7 allows the downstream DNA to bend or kink

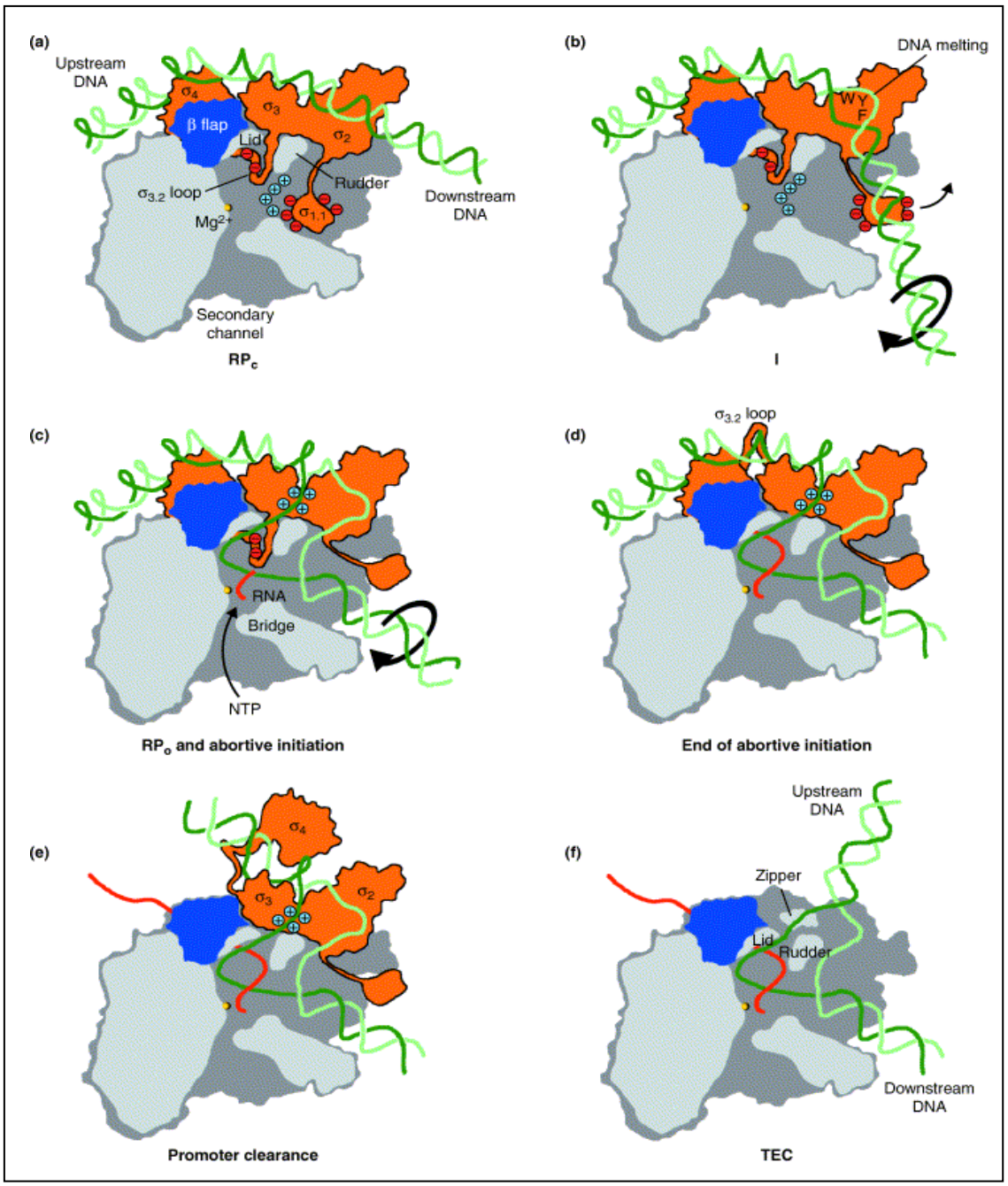


Figure 1.6. Transcription initiation by bacterial RNA polymerase.

The figure shows schematic diagrams depicting structural transitions during the steps of transcription initiation. Cross-sectional views of the RNA polymerase-promoter DNA complexes at different stages are shown: (a) RNA polymerase-promoter closed complex (RP_C) (b) intermediate complex (I), (c) RNA polymerase-promoter open complex (RP_O) and abortive initiation, (d) end of abortive initiation, after displacement of the σ 3.2 loop (e) promoter clearance and (f) elongation complex (TEC). RNA polymerase β flap, blue; σ, orange; rest of RNAP, grey; catalytic Mg^{2+}, yellow sphere; promoter DNA-template strand, dark green; nontemplate strand, light green; and the RNA transcript, red. The figure has been adapted from Murakami and Darst (2003).

across the entrance of the active-site channel leading to the entrance of the ~20 bp downstream DNA duplex (ahead of the +1 start site) into the DNA-binding clamp. At this point, the fully unwound duplex and transcription bubble stretches from -11 to +2, thus occupying the major channel. This sequence of events allow the correct placement of the +1 base of the template strand in the catalytic centre, which enables the initiation phase of transcription (Naryshkin *et al.*, 2000; Murakami *et al.*, 2002b; Murakami and Darst, 2003). However, the process of DNA entry to RNAP channel is accompanied by simultaneous exit of $\sigma_{1.1}$, which occupies the main RNAP channel in the RP_C (Mekler *et al.*, 2002). On formation of RP_O, the RNAP active site provided with NTP substrates through the secondary channel, starts catalyzing the synthesis of the RNA chain. The initial short transcript encounters the $\sigma_{3.2}$ loop that blocks the RNA exit channel, resulting in multiple rounds of abortive initiation. To be retained in elongating RNAP, nascent RNAs must successfully compete with region of $\sigma_{3.2}$ (Young *et al*., 2002). If RNA transcripts lose the competition, they are ejected as abortive transcripts; if they win, region $\sigma_{3.2}$ is ejected and the transcript is successfully elongated. Eventually, the RNA chain elongates to a length of about 12 nucleotides, which is sufficient to fill the RNA-DNA hybrid and upstream RNA exit channel completely under the β flap, thereby displacing the $\sigma_{3.2}$ and marking the end of abortive initiation as well as allowing the transcription to proceed into the elongation phase (Murakami and Darst, 2003) (Figure 1.6).

1.4.5 Small ligands

Small ligands like guanosine 3', 5'-biphosphate (ppGpp) or ATP (adenosine triphosphate) provide alternative mechanisms by which bacteria can respond quickly and efficiently to environmental changes and thus act as transcriptional regulators. For example, increased levels of ATP in a bacterial cell helps the promoters of housekeeping genes to capture more RNAP at exponential growth phage (Gaal *et al*., 1997) whereas the availability of intracellular levels of ppGpp shut down the transcription process rapidly by destabilizing open complexes if protein synthesis slows down due to amino acid limitation (Chatterji and Ojha, 2001). It seems that ppGpp controls expression of the translational machinery in response to sudden starvation, while ATP availability controls expression in response to growth rate. Apart from inhibition of transcription of rRNA and tRNA genes, ppGpp stimulates expression of proteins required for amino acid biosynthesis and transport (Paul *et al*., 2005). The overall effect of ppGpp action is thus to increase amino acid pools in the cell.

The activity of ppGpp upon RNAP is modulated by DksA. DksA is a small α-helical protein that consists of 151-amino-acids (Perederina *et al*., 2004). This protein has three distinct structural parts. The globular portion or G domain is composed of N- and C- terminal regions (residues 7-33 and 110-134). The central domain, a 75 residue coiled coil (CC), consists of two long N- and C-terminal α helices (residues 35-68 and 75-109) connected by a linker at the tip of the CC that comprises two residues (69 and 70) in an extended conformation and an α-helical turn (α turn, residues 71 and 75). The third part is a C terminal α helix (residues 135-151) that is loosely bound with rest of the protein. Unlike conventional transcription factors, DksA binds directly to RNAP and modifies the kinetic properties of the promoter complex, sensitizing it to changes in the concentrations of ppGpp and the initial NTP in the transcript. Hence, it can enhance direct effects of ppGpp on the negative control of rRNA promoters (Paul *et al.*, 2004). Structural analysis of the RNAP-ppGpp complex revealed that ppGpp binds in the secondary channel close to the RNAP active site, where they likely coordinate the catalytic Mg^{+2} ion (Artsimovitch *et al*., 2004). Upon binding to RNAP, the DksA CC extends into the secondary channel toward the ppGpp binding site and coordinates one of the ppGpp bound Mg^{2+} ion through the invariant acidic side chains at the CC tip, thereby stabilizing the RNAP-ppGpp complex (Perederina *et al*., 2004).

1.4.6 Transcription factors

The concept of transcriptional regulation by transcription factors started from the early paradigm of the operon model, which was ingeniously proposed by Jacob and Monod (1961). An operon has three parts: promoter, operator and structural gene(s). In addition, there are associated regulatory genes that are located at some distance from the operon on the DNA. Regulatory genes produce repressor or activator proteins that function as the on-off switch for a gene. The operon model explained gene regulation by direct repression, which today is known to be only one kind of the many regulatory concepts. More than 300 hundred genes in *E. coli* genome encode proteins that act as transcription factors and are supposed to be bound to promoters to scale up or down the expression of a large number of genes in response to environmental signals (Perez-Rueda and Collado-Vides, 2000). Most of these proteins have sequence specific DNA binding sites and are therefore, directed to specific promoters. Some proteins act as global regulators and interact at scores of promoters, whereas others relate to just one or two promoters. For example, the *E. coli* cyclic adenosine monophosphate (AMP)

receptor protein, CRP, interacts at nearly 200 different promoters and regulates their efficiency in response to certain stress conditions. It has been estimated that seven transcription factors (CRP, FNR, IHF, Fis, ArcA, Narl and Lrp) have a role in regulation of >50% promoters, whereas ~60 transcription factors regulate single promoters (Martinez-Antonio and Collado-Vides, 2003).

Transcription factors control the expression of genes in response to environmental signals. Thus, the action of transcription factors must be regulated. This can be done by small ligands that can alter the DNA binding affinity of a transcription regulator or by covalent modification, which can also alter the affinity of a transcription regulator to its DNA target. Moreover, in some cases, the activity of a transcription factor is solely controlled by its cellular concentration. In other cases, the activity of a transcription regulator is controlled by binding to other regulatory proteins (Browning and Busby, 2004).

A transcription factor can act as an activator or repressor upon binding to a promoter. Some transcription factors function purely as activators or repressors, whereas others can act as either according to the architecture of the target promoter (Perez-Rueda and Collado-Vides, 2000). For instance, CRP functions as an activator in the case of the *gal* P1 promoter and simultaneously as a repressor in the case of the *gal* P2 promoter (Musso *et al*., 1977). Activators improve the performance of a promoter by improving its receptivity to RNAP. On the other hand, repressor proteins reduce transcription initiation at target promoters. Most activators function by binding to target promoters before acting on RNAP. However, several transcription regulators, for example MarA and SoxS, provide alternative mechanisms in which they interact with free RNAP before binding to promoter DNA (Martin *et al*., 2002).

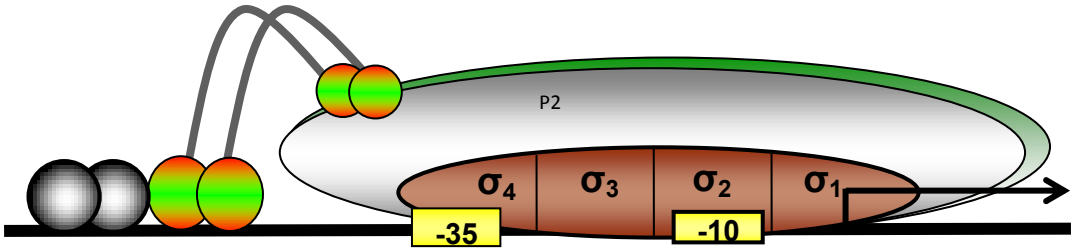
At many promoters, regulation of transcription is simple, mediated just by a single factor. However, in other cases, more than one transcription factors are involved in influencing the activity of a single promoter with each factor relying on an environmental signal. In the case of simple activation, most activators function by binding to a target that is located upstream or overlapped with the -35 element and recruit RNAP to the promoter by interacting with the RNAP subunits. The first mode is known as Class I activation whilst the latter is referred to as Class II activation (Browning and Busby, 2004). In the case of Class I activation, the activator recruits the RNAP to the promoter by directly interacting with the RNAP αCTD

(Ebright, 1993) (Figure 1.7A). In case of Class II activation, RNAP recruitment to the promoter is facilitated by the interaction between the activator and domain 4 of RNAP σ subunit (Dove *et al*., 2003) (Figure 1.7B).

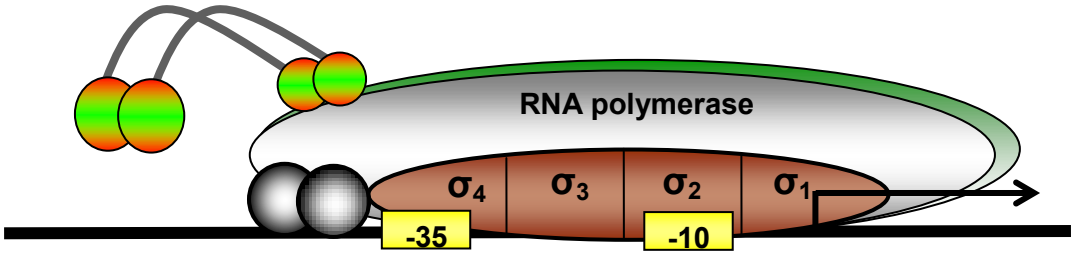
Though most transcription activators bind sites that are upstream of or overlapping the -35 elements of the promoters, a number of transcription factors bind to sites that are located between promoter -10 and -35 elements and activate the promoters. In these cases, the spacer DNA sequences between the promoter -35 and -10 elements recognised by the RNAP σ factor are greater than the optimal 17 base pairs. For instance, members of the MerR family of transcription factors bind to the suboptimal spacing sequence of the target promoters and twist and/or bend the spacer DNA to reorientate the -35 and -10 elements so that they can be bound by the RNAP σ subunit (Brown *et al.,* 2003; Hobman, 2007) (Figure 1.7C). MerR family regulators include metal-responsive regulators MerR, PbrR, CueR, CadR; dye-responsive regulators BmrR, BltR; and stress-responsive regulator SoxR (Ahmed *et al*., 1995; Markham *et al*., 1996; Caslake *et al*., 1997; Hidalgo and Demple 1997; Borremans *et al*., 2001; Lee *et al*., 2001; Adaikkalam and Swarup, 2002). A MerR family transcription activator has an N-terminal helix-turn-helix DNA binding region and a C-terminal effector-binding region. The signature of the family is amino acid similarity in the first 100 amino acids, including a helix-turn-helix motif followed by a coiled-coil region (Brown *et al*., 2003; Newberry and Brennan, 2004).

The MerR transcriptional activator is the best-studied member of MerR family regulators. It is a regulator of Gram-negative mercury resistance (*mer*) operons found on transposable elements Tn*21* and Tn*501* (Brown *et al*., 1983; 2003; Barrineau *et al*., 1985). This regulator acts as an activator of the *mer* genes in the presence of Hg (II) salts and as a weak repressor in the absence of Hg (II). Genetic studies revealed that the 19 base pairs suboptimal spacer region between *mer* promoter -10 and -35 elements is essential for normal induction of the promoter by the MerR (Parkhill and Brown, 1990). A model for the mechanism of action of MerR at the *mer* operator is now well understood. In the absence of Hg (II) and MerR, RNAP transcribes from the *merR* promoter, which results in an increased concentration of MerR. The MerR then binds to the *mer* operator and represses the *merR* promoter by bending the DNA and unwinding the operator sequence. In the absence of Hg (II), the MerR is thus bound to

A: Class I activation



B: Class II activation



C: Activation by conformation change

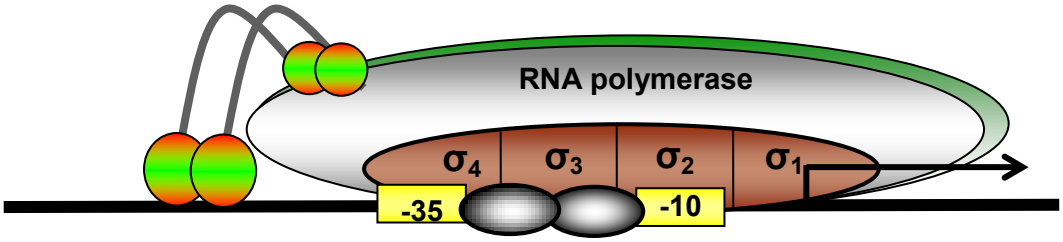


Figure 1.7. Activation at simple promoters.

A. The activator binds to an upstream site and contacts the αCTD of RNA polymerase, thereby recruiting the polymerase to the promoter.
B. The activator binds to a target that is adjacent to the promoter -35 element, and interacts with domain 4 of σ^{70}.
C. The activator binds in abnormal spacer sequence between promoter -10 and -35 elements elements and realigns promoter elements so that the RNA polymerase holoenzyme can bind to the promoter efficiently. The figure has been adapted from Browning and Busby (2004).

DNA in the repressor confirmation, maintaining repression of the promoter. Binding of Hg (II) to one of two binding sites on the MerR causes a conformational change to put MerR in an activating conformation. Due to the tight binding at the centre of the operator, this causes DNA distortion at the centre of the operator, giving ~33° unwinding of DNA and straightening of helix backbone. This results in reorientation of the -35 and -10 sequences and allows them to interact productively with the RNAP σ^{70} subunit to form an open transcriptional complex and transcription is initiated (Ansari *et al*., 1992, 1995).

Repressor proteins reduce transcription initiation at target promoters. Like activators, they can convey their regulatory role either in simple or complex ways (Browning and Busby, 2004). In case of simple repression, the repressor can use three mechanisms (Figure 1.8). Steric hindrance of RNAP binding to the promoter is the simplest mechanism of repression. In this case, the repressor binds in or close to, the core elements and may prevent RNAP recruitment or may interfere post-recruitment steps in transcription initiation. For instance, binding of Lac-repressor at *lac* promoter interferes with the recruitment of the RNAP to the promoter (Muller-Hill, 1998). In some cases, multiple repressor molecules may bind to promoter-distal sites that cause looping, which shuts off transcription in the looped domain. Repression of the *gal* promoter by the GalR repressor follows this mechanism (Choy and Adhya, 1996). Simple repression can also occur indirectly. In this case, a repressor can act as an anti-activator. The best example is the repression by CytR of CRP-dependent promoter, where repressor, CytR interacts with the activator, CRP and thus modulates the transcription of the CRP-dependent promoter (Shin *et al*., 2001).

1.4.7 Nucleoid-associated proteins

Bacterial chromosomes are highly compacted by supercoiling and interactions with proteins and RNAs. Folding of bacterial chromosomes by nucleoid-associated proteins affects the distribution of RNAP between promoters and thus controls its transcription. In *E. coli*, a number of nucleoid-associated proteins like Fis (factor for inversion stimulation), IHF (integration host factor), H-NS and HU (histone-like nucleoid-structuring proteins), StpA (a H-NS homologue), Dps (DNA-protein from starved cells), modulate gene expression either positively or negatively (Azam and Ishihama, 1999; Browning *et al*., 2010). In addition to a role in genome compaction, these proteins can act solely as transcription regulators.

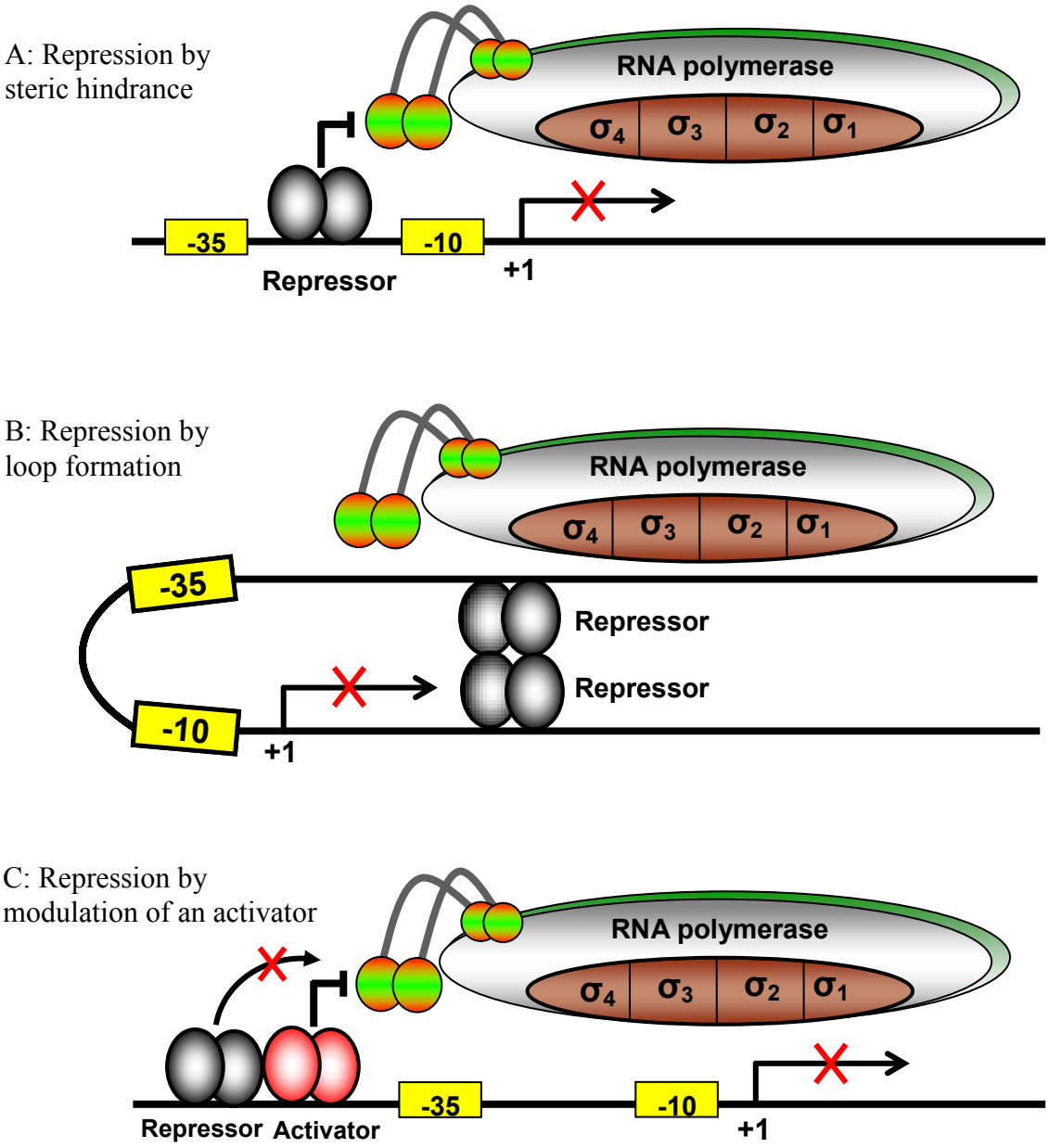


Figure 1.8. Mechanisms of repression.

A. The repressor-binding site overlaps core promoter elements and blocks recognition of the promoter by the RNA polymerase holoenzyme.
B. Repressors bind to distal sites and interact by looping, repressing the intervening promoter.
C. The repressor binds to an activator and prevents the activator from functioning by blocking promoter recognition by the RNA polymerase holoenzyme. The figure has been adapted from Browning and Busby (2004).

H-NS, encoded by *hns*, belongs to a family of small abundant (20,000 molecules per cell) proteins in Gram-negative bacteria (Grainger *et al*., 2006; Dorman, 2007). This protein acts as a global repressor in bacterial gene expression (Stoebel *et al*., 2008; Browning *et al*., 2010). Genome-wide analysis using whole-genome transcriptomic methods including DNA microarray technique and chromatin immunoprecipitation on chip (ChIP-on-chip) reveals an extensive regulon of H-NS in enteric bacteria (Grainger *et al.*, 2006; Navarre *et al.*, 2006; Oshima *et al.*, 2006; Wade *et al.*, 2007). H-NS or H-NS like proteins consist of a C-terminal DNA binding domain and an N-terminal oligomerization domain that are separated by a flexible linker sequence (Bertin *et al*., 1999; Dorman *et al*., 1999; Bloch *et al*., 2003; Stoebel *et al*., 2008; Arold *et al*., 2010). H-NS acts as a transcription silencer and preferentially binds A+T rich DNA sequences irrespective of chromosome location (Navarre *et al.*, 2006). Horizontally acquired foreign DNAs, particularly various pathogenic islands in Gram-negative bacteria contain a substantially increased level of A+T content compared with that of the resident genome and are found to be silenced by H-NS (Hacker *et al*., 1997; Elliott *et al*., 1998; McClelland *et al*., 2001; Dorman and Kane, 2009). H-NS mediated silencing of the foreign genes prevents inappropirte expression and thus, protect the cell from detrimental consequences of invading DNA.

Unlike classical regulators, H-NS does not recognise a particular DNA sequence (Lucchini *et al*., 2006). However, different *in vitro* experiments suggest that H-NS interacts with a A+T rich DNA region that is prone to intrinsic curvature (Yamada *et al*., 1991; Rimsky *et al*., 2001). Moreover, oligomerization properties of H-NS molecules suggest that it can interact with an extended region of curved DNA and may develop DNA-protein-DNA bridges both between separate DNA molecules and between different portions of the same DNA molecule (Stoebel *et al*., 2008). However, Lang *et al*. (2007) proposed that H-NS preferentially binds to a sequence of 5'-TCGATATATT-3' located in A+T rich DNA region of a target gene. It is likely that H-NS nucleates at the proposed or related sites, then spreads along the DNA sequences laterally and forms H-NS filaments that possibly lead to formation of DNA-H-NS-DNA bridges.

Many regulators including SlyA, RovA, ToxT and Ler have been shown to act as antisilencers by displacing H-NS from the target site(s) of respective genes (reviewed by Stoebel *et al*.,

2008) and thus, provide bacteia selective advantages. SlyA protein is usually an activator that regulates many genes in both *E. coli* and *Salmonella* and is related with the MarR protein in *E. coli* (Ellison and Miller, 2006). This protein acts as an anti-repressor in the expression from *hlyE*, the haemolysin gene by counteracting H-NS- mediated repression. Usually H-NS prevents RNAP to direct transcription from the *hlyE* promoter, whereas SlyA permits RNAP to bind and prevents H-NS binding (Lithgow *et al*., 2007). Moreover, based on the relative abundance, both H-NS and SlyA proteins can show mutual antagonism at the target promoter site of the *hlyE* gene in such a way that H-NS can displace SlyA protein if the concentration of H-NS is higher than that of the SlyA protein and *vice versa* (Lithgow *et al*., 2007). H-NS also represses horizontally acquired genes that encodes major toxins, CTX and Tcp in *Vibrio cholerae* responsible for human cholera disease (Skorupski and Taylor, 1997). H-NS binds A+T rich promoter sequences and represses the expression from the promoters (Nye *et al.*, 2000). However, the ToxT regulatory protein, an AraC like DNA binding protein, negates the repression (Yu and DiRita, 2002). It is likely that ToxT plays dual roles: displacing the H-NS from the promoters and activates the promoters, possibly by directly interacting with RNAP (Yu and DiRita, 2002).

1.5 Post-transcriptional regulation

from a gene to its protein product. Though the regulation at transcription level is pivotal for any gene expression, various forms of post-transcriptional control can also affect bacterial gene expression. These forms of regulation are independent of changes in transcription. In post-transcriptional regulation, messenger RNA (mRNA) and proteins are the regulatory targets. Examples are regulation of translation initiation, changes in mRNA or protein stability or availability, or programmed frameshifting.

1.5.1 Regulation of translation initiation

mRNAs are used as templates to form proteins during translation when the templates are read according to the genetic code that relates base sequences to amino acid sequences. In fact, the process mediates through interplay of many macromolecules including mRNAs, transfer RNAs (tRNA) and a number of proteins. Most importantly, molecular machines (known as ribosomes) coordinate this interplay and catalyze the process. Usually, an *E. coli* cell in exponential growth contains ~20000 ribosomes (Forsman *et al*., 1990; Stryer, 1995). Each

ribosome consists of RNAs and proteins and has a size of ~2700 kDa with a sedimentation coefficient of 70S. This 70S ribonucleoprotein can be disassembled to a large 50S subunit and a small 30S subunit. The 50S subunit is made up of 34 different proteins (L1-L34) and 2 species of RNA molecules (23S and 5S). The 30S subunit possesses 21 different proteins (S1-S21) and a single RNA species named 16S (Wittmann, 1982; Wittmann *et al*., 1982; Wittmann, 1983).

The translation process starts at the 5' end segment of bacterial mRNA with recognition of a translation initiation region (TIR) by the 30S ribosomal subunit together with initiation factors and initiator tRNA. This event is then followed by the attachment of the large 50S subunit and release of initiation factors. In most bacteria, the TIR contains two key sequences, a translation initiation codon and a ribosomal-binding sequence termed the Shine-Dalgarno (SD) sequence (Shine and Dalgarno, 1974; 1975). The SD sequence is a purine-rich nucleotide sequence (typically GGAGG), that is usually located ~10 nucleotides upstream of the initiation codon. The SD sequence base pairs directly to a complementary sequence (CCUCC) near the 3’ end of the 16S ribosomal RNA. The initiation codon, on the other hand, base pairs to the anti-codon of the initiator tRNA. The most common translation initiator codon in mRNA is AUG. However, translation can also initiate less frequently at GUG, and rarely at TTG, AUU, CUG, AUA or AUC (Kozak, 1983; Schneider *et al*., 1986; Romero and García, 1991; Chalut and Egly, 1995; Van Etten and Janssen, 1998; O'Donnell and Janssen, 2001). Alterations of either the SD sequence or the initiation codon affect the translation process. Even, the distance between these two elements is crucial, can markedly affect the translation efficiency (Chen *et al*., 1994). Moreover, bases upstream of the SD sequence and downstream of the initiator codon can function as translation enhancers (Qing *et al*., 2003; Vimberg *et al*., 2007).

Secondary structures of the 5' untranslated region of mRNA molecules nearer to the translation initiation region often provide mechanisms for regulation of translation. This *cis*-acting structured segment can down-regulate the translation efficiency by competing with the 30S ribosome for the SD sequence (de Smit and van Duin, 1994a,b). Moreover, down-regulation can also arise when a *trans*-acting repressor prevents the 30S ribosomal subunits from binding to the SD sequence (Moine *et al*., 1988; Draper *et al*., 1998). In many

cases, secondary structures near the initiation codon, in conjunction with repressor proteins, can stall or entrap ribosomes resulting in unproductive translation preinitiation complexes. For instance, translation of *rpsO* mRNA is repressed by the S15 ribosomal protein by an entrapment mechanism. The model for this regulation suggests that, though the SD sequence can base pair to the ribosome, the ribosome is still stalled because the translation initiator codon is hindered by the pseudoknot structure of the mRNA that makes the initiation codon inaccessible for binding to the initiator tRNA (Marzi *et al*., 2007).

In many cases, 5' untranslated regions of mRNA molecules contain characteristic genetic elements known as riboswitches that can directly sense various metabolites without involving proteins (Mandal *et al*., 2003; Breaker, 2010). Riboswitches usually contain two functional components: a metabolic binding aptamer domain and an expression platform. The size of an aptamer can range from 35 to 200 nucleotides. The binding of metabolites to the aptamer induces allosteric changes in riboswitches causing downstream structural changes and affects expression of adjacent genes. In contrast to the conserved aptamer, the expression platform varies widely in size, sequences and structures allowing regulation at different levels from transcriptional elongation to translation initiation. This form of post-transcriptional regulation has been reported to control the genes of many essential metabolic pathways including biosynthesis of cobalamin transport protein, thiamine, riboflavin or riboflavin transport protien, lysine, adenine and glycine (Mironov *et al*., 2002; Nahvi *et al*., 2002; Winkler *et al*., 2002a,b; Sudarsan *et al*., 2003; Mandal and Breaker., 2004; Lemay and Lafontaine, 2007).

In addition to secondary structure, many bacterial mRNA molecules may contain short ORFs in the 5' untranslated long leader region. Translation of these leader ORFs can cause changes in the structure of the downstream region and subsequently can modulate the translation of downstream genes. These leader ORF-dependent regulatory mechanisms depend on the coupling between transcription and translation and examples can be found in regulation of many amino acid biosynthesis operons including those for tryptophan, leucine, isoleucine, and histidine (reviewed by Yanofsky, 1981; Vitreschak *et al*., 2004). Leader peptide-mediated post-transcriptional regulation was first found in tryptophan biosynthesis and it has been intensively studied (Yanofsky, 1981; 2000). The tryptophan operon, consisting of five contiguous *trp* genes, contains a 161 nucleotide leader sequence. The genes of the *trp* operon

are transcribed when tryptophan is limiting. However, once RNAP initiates a *trp* mRNA molecule, it does not always complete the full transcript as the *trp* leader sequence contains an attenuator. In the presence of high levels of tryptophan, *trp* transcription stops at the attenuator and yields an immature 139 nucleotides long RNA product. However, certain features in the leader sequence allow the attenuator to be passed by RNAP when tryptophan is limiting. This feature is an ORF encoding a short leader peptide of 14 amino acids and this ORF is preceded by a strong-ribosome binding site. The important feature of this leader peptide is to have two adjacent tryptophan codons. The function of these two codons is to stop a ribosome attempting to translate the leader peptide. Therefore when tryptophan is limiting, translation is stalled at the tryptophan codons long enough for the succeeding segment of the transcript to form a secondary structure that allows the transcribing RNAP molecule to proceed through the attenuator ensuring the full length transcription of the *trp* genes.

1.5.2 Messenger RNA turnover

mRNA stability is co-related with the level of expression of a particular gene. In a bacterial cell, activities of various exo- and endoribonucleases can limit the lifetime of mRNA molecules or impede the ability of an mRNA molecule to participate in the initiation of synthesis to generate the normal functional protein product and thus, allows timely changes in gene expression within a cell (Kennell, 2002). Exonucleases catalyse sequential cleavage of mononucleotides from either a free 5′ or 3′ terminus, or both. Generally, exonucleases in *E. coli* digest RNA in a 3' to 5' direction (Deutscher, 1993), though 5' to 3' exonuclease activity has also been reported to play a role in rRNA maturation and mRNA stability (Mathy *et al*., 2007). Endoribonucleases cleave the polynucleotide internally and can increase the decay rate by the removal of stabilization elements (stem-loop structures) that impede processive degradation by exoribonucleases (Higgins *et al*., 1992).

Endoribonuclease E (RNase E) is a key enzyme involved in RNA metabolism participating in the degradation, processing and maturation of variety of RNAs in bacteria. This enzyme is found to be involved in the post-transcriptional processing of EHEC *LEE4* operon in the LEE pathogenicity island (Roe *et al*., 2003; Lodato and Kaper, 2009). Transcription of the *LEE4* operon initiates upstream of the first gene *sepL* and terminates at the last gene, *espF*. However, RNase E is involved in processing the transcript to produce a partial transcript corresponds to

espADB transcript. Ribonuclease protection assays showed that the processing event occurs in the A+T-rich intergenic region between *sepL* and *espA*. 3' RACE mapped the 3' end of the *espADB* transcript that is located inside the coding region of *cesD2* and followed by stem-loop structure, predicted to be a Rho-independent transcription terminator (Lodato and Kaper, 2009). Interestingly, translation itself stabilizes mRNA. During translation, translating ribosomes can protect mRNA decay by shielding potential endoribonuclease sensitive sites in the mRNA (Arraiano *et al*., 2010) or a translational event can remove secondary structures, which may be necessary for initiation of the endoribonucleatic activity (Cole and Nomura, 1986; Nilsson *et al*., 1987).

1.5.3 Regulation by small RNAs

Small regulatory RNAs play many important roles in modifying mRNA stability and translatability (Waters and Storz, 2009). Usually, regulatory RNAs base pair with target mRNAs and directly modulate mRNA stability and translation. The best example is the DsrA small RNA that can regulate mRNA translation either positively or negatively. It binds the 5'-untranslated leader region of *rpoS*, removing a hairpin-like secondary structure, associated with the sequesteration of the SD sequence of *rpoS*. This permits access of ribosomes and enhances the mRNA stability and translation (Brown and Elliott, 1997; Majdalani *et al*., 1998). In case of *hns* mRNA, DsrA pairs just beyond the translation initiation codon and reduces the half-life of *hns* possibly by exposing or creating a RNase sensitive site within *hns* or by blocking translation (Lease and Belfort, 2000). In either case, a role for Hfq is crucial. Hfq is considered to be an RNA chaperone that can bind, stabilize and bring small RNAs to their targets (Zhang *et al*., 2003; Geissmann and Touati, 2004; Brennan and Link, 2007; Sittka *et al.*, 2008).

The Hfq protein has two RNA binding sites. The proximal site binds small RNAs and mRNAs, whereas the distal site binds the poly (A) tail. Apart from a role in stabilizing small RNAs and helping them to bind preferentially to single-stranded A+ U-rich regions of target mRNAs, it interacts with ribonucleases that are involved in mRNA decay (Brennan and Link, 2007). Hfq in conjunction with small RNAs play important roles in post-transcriptional regulation of virulence genes in different pathogens including *S. enterica* serovar Typhimurium (Sittka *et al.*, 2007; Wilson *et al.*, 2007), *S. sonnei* (Mitobe *et al.*, 2008), *Y.*

enterocolitica (Nakao *et al.*, 1995) and uropathogenic *E. coli* (Kulesus *et al.*, 2008). Recently, Hansen and Kaper (2009) reported that Hfq is involved in post-transcriptional regulation of genes in the LEE pathogenicity island in EHEC, though the small RNA involved in the system remains to be determined. The regulation is indirect, mediating through destabilizing mRNA of the *grlRA* operon that encodes regulators, GrlA and GrlR. Since GrlA positively regulates expression of Ler, which functions as an activator in most of the LEE gene expression, Hfq-mediated post-transcriptional control of GrlRA thus affects the LEE genes.

1.6 An overview of regulation of the LEE genes

As for other pathogenesis-related genes, the expression of the LEE genes is co-ordinately controlled by a complex network of signals and transcription factors in response to external stimuli (reviewed by Mellies *et al.*, 2007; Tree *et al*., 2009). The promoters of the LEE region are not simple and their activities are influenced by many different transcription factors. The *LEE1* operon in EHEC has two promoters, one of which is a proximal promoter whose transcription start site is located 32 base pairs upstream from the predicted *ler* translation start site (Sperandio *et al*., 1999). This promoter is present in EHEC and absent in EPEC (Sperandio *et al*., 1999; Elliott *et al*., 2000). The second one is a distal promoter whose transcription start site is located 163 base pairs upstream from the predicted *ler* translation start site and is present in both EHEC and EPEC (Mellies *et al*., 1999; Sperandio *et al*., 2002). Both *LEE2* and *LEE3* seem to have two overlapping promoters. The transcription start point of the *LEE2* operon is located 110 base pairs 5' to the translation start site of the *espZ* whereas transcription start point of the *LEE3* is located 60 base pairs upstream of the translation start site of the *cesL* (Mellies *et al*., 1999) in EPEC. The *LEE4* and *LEE5* operons are predicted to have simple promoters. The transcription start point of the *LEE4* is located upstream from the translation start site of the *sepL* (Roe *et al*., 2003). Transcription of the *LEE5* operon starts 86 nucleotides upstream of the *tir* start codon in EPEC (Sanchez-SanMartin *et al*., 2001).

Expression of the LEE genes is under the control of multiple environmental signals and host regulators. The first gene of the *LEE1* operon encodes a protein known as Ler (LEE encoded regulator) (Mellies *et al*., 1999) that functions as an activator of expression of most of the LEE genes (reviewed by Spears *et al*., 2006; Mellies *et al*., 2007; Tree *et al*., 2009). Ler is

often regarded as the master regulator of LEE gene expression and thus, the regulatory region of the *LEE1* operon is the main target for many of the regulatory inputs (Figure 1.9).

Various environmental signals appear to be necessary for the expression of the LEE genes. For instance, LEE genes are thermoregulated (Kenny and Finlay, 1995; Rosenshine *et al*., 1996). Cells, growing at 27-30°C show reduced level of LEE gene expression, whereas growth at 37°C stimulates LEE gene expression. LEE gene expression is regulated in response to signals such as calcium, iron, osmolarity, bicarbonate or pH, and starvation or growth arrest (in the SOS response) (Kenny *et al*., 1997a; Abe *et al*., 2002; Nakanishi *et al*., 2006). Quorum sensing, a cell-to-cell communication mechanism through which a bacterium can sense other bacteria in a given environment by the production of autoinducer (AI-3), also plays an important role in LEE gene expression (Sperandio *et al*., 1999; Kendall *et al*., 2010). Regulation depends on a wide range of both LEE and non-LEE encoded regulatory factors. Hence, the coordinate transcription of genes in the *LEE* operons is blocked by repressor proteins and activated by stimulatory proteins.

Global transcriptional regulators including H-NS and its homologue Hha, IHF, and Fis that affect bacterial chromosomal structure, appear to control the LEE gene expression. The 15.6 kDa nucleoid-associated protein, H-NS represses transcription of the LEE genes (reviewed by Clarke *et al*., 2003; Mellies *et al*., 2007). Umanski *et al*. (2002) reported that H-NS represses *ler* expression at 27°C, but not at 37°C, whereas it can repress *LEE2-5* operons at both 27°C and 37°C. This suggested that H-NS plays a role in the thermoregulated expression of the LEE operons. After induction at 37°C, Ler activates other LEE operons. *In vitro* experiments including electro mobility shift assays and footprinting analyses using purified H-NS, revealed that H-NS can bind and protect extended regions of *LEE2*, *LEE3* and *LEE5* regulatory sequences and thus directly controls transcription (Bustamante *et al*., 2001; Haack *et al*., 2003). However, the precise mechanism of H-NS mediated silencing of LEE genes remains to be determined, though several studies reported that H-NS silences expression from promoters by binding over extended AT-rich regions and is capable of bridging or looping DNA (Dorman, 2004; Dame *et al*., 2005; Browning *et al*., 2010) (see details in section 1.4.7).

Another nucleoid-associated protein, Hha directly binds to the *LEE1* regulatory region and

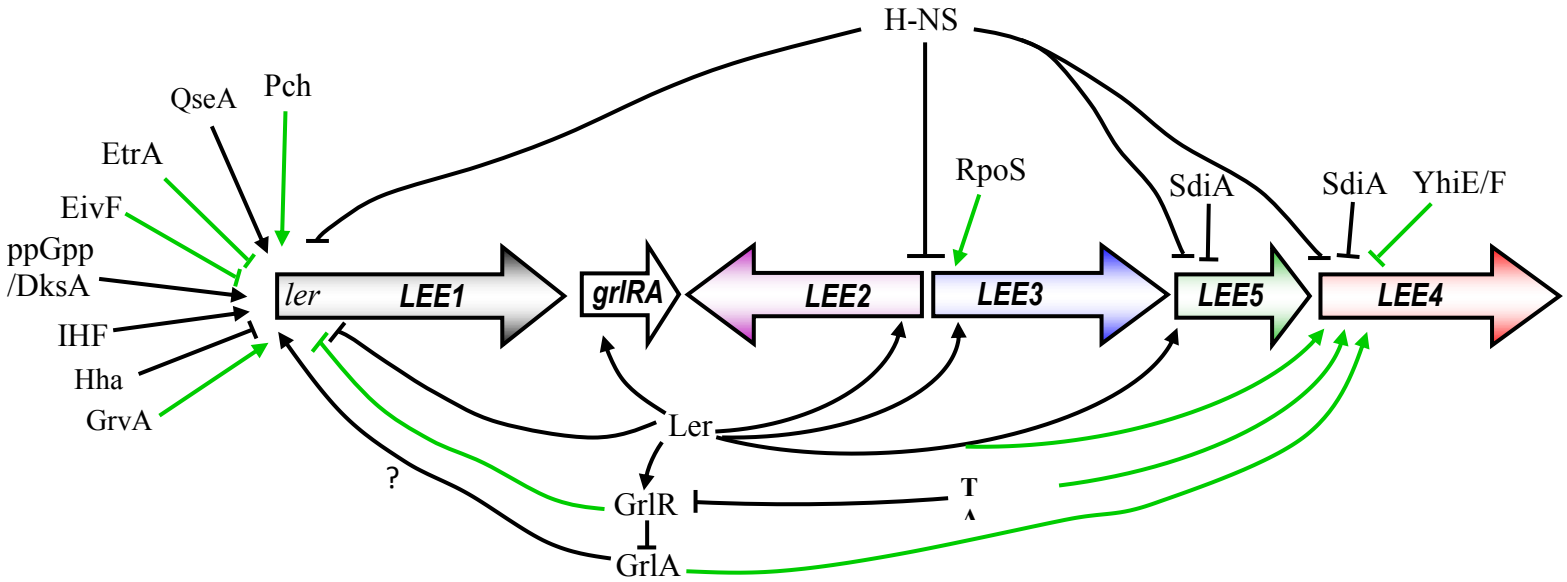


Figure 1.9. Schematic representation of the regulation of LEE genes.

The figure summarizes regulation of locus of enterocyte effacement (LEE) genes in enterohaemorrhagic *E. coli* O157:H7. Transcriptional activators and repressors are shown by pointed and blunt arrows, respectively. Black lines show regulators that interact directly with their target promoters, whereas green lines show regulators that interact indirectly with their targets or those that have not been shown to bind their targets *in vitro*. This figure has been adapted from Kendall *et al*. (2010).

exerts a negative role in the *ler* gene expression (Sharma and Zuerner, 2004). It represses *E. coli* haemolysin gene expression, where it oligomerizes with H-NS protein before binding to a specific regulatory sequence in the target gene (Madrid *et al*., 2002). Nucleoid associated IHF protein is reported to bind DNA sequence located upstream of the *LEE1* functional promoter region, and positively regulates the *ler* expression. It is predicted that IHF-mediated activation of the *ler* expression involves direct interaction between IHF and RNAP (Friedberg *et al*., 1999). However, in many cases, IHF is reported to bend the DNA to facilitate interaction of RNAP with another activator bound upstream (Santero *et al*., 1992; Engelhorn and Geiselmann, 1998). In the case of EHEC or EPEC flagellar biosynthesis, IHF plays the opposite role functioning as a repressor (Yona-Nadler *et al*., 2003). Additionally, nucleoid associated Fis protein has been shown to activate expression of many LEE genes via upregulation of *ler* gene expression in EPEC (Goldberg *et al*., 2001), but its role in EHEC remains to be determined.

As mentioned earlier, quorum sensing plays an important role in LEE gene expression in both EPEC and EHEC. Quorum sensing-activated QseA protein acts as a positive regulator for *ler* expression and thus for other LEE genes (Sperandio *et al*., 1999; Sperandio *et al*., 2001; Sperandio *et al*., 2002; Sharp and Sperandio, 2007; Habdas *et al*., 2010). Moreover, it can also upregulate expression of the *grlRA* and many non-LEE encoded virulence factors in EHEC (Russell *et al*., 2007; Kendall *et al*., 2010). Usually, QseA responds to quorum sensing signals through the two-component system, QseC and QseB, which has been shown to regulate flagellar biosynthesis (Clarke and Sperandio, 2005). Moreover, another two-component system, QseEF, encoded by the *qseE and qseF* genes, is reported to activate the gene encoding $EspF_u$, a non-LEE encoded T3S effector that plays a crucial role in the rearrangement of host cell cytoskeleton in EHEC. However, QseEF do not have any regulatory role in LEE gene expression (Reading *et al*., 2007).

A number of host specific regulators influence expression of the LEE genes. In the EPEC 2348/69 strain, the EAF plasmid encodes a protein PerC that positively regulates *ler* expression (Porter *et al*., 2004). However, expression of the *per* genes, including *perC*, is downregulated by GadX, an activator of the glutamate decarboxylase genes involved in acid tolerance (Shin *et al*., 2001). EHEC lacks the EAF plasmid but possesses a number of *perC*

homologues (*pch*), encoded on horizontally acquired prophage sequences in the EHEC chromosome. Some of these proteins (PchA, PchB and PchC) positively regulate *ler* expression (Iyoda and Watanabe, 2004), though the precise mechanisms remain unclear. Genome wide analysis reveals that PchA acts as a global regulator, modulating expression of a number of genes including non-LEE genes associated with EHEC virulence (Abe *et al*., 2008). Another two regulatory factors, EtrA and EivF, encoded by a second cryptic T3S system, have been shown to be involved in negative regulation of the EHEC LEE genes (Zhang *et al*., 2004). Moreover, the RcsD–RcsC–RcsB His–Asp phosphorelay system affects the LEE gene expression in both positive and negative manners indirectly through controlling *ler* expression. RcsB-mediated indirect activation of *ler* expression needs the EHEC specific regulator, GrvA, which has been thought to activate the *ler* expression independently. However, expression of the *grvA* is positively regulated by RcsB. RcsB has been shown to downregulate expression of the EHEC specific *pch* genes and thus thought to exert a negative role in LEE gene expression by influencing *ler* expression (Tobe *et al*., 2005).

The ClpX protease, which degrades damaged and incomplete proteins, has also been found to increase expression of the LEE genes by downregulating expression of GrlR and interacting with a stationary-phase sigma factor (RpoS) (Iyoda and Watanabe, 2005). In contrast, ClpXP negatively regulates flagellar biosynthesis in both EHEC and *S. enterica* serovar Typhimurium (Tomoyasu *et al*., 2003; Iyoda and Watanabe, 2005). Also, the small non-coding RNA, DsrA, in conjunction with stress sigma factor (RpoS) activates *ler* expression in EHEC (Laaberki *et al*., 2006). RpoS has also been shown to activate the *LEE3* operon in the *E. coli* K-12 background (Sperandio *et al*., 1999).

The small signalling-molecule ppGpp, in conjunction with DksA, can modulate bacterial gene expression (see details in section 1.4.5). The stringent response system, triggered by starvation of nutrients or growth arrest, causes an increase in the ppGpp concentration in EHEC, which in association with DksA, upregulates the expression of LEE genes via direct activation of the virulence regulatory genes, *ler* and *pch* (Nakanishi *et al*., 2006). The YhiE and YhiF regulatory proteins, that belong to the of LuxR family, have been shown to act as negative regulators for LEE-encoded *espA*, *espB*, and *espD* expression at the RNA level. However, this regulation is not mediated through the *ler* gene (Tatsuno *et al*., 2003).

Moreover, overproduced SdiA, an *E. coli* homologue of quorum sensing regulator LuxR, interacts with homoserine lactone or HSL type molecules and downregulates expression of EspD and Tir in EHEC. However, EPEC does not have the same repression (Kanamaru *et al*., 2000).

In addition to Ler, two other regulators encoded within the LEE are GrlR (global regulator of LEE repressor) and GrlA (global regulator of LEE activator) (Deng *et al*., 2004). These two proteins appear to influence the expression of the LEE genes (Deng *et al*., 2004; Lio and Syu, 2004; Barba *et al*., 2005; Jimenez *et al*., 2010).

1.6.1 Activities of GrlR/GrlA as transcription regulators

GrlR and GrlA consist of 123- and 137-amino acids respectively and are encoded by a putative bicistronic operon *grlRA* that is located between the *etgA* gene and the *LEE2* operon (Figure 1.2). Yeast two-hybrid experiments revealed that the GrlR interacts with itself and with its bicistronic partner GrlA (Creasey *et al*., 2003). Both *grlR* and *grlA* are transcribed as a single unit and the transcription start site is located upstream of *grlR* gene (Mellies *et al*., 1999; Barba *et al*., 2005). Recently, Tauschek *et al*. (2010) reported that transcription of *grlA* gene is controlled by a σ^{70}-dependent promoter located in the intergenic region between *grlR* and *grlA* in *C. rodentium*. However, the intergenic DNA sequence between *grlR* and *grlA* genes is poorly conserved between *C. rodentium*, EHEC and EPEC.

Both GrlR and GrlA proteins are highly conserved among all AE pathogens (Elliott *et al*., 1998; Perna *et al*., 1998; Deng *et al*., 2001; Barba *et al*., 2005) but are not clearly related to any known protein. GrlA has 37% identity to *Salmonella* GrlA homologue (Sgh) and 23% identity to CaiF of *Enterobacteriaceae* (Deng *et al*., 2004). Moreover, it has similarity to several other putative transcriptional regulators including YheC, a putative regulator, encoded by R721 plasmid in *E. coli* K-12 (Pallen *et al*., 2005a). GrlR, on the other hand, has been reported to have only one homologue that is located next to the GrlA homologue in *S. bongori* (Pallen *et al*., 2005a). CaiF protein, which has a higher sequence similarity with GrlA, is encoded by the *caiF* gene, which lies downstream of the *cai* operon and in the opposite orientation and acts as a specific DNA binding regulatory protein for the activation of carnitine metabolism (Eichler *et al.*, 1996; Buchet *et al*., 1999). Deng *et al*. (2004) predicted

that GrlA has a helix-turn-helix DNA binding motif at the N-terminal domain where most of the similarity with the CaiF and the *Salmonella* GrlA homologue (Sgh) is found. The C-terminal region, on the other hand, is rich in basic side chains (nine arginines, seven lysines) and supposed to interact with its bicistronic partner, GrlR (Jobichen *et al*., 2007) (Figure 1.10).

Structural-based mutational analysis identified a surface exposed EDED motif in GrlR containing negatively charged residues Glu^{46}, Asp^{47}, Glu^{48} and Asp^{49} (^{46}EDED^{49}) that is important for the recognition of the positively charged segment of GrlA by GrlR. To confirm whether motif ^{46}EDED^{49} is crucial for the interaction between GrlR and GrlA, Jobichen *et al.* (2007) performed pull down assays. The experiment revealed that wild type GrlR bound to GrlA whilst GrlR derivatives containing alanine substitutions in the ^{46}EDED^{49} motif did not bind to GrlA. This suggested that negatively charged residues Glu^{46}, Asp^{47}, Glu^{48} and Asp^{49} are important for GrlR-GrlA interaction.

GrlR and GrlA act as negative and positive regulators, respectively for *ler* transcription (Deng *et al*., 2004; Lio and Syu, 2004; Barba *et al*., 2005; Iyoda *et al*., 2006; Huang and Syu, 2008; Saitoh *et al.*, 2008; Jimenez *et al*., 2010). The expression of *ler* is strongly activated by a positive regulatory loop formed by Ler and GrlA. GrlR apparently conveys its negative regulatory role through its interaction with GrlA and prevents the accumulation of Ler (Creasey *et al*., 2003; Iyoda and Watanabe, 2005; Jobichen *et al*., 2007).

Besides the regulation of the LEE genes, it is reported that GrlR/GrlA regulates transcription of at least six non-LEE-encoded effectors (Deng *et al*., 2004). GrlA activates the expression of the Ehx enterohaemolysin (Saitoh *et al.,* 2008). Ehx is encoded by the *ehxCABD* genes, which are carried on the large plasmid, pO17 of EHEC O157: Sakai. The expression of enterohaemolysin (Ehx) protein of EHEC O157 was examined by characterizing the *grlR* mutant phenotype. The GrlR deletion significantly induced Ehx activity of EHEC O157 on plates containing defibrinated sheep erythrocytes. Moreover, *ehx* expression was not induced by the *grlR grlA* double mutants but strikingly increased by over expression of *grlA*. These results suggested that GrlA acts as a positive regulator for *ehx* transcription in EHEC. GrlA plays a negative role in the expression of flagellar-genes in EHEC. Flagellar operon, *flhD* contains genes that are responsible for formation of flagella. In an effort to know whether

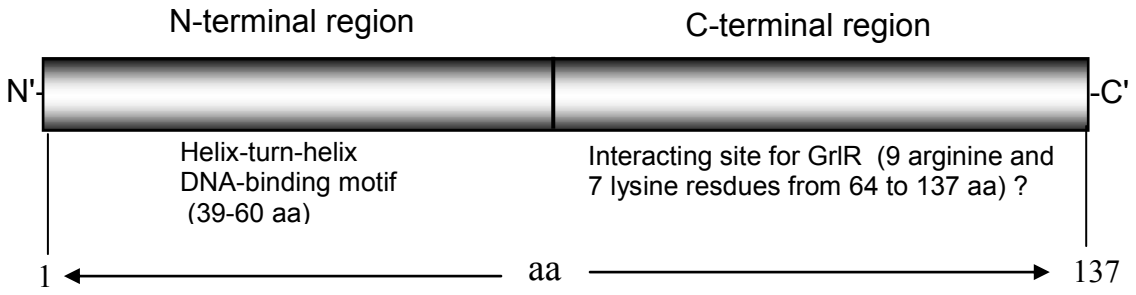


Figure 1.10. Schematic representation of regions of GrlA important for regulatory function.

The figure showing different regions of GrlA, has been drawn on the basis of the prediction of helix-turn-helix DNA binding motif in N-terminal region (Deng *et al.*, 2004), and GrlR interacting region located in the C-terminal region (Jobichen *et al*., 2007). In GrlA that consists of 137 amino acids (aa), predicted DNA binding motif is located between amino acid residues (aa) 39 and 60 in N-terminal region whilst acidic GrlR is thought to interact with basic residues (9 arginines and 7 lysines) of C-terminal region ranging from 64-137 aa .

regulatory genes for T3SS can regulate the expression of flagellar-genes in EHEC, Iyoda *et al.* (2006) assayed the transcription level of master flagellar-operon, *flhD* in the *grlR* mutant background and found reduced level of transcription of the operon. This suggested that GrlA functions as a repressor in the expression of flagellar-genes.

Though it is obvious that GrlA has a positive role in the expression of the LEE genes through inducing the *ler* expression, it is a matter of dispute whether GrlA activates *LEE1* by binding to the regulatory region independently or it needs a co-factor to bind the DNA or it activates indirectly by counteracting any repressor. Barba *et al* (2005) showed that GrlA activates transcription of the *C. rodentium LEE1* operon in both *C. rodentium* and *E. coli* K-12 backgrounds. This suggested that GrlA-dependent activation of *C. rodentium LEE1* transcription does not require any additional *C. rodentium*-specific genes. However, transcription of the EHEC *LEE1*, although activated by GrlA in an EHEC background (Iyoda *et al*., 2006) is not activated by GrlA in an *E. coli* K-12 background (Russell *et al*., 2007). These data suggested that EHEC *LEE1* activation by GrlA is indirect and requires an additional transcription factor, which is absent in *E. coli* K-12. In contrast to Russell *et al*. (2007), Huang and Syu (2008) showed that purified GrlA-Gst fusion protein binds the EHEC *LEE1* regulatory region. This result first suggested that GrlA could activate expression from the EHEC *LEE1* regulatory region independently. However, the study failed to identify the GrlA target in the *LEE1* regulatory region.

Previous studies, concerning the regulatory role of GrlA failed to find any obvious binding site in the regulatory region of putative target genes of GrlA. Moreover, alignment of the upstream regulatory regions of the putative target genes viz., *ler*, *ehxC* and *flhD* of GrlA failed to identify any clear common regulatory sequences. Together, this suggests that GrlA may activate putative target genes indirectly. It is important to note that expression of target genes of GrlA is negatively controlled by several regulators. For instance, under repressive conditions, both *ler* and *ehxC* are silenced by H-NS (Godessart *et al*., 1988; Nieto *et al.*, 1997; Umanski *et al*., 2002; Saitoh *et al*., 2008).

In order to determine the mechanism of GrlA-dependent activation of *LEE1* promoter, both Barba *et al*. (2005) and Jimenez *et al*. (2010) performed a deletion analysis of *LEE1* promoter

analysis and found that GrlA activates the expression from a *LEE1* promoter fragment that lacks the upstream sequence of the promoter -35 element and carries promoter elements with the downstream sequence. The basal activity of this fragment was found to be slightly higher in the *Δhns E. coli* K-12 background than wild type background but did not reach the levels in wild type *E. coli* K-12 carrying plasmid encoding GrlA. These data suggested that GrlA, in part, counteracts H-NS repression but it is also essential for the efficient activation of the *ler* promoter, even in the absence of H-NS. To test this further, they measured the promoter activity in *Δhns E. coli* K-12 background containing plasmid-expressing GrlA, and found that GrlA induces the promoter activity significantly compared to the activity observed in the *E. coli* K-12 *hns* mutant strain carrying only the vector. From these results, it was suggested that, in addition to H-NS, another factor could also partially repress *ler* expression, and thus GrlA could counteract the total repression exerted by more than one negative regulator. Alternatively, GrlA could counteract the H-NS-mediated repression but also promote the interaction of the RNAP with the *ler* promoter.

To nail this down further, Jimenez *et al*. (2010) performed *in vitro* binding assays with purified MBP-GrlA fusion protein and *LEE1* promoter fragment. Like Huang and Syu (2008), they reported that GrlA binds to the *LEE1* regulatory region. They also explored their *in vitro* binding system to check whether the predicted helix-turn-helix DNA binding motif located in the N-terminal region (Deng *et al*., 2004) is functional. They substituted the residues in the predicted DNA binding motif with alanine and found that some substitutions reduce the ability of GrlA to activate *LEE1* promoter *in vivo*, and subsequently the binding ability of GrlA to the *LEE1* promoter *in vitro*. However, like Huang and Syu (2008), this study failed to define GrlA binding target at the *LEE1* promoter.

Both the GrlA and GrlR regulatory proteins are required to optimize *LEE1* promoter activity (Jobichen *et al*., 2007). Generally, free GrlA increases *LEE1* promoter efficiency and thus induces *ler* transcription, which in turn activates the expression of all LEE genes, including the *grlRA* operon. On the other hand, to prevent the detrimental concentrations of Ler, GrlR interacts with GrlA and inhibits GrlA function, suggesting that GrlR acts as an anti-GrlA factor. However, an important question is open to all that whether one ore more additional factors are involved in modulating the anti-GrlA activity of GrlR in response to

environmental signals. Structural studies revealed that GrlR consists of a typical β-barrel fold with eight β-strands containing an internal hydrophobic cavity and a plug-like loop on one side of the barrel (Jobichen *et al.,* 2007). This highly hydrophobic β-barrel cavity is supposed to be involved in recognizing a small hydrophobic ligand. Jobichen *et al*. (2009) further explored the structure of GrlR, and determined that GrlR shares lipid-binding properties similar to that of lipocalin structural homologue, beta-lactoglobulin. Lipocalins are a broad family of lipid-binding proteins identified in both eukaryotes and Gram-negative bacteria.

1.6.2 Ler and its regulation

As mentioned earlier, the LEE genes are silenced by H-NS. After being expressed from the first gene of the *LEE1* operon under appropriate conditions, Ler negates the H-NS-mediated repression and thus activates expression of genes of *LEE2*-5 operons (Mellies *et al*., 1999; Elliott *et al*., 2000; Sperandio *et al*., 2000; Bustamante *et al*., 2001; Umanski *et al*., 2002; Haack *et al*., 2003). Moreover, it has been shown that Ler auto-regulates its own transcription (Berdichevsky *et al*., 2005; Yerushalmi *et al*., 2008). However, the precise mechanism of Ler function remains to be determined.

Ler is predicted to have a molecular weight of 15.1 kDa (Mellies *et al*., 2007). This protein shares 24% identity and 44% similarity to 15.5 kDa H-NS protein of *S. enterica* Typhimurium (Sperandio *et al*., 2000). Like H-NS as described in section 1.5.7, Ler consists of a C-terminal DNA-binding domain and an N-terminal coiled-coil oligomerization domain, which are connected by a linker region (Yerushalmi *et al*., 2008; Mellies *et al*., 2011) (Figure 1.11). Taken all together, it is believed that Ler is a member of H-NS family nucleoid-associated proteins. However, despite the similarity in domain structure, Ler protein differs from H-NS family members because all other members of H-NS protein family silence transcription of target genes, whereas Ler usually acts as a transcription activator. Recently, Mellies *et al*. (2011) revealed some distinct differences in domain structures between these two regulators. They found that length of the N-terminal coiled-coil domain of Ler is shorter than that of H-NS. The N-terminal region of the Ler contains two α-helices, whereas region of the H-NS has three α-helices. These structural differences lead to differences in biochemistry. For instance, in solution, Ler oligomerizes more readily than H-NS. It shows higher binding affinity to target promoters than H-NS. Moreover, the Ler-H-NS-DNA complex migrates to a

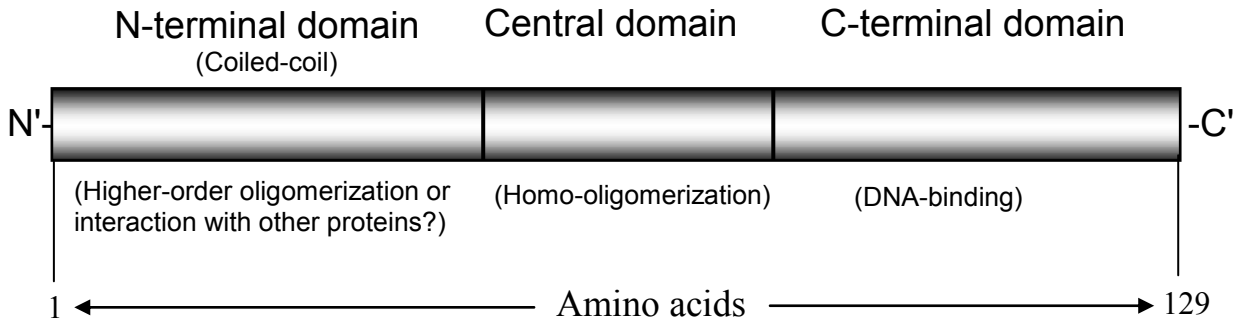


Figure 1.11. Schematic representation of different domains of Ler.

The figure shows three predicted domains of 129-peptides long Ler protein: N-terminal domain that functions either as a higher-order oligomerization domain or as a domain that interacts with other proteins, a central domain that functions as a homo-oligomerization domain and a C-terminal DNA binding domain. The three domains are essential for DNA binding and transcriptional regulation. Figure adapted from Yerushalmi *et al*. (2008).

distinct position in a band-shift assay suggesting that the Ler and H-NS get different shapes when they bind to the DNAs. Finally, a toroidal Ler-DNA structure that is distinct from the H-NS-DNA structure was found. This toroidal structure is believed to be associated with Ler-mediated activation of gene expression. Therefore, though the Ler belongs to the H-NS family, it functions differently (Mellies *et al*., 2011).

Base substitution in the putative oligomerization domain in the N-terminus of Ler eliminated the ability of the protein to bind to DNA and the ability to activate expression of a *LEE2-lacZ* fusion, suggesting that Ler forms oligomers in order to activate transcription (Sperandio *et al*., 2000). Moreover, biochemical assays revealed that Ler protects extended regions of *LEE1*, *LEE2* and *LEE5* operon regulatory sequences supporting that Ler binds DNA in multimerized form (Sperandio *et al*., 2000; Haack *et al*., 2003; Berdichevsky *et al*., 2005).

The mechanism of Ler-mediated activation of different LEE genes is still open to question. To date, there is no report of identification of the determinants at target promoters for Ler. In all cases both genetic and biochemical assays revealed that Ler interacts with an extended region of highly A/T rich regulatory sequences of the *LEE* operons which are also targets for H-NS (Sperandio *et al*., 2000; Haack *et al*., 2003; Berdichevsky *et al*., 2005; Mellies *et al*., 2011). Ler-dependent hypersensitive sites in DNase I footprinting assays appear in regions known to be protected by H-NS. Therefore, the mechanism of Ler-mediated regulation is thought to involve the displacement of H-NS by Ler like antagonism of H-NS repression by SlyA of haemolysin, *hlyE* in *E. coli* (Stoebel *et al*., 2008) (Figure 1.12).

Ler acts as a global regulator since it can regulate genes outside of the LEE pathogenicity island (Elliott *et al*., 2000; Deng *et al*., 2004; Abe *et al*., 2008). Most of the non-LEE-encoded genes, reported to be controlled by the Ler, are associated with the attaching and effacing phenotype. For instance, the Ler regulon includes EspFu/TccP, which is required for actin polymerization at the cytoplasmic domain of Tir in EHEC (Campellone *et al*., 2004) and NleA/EspI, which is necessary for *C. rodentium* to colonise the mouse colon (Mundy *et al*., 2004). Moreover, Ler is believed to control some genes found in non-pathogenic *E. coli* strain K-12 (Abe *et al*., 2008). However, they reported that the Ler has limited number of target genes outside of the LEE region compared with another activator of the LEE genes, PchA in

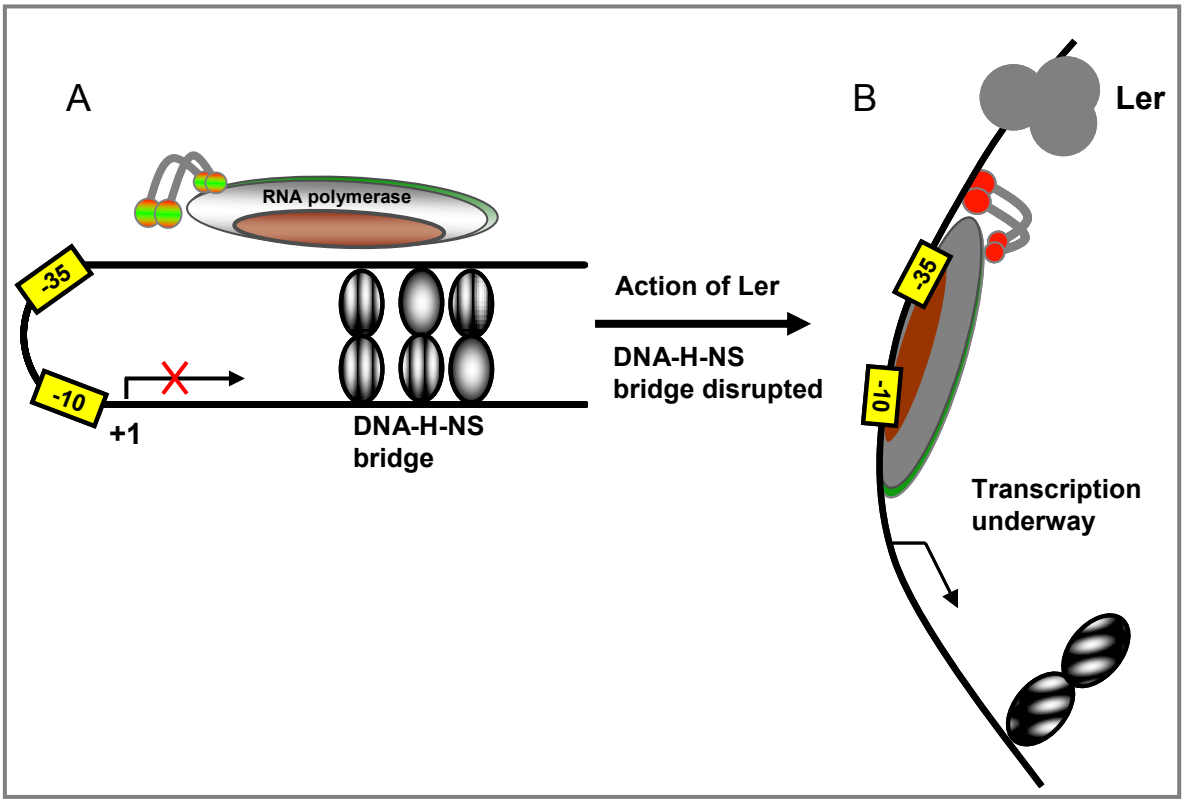


Figure 1.12. Schematic representation of action of Ler on a promoter that is repressed by H-NS.

A. H-NS binds DNA sequences both upstream and downstream of the target promoter and forms a bridge, which prevents transcription from the promoter.
B. Ler overcomes the H-NS mediated repression by displacing or antagonising the H-NS and permits transcription from the promoter. The figure has been adapted from Stoebel *et al*. (2008).

EHEC. Therefore, the action of Ler is confined mainly on the LEE genes. This can be explained by the closed proximity of the *ler* and its interacting counterparts in the LEE region. Recently, it has been reported that interacting proteins synthesized in the same vicinity, can facilitate interaction (Montero Llopis *et al*., 2010).

It is clear that a number of regulatory inputs control the expression of the Ler, which in turn enhances expression of genes associated with the T3S apparatus, translocation secretors and effector molecules. However, when assembly of T3SSs is complete, Ler is no longer necessary and hence, it needs to be controlled. The action of Ler is found to be controlled in two ways. As mentioned earlier, Ler represses its own transcription (Berdichevsky *et al*., 2005; Yerushalmi *et al*., 2008). Moreover, the action of Ler appears to be regulated post-transcriptionally by a T3S secretion chaperone (Tsai *et al*., 2006; Younis *et al*., 2010).

It has been shown that SepL and its binding partner SepD control the secretion of EspADB molecules that form a translocon through which effectors are delivered to host cell (O'Connell *et al*., 2004; Deng *et al*., 2005). Therefore, these two proteins largely determine the secretion specificity in the system (reviewed by Tree *et al*., 2009; Bhatt *et al*., 2011). Recently, Younis *et al*. (2010) reported that the LEE-encoded protein, CesL functions as a class 1 chaperone (chaperons for effectors) for the SepL protein. Additionally, as a T3S effector, SepL is reported to have a secretion signal at its N-terminus. Tsai *et al*. (2006) showed that the CesL (L0036 or Mpc in their report) binds the Ler protein and down-regulates its activation capability. Therefore, it is believed that at post-transcriptional level, secretion of SepL releases its chaperone CesL, which in turn binds the Ler and restricts its activity. However, this partner switching approach that couples protein secretion to gene expression has been reported for flagellar and many other non-flagellar T3S systems (Hughes *et al*., 1993; Darwin and Miller, 2001; Francis *et al*., 2001; Parsot *et al*., 2005; Urbanowski *et al*., 2005; Brutinel and Yahr, 2008).

1.7 Aims and an outline of the project

The overarching aim of this work was to understand how the expression of the LEE genes is triggered. Since LEE expression is coordinated by the Ler protein, most of the work has focussed on the regulatory region of the *LEE1* transcription unit. The approach that was taken

was to study this region in an *E. coli* K-12 genetic background with a view to building up a step-by-step picture of its regulation.

The first aim was to define the key functional promoter elements of the *LEE1* regulatory region and to do this, both genetic and biochemistry methods were exploited. Having done this, as described in the first part of Chapter 3, it was noticed that the *LEE1* leader region is unusually long. This led to further aim that was to investigate the consequence of the long *LEE1* leader sequence. Genetic analyses revealed a mini-gene in the leader region and its function. This part of the work is described in Chapter 5.

During the investigation of the key functional *LEE1* promoter elements, a secondary cryptic promoter, which overlapped with the main promoter, was discovered. Having observed this, the aim was set to identify the functional elements of this secondary promoter and to determine its role. This part of the work is described in Chapter 4 and has been published (Appendix II).

Although many transcriptional regulators are known to affect expression of the *LEE1* operon, their effects at the *LEE1* promoter have not been studied in detail and binding sites have not been identified. Hence, a further aim of my work, having defined the *LEE1* promoter elements, was to identify binding targets for different transcription factors. The last part of Chapter 3 describes an experimental work to investigate the action of GrlA. To do this, the reported GrlA-mediated activation of the promoter was reconstituted in the *E. coli* K-12 system. Different genetic analyses revealed that GrlA binds to the *LEE1* promoter between its -35 and -10 elements and compensate non-optimal spacer length. This part of the work has been published (Appendix I).

A final aim was to further understanding of how Ler activity is regulated and thus, the work, described in Chapter 6, focuses on transcriptional autoregulation of *ler* expression by Ler protein. In parallel, effects of Ler at the *LEE2* regulatory region were measured.

Taken together, the results in this thesis advance our understanding of promoter architecture and regulation in EHEC.

Chapter 2

Materials and Methods

2.1 Suppliers

Unless otherwise stated, all chemicals were purchased from Sigma-Aldrich, BDH or Fisher Scientific. All radionucleotides were purchased from Perkin Elmer. Oligodeoxyribo -nucleotides were supplied by Alta Bioscience, University of Birmingham. Restriction endonucleases, Calf alkaline phosphatase, T4 DNA ligase and T4 polynucleotide kinase were purchased from New England Biolabs. Biotaq DNA polymerases and dNTP mix were supplied by Bioline. All enzymes were used according to the manufacturer's instructions, and in the buffers provided by the suppliers. Glycogen was purchased from Invitrogen. Holo RNAP-σ^{70} was purchased from Epicentre Biotechnologies (Madison, Wisconsin).

2.2 Buffers, solutions and reagents

Solutions for use in bacterial growth or DNA manipulations were autoclaved for 20 mins at 120°C and 15 psi, or filter-sterilised using 0.2 μm filters.

2.2.1 Gel electrophoresis of DNA and proteins

General

DNA loading dye: 0.025% Bromophenol blue; 0.025% Xylene cyanol FF; 20% Glycerol; 10 mM Tris, pH 7.5; 1mM EDTA

Ethidium bromide: 10 mg/ml (Bio-Rad)

DNA markers: 100 bp and 1 kb DNA ladders were purchased from New England Biolabs, and were diluted 6 fold in DNA loading dye

Protein molecular weight markers: SeeBlue® Plus2 Protein pre-stained standard consisting of 10 protein bands in the range of 4-250 kDa was purchased from Invitrogen

Agarose gel electrophoresis

50 x TAE Buffer: 2 M Tris acetate, 100 mM Na_2EDTA (National Diagnostics). Diluted to 1 x for use as running buffer

Agarose solutions: 0.8% and 1.3% in 1 x TAE buffer. Heated to 100°C for 2-3 mins in a microwave to dissolve agarose

Polyacrylamide gel electrophoresis

5 x TBE buffer: 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA (National Diagnostics). Diluted to 1 x or 0.5 x for use as running buffer

Stock acrylamide solution: 30% (w/v) acrylamide, 0.8% bisacrylamide stock solution (ProtoGel) (National Diagnostics)

TEMED: N, N, N', N'-Tetramethylethylenediamine

7.5% Acrylamide working solution: 125 ml 30% (w/v) acrylamide, 0.8% bisacrylamide stock solution; 100 ml 5 x TBE; 20 ml glycerol; made up to 500 ml with distilled water

Sequencing acrylamide gel solution: SequaGel Sequencing System was prepared according to the manufacturer's (National Diagnostics) instructions

Fixing solution: 10% (v/v) methanol, 10% (v/v) acetic acid

Denaturing gel electrophoresis of proteins

Gels: NuPAGE Novex 4-12 % Bis-Tris gels (Invitrogen)

SDS PAGE loading buffer: 2 g SDS, 20 ml glycerol, 5 mg bromophenol blue, made up to 92 ml using 0.1 x stacking gel buffer. Prior to use, 87 µl β-mercaptoethanol was added to a 1 ml aliquot

Stacking gel buffer: 15.15 g Tris, 10 ml 10 % SDS, 500 µl TEMED, pH 6.8. Made up to 100 ml in distilled water

Coomassie blue stain solution: 50 % (v/v) methanol, 10 % (v/v) acetic acid, 2g/l Coomassie brilliant blue

Fast destain solution: 40 % (v/v) methanol, 10 % (v/v) acetic acid

Shrink solution: 48 % (v/v) methanol, 2% (v/v) glycerol

2.2.2 Extraction and purification of DNA fragments

Phenol/Chloroform/Isoamyl alcohol: phenol/chloroform/isoamyl alcohol (composition 25/24/1 v/v), pH 8.0 purchased from Fisher scientific

TEN buffer: 60 mM Tris (pH 8.0), 5 mM EDTA, 50 mM NaCl

TE buffer: 10 mM Tris (pH 8.0), 0.5 mM EDTA

2.2.3 DNA transformation in *E. coli*

Calcium chloride treated competent cells

Calcium chloride: 100 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5 (autoclaved)

Freeze-thaw buffer: 100 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5, 15% glycerol (autoclaved)

Rubidium chloride treated competent cells

TFB1 buffer: 30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100mM RbCl, 15% glycerol, pH 5.8 with 1 M acetic acid, filter-sterilized (0.2μM) and stored at room temperature

TFB2 buffer: 10 mM PIPES, pH 6.5, 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol, pH 6.5 with 1 M KOH, filter-sterilized (0.2μM) and stored at room temperature

2.2.4 DNA sampling

Extraction buffer: 50 mM Tris–HCl pH 7.5, 100 mM NaCl, 10% sucrose, 2 mM $MgCl_2$, 0.1% (v/v) Triton X-100, 200 μg/ml phenylmethylsulfonyl fluoride, 4 μg/ml pepstatin
Protease inhibitor cocktail tablet (Roche diagnostics)
RNase A (300 μg/ml)
Lysozyme (400 μg/ml)

Wash buffer: 20 mM HEPES pH 7.4, 100 mM NaCl, and 0.1% (v/v) Tween 20

Elution buffer: 0.3 M NH_4OH, 0.5 M EDTA, pH 8.0

2.2.5 Pull down assays

Extraction buffer: 20 mM HEPES [pH 7.4], 150 mM NaCl, 2 mM MgCl2, 0.1% Tween 20, 200 μg/ml phenylmethylsulfonyl fluoride, 4 μg/ml pepstatin, protease inhibitor cocktail (1 tablet; Roche Diagnostics)
DNase I (20 μg/ml)
RNase A (300 μg/ml)
Lysozyme (200 μg/ml)
Dynabeads (M270-epoxy; Dynal) coated with rabbit immunoglobulin G

Wash buffer: 20 mM HEPES pH 7.4, 150 mM NaCl, 0.1% Tween 20

Elution buffer: 0.5 M NH_4OH, 0.5 M EDTA pH 8.0

2.2.6 $KMnO_4$ footprinting

[γ^{32}P]-ATP: 7000 Ci/mmol, 100 mCi/ml (Perkin Elmer)

Sephadex G-50 suspension: 5 g Sephadex G-50 (Pharmacia Biotech) autoclaved in 100 ml TE buffer. Washed three times in 150 ml TE buffer and finally suspended in 50 ml TE buffer. Stored at 4ºC

Denaturing gel loading buffer: 40% deionised formamide, 5 M urea, 5 mM sodium hydroxide, 1 mM EDTA, 0.025% bromophenol blue, 0.025% xylene cyanole FF

10x HEPES/glutamate buffer: 200 mM HEPES pH 8.0, 50 mM $MgCl_2$, 500 mM potassium glutamate, 10 mM DTT

$KMnO_4$ solution: 31 mg $KMnO_4$ per 1 ml solution with distilled water

Stop solution: 3 M ammonium acetate, 0.1 mM EDTA, 1.5 M β-mercaptoethanol

2.2.7 *In vitro* transcription assays

[α^{32} P]-UTP: 300 Ci/mmol, 10 mCi/ml (Perkin Elmer)

10 x transcription buffer: 400 mM Tris Acetate pH 7.9, 100 mM $MgCl_2$, 10 mM DTT, 1 M KCl

10 mg/ml BSA

Transcription stop solution/loading buffer: 95% (v/v) deionised formamide, 10 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF

2.2.8 β-galactosidase assays

Z-buffer: 0.75 g KCl, 0.25 g $MgSO_4.7H_2O$, 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_2O$ and 2.70 ml β-mercaptoethanol, made up to 1.0 l with distilled water and autoclaved

Note that β-mercaptoethanol was added after autoclaving, immediately prior to use

ONPG: ortho-nitrophenol-β-D-galactopyranoside dissolved in Z-buffer (0.8 g/l), prepared fresh prior to use

Sodium carbonate: 1 M Na_2CO_3

Toluene

Sodium deoxycholate: 1% (w/v) solution in distilled water

2.3 Bacterial growth media

2.3.1 Liquid media

All liquid media were made up in distilled water, sterilised by autoclaving for 20 mins at 120°C and 15 psi, and stored at room temperature.

LB medium: 20 g/l tryptone, 10 g/l yeast extract, 10 g/l NaCl, made up in distilled water.

10 x M9 salts: 60 g/l Na_2HPO_4, 30 g/l KH_2PO_4, 5 g/l NaCl, 10 g/l NH_4Cl, 10 mg/l Biotin, 10 mg/l thiamine, pH 7.4

M9 salts or minimal medium: The constituents of M9 salts medium were made up individually, autoclaved, and fresh medium was prepared aseptically when required.
For 100 ml medium: 10 ml 10 x M9 salts, 200 µl 1 M $MgSO_4$, 100 µl 0.1 M $CaCl_2$, 90 ml sterile distilled water. Supplemented where required with 10 mM or 50 mM L-lysine.

DMEM medium: Dulbecco's modified Eagle's medium, pH 7.4, adjusted with HEPES buffer, purchased from Sigma-Aldrich

SOC medium: A rich media used in the recovery step of *E. coli* competent cell transformations. Use of SOC maximizes the transformation efficiency of competent cells. Ready-made solution was purchased from Sigma-Aldrich.

2.3.2 Solid media

All solid media were made up in distilled water and sterilised by autoclaving for 20 mins at 120°C and 15 psi. Media were cooled to ~50°C before addition of antibiotics or other supplements. Agar plates were poured under sterile conditions (approximately 25 ml agar per Petri dish) and stored at 4°C.

Nutrient agar: 23 g/l nutrient agar (Difco)
Lactose MacConkey agar: 50 g/l MacConkey lactose agar (Difco)

2.3.3 Antibiotics

All stock solutions were made up as indicated below and sterile filtered through 0.2 µm syringe filters.

Ampicillin: 40 mg/ml in sterile distilled water (stored at -20°C)
Tetracycline: 10 mg/ml in ethanol (stored at -20°C)
Kanamycin: 50 mg/ml in sterile distilled water (stored at -20°C)
Chloramphenicol: 25 mg/ml in ethanol (stored at -20°C)

To select for strains carrying resistance genes, antibiotics were added to liquid or solid media after autoclaving at the following final concentrations: 80 μg ml^{-1} ampicillin, 35 μg ml^{-1} chloramphenicol, 50 μg ml^{-1} kanamycin, and 35 μg ml^{-1} tetracycline.

2.4 Bacterial strains and plasmids

2.4.1 Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 2.1. The bacterial strains were restreaked before use onto freshly prepared agar plates and incubated at 37°C overnight. The strains were stored at 4°C for up to one month. Glycerol stocks were made from a culture and stored at -80°C. Overnight cultures were grown by inoculating a single fresh colony into 5-10 ml LB medium supplemented with antibiotics as appropriate and incubating at 37°C with shaking for 16-18 hrs. When sub-culturing was necessary, required volume of medium was inoculated with 0.01-0.02 volume overnight culture. Growth of cultures was monitored by measuring the optical density at 650 nm using Helios Gamma Spectrophotometer (Thermo Fisher Scientific Inc).

2.4.2 Plasmids

Plasmids used in this study are listed in Table 2.2. Plasmid maps are shown in the Figures 2.1 to 2.7. In brief, plasmid pRW224 carrying different fragments of *LEE1* and *LEE2* regulatory sequences as transcription fusions to *lacZ* was used to quantify the promoter activities by measuring the amount of β-galactosidase. Plasmid pJW15Δ100 was used to express genes of interest under the control of *melR* promoter. Plasmid pACYC184 was used to clone target genes under their own regulatory sequences. Plasmid pRW902 was used to clone DNA sequences of interest to facilitate *in vivo* identification of bound protein on it following the DNA sampling method. Plasmid pRW225 was used to clone the DNA sequence of interest as translational fusions to *lacZ*. Plasmid pSR was used in *in vitro* potassium permanganate analysis and transcription assays.

2.5 Gel electrophoresis

2.5.1 Agarose gel electrophoresis of DNA

Agarose gels were used to analyse and purify DNA fragments greater than 500 bp in length.

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

Strains	Description	References
M182	*E. coli* K-12 strain, *lacX74 galK galU strA*	Casadaban and Cohen (1980)
MG1655 *Δhns*	*E. coli* K-12 MG1655 *Δhns rpsL*	Grainger *et al*. (2008)
Sakai 813 Δ*Stx*	A derivative of EHEC strain O157:H7 Sakai, Δ*Stx1 and* Δ*Stx2*	Given by Chihiro Sasakawa
EDL933	A derivative of EHEC strain O157:H7 EDL933, TUV93-0, Δ*Stx1 and* Δ*Stx2*	Given by Arthur Donohue -Rolfe
MG1655 *lacI* 3xFLAG	*E. coli* K-12 MG1655 3xFLAG-tagged *lacI*	Butala *et al.* (2009)
Sakai *grlR*	A derivative of EHEC Sakai Δ*Stx*, the alanine codon 37 in *grlR* changed to TAG amber stop codon	Zhang and Pallen (unpublished)[1]
Sakai *grlA*	A derivative of EHEC Sakai Δ*Stx*, the alanine codon 55 in *grlA* changed to TAG amber stop codon	Zhang and Pallen (unpublished)[1]
EDL933 Ler-KO	A derivative of EHEC strain EDL933, Δ*ler*	Bingle and Pallen (unpublished)[2]
Sakai *rpoC*::*4PrA*	A derivative of EHEC O157 Sakai Δ*Stx, rpoC*::*4PrA*	Lee *et al*. (2008)

[1]The Sakai strain mutants (both Sakai *grlR* an Sakai *grlA*) were constructed by Lihong Zhang in Mark Pallen's laboratory by replacing alanine codon 55 in *grlA* and alanine codon 37 in *grlR* with a TAG amber stop codon, using the gene gorging method (Herring *et al.*, 2003). In each case, a BfaI restriction site (CTAG) was simultaneously created by mutating the preceding base to a C. This site was exploited to identify mutant colonies that were then confirmed by DNA sequencing. Both mutants were checked to contain no other gene deletion by genomic hybridization to a microarray.

[2]The EDL933 *ler* KO strain was constructed by Lewis Bingle in Mark Pallen's laboratory. An inframe (i.e. non-polar) deletion of the *ler* gene was made using the Gene Doctoring system (Lee *et al*., 2009) first by replacement of the *ler* coding sequence with the kanamycin resistance gene cassette and then removal of this cassette via flanking FLP recombination sites. The resultant *ler*-KO locus encodes the first and last 9 amino acids of Ler sandwiching a central "scar region" derived from the FLP recombinase sites encoding 29 (non-Ler) amino acids. The strain was checked for unwanted secondary deletions by microarray comparative genomic hybridisation.

Table 2.2: Plasmids used in this study (continued on pages 65- 67)

Plasmids	Description	References
pRW224-U9	Derivative of pRW50, which allows cloning of promoter fragments as transcription or translation fusions to *lac*Z carrying pUC9 linker between the EcoRI and HindIII sites (Tet^R) (Figure 2.1)	Islam *et al*. (2011)
LEE10-568/pRW224	Derivative of pRW224 carrying an EcoRI-HindIII *LEE1* regulatory region fragment (-568 to -19, numbered with respect to the translation start site of *ler*) as a transcription fusion to *lacZ* (Figure 2.10, 2.11)	This study
LEE10-315/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-315 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-275/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-275 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-235/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-235 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	
LEE10-215/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-215 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-203/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-203 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-195/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-195 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-155/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-155 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE10-115/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-115 to -19) as a transcription fusion to *lacZ* (Figure 2.10, 2.11)	This study
LEE10-75/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-75 to -19) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study

Table 2.2: Plasmids used in this study (continued)

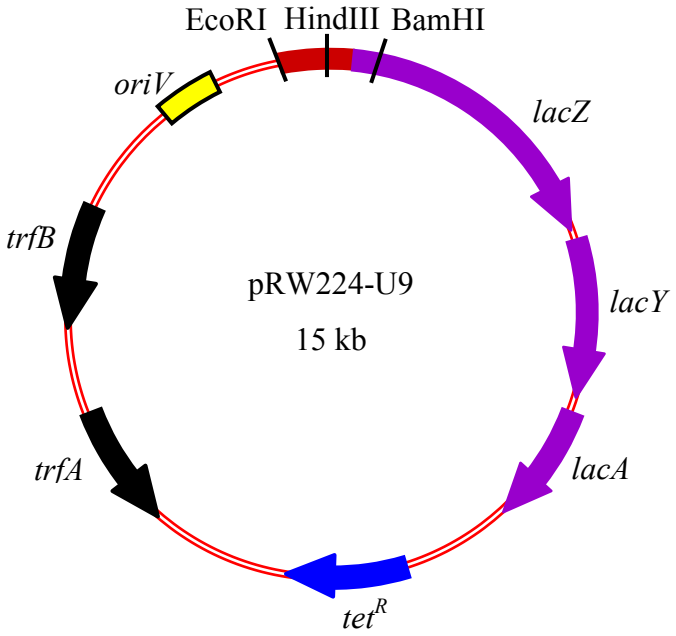
Plasmids	Description	References
LEE20-568/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-568 to -158) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE20-315/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-315 to -158) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE20-275/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-275 to -158) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE20-203/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-203 to -158) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE30-275/pRW224	Derivative of LEE10-568/pRW224 carrying an EcoRI-HindIII fragment (-275 to -114) as a transcription fusion to *lac*Z (Figure 2.10, 2.11)	This study
LEE2-220/pRW224	Derivative of pRW224 carrying an EcoRI-HindIII *LEE2* regulatory region fragment (-220 to +109, numbered with respect to the transcript start site of the *LEE2* promoter) as a transcription fusion to *lac*Z (Figure 2.12, 2.13)	This study
LEE2-125/pRW224	Derivative of LEE2-220/pRW224 carrying an EcoRI-HindIII fragment (-125 to + 109) as a transcription fusion to *lacZ* (Figure 2.12, 2.13)	This study
LEE2-100/pRW224	Derivative of LEE2-220/pRW224 carrying an EcoRI-HindIII fragment (-100 to +109) as a transcription fusion to *lacZ* (Figure 2.12, 2.13)	This study
pSR	High copy number pBR322 derivative, used for cloning EcoRI-HindIII promoter fragments upstream of the λ *oop* terminator (Amp^R) (Figure 2.2)	Kolb *et al*., (1995)
LEE20-275/pSR	Derivative of pSR containing LEE20-275 fragment between EcoRI-HindIII sites (Figure 2.14, 2.15)	This study

Table 2.2: Plasmids used in this study (continued)

Plasmids	Description	References
LEE20-275 98C/pSR	Derivative of pSR containing LEE20-275 98C fragment between EcoRI-HindIII sites (Figure 2.14, 2.15)	This study
LEE20-275 98C 64G/pSR	Derivative of pSR containing LEE20-275 98C 64G fragment between EcoRI-HindIII sites (Figure 2.14, 2.15)	This study
pACYC184	Cloning vector constructed by ligating restriction fragments from pSC101, Tn9 and p15A. Used for cloning HindIII-SalI DNA fragments (Tet^R and Cm^R) (Figure 2.3)	Chang and Cohen (1978)
pACYC184ΔHN	Derivative of pACYC184 in which HindIII-NruI fragment has been deleted, chloramphenicol resistance and tetracycline sensitive	Mitchell *et al*. (2007)
pSI01	Derivative of pACYC184 carrying *grlR+A+* genes cloned between HindIII and SalI sites (Figure 2.16, 2.17)	This study
pSI02	Derivative of pSI01 in which there is an in-frame deletion of *grlR* gene, carrying *grlR-A+* cloned between HindIII and SalI sites (Figure 2.16, 2.17)	This study
pSI03	Derivative of pSI01 in which *grlA* gene was removed, carrying *grlR+A-* between HindIII and SalI sites (Figure 2.16, 2.17)	This study
pRW902	Derivative of pRW50. DNA sequence of interest was cloned into the EcoRI-HindIII sites, immediate downstream of 5 LacI operators with two flanking target sites for SceI. Used to pull down the DNA sequences together with bound protein on it during DNA sampling (Tet^R) (Figure 2.4)	Butala *et al*. (2009)
LEE20-275/pRW902	Derivative of pRW902 carrying LEE20-275 promoter fragment (Figure 2.18) between EcoRI and HindIII sites	This study
LEE30-275/pRW902	Derivative of pRW902 carrying LEE30-275 promoter fragment (Figure 2.18) between EcoRI and HindIII sites	This study
LEE30-275 151T/ pRW902	Derivative of pRW902 carrying LEE30-275 151T promoter fragment (Figure 2.18) between EcoRI and HindIII sites	This study

Table 2.2: Plasmids used in this study (continued)

Plasmids	Description	References
pJW15Δ100	Derivative of pJW15 (Williams *et al*., 1994) carrying *melR* promoter with truncated *melR* gene. Used to clone promoterless DNA fragments into NsiI-HindIII sites under the control of *melR* promoter (Amp^R). (Figure 2.5)	Wade *et al*. (2000)
pSI04	Derivative of pJW15Δ100 carrying *ler* gene as a NsiI-HindIII fragment under the control of *melR* promoter (Figure 2.19, 2.20)	This study
pSI05	Derivative of pJW15Δ100 carrying *pchC* gene as a NsiI-HindIII fragment under the control of *melR* promoter (Figure 2.19, 2.20)	This study
pSI06	Derivative of pJW15Δ100 carrying *cesD* gene as a NsiI-HindIII fragment under the control of *melR* promoter (Figure 2.19, 2.20)	This study
pSI07	Derivative of pJW15Δ100 carrying *cesT* gene as a NsiI-HindIII fragment under the control of *melR* promoter (Figure 2.19, 2.20)	This study
pSI08	Derivative of pJW15Δ100 carrying *cesL* as a NsiI-HindIII fragment under the control of *melR* promoter (Figure 2.19, 2.20)	This study
pRW225	Derivative of pRW224, which allows cloning of promoter fragment in EcoRI-HindIII or EcoRI-BamHI as translational fusion to *lacZ* (Tet^R) (Figure 2.6)	Kerry Hollands (unpublished)
LEE30-225/pRW225	Derivative of pRW225 carrying LEE30-275 promoter fragment between EcoRI and BamHI sites (Figure 2.21)	This study
LEE151/pRW225	Derivative of pRW225 carrying LEE151 fragment (Figure 2.22) between EcoRI and HindIII sites	This study
LEE150/pRW225	Derivative of pRW225 carrying LEE150 fragment (Figure 2.22) between EcoRI and HindIII sites	This study
LEE150-1/pRW225	Derivative of pRW225/LEE150 in which small ORF was replaced with KpnI site (Figure 2.23)	This study
pACBSR-DL1	Mutagenesis plasmid for use in DNA sampling (Cm^R) (Figure 2.7)	Butala *et al*. (2009)



EcoRI pUC9 linker HindIII

ATTCGCGAGAGCCTTGAGTCCACGCTAGATCTGAATTCCCGGGGATCCGTCGACCTGCAGCCAAGCTT

SD sequence BamHI *lacZ* →

AATGGAGCGAATTATGAGAGTTCTGGTTACCGCCAAGCTCCGGGATCCCGTCGTTTTACAACGTCGTG

Translation initiation codon

ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGC

Figure 2.1. Plasmid map of the *lac* fusion vector pRW224-U9.

pRW224-U9 carries pUC9 linker between EcoRI and HindIII sites. It is a derivative of plasmid pRW50 (Lodge *et al*., 1992) from which *trpBA* has been removed and whose cloning site has been modified such that promoter fragments can be cloned as either transcription or translation fusions to *lacZYA*. The sequence of the cloning site is shown in the lower part. EcoRI-HindIII promoter fragments are cloned as transcription fusions, where translation of *lacZ* is directed by a translation initiation region (TIR), i.e., Shine-Dalgarno (SD) sequence and start codon. EcoRI-BamHI fragments are cloned as translation fusions to *lacZ*, and in this case, *lacZ* translation is initiated by the TIR on the cloned fragment.

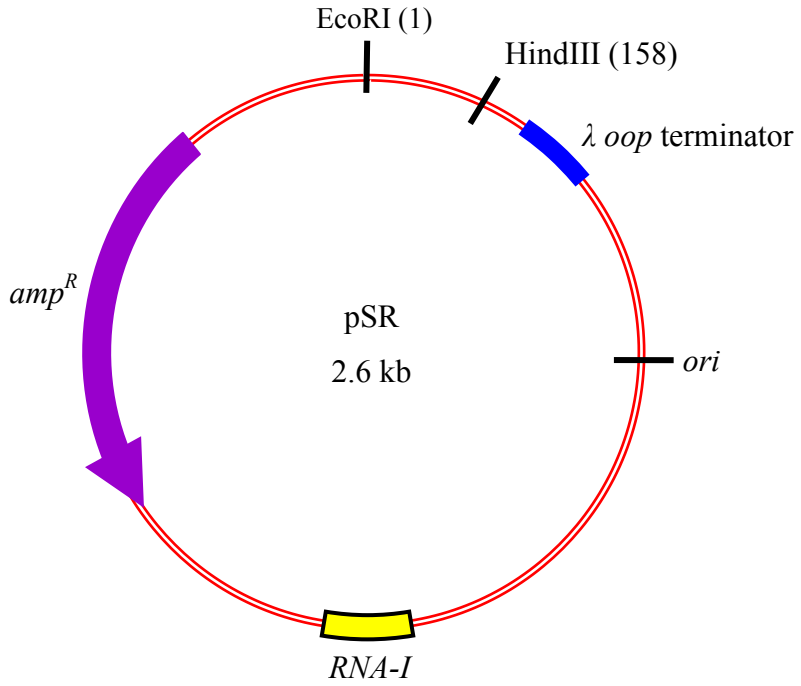


Figure 2.2. Plasmid map of pSR.

pSR-TB10 carries an EcoRI-HindIII *melR* promoter fragment. EcoRI-HindIII promoter fragments are cloned upstream of the λ *oop* terminator. pSR is used in *in vitro* transcription reactions, in which transcription initiates at the cloned promoter and terminates at the λ terminator, to produce a discrete transcript of a defined length. Also shown are the *RNA-I* gene, which produces a control transcript during *in vitro* transcription, the ampicillin resistance gene (*amp*^{R}), which produce ampicillin resistance marker (Amp^R), and the origin of replication (*ori*).

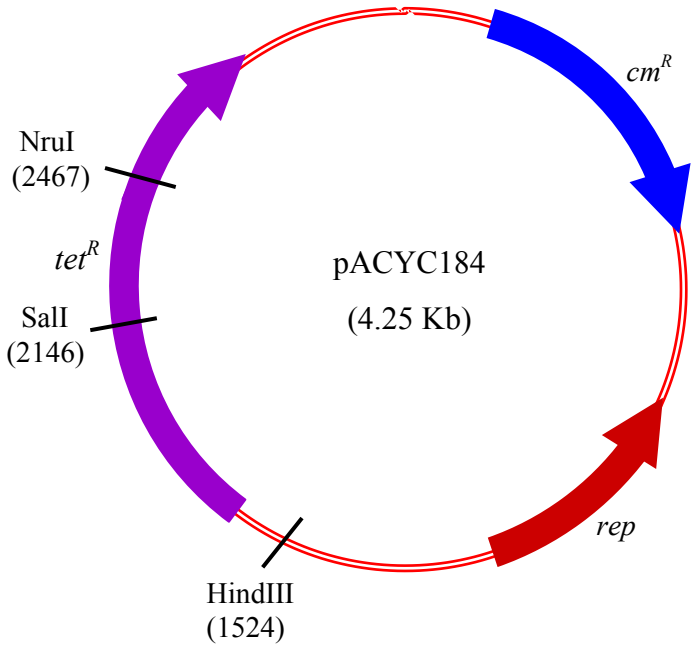


Figure 2.3. Plasmid map of pACYC184.

pACYC184 is a small and low copy number *E. coli* plasmid. The cloning vector contains the replicon *rep* responsible for the replication of the plasmid. Also shown are tetracycline resistance gene (tet^R), and chloramphenicol resistance gene (cm^R). The plasmid contains single or multi-cleavage sites for a number of commonly employed site-specific endonucleases.

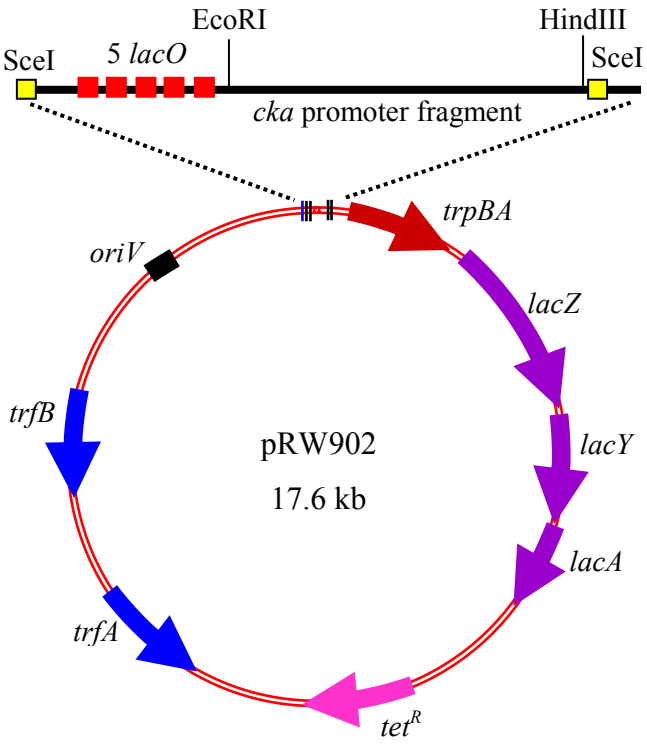


Figure 2.4. Plasmid map of pRW902.

pRW902 carries an EcoRI-HindIII *cka* promoter region. EcoRI-HindIII promoter fragments are cloned immediately downstream of five *lac* operator sites with two flanking 18-bp target sites for the yeast meganuclease I-SceI. Also shown are reporter gene *lacZ*, the tetracycline resistance gene (tet^R) and the origin of replication (*oriV*).

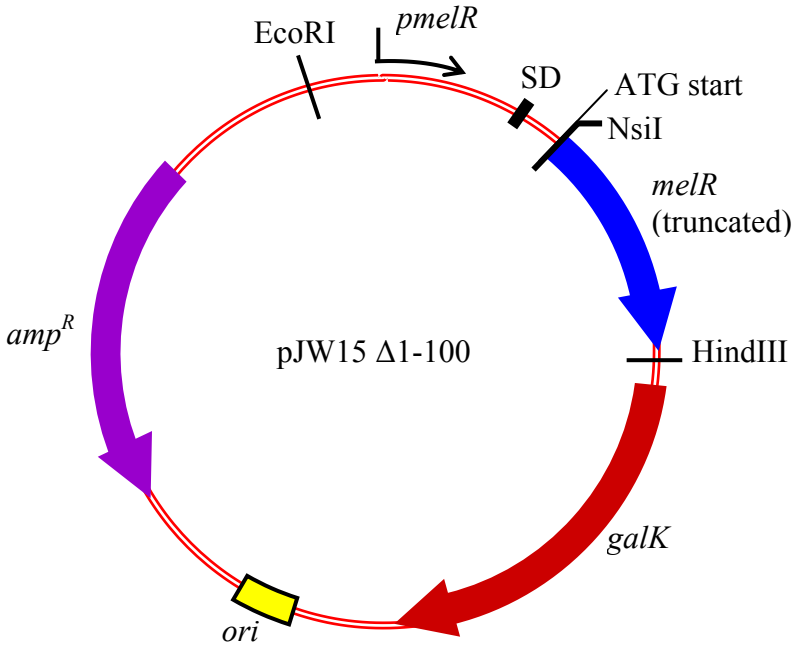
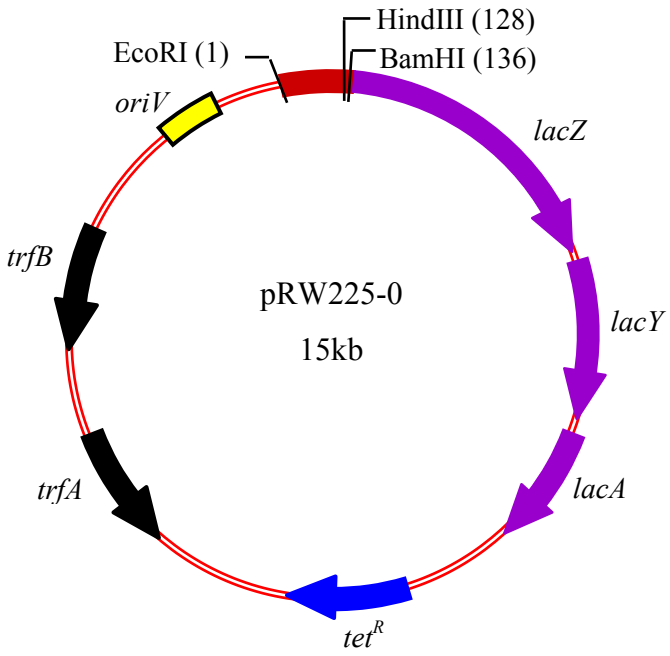


Figure 2.5. Plasmid map of pJW15Δ100.

pJW15Δ100 is a pAA121 derivative multicopy plasmid carrying *melR* promoter. The plasmid contains a truncated *melR* gene (Δ1-100 nucleotides) between NsiI and HindIII sites. Promoterless genes of interest are cloned between NsiI and HindIII sites under the control of *melR* promoter. The ATG in NsiI site (5'-ATGCAT-3') is used as the translation start codon, therefore ATG start in the gene of interest is removed. Also shown are the ampicillin resistance gene (amp^R), the *galK* gene and the origin of replication (*ori*).



EcoRI

ATTCGCGAGAGCCTTGAGTCCACGCTAGATCT**GAATTC**TGACCTGACGTACCGACCCCATATCGTGACGCCGCTG

CTGTTGTACTAACCAAACAGGTTCCCCCTGCCATTCGCCGATTTGTAGTGCGCGCTGACCCACAAGATCGAAATT

HindIII BamHI *lacZ* →

TGCCCGTTCG**AAGCTT**CC**GGATCC**CGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAA

Figure 2.6. Plasmid map of the *lac* fusion vector pRW225.

pRW225 is a derivative of pRW224, which allows cloning of EcoRI-HindIII promoter fragments as well as EcoRI-BamHI fragments as translation fusions to *lacZ*. The sequence of the cloning site is shown in the lower part. The plasmid carries a promoterless "stuffer" sequence between the EcoRI and HindIII sites. The plasmid was produced by cloning a promoterless EcoRI-BamHI fragment, derived from upstream of the *rsd* promoter and carrying a HindIII site, between the EcoRI and BamHI sites in pRW224. With the exception of the altered cloning site, pRW225 is identical to pRW224 (see Figure 2.1).

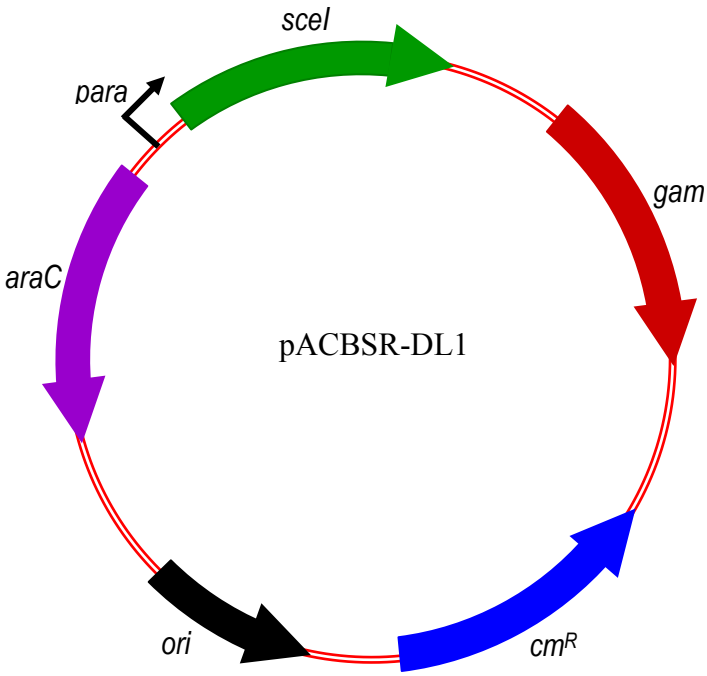


Figure 2.7. Plasmid map of pACBSR-DL1.

pACBSR-DL1 is used in DNA sampling experiment in which DNA fragment cloned between two SceI sites in pRW902 (Figure 2.4) are cut *in vivo* by I-SceI enzymes encoded by pACBSR-DL1. It is a derivative of pACBSR (Herring *et al*., 2003) in which the internal SphI-SphI DNA fragment encoding the bacteriophage lambda Red genes (*exo* and *bet* genes) was replaced with a smaller SphI-SphI fragment encoding the Gam protein. It encodes I-SceI and the bacteriophage lambda Gam protein, under the control of an arabinose inducible *araBAD* promoter (p*ara*). The *araC* gene, which encodes an activator of *araBAD*, the chloramphenicol resistance gene (cm^R) and the origin of replication (*ori*) are also indicated.

Molten 0.8% or 1.3% solutions of agarose in 1 x TAE buffer were cooled to approximately 50°C and poured onto the surface of a glass plate or onto a gel-casting tray. A gel of about 2-3 mm in thickness was made in this way. Sample wells, which could hold 10-100 µl samples, were created by placing a comb above the surface of the plate or the gel-casting tray. Once solidified, gel was immersed in 1 x TAE buffer in a horizontal electrophoresis tank. DNA samples for electrophoresis were mixed in a 1:1 ratio with DNA loading dye and loaded into the wells. The samples were run at 10 V/cm for 30 to 45 mins. Gels were stained in a solution of 0.5 µg/ml ethidium bromide for 20 mins, viewed and photographed under ultraviolet light of 300 nm using gel documentation system (Bio-Rad). However, to reduce the damage to DNA fragments, which were to be purified, gels were viewed under ultraviolet light of 360 nm.

2.5.2 Polyacrylamide gel electrophoresis of DNA

Polyacrylamide gels were used to analyse DNA fragments of 50 to 1000 bp in length. Polyacrylamide gels contained 7.5% (w/v) stock acrylamide, 4% glycerol and 1 x TBE, and were polymerised by adding 0.01 volumes of 10% (w/v) ammonium persulphate (freshly made solution) and 0.001 volumes TEMED. Gels were cast between small glass plates (10x10 cm) or large glass plates for fragment preparation (18x18 cm). Combs were inserted into the top of the gel to create sample wells. The gels were assembled into vertical electrophoresis apparatus. DNA samples for electrophoresis were mixed in a 1:1 ratio with DNA loading dye and loaded into the sample wells. Gels were run in 1 x TBE at 30-40 mA for 30 mins to 4 hrs, then stained in ethidium bromide solution and visualized under UV light as described in section 2.5.1.

2.5.3 Sequencing polyacrylamide gel electrophoresis of DNA and RNA

For analysis of DNA footprinting or *in vitro* transcription reactions, samples were run on thin denaturing 5.5-6% acrylamide sequencing gels (40 cm x 30 cm x 0.4 mm), prepared using the SequaGel sequencing system as described in the manufacturer's instructions. Gels were pre-run in 1 x TBE buffer at 60 W for approximately 1 hr prior to loading. DNA samples in loading buffer were heated to 90º for 2 mins before loading onto the gel. Gels were run in 1 x TBE buffer at 60 W for 2.5-3 hrs, fixed for 10 mins in a 10% methanol/ 10% acetic acid solution. Gels were transferred to Whatman filter paper and dried under vacuum at 80ºC for

30-40 mins. Dried gels were exposed to a Fuji Imaging Phosphor screen for 1 to 16 hrs, the phosphor screen was scanned using a Bio-Rad Molecular Imager FX and images were analysed using QuantityOne software (Bio-Rad).

2.5.4 Denaturing gel electrophoresis of proteins

During DNA-sampling and pull-down assays, proteins were separated and analysed using NuPAGE 12% Bis-Tris pre-cast gels (Invitrogen). Proteins were mixed with an equal volume of SDS-PAGE loading buffer. Samples were heated to 95°C for 5 mins, vortexed for 10 s, and centrifuged for 15 s at ~18000 x *g*, then loaded onto the gel. Gels were run in 1 x MES SDS running buffer (Invitrogen) at 140 V for approximately 1 hr. Gels were stained by incubating for 30 mins with agitation in Coomassie blue stain solution or by silver staining following SilverQuest™ Microwave Silver Staining Protocol (Invitrogen).

2.6 Extraction and purification of DNA fragments

2.6.1 Phenol/chloroform extraction of DNA

In order to remove contaminating proteins, solutions of DNA were mixed with an equal volume of phenol/chloroform/isoamylalcohol solution, vortexed for 15 s and centrifuged for 3 mins at ~18000 x *g* to separate the aqueous and organic phases. The DNA containing aqueous layer was transferred to a fresh eppendorf tube. An equal volume of water or TE buffer was then added to the remaining organic phase, which was then vortexed and centrifuged for 3 mins at ~18000 x *g*. The aqueous layer was combined with that removed previously and concentrated by ethanol precipitation as detailed below.

2.6.2 Ethanol precipitation of DNA

Ethanol precipitation was used to concentrate DNA solutions and to allow transfer of DNA samples to a different buffer. 1/10 volume of 3 M sodium acetate (pH 5.2) and 2.5-3 volumes of ice-cold 100% ethanol were added to the DNA solution. For samples containing DNA fragments smaller than 500 bp, 1 µl 20 mg/ml glycogen was also added. The samples were then incubated at -20°C for at least 30 mins and centrifuged for 15 mins at 4°C at ~18000 x *g*. The supernatant was removed; the pellet was washed with 1 ml ice-cold 70% ethanol and centrifuged for 10 mins at room temperature. The supernatant was removed and the pellet was

dried for 10-15 mins in a vacuum dryer. The dry pellet was resuspended in TE buffer, sterile distilled water or loading buffer, as indicated.

2.6.3 Purification of DNA using QIAquick PCR purification kit

DNA purification after PCR or restriction digestion reactions was carried out using the QIAquick PCR purification kit (Qiagen), following manufacturer's instructions. Purified DNA was eluted from QIAquick columns in 50 µl sterile distilled water.

2.6.4. Extraction of DNA fragments from agarose gels

DNA samples were mixed in 1:1 ratio with gel loading dye and run on 0.8%-1.3% agarose gels as described in section 2.5.1. DNA bands for extraction were excised from the gel and eluted using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. Extracted DNA fragments were eluted from the QIAquick columns in 50 µl sterile distilled water.

2.6.5 Electroelution of DNA fragments from polyacrylamide gels

The process of electroelution (Maniatis *et al.,* 1982) was used to purify DNA fragments after PCR, restriction digestion or other enzymatic modifications. DNA samples were mixed in a 1:1 ratio with gel loading dye and run on a 7.5% polyacrylamide gel as described in section 2.5.2. DNA bands for extraction were excised from the gel and placed into 6.3 mm dialysis tube (Medicell International Ltd.), sealed with clips and filled with 0.1 x TBE buffer. Electroelution was carried out at 30 mV for 20-30 mins. After that the buffer was removed from the dialysis bag and transferred to a microcentrifuge tube. The bag was rinsed out with 100 µl sterile distilled water, which was also added to the microfuge tube. The DNA fragment was extracted from the buffer by phenol/chloroform extraction and ethanol precipitation method as described in sections 2.6.1 and 2.6.2.

2.6.6 Sephadex G-50 spin columns

Sephadex G-50 spin columns were used to remove unincorporated radioisotopes from radio labelled DNA fragment during *in vitro* footprinting or transcription assays. 5 g of Sephadex G-50 was autoclaved in ~100 ml TE buffer and allowed to settle. Excess TE was decanted and a further 150 ml of fresh TE mixed with the Sephadex. The larger Sephadex particles were

allowed to settle for 5 mins, the TE buffer decanted and a further 150 ml added. This wash procedure was repeated a further two times to remove the under-sized particles of Sephadex, leaving a homogeneous G-50 suspension. The final TE wash was decanted and the particles resuspended in 100 ml fresh TE buffer (50% G-50 suspension) and stored at 4°C.

400 µl of settled 50% G-50 suspension was added to Micro Bio-spin columns (Bio-Rad) and centrifuged for 2 mins at ~1300 x *g* to pack the column. 20 µl of DNA sample (or one-tenth column volume) was applied to the top of the column, and the column assembly centrifuged for a further 2 mins at ~1300 x *g*. The purified DNA fragment that eluted during centrifugation was collected into a microcentrifuge tube.

2.7 Isolation of chromosomal and plasmid DNA from bacteria

2.7.1 Small-scale preparation of genomic DNA

Genomic DNA was isolated in a small-scale from 1 ml overnight bacterial culture ($A_{600} \leq 4.0$) using illustra bacteria genomicPrep Mini Spin Kit (GE Healthcare) following manufacturer's instruction. In brief, 1 ml overnight bacterial culture was pelleted and culture medium was removed. Cells were lysed by the detergent and salt in combination with Proteinase K to release genomic DNA into solution from bacterial cells. RNA was removed by RNase A. The genomic DNA was then bound onto a silica column in the presence of the chaotrope. Protein and other contaminants from membrane-bound genomic DNA were removed by using chaotropic salt. Ethanolic wash buffer was used to remove salts and other contaminants and dry the silica membrane at the same time. Finally, genomic DNA was eluted in 200µl low ionic strength buffer, collected in sterile microfuge tube and stored at -20ºC.

2.7.2 Small-scale preparation of plasmid DNA ("mini-prep")

QIAprep Spin Miniprep kit (QIAGEN) was used to isolate plasmids from bacteria in small-scale following manufacturer's instructions. A brief outline of the protocol followed from the handbook provided with the kit is given below. A single bacterial colony from a freshly restreaked plate was used to inoculate 5 ml of LB containing appropriate antibiotic(s) in a 25 ml sterile flask. The culture was incubated aerobically overnight at 37°C with vigorous shaking. Three 1.5 ml aliquots of this culture were transferred to microcentrifuge tubes and harvested by centrifugation at ~20000 x *g* for 1 min. The bacterial pellet was resuspended in

250 µl Buffer P1 (RNase A solution was added to this solution prior to use). An aliquot of 250 µl of Buffer P2 was added and then the microfuge tube inverted 4-6 times. Immediately following this 350 µl of Buffer N3 was added and the tubes were inverted 4-6 times as before. The tubes were then centrifuged at ~20000 x *g* for 10 mins until a white pellet had formed.

The supernatant was then decanted into a QIAprep spin column. The column and collection tube were centrifuged for 60 s and the flow through discarded. The column was then washed by the addition of 0.5 ml Buffer PB and this was again centrifuged for 60 s. The QIAprep column was then washed by the addition of 0.75 ml Buffer PE and was centrifuged again. The flow through was once again discarded. The column was then centrifuged again to remove any final traces of PE Buffer. The QIAprep column was then placed in a clean microcentrifuge tube and 50 µl (for high copy number plasmids) or 30 µl (for low copy number plasmids) of Buffer EB was added to the centre of the column. The tubes were then centrifuged one final time for 60 s at ~20000 x *g* to collect plasmid DNA. The isolated plasmid DNA was then stored at -20°C.

2.7.3 Large-scale plasmid DNA preparation ("maxi-prep")

QIAfilter Plasmid Maxi Kit (QIAGEN) was used to prepare large scale plasmid DNA from bacterial cells according to the manufacturer's instruction. In brief, a single colony from a freshly restreaked selective plate was inoculated as a starter culture in 5 ml LB medium supplemented with appropriate antibiotic and incubated for approximately 8 hrs at 37°C with vigorous shaking. 100 ml LB medium supplemented with required antibiotic in a 500 ml flask was inoculated with 200 µl of starter culture (dilution 1/500) and incubated at 37°C for 16 hrs with vigorous shaking.

Cultures were taken into two 50 ml falcon tubes and subjected to centrifugation at ~3400 x *g* for 30 mins at 4°C. Bacterial pellets were resuspended in resuspension buffer containing RNase A, lysed with NaOH/SDS and neutralized by potassium acetate. Cell debris, genomic DNA, protein and SDS were removed by filtration with QIAfilter cartridge and the lysates were passed through an anion-exchange resin by gravity flow under appropriate low salt and pH condition that helps bind the plasmid DNA to the resin. The plasmid DNA bound onto the resin were then washed with medium salt to remove contaminants (protein, RNA, or low

molecular weight impurities), eluted in a high salt buffer, concentrated and desalted by isopropanol and 70% ethanol precipitation and air-dried. Finally, purified plasmid DNA was redissolved in appropriate volume of TE buffer and stored at -20°C.

2.8 Bacterial transformations

2.8.1 Preparation of competent cells using the $CaCl_2$ method

50 ml LB was inoculated with 1 ml of an overnight culture of the strain to be transformed, and incubated at 37°C until the culture reached mid-logarithmic phase (OD_{650} = 0.4-0.6). The culture was then transferred to a sterile 50 ml Falcon tube, incubated on ice for 10 mins and subjected to centrifugation at ~2700 x *g* at 4°C. The pellet was then resuspended in 25 ml ice-cold 0.1 M $CaCl_2$, incubated for 10 mins on ice, and the cells were harvested by centrifugation for 5 mins at ~2700 x *g*. Following removal of supernatant, the pellet was resuspended in 3.3 ml ice-cold buffer containing 0.1 M $CaCl_2$ and 15% glycerol, and the cells were kept on ice for 24 hrs before use. For storage, each of 200 µl aliquots in sterile microfuge tube was stored at -70°C.

2.8.2 Preparation of competent cells using the RbCl method

The method was used to maximize the transformation efficiency of competent cells, needed for selection of hundreds of transformants during random mutagenesis. 100 ml LB in a 250 ml flask was inoculated with 1 ml of an overnight culture of the strain to be transformed, and incubated at 37°C until the culture reached mid-logarithmic phase (OD_{650} = 0.4-0.6). The culture was then transferred to 2 sterile 50 ml falcon tubes, kept on ice for 10 mins, and centrifuged for 5 mins at ~3400 x *g* at 4°C. Following removal of supernatant, contents of one tube was resuspended in ice-cold 40 ml (1/2.5 volume) of TFB1 (see section 2.2.3) buffer and transferred to the second tube and resuspended the cells. Cells were kept on ice for 10 mins and centrifuged ~3400 x *g* for 5 mins at 4°C. The pellet was resuspended in 4 ml (1/25 volume) ice-cold TFB2 buffer (see section 2.2.3) and kept on ice for an hr. For storage, each of 200 µl aliquots in sterile microfuge tube was stored at -70°C.

2.8.3 Transformation of competent cells with plasmid DNA

50-100 µl competent cells were mixed with 1-3 µl plasmid DNA on ice, and incubated 45-60 mins on ice. Cells were then heat-shocked at 42°C for 1 min and kept on ice for 2 mins. 1 ml

LB or SOC medium was added to the heat-shocked cells, and they were then incubated for 1 hr at 37°C. After 1 min centrifugation at ~18000 x *g*, the pellet was resuspended in approximately 100 µl of the supernatant, plated onto nutrient agar or MacConkey agar supplemented with required antibiotic(s), and incubated overnight at 37°C.

2.8.4 Transformation of bacterial cells using electroporation

Cell preparation

500 µl overnight cultures of bacterial cells were inoculated in 50 ml LB medium supplemented with appropriate antibiotics if necessary. When cells reach mid-exponential growth phase (OD_{650} = 0.5), they were collected in a 50 ml falcon tube and subjected to centrifugation at ~3400 x *g* at 4°C for 10 mins. Pellets were resuspended in ice-cold 25 ml of 10% glycerol and the contents were centrifuged at ~3400 x *g* for 10 mins at 4°C. Pellets were resuspended in 1 ml of ice-cold 10% glycerol, transferred to a 1.5 ml eppendorf tube and centrifuged at ~18000 x *g* for 1 min. Following removal of supernatant, pellets were finally resuspended in 500 µl ice-cold 10% glycerol and kept on ice.

Electroporation

Prior to electroporation, SOC medium was preheated to 37°C, P1000 pipette was set to 1 ml and P100 to 40 µl, 1 mM cuvettes were placed on ice, electroporator was set to 1.8 kvolts, post-electroporation tubes were kept in a rack, and dry orbital shaker was turned on 37°C. Then 1 µl plasmid DNA to be transformed was transferred in a tube and placed on ice. 40 µl cells were added to the DNA on ice and mixed by pipetting up and down twice and immediately transferred to an electroporation cuvette and subjected to pulse in electroporator, immediately followed by addition of 1 ml of 37°C SOC. The content was transfer to a 1.5 ml tube and placed at 37°C in a heat block. When all the samples had been electroporated, the tubes were transferred to a rack and placed them in the dry orbital shaker with vigorous shaking for an hr. After 1 min centrifugation at ~18000 x *g*, the pellet was resuspended in approximately 100 µl of the supernatant, plated onto nutrient agar or MacConkey agar supplemented with required antibiotic(s), and incubated overnight at 37°C.

2.9 Recombinant DNA techniques

2.9.1 Routine PCR

Polymerase chain reaction (PCR) is used to amplify specific region of a DNA strand. It consists of a mixture of template DNA, two oligonucleotide primers, DNA polymerase and deoxynucleoside triphosphates (dNTPs). The two oligonucleotide primers are designed so that they anneal to complementary DNA strands, either side of the target DNA to be amplified, with their 3' ends facing towards each other. The PCR reaction amplifies DNA by cycling three different steps. The first step is denaturing step in which the double-stranded DNA template is denatured and separated to single stranded templates. The second step is an annealing step, which allows the primers to anneal to complementary sequences of the single -stranded DNA templates. The third step is an elongation step, which allows the DNA polymerase to synthesise complementary DNA strands from 3' end of each primer.

PCR was carried out using Finnzymes' Phusion™ High-Fidelity DNA polymerase (New England BioLabs® Inc.) or Biotaq polymerase (Bioline) in the buffer supplied with the relevant enzyme. Reaction conditions for different enzymes were optimised according to manufacturer's instructions. Phusion polymerase was routinely used for PCR throughout the study whilst Biotaq polymerase was used mainly for error-prone PCR. Phusion DNA polymerase enzyme possesses both 5'→3' DNA polymerase activity and 3'→5' exonuclease activity and thus it can clone the DNA fragment with more accuracy and speed. Optimum reaction conditions of this enzyme differ from standard enzyme protocols. PCR reactions were performed in a 50 µl reaction mix containing 10 µl of 5 x Phusion HF buffer containing 1.5 mM $MgCl_2$, 0.5 µM of each of the two primers , 0.8 mM dNTPs (200 µM each) (Bioline), 10-50 ng of template DNA, 1 unit of enzyme and a suitable amount of sterile distilled water. DNA amplification was performed in an oil-free thermal cycler (GeneAmp® PCR System, Applied Biosystems). PCR cycling conditions are shown in the Table 2.3.

2.9.2 Colony PCR

Colony PCR was used to amplify DNA fragments from the chromosome, to screen for chromosomal insertions/deletions and to check the presence of cloned inserts in plasmids. Template DNA was prepared by suspending one fresh colony in 100 µl sterile distilled water, heating the cell suspension to 100°C for 10 mins, and removing cell debris by centrifugation for 1 min at ~18000 × *g*. 10 µl template preparation was used in a 50 µl PCR as described in section 2.9.1.

Table 2.3: PCR cycle

Temperature	Time		Purpose
98 °C	30 s		"hot start"
98 °C	10 s		Melting
T_A °C	30 s	30-35 cycles	Annealing
72 °C	X s		Extension
72 °C	5 mins		Final extension

The table shows thermal profile of a PCR reaction using Phusion High Fidelity DNA polymerase. An annealing temperature (T_A°C) of 3°C above the melting temperature (T_m) of the lower T_m primer was used. The extension time (X s) was calculated based on PCR product length and complexity of the amplicon. For genomic DNA, X = 30 s/kb, whilst for plasmid DNA, X= 15 s/kb was used.

2.9.3 Error-prone PCR

Error-prone PCR is a modification of a standard polymerase chain reaction method. It takes advantage of the inherently low fidelity of Taq DNA polymerase and thus introduces random mutations into a defined segment of a DNA. Taq DNA polymerase can misincorporate with a frequency as high as 0.02% per position (Eckert and Kunkel, 1991). The error rate can be further increased by skewing the relative dNTP concentrations and using a high concentration of Mg^{2+} in the reaction mix (Leung *et al*., 1989).

Mutagenic or error-prone PCR reactions were performed in a 50 µl reaction mix containing final concentration of 1.5-5 mM $MgCl_2$ (Bioline) , 1 µM each of the two oligos, 0.4-1.0 mM dNTPs (Bioline), ~ 10 ng of template DNA, 5 µl of 10 x NH_4 buffer (no $MgCl_2$), 1 unit of Biotaq polymerase (Bioline) and a suitable amount of sterile distilled water. DNA amplification was performed in an oil- free thermal cycler (GeneAmp® PCR System, Applied Biosystems). The PCR profiles were: pre-amplification denaturation at 95°C for 5 mins, followed by 40 cycles of 10 s denaturation at 95°C, 1 min annealing at 60°C, and 2 mins extension at 74°C. After the last cycle, a final step of 5 mins at 74°C was added to allow complete extension of the all amplified fragments.

2.9.4 Site directed mutagenic PCR

Site directed mutagenic PCR was used to introduce point mutations in a DNA sequence of interest. To do this, a downstream or upstream oligonucleotide primer containing the desired base change together with an upstream or downstream primer were used to amplify the DNA fragment carrying desired mutations. A number of oligonucleotide primers used for site directed mutagenesis are listed in the Table 2.4.

2.9.5 Megaprimer PCR

Megaprimer PCR (Perrin and Gilliland, 1990), a polymerase chain reaction based method was used to introduce a desired mutation in an internal position, which was not adjacent to a suitable restriction site in a template DNA. This method involved two or more rounds of PCR that utilize two 'flanking' primers and one internal mutagenic primer containing the desired base substitution(s). The first PCR was performed using the mutagenic internal primer and the first flanking primer to amplify megaprimer with desired base change from template DNA

Table 2.4: DNA oligonucleotides used in this study (continued on pages 86-92)

Name	Sequence (5'-3')	Use
Oligonucleotides used for sequencing and PCR screening of inserts in plasmid vectors		
D10520	CCCTGCGGTGCCCCTCAA C	Anneals upstream of EcoRI site in pRW224 or pRW225 or pRW902. Used for sequencing and amplification of inserts in these vectors
D53463	GGGGGATGTGCTGCAAGG CG	Anneals within the *lacZ* coding sequence, downstream of the BamHI or HindIII site in pRW224, pRW225 or pRW902. Used for sequencing and amplification of inserts in these vectors
D3407	GTAATACGACTCACTATA GGGC	Anneals upstream of NsiI site in pJW15Δ100. Used for sequencing and amplification of inserts in pJW15Δ100
D4600	GTAGTCGGTGTGTTCAC	Anneals downstream of HindIII site in pJW15Δ100. Used for sequencing of inserts in pJW15Δ100
D63048	CTCTTCAAATGTAGCACC TGAAG	Anneals upstream of HindIII site in pACYC184. Used for sequencing and amplification of inserts in pACYC184
D63049	GAAGACAGTCATAAGTGC GGCGAC	Anneals downstream of SalI site in pACYC184. Used for sequencing and amplification of inserts in pACYC184
D5431	ACCTGACGTCTAAGAAAC C	Anneals upstream of EcoRI site in pSR. Used for sequencing and amplification of inserts in pSR
D56033	AGCGTTCTG AACAAATC	Anneals downstream of HindIII site in pSR. Used for sequencing and amplification of inserts in pSR
Oligonucleotides used for deletion analysis of* LEE1 *operon regulatory region		
D61221	GCAGAATTCTGCACCCGT TCCAGG	Upstream primer for amplification of EcoRI -HindIII LEE10-568 fragment
D61222	GCAAAGCTTGCTTTAATA TTTTAAGC	Downstream primer for amplification of EcoRI-HindIII LEE10-568 fragment

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
D68822	GCAGAATTCAGAGAAACGCTTAACTAAAT	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-315 fragment
D63949	GCAGAATTCCGTTTGTTAACGAGATGATTTTCTTC	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-275 fragment
D63948	GCAGAATTCTAAATGGATTTTAAAAATATATG	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-235 fragment
D64557	GCAGAATTCATGATTTTTTTGTTGACA	Upstream primer used in PCR reaction with D61222 to generate EcoRI-HindIII LEE10-215 fragment
D68959	GCAGAATTCTTGACATTTAATGATAATGT	Upstream primer used in PCR reaction with D61222 to generate EcoRI-HindIII LEE10-203 fragment
D63947	GCAGAATTCTAATGATAATGTATTTTACACATTAG	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-195 fragment
D63697	GCAGAATTCTAACATTTTAAGGTGGTTGTTTGATG	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-155 fragment
D63696	GCAGAATTCTAATTTGATAGATAAACGTTATCTCAC	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-115 fragment
D63695	GCAGAATTCTTTGATTAATTGTTGGTCCTTCCTG	Upstream primer used in PCR reaction with D61222 to produce EcoRI-HindIII LEE10-75 fragment
D64100	GCAAAGCTTATTCTCTTTTTTCTAATG	Downstream primer used in PCR reaction with D61221, D68822, D63949 or D68959 to produce EcoRI-HindIII LEE20-568, LEE20-315, LEE20-275 or LEE20-203 fragments, respectively. Both D63949 and D64100 were also used in error-prone PCR reaction during random mutagenesis of LEE20-275 fragment

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
D64101	GCAAAGCTTTAGGACACATCTATTTCA	Downstream primer used in PCR reaction with upstream D63949 primer to produce EcoRI-HindIII LEE30-275 fragment. Both D64101 and D63949 were also used in error-prone PCR reaction during random mutagenesis of LEE30 -275 fragment
Oligonucleotides used for deletion analysis of* LEE2 *operon regulatory region		
D62263	GCAGAATTCACGTAACAAAAACATTATTATC	Upstream primer for amplification of EcoRI-HindIII LEE2-220 fragment.
D62265	GCAAAGCTTTGATCTTTCTCCTTTTGTC	Downstream primer for amplification of EcoRI-HindIII LEE2-220 fragment
D62264	GCAGAATTCCGTTTCTTTTAACTAAAAG	Upstream primer used in PCR reaction with downstream D62265 to produce EcoRI-HindIII LEE2-125 fragment.
D62865	GCAGAATTCCCCAATAATCTTAAAAACTC	Upstream primer used in PCR reaction with downstream D62265 to produce EcoRI-HindIII LEE2-100 fragment.
Oligonucleotides used for mutagenesis of LEE20-275 and LEE20-203 promoter fragments		
D64737	GCAAAGCTTATTCTCTTTTTTCTAATGTGTAAAATACATTATCATTAAATGTDAACAAAAAAATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 75C
D64736	GCAAAGCTTATTCTCTTTTTTCTAATGTGTAAAATACATTATCATTAAATHTCAACAAAAAAATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 77A and LEE20-275 77G
D65245	GCAAAGCTTATTCTCTTTTTTCTAATGTGTAAATACATTATC	Downstream primer used in PCR reaction with D63949 and D68959 to produce LEE20-275 Δ94T and LEE20-203 Δ94T, respectively
D64735	GCAAAGCTTATTCTCTTTTTTCTAATGTGTABAATACATTATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 96C or LEE20-275 96G

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
D68520	GATAATGTATTTTCCA CATTAGA	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE20-275 98C and LEE20-203 98C
D65244	GCAAAGCTTATTCTCT TTTTTCTAATGTHTAA AATACATTATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 99A and LEE20 -275 99G
D65243	GCAAAGCTTATTCTCT TTTTTCTAATHTGTAA AATACATTATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 101G
D65242	GCAAAGCTTATTCTCT TTTTTCTAAVGTGTAA AATACATTATC	Downstream primer used in PCR reaction with D63949 to produce LEE20-275 102C and LEE20 -275 102G
D69576	TTGACACTTAATGATA AT	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 79C
D69715	TTGACATGTAATGATA ATGTATT	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 80G
D68262	GACATTTAATGCTAAT GTATTT	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 86C
D70042	TGACATTTAATGAAAA TGTATTTTA	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 87A
D68263	GACATTTAATGATCAT GTATTTT	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 88C
D69717	TAATGATAATATATTT TACACA	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE20-203 91A
D69323	TAATGATAATCTATTT TACA	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 91C
D70041	TAATGATAATGAATTT TACACA	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE20-203 92A

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
D69324	TGTATTTTACAGATTAG AAAAAAGA	Upstream primer used in PCR reaction with D54363 to produce megaprimer for construction of LEE20-203 101G
D69325	ATTTTACACAATAGAAA AAAGA	Upstream primer used in PCR reaction with D54363 to produce megaprimer for construction LEE20-203 103A
D69718	ATTTTACACATGAGAAA AAAGAGAA	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE20-203 104G
D69894	TTGACATTAATGATAAT GTATTTTACA	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 Δ79T
D71441	TTGACATAATGATAATG TATTTTACA	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 Δ79/80
D69893	TTGACATTTTAATGATA ATGTATTTTACA	Upstream primer used in PCR reaction with D64100 to produce megaprimer for construction of LEE20-203 InsT (78-79)
D68075	GCAAAGCTTATTCTCTT TTTTCTAA	Downstream primer used in PCR reaction with D68959 to generate LEE20-203 InsT (93-94) fragment with pRW224 carrying LEE20-275 InsT (93-94) as a template
Oligonucleotides used for cloning* grlRA *and derivatives		
D62895	GCAAAGCTTTTTTACGT TGTTACTCAATATTATT AATCAG	Upstream primer for amplification of HindIII-SalI *grlR+A+* fragment
D62897	GCAGTCGACCTAACTCT CCTTTTTCCGCC	Downstream primer for amplification of HindIII -SalI *grlR+A+* fragment
D63698	GCAGTCGACTCGACATA AAAAACATAC	Downstream primer used in PCR reaction with D62895 to generate HindIII-SalI *grlR+A-* fragment
D63209	GCAGGATCCAATGCTAT AGATGCCATC	Downstream primer used in PCR reaction with D62895 to generate upstream BamHI DNA fragment for construction of an in-phase deleted mutant *grlR-A+*

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
D63210	GCAGGATCCACAGGAATGCCACAAGTT	Upstream primer used in PCR reaction with D62897 to generate downstream BamHI DNA fragment for construction of in-phase deleted mutant *grlR-A+*. Both D63210 and D62897 were also used in error-prone PCR reaction to analyze GrlA-dependent activation by suppression genetics
***Oligonucleotides used for mutagenesis of* grlA**		
D71098	TCGGTATAAACGCGAGGAGAGCATC	Codes for LA substitution at codon 52 of *grlA*
D71099	GTATAAACCTGGCGAGAGCATCATTTA	Codes for RA substitution at codon 53 of *grlA*
D71100	GTATAAACCTGAGGGCAGCATCATTTATTATAAC	Codes for RA substitution at codon 54 of *grlA*
D71101	GAGGAGAGCAGCATTTATTATAACT	Codes for SA substitution at codon 56 of *grlA*
D71102	GAGAGCATCAGCTATTATAACTTATATATC	Codes for FA substitution at codon 57 of *grlA*
D71103	AGCATCATTTGCTATAACTTATATATC	Codes for IA substitution at codon 58 of *grlA*
D71104	CATCATTTATTGCAACTTATATATC	Codes for IA substitution at codon 59 of *grlA*
D71105	CATTTATTATAGCTTATATATCGAG	Codes for TA substitution at codon 60 of *grlA*
D71106	TATTATAACTGCTATATCGAGAAGAAAAG	Codes for YA substitution at codon 61 of *grlA*
D71107	TATAACTTATGCATCGAGAAGAAAAG	Codes for IA substitution at codon 62 of *grlA*
D71108	AACTTATATAGCGAGAAGAAAAG	Codes for SA substitution at codon 63 of *grlA*

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
Oligonucleotides used for amplification of genes cloned into pJW15Δ100		
D61652	GCAATGCATCGGAGATTATTTATTATG	Upstream primer containing NsiI site for amplification the *ler* gene
D61223	GCAAAGCTTTTAAATATTTTTCAGCGG	Downstream primer containing HindIII site for amplification of the *ler* gene
D61658	GCAATGCATCTACATGATCACCTGGCAG	Upstream primer containing NsiI site for amplification of *pchC* gene
D61659	GCAAAGCTTTTAGCATTTTTTTGACCG	Downstream primer containing HindIII site for amplification of the *pchC* gene
D64149	GCAATGCATAGCAGGAAATTTAGCTCTC	Upstream primer containing NsiI site for amplification of the *cesD* gene
D64150	GCAAAGCTTTTACTCTGTATTACCTAAC	Downstream primer containing HindIII site for amplification of the *cesD* gene
D64151	GCAATGCATTCATCAAGATCTGAACTTTTA	Upstream primer containing NsiI site for cloning *cesT* gene
D64152	GCAAAGCTTTTATCTTCCGGCGTAATA	Downstream primer containing HindIII site for amplification of *cesT* gene
D65247	GCAATGCATAATCTTTTAGTTAAAAGAAACG	Upstream primer containing NsiI site for cloning *cesL* gene
D65248	GCAAAGCTTTCATGATGTCATCCTGCGAA	Downstream primer containing NsiI site for cloning *cesL* gene
Oligonucleotides used for amplification of promoter fragments cloned into pRW225		
D65474	GCAGGATCCCAAGCTTTAGGACACATC	Downstream primer used in PCR reaction with D63949 to amplify EcoRI -BamHI LEE30-275 fragment
D65812	GCAAAGCTTCATGCTTTAATATTTTAAGCT	Downstream primer used in PCR reaction with D61221 containing HindIII site for amplification of LEE151
D65811	GCAAAGCTTCATAATAAATAATCTCCG	Downstream primer used in PCR reaction with D61221 containing HindIII site for amplification of LEE150

Table 2.4: DNA oligonucleotides used in this study (continued)

Name	Sequence (5'-3')	Use
Oligonucleotides used for mutagenesis of LEE30-275 and LEE150 promoter fragments		
D65813	GGTGGTTGTTTGTTGAAA TAGATGT	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE30-275 144T fragment
D65814	GTTTGATGAAATTGATGT GTCCTAA	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of EE30-275 151T fragment
D66032	GCAGGTACCAGATGTGTC CTAATTTGATAG	Upstream primer used in PCR reaction with D65811 to generate downstream KpnI fragment for construction of LEE150-1 fragment
D66033	GCAGGTACCTCAAACAAC CACCTTAAAATG	Downstream primer used in PCR reaction with D61221 to generate upstream KpnI fragment for construction of LEE150-1 fragment
D66176	TCTGGTACCTGAAACAAC CAC	Downstream primer used in PCR reaction together with D61221 to generate LEE150-1 143C fragment. This oligo was also used with D10520 to produce megaprimer for construction of LEE30-275-1 143C fragment
D66271	GTGGTTGTTTGAGGAAAT AGATGTGT	Upstream primer used in PCR reaction together with D65811 or D53463 to produce megaprimers for construction of LEE150 145G and LEE30-275 145G, respectively
D66270	TAAGGTGGTTGTTTCAGG AAATAGATGTGTC	Upstream primer used in PCR reaction with D65811 or D53463 to produce megaprimers for construction of LEE150 143C 145G or LEE30 -275 143C 145G, respectively
D66870	GAAATATATGTGTCCTAA AGCTT	Upstream primer used in PCR reaction with D53463 to produce megaprimer for construction of LEE30-275 152T fragment
D67046	GTGGTTGTTTCATGAAAT AGATG	Upstream primer used in PCR reaction with D65811 or D53463 to produce megaprimers for construction of LEE150 143C or LEE30-275 143C fragments, respectively

Restriction sites incorporated in the oligos are underlined

containing the sequence of interest. The product of this first PCR, the 'megaprimer', was purified by electroelution from polyacrylamide gel (see section 2.6.5) and used, along with the second flanking primer, as a primer for a second round PCR. The final PCR product that contained the desired mutation in the particular DNA sequence was purified using QIAquick PCR purification kit (Qiagen) or Qiagen gel extraction kit (see sections 2.6.3 and 2.6.4, respectively), double-digested with appropriate restriction enzymes as described below and cloned into the desired plasmid. The basic megaprimer procedure is shown schematically in Figure 2.8 and a list of mutagenic internal primers used during megaprimer PCR is shown in the Table 2.4.

2.9.6 Restriction digestion of DNA

50 µl DNA (purified PCR product or plasmid miniprep) was digested using 2 µl each restriction enzyme (all from New England BioLabs® Inc.) in a final volume of 60 µl of the appropriate buffer, as determined using the New England Biolabs double digest finder (http://www.neb.com/nebecomm/DoubleDigestCalculator.asp). Restriction digests were incubated for 3 hrs at 37°C. Digested plasmid DNA for use in cloning was treated with alkaline phosphatase to remove terminal 5' phosphate groups, in order to prevent religation of vector DNA. Where this was required, 3µl calf alkaline phosphatase (CAP) (New England BioLabs® Inc.) was added to the restriction digest mix after the 3 hrs incubation, and incubated for a further 1 hr at 37°C. Digested and CAP treated DNA was purified by PCR purification kit or agarose gel extraction kit or electroelution from 7.5% polyacrylamide gel (see sections 2.6.3- 2.6.5).

2.9.7 Ligation

Concentration of both vector and insert DNA was estimated by agarose gel electrophoresis along with molecular weight standards of a known concentration and tested various vector: insert DNA ratios in order to find the optimum ratio for a particular vector and insert. 1:1 or 1:3 molar ratio of vector: insert DNA has been found to work well and used throughout the study. Ligation reactions were performed in a 20 µl reaction mix containing 1-3 µl restriction digested insert DNA, 3-6 µl digested and alkaline phosphate-treated vector DNA, 1 µl T4 DNA ligase (New England BioLabs® Inc.), 2 µl 10x T4 DNA ligase buffer and a suitable amount of sterile distilled water. Ligation mixes were incubated for 5 mins at room

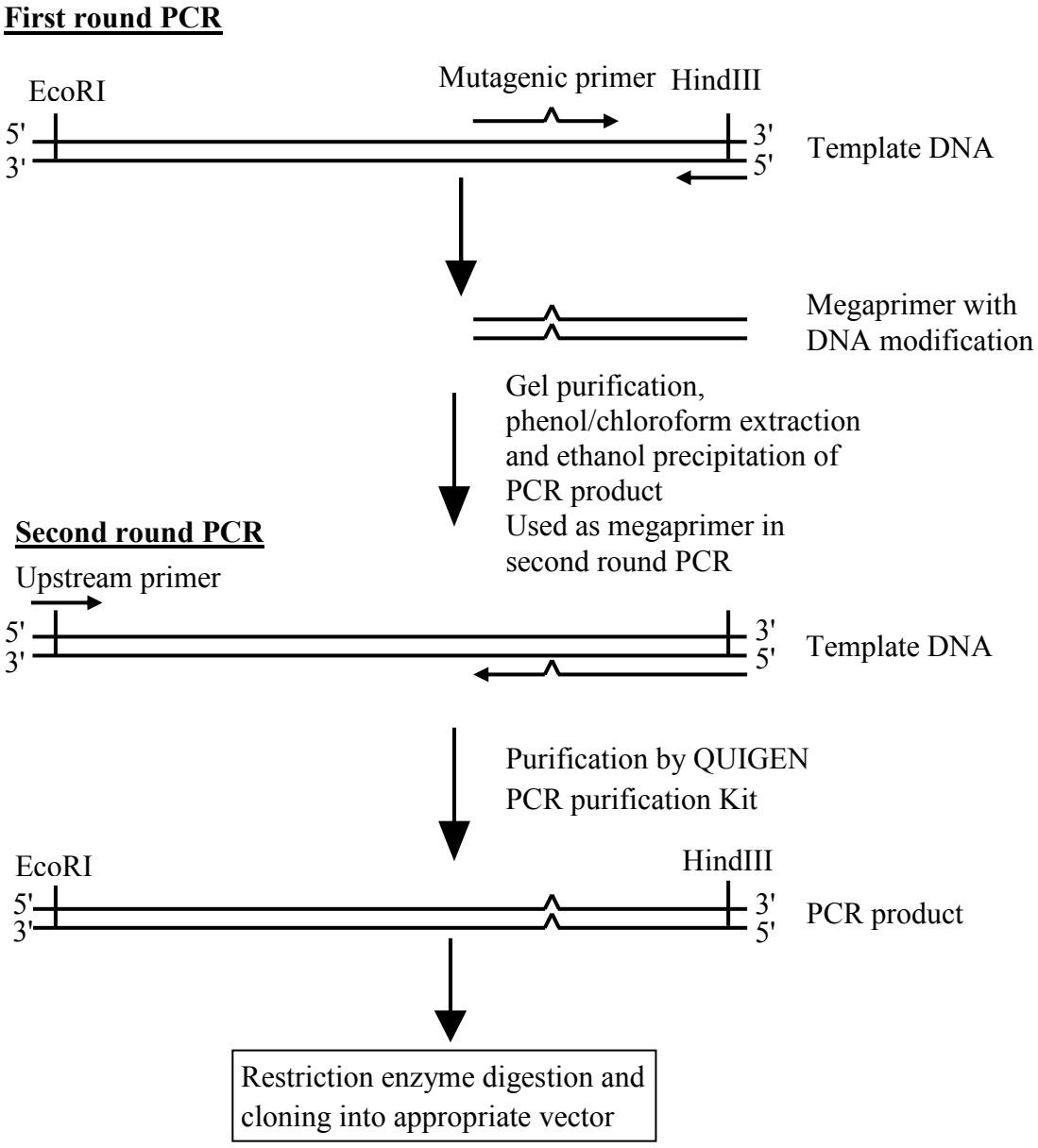


Figure 2.8. Introduction of point mutations into EcoRI-HindIII promoter fragments by megaprimer PCR.

A mutagenic primer, carrying the desired point mutation(s), was used together with the down stream flanking primer, to amplify a megaprimer by PCR from the plasmid carrying the promoter of interest. The megaprimer carrying the mutant(s) of interest was then used together with the upstream flanking primer, to generate a mutated EcoRI-HindIII promoter fragment, which was subsequently cloned into the plasmid. Note that if the position of the mutation(s) was closer to the EcoRI site, the megaprimer was synthesized using a mutagenic primer and the downstream flanking primer. The megaprimer was then used with the upstream primer, in a second round PCR to produce the mutated fragment.

temperature. 100 µl *E. coli* M182 competent cells were transformed with ligation mix. In order to select transformants, cells were plated onto nutrient agar or MacConkey agar supplemented with appropriate antibiotics. In order to screen candidate plasmids for the presence of an insert of the expected size, candidate transformants were grown overnight and plasmid minipreps were prepared from each. Plasmids were digested with appropriate restriction enzymes to excise the cloned fragment and run on 1.3% agarose gels or 7.5% PAGE gels (see sections 2.5.1, 2.5.2) to check for the presence of an insert of the expected size. As an alternative, primers which anneal to the plasmid either side of the insert site were used to screen candidates by colony PCR. Candidates testing positive for an insert were sequenced.

2.9.8 Sequencing

Plasmid-to-profile sequencing was carried out by the Functional Genomics and Proteomics Laboratory, University of Birmingham, UK. For sequencing of plasmid templates, 6.8 µl plasmid miniprep was mixed with 3.2 µl of 1 µM sequencing primer. For sequencing of PCR products, 3 µl purified PCR product was mixed with 3 µl of 1 µM sequencing primer and 4 µl of sterile distilled water. Primers used for sequencing inserts in plasmids are listed in Table 2.4.

2.10 Construction of pBAD-SG plasmid

For DNA sampling experiment, pBAD-SG (Amp^R) plasmid was constructed and used as an alternative of pACBSR-DL1 plasmid (Cm^R) particularly when pSI02 (Cm^R) was needed to co-transform in *E. coli* cells. The pBAD-SG was generated by introducing the sequence encoding both the SceI and the Gam proteins from pACBSR-DL1 to the MulI-EcoRI restriction sites in the pBAD plasmid (Guzman *et al*., 1995). The resulting plasmid could encode I-SceI meganuclease and Gam protein, which are necessary to obtain discreate and stable DNA fragments in *E. coli* cells during DNA sampling experiment.

The construction of pBAD-SG is shown schematically in Figure 2.9. The MluI-EcoRI insert fragment was derived by PCR from pACBSR-DL1 plasmid (Figure 2.7). PCR was carried out using an upstream primer D70504 (5'-AAAAACGCGTAACAAAAGTGTCTATAATCA

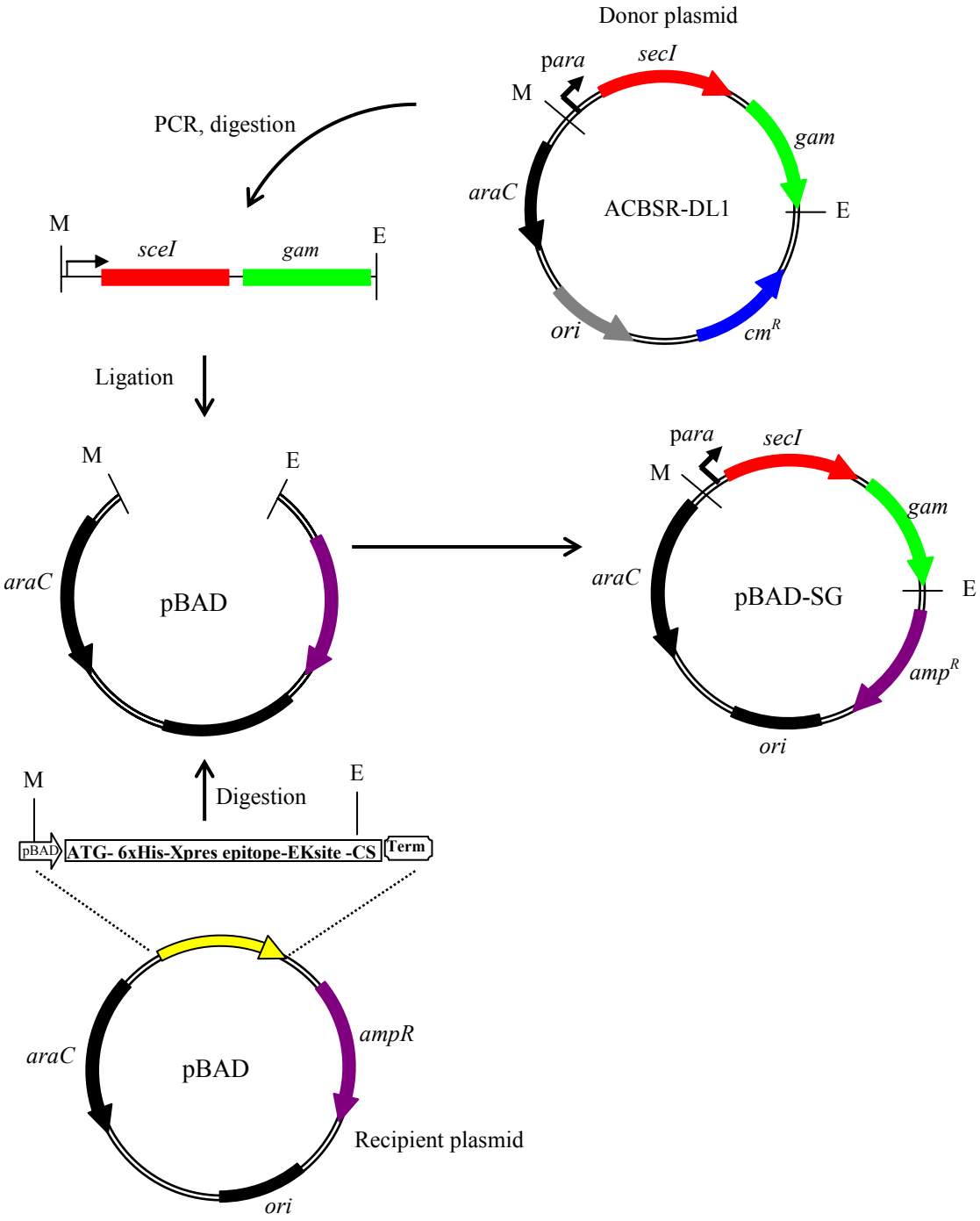


Figure 2.9. Construction of plasmid pBAD-SG.

Upstream primer D70504 that anneals upstream of sequence encoding SceI and downstream primer D70353 that anneals downstream of the sequence encoding Gam protein were used to amplify MluI-EcoRI fragment encoding SceI and Gam proteins using pACBSR-DL1 as template. Resulting fragment was digested with MluI-EcoRI restriction enzymes (M-E) and cloned into the MluI-EcoRI digested and dephosphorilated pBAD plasmid. The resulting plasmid is pBAD-SG containing sequences encoding SceI and Gam proteins under the control of an arabinose inducible *ara*BAD promoter (p*ara*).

-3'), which anneals upstream of the sequence encoding SceI and a downstream primer D70353 (5'-GACGAATTCGCCAAAACAGCCAAGCTTGC-3') that anneals downstream of the sequence encoding Gam protein. The PCR product was purified using the QIAquick PCR purification kit (see section 2.6.3), eluted in 50 µl sterile distilled water and digested using MluI and EcoRI. The MluI and EcoRI fragment was then purified by electroelution from a 7.5% polyacrylamide gel (see section 2.6.5).

To prepare pBAD vector DNA, pBAD/His was digested with MluI and EcoRI, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation (see sections 2.6.1, 2.6.2). The MluI and EcoRI insert fragment was ligated into pBAD, strain M182 was transformed with the ligation mix using the $CaCl_2$ method, and transformants were selected on nutrient agar supplemented with ampicillin. Plasmids were isolated from candidate colonies, digested with both MluI and EcoRI restriction enzymes and run on 1.3% agarose gel to check inserts. Positive candidates were sequenced using primers D70504 and D70353.

2.11 Cloning of promoter fragments and/or genes

A full list of oligonucleotide primers used to amplify different DNA fragments is shown in Table 2.4. All the promoter fragments or target genes were amplified from the *E. coli* O157: Sakai genomic DNA by PCR. However, when nested or in-phase deletion derivatives or point mutations at a specific promoter fragment were made, the plasmid containing the starting promoter fragment was used as template. In most cases, PCR primers were designed such that the forward and reverse primers incorporate restriction enzyme sites upstream and downstream of the amplified target region, respectively. These restriction sites in the DNA fragments are specific to the restriction sites of plasmid DNA and facilitate the fragments to be cloned within the plasmid.

2.11.1 Construction of pRW224 derivatives

Vector DNA was prepared from plasmid pRW224-U9 (see plasmid detail in Figure 2.1). Miniprep plasmid DNA was digested with EcoRI and HindIII, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation method (see sections 2.6.1, 2.6.2). PCR amplified target DNA fragments were digested with the same

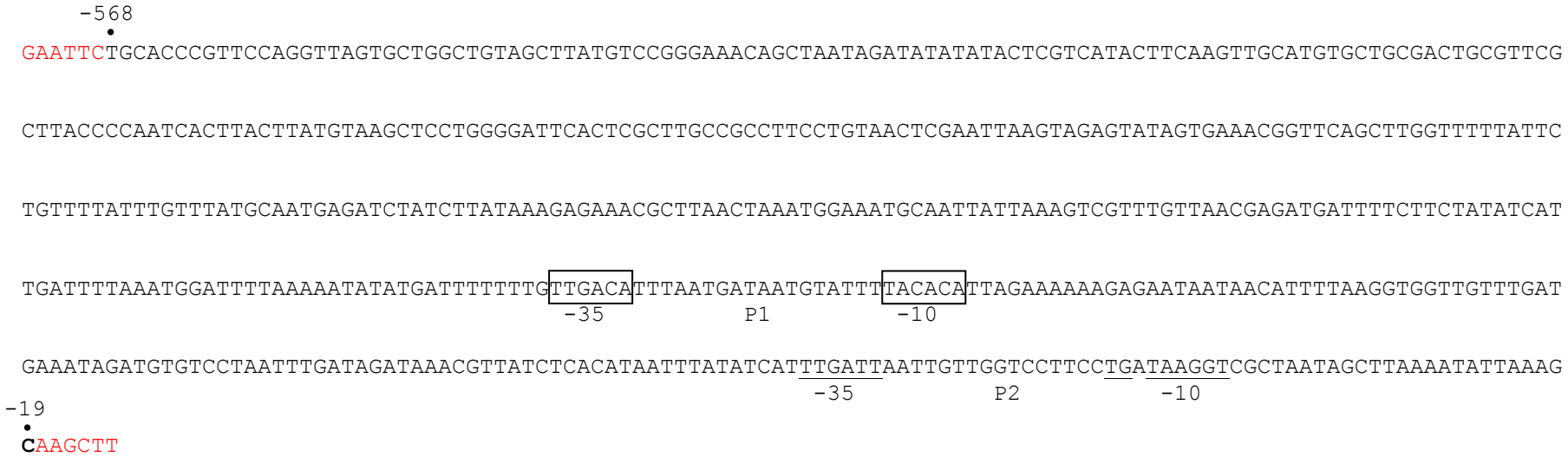
enzymes and purified using PCR purification kit or agarose gel extraction kit or electroelution from 7.5% polyacrylamide gel (see sections 2.6.3-2.6.5). Insert DNA was ligated into EcoRI -HindIII digested vector for 15 mins at room temperature. *E. coli* M182 competent cells were transformed with the ligation mix and transformants were selected for on MacConkey lactose agar supplemented with tetracycline. Screening of candidates for inserts was carried out by digesting plasmid DNA with EcoRI and HindIII. Positive candidates were sequenced using primers D10520 and/or D53463 which anneal to the plasmid DNA sequence flanking the insert site (Table 2.4).

LEE1 operon regulatory region is located upstream of the *ler* gene. PCR was used to amplify the LEE10-568 fragment using the D61221 and D61222 oligos and genomic DNA from the O157:H7 Sakai 813 strain. This fragment carries the base sequence from position -568 to position -19 upstream of the functional ATG start codon of the *ler* gene (Yerushalmi *et al.*, 2008). The DNA sequence of the LEE10-568 fragment is shown in the Figure 2.10. The fragment was designed with flanking EcoRI and HindIII sites respectively located upstream and downstream of the *LEE1* regulatory region, with the HindIII site positioned such that the fragment excluded the *ler* gene SD sequence. The resulting product was restricted with EcoRI and HindIII and cloned into pRW224 (Figure 2.11).

In order to determine the relative activity of *LEE1* P1 and P2 promoters, different fragments in the LEE10 and LEE20 or LEE30 sets of nested deletions were made using primers listed in the Table 2.4. Each fragment is described by the location of its upstream end, in terms of the number of bases upstream from the functional *ler* ATG codon. DNA sequence of each of the nested deleted fragments is shown in Figure 2.10. All of these fragments were cloned into the EcoRI-HindIII sites of the pRW224 as transcription fusion to *lacZ* (Figure 2.11). The resulting plasmids are recorded in Table 2.2.

To identify functional elements of the *LEE1* P1 promoter or to determine the GrlA target site in the *LEE1* promoter sequence, different mutations were introduced into the LEE20-275 and LEE20-203 fragments, respectively using error prone PCR (Leung *et al*., 1989), or by using megaprimer PCR (Perrin and Gilliland, 1990) as in Chismon *et al*. (2010). All the primers are

LEE10-568



LEE10-315

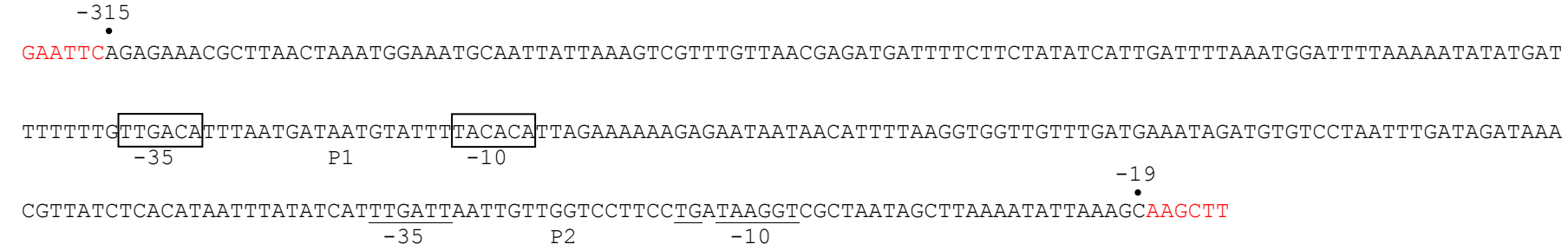
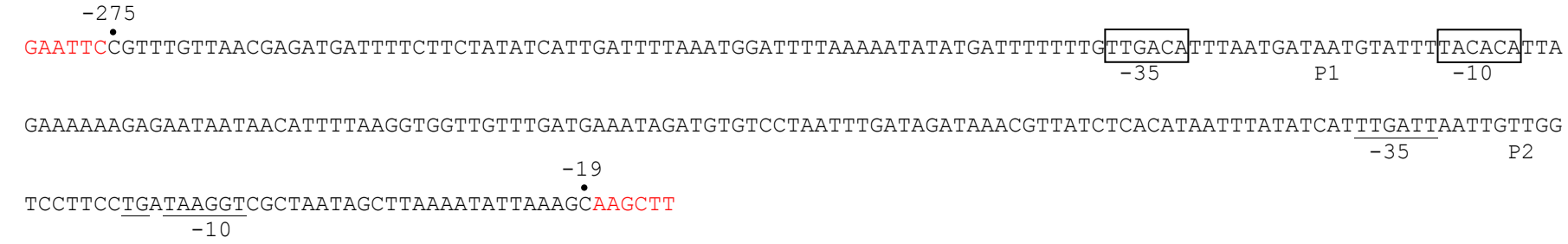
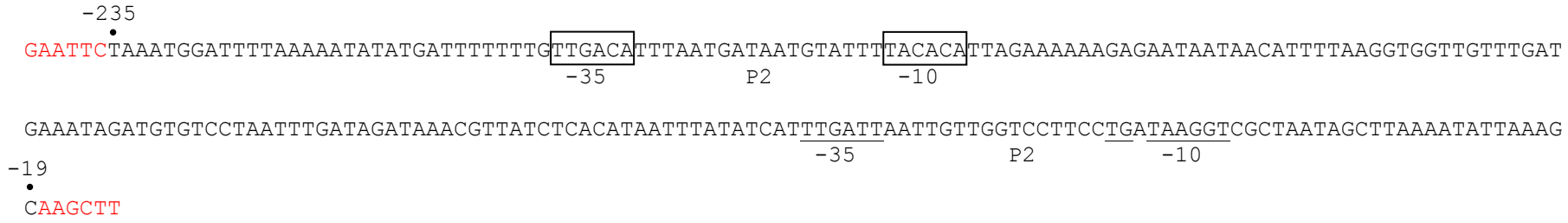


Figure 2.10. DNA sequences of the derivatives of *LEE1* promoter regulatory region (continued on pages 100-103).

LEE10-275



LEE10-235



LEE10-215

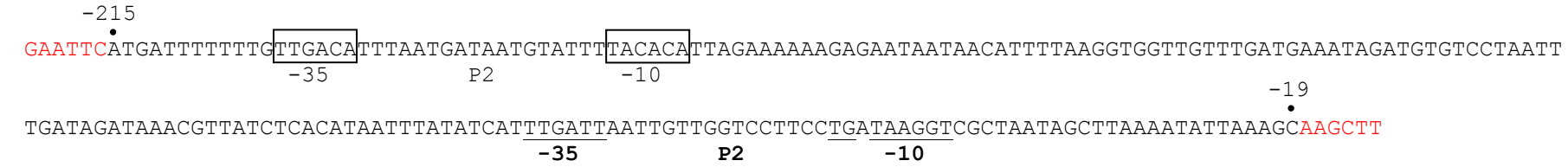


Figure 2.10. DNA sequences of the derivatives of *LEE1* promoter regulatory region (continued).

LEE10-203

```
     -203
GAATTCTTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAATAACATTTTAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAATTTGATAGATAAAC
       -35           P1          -10
                                                                                -19
GTTATCTCACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTGATAAGGTCGCTAATAGCTTAAAATATTAAAGCAAGCTT
                          -35           P2          -10
```

LEE10-195

```
     -195
GAATTCTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAATAACATTTTAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAATTTGATAGATAAACGTTATCTC
                                                                          -19
ACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTGATAAGGTCGCTAATAGCTTAAAATATTAAAGCAAGCTT
                  -35         P2           -10
```

LEE10-155

```
     -155
GAATTCTAACATTTTAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAATTTGATAGATAAACGTTATCTCACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTG
                                                                                  -35           P2
                                -19
ATAAGGTCGCTAATAGCTTAAAATATTAAAGCAAGCTT
  -10
```

LEE10-115

```
     -115                                                                                                    -19
GAATTCTAATTTGATAGATAAACGTTATCTCACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTGATAAGGTCGCTAATAGCTTAAAATATTAAAGCAAGCTT
```

Figure 2.10. DNA sequences of the derivatives of *LEE1* promoter regulatory region (continued).

LEE10-75

```
     -75                                                        -19
      •                                                          •
GAATTCTTTGATTAATTGTTGGTCCTTCCTGATAAGGTCGCTAATAGCTTAAAATATTAAAGCAAGCTT
        -35        P2            -10
```

LEE20-568

```
    -568
     •
GAATTCTGCACCCGTTCCAGGTTAGTGCTGGCTGTAGCTTATGTCCGGGAAACAGCTAATAGATATATATACTCGTCATACTTCAAGTTGCATGTGCTGCGACTGCGTTCG

CTTACCCCAATCACTTACTTATGTAAGCTCCTGGGGATTCACTCGCTTGCCGCCTTCCTGTAACTCGAATTAAGTAGAGTATAGTGAAACGGTTCAGCTTGGTTTTTATTC

TGTTTTATTTGTTTATGCAATGAGATCTATCTTATAAAGAGAAACGCTTAACTAAATGGAAATGCAATTATTAAAGTCGTTTGTTAACGAGATGATTTTCTTCTATATCAT

                                                                                    -158
                                                                                     •
TGATTTTAAATGGATTTTAAAAATATATGATTTTTTTGTTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAAGCTT
                                      -35          P1           -10
```

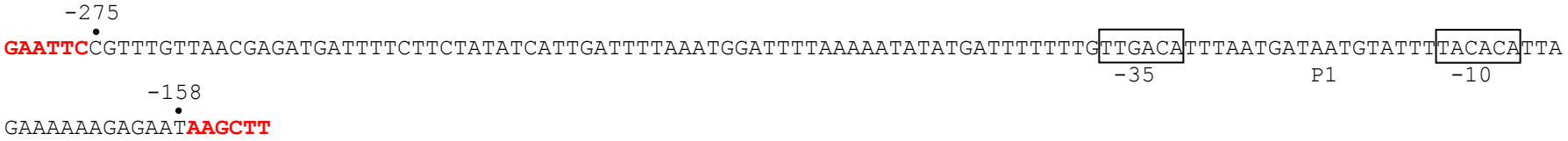
LEE20-315

```
    -315
     •
GAATTCAGAGAAACGCTTAACTAAATGGAAATGCAATTATTAAAGTCGTTTGTTAACGAGATGATTTTCTTCTATATCATTGATTTTAAATGGATTTTAAAAATATATGAT

                                                    -158
                                                     •
TTTTTTGTTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAAGCTT
        -35        P2            -10
```

Figure 2.10. DNA sequences of the derivatives of *LEE1* promoter regulatory region (continued).

LEE20-275



LEE20-203

LEE30-275

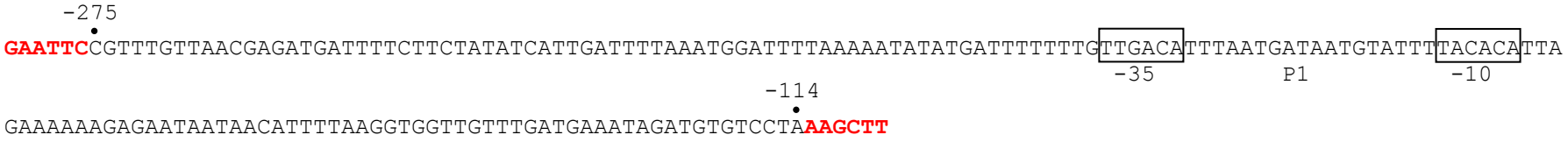


Figure 2.10. DNA sequences of the derivatives of *LEE1* promoter regulatory region (continued).

The figure shows the DNA sequences of different nested deleted fragments of *LEE1* regulatory region. EcoRI and HindIII sites are red coloured. Predicted P1 promoter -35 and -10 elements are boxed and predicted P2 promoter -35 and -10 elements are underlined (Sperandio *et al*., 2002). The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene reported by Yerushalmi *et al*. (2008).

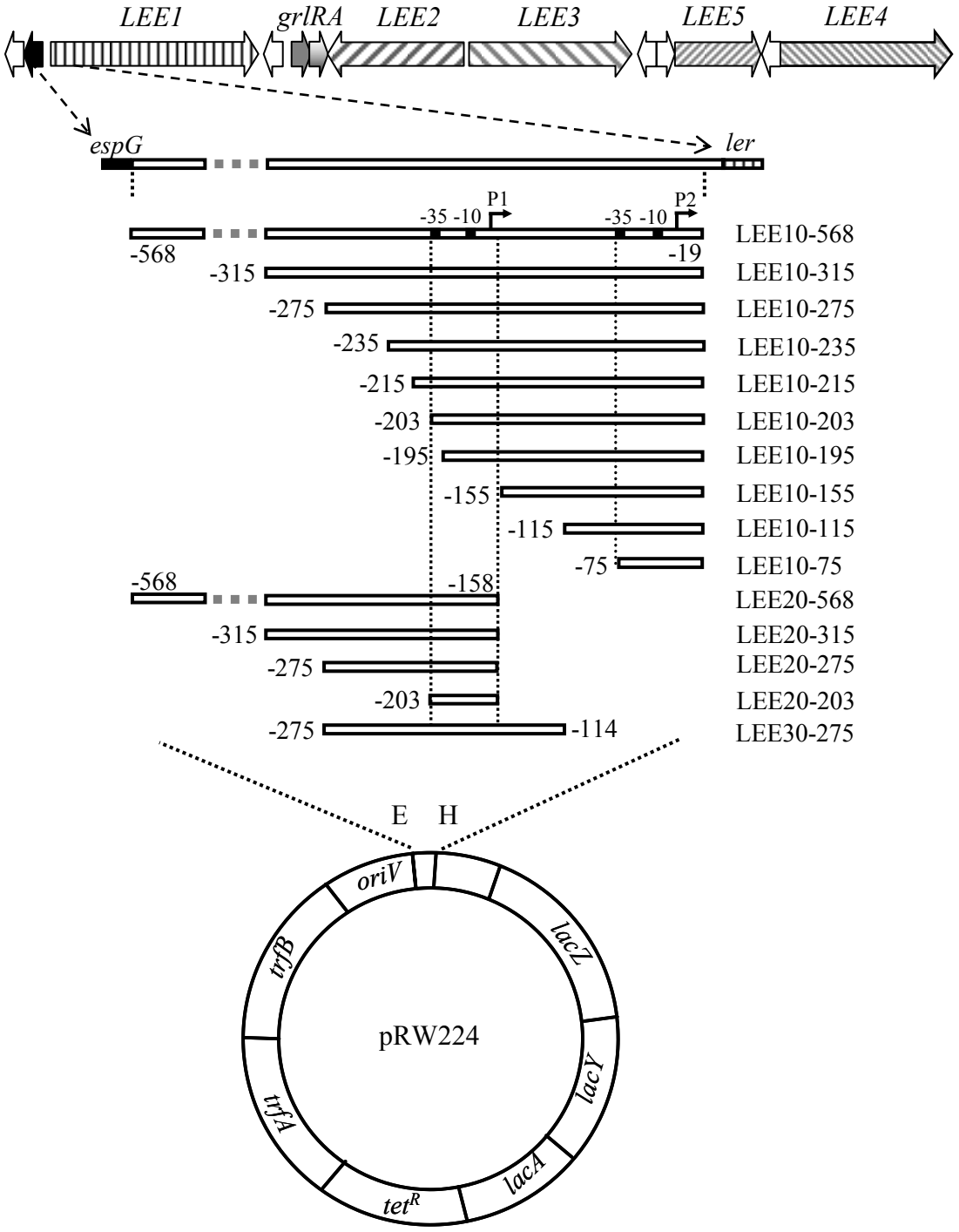


Figure 2.11. Nested deletions of the *LEE1* promoter regulatory region.

The upper line shows the organisation of the different LEE transcription units. The lower part shows an expanded sketch of the *LEE1* regulatory region and illustrates the LEE10-568 fragment and nested deleted derivatives cloned into the EcoRI-HindIII sites (E-H) of pRW224 *lac* expression vector (see plasmid detail in Figure 2.1). The locations of the two *LEE1* regulatory region promoters P1 and P2 are indicated. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene reported by Yerushalmi *et al*. (2008).

listed in Table 2.4. The different bases at the *LEE1* P1 promoter in the LEE20-275 fragment are numbered 1-118, as shown in Figure 3.3. This numbering system was used to describe different P1 promoter mutations, in both the LEE20-275 and LEE20-203 fragments (Figure 3.3, 3.11). Note that the customary convention of numbering bases with respect to the transcript start point was not followed because of uncertainty about its location.

In order to understand regulation of expression from *LEE2* promoter, different *LEE2* operon regulatory region fragments were cloned into pRW224. Each fragment is described by the location of its upstream end, in terms of the number of bases upstream from the transcription start site reported by Mellies *et al*. (1999) (Figure 2.12). Primers used for amplification of the *LEE2* operon regulatory region are listed in the Table 2.4. LEE2-220 (-220 to +109) fragment was PCR amplified using primers D62263 and D62265 and genomic DNA of Sakai: O157 as template. The resulting product was restricted with EcoRI and HindIII, and cloned into pRW224. Two nested deletions were made of starting fragment LEE2-220. LEE2-125 (-125 to +109) was amplified using LEE2-220/pRW224 as template and primers D62865 and D62265. The resulting product was EcoRI-HindIII digested and cloned into pRW224 to give LEE2-125/pRW224. LEE2-100 (-100 to +109) fragment was made using LEE2-125/pRW224 as template and primers D62264 and D62265 (Figure 2.13, Table 2.2).

2.11.2 Construction of pSR derivatives

The cloning vector pSR was used to clone promoter fragments for *in vitro* $KMnO_4$ footprinting and transcription assays. Plasmid miniprep DNA was digested with EcoRI and HindIII, treated with alkaline phosphatase and purified by gel extraction from a 0.8% agarose gel. PCR amplified target promoter fragments were digested with the same enzymes and extracted from 1.3% agarose gel. EcoRI-HindIII promoter fragments were then ligated into the EcoRI-HindIII digested pSR vectors for 15 mins at room temperature and *E. coli* M182 competent cells were transformed with the ligation mix. Cells were plated onto nutrient agar supplemented with ampicillin and incubated overnight at 37°C to select for transformants. Candidates were screened for the presence of an insert of the expected size by restriction digestion with EcoRI and HindIII, and candidate plasmids testing positive for an insert were sequenced using primers D5431 and/or D56033, which anneal up or downstream of the insert site, respectively (Table 2.4).

LEE2-220

-220

GAATTCACGTAACAAAAACATTATTATCAATTATTTCAATAAAAAAAGAAAACCCATTAAAATCAATTTCAACACGGTTATCAAATAAATAAAAATCATTTCCCAATAATC

TTAAAAACTCTTCAACGTTTCTTTTAACTAAAAGATTCATCTGCAGGCTCTGAAGTAAGTATACGTGTGAATATTATTGCTTGAAAAAGCGTATTGGATAATATATACAGT

ATATGTACTTTATGTACATTGCAACTATTTAAATAATATAATAGTTGCCTATGGGATAATTTGGTTATTTATAACCAGGTAAATGTCTAAATTAGACAAAAGGAGAAAGAT

+109

CAAAGCTT

LEE2-125

+125

GAATTCCCCAATAATCTTAAAAACTCTTCAACGTTTCTTTTAACTAAAAGATTCATCTGCAGGCTCTGAAGTAAGTATACGTGTGAATATTATTGCTTGAAAAAGCGTATT

GGATAATATATACAGTATATGTACTTTATGTACATTGCAACTATTTAAATAATATAATAGTTGCCTATGGGATAATTTGGTTATTTATAACCAGGTAAATGTCTAAATTAG

+109

ACAAAAGGAGAAAGATCAAAGCTT

LEE2-100

+100

GAATTCCGTTTCTTTTAACTAAAAGATTCATCTGCAGGCTCTGAAGTAAGTATACGTGTGAATATTATTGCTTGAAAAAGCGTATTGGATAATATATACAGTATATGTACT

+109

TTATGTACATTGCAACTATTTAAATAATATAATAGTTGCCTATGGGATAATTTGGTTATTTATAACCAGGTAAATGTCTAAATTAGACAAAAGGAGAAAGATCAAAGCTT

Figure 2.12. DNA sequences of the *LEE2* promoter regulatory region.

EcoRI and HindIII sites are coloured red. Predicted -35 and -10 promoter elements are boxed and transcription start site is indicated by bent arrow (Mellies *et al*., 1999). Nucleotide positions are numbered with respect to the transcription start site. Starting with LEE2-220 fragment, LEE2-125 and LEE2-100 were generated by deletion of upstream sequences from positions -125 and -100, respectively.

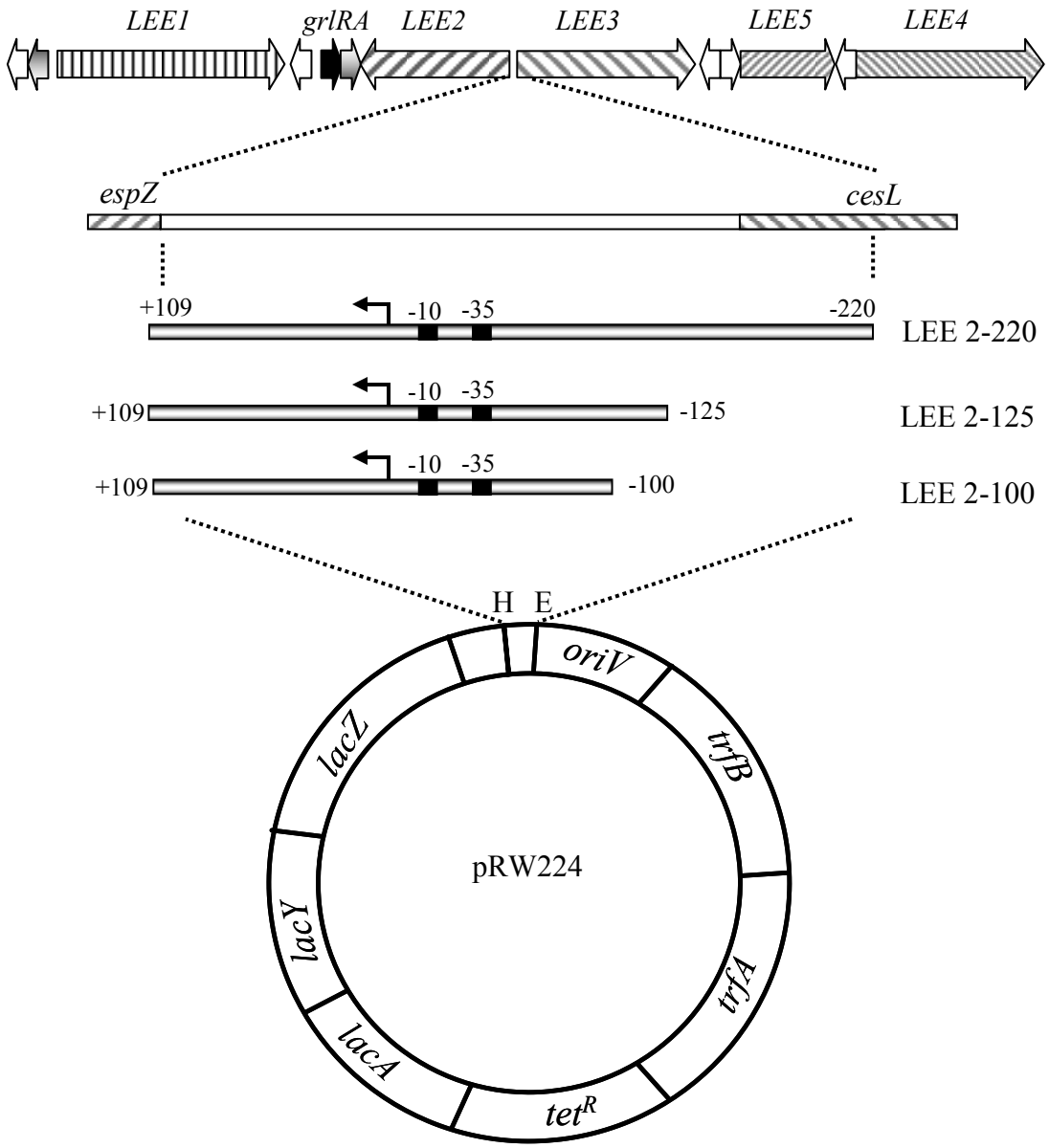


Figure 2.13. Nested deletions of the *LEE2* promoter regulatory region.

The upper line shows the organisation of the different LEE transcription units. The lower part shows an expanded sketch of the *LEE2* regulatory region and illustrates the LEE2-220 fragment and nested deleted derivatives cloned into the EcoRI-HindIII sites (E-H) of pRW224 *lac* expression vector (see plasmid detail in Figure 2.1). The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream and downstream from the transcription start point (Mellies *et al*., 1999). Predicted promoter -35 and -10 elements are shown by black boxes and transcription star site are indicated by bent arrow.

For both *in vitro* $KMnO_4$ footprinting analysis and *in vitro* transcription assays, LEE20-275 and two derivatives such as LEE20-275 98C and LEE20-275 98C 64G (Figure 2.14) were cloned into EcoRI-HindIII sites of pSR. Fragments were amplified from pRW224 containing LEE20-275 or LEE20-275 98C or LEE20-275 98C 64G using primers D63949 and D64100 and cloned into EcoRI-HindIII sites of pSR resulting in plasmids LEE20-275/pSR, LEE20 -275 98C/pSR and LEE20-275 98C 64G/pSR, respectively (Table 2.2). Linear fragments were produced following PstI and BamHI restriction digestion during performing both *in vitro* $KMnO_4$ footprinting analysis and *in vitro* transcription assays (Figure 2.15).

2.11.3 Construction of pACYC184 derivatives

Vector DNA was prepared from pACYC184 plasmid (see plasmid detail in Figure 2.3). Plasmid miniprep DNA was digested sequentially using HindIII and SalI restriction enzymes, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation method (see section 2.6). PCR amplified target DNA fragments were digested with the same enzymes and extracted from 1.3% agarose gel. Insert DNA was ligated into HindIII-SalI digested pACYC184 vector for 15 mins at room temperature. *E. coli* M182 competent cells were transformed with the ligation mix and transformants were selected for on nutrient agar supplemented with chloramphenicol. Screening of candidates for an insert was carried out by digesting plasmid DNA with HindIII and SalI. Positive candidates were sequenced using primers D63048 and D63049, which anneal up- and downstream of the insert site, respectively (Table 2.4).

In order to determine the effect of GrlR/GrlA on the activity of the *LEE1* promoter, the *grlRA* operon and derivatives were cloned into pACYC184. The DNA sequence of *grlRA* operon is shown in Figure 2.16. Maps of *grlRA* operon and the derivatives generated upon it by deletion mutation are shown in Figure 2.17. PCR was used to amplify a HindIII-SalI fragment carrying the *grlRA* operon using the D62895 and D62897 oligos (Table 2.4) and genomic DNA from the O157:H7 Sakai 813 strain. The resulting product was restricted with HindIII and SalI and cloned into pACYC184 to give pSI01 (Table 2.2). To construct pSI02, which is a derivative of pSI01 carrying a large in-frame deletion in *grlR*, PCR amplification was done using primers D63209 and D63210 (Table 2.4) and pSI01 as a template. The resulting product was cut with BamHI and circularised by ligation to give pSI02 (Table 2.2). To construct pSI03,

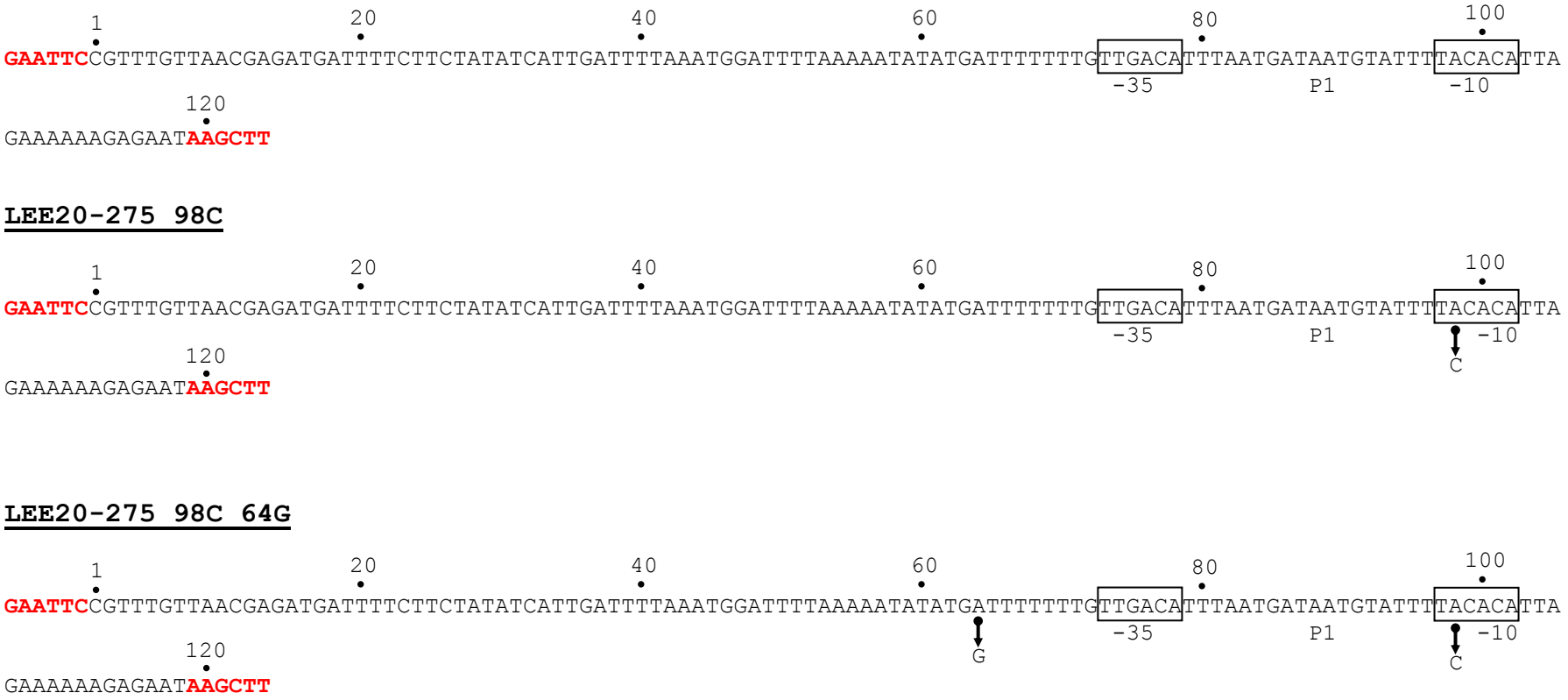


Figure 2.14. DNA sequences of the LEE20-275 promoter fragment and derivatives.

EcoRI and HindIII sites are red coloured. Nucleotide positions are numbered, 1-120, starting with the first cloned *LEE1* regulatory region base (that is 275 base pairs upstream from the *ler* translation start codon). Base substitution at 98 position in LEE20-275 98C fragment and substitutions at 64 and 98 positions in LEE20-275 98C 64G are indicated by the arrows.

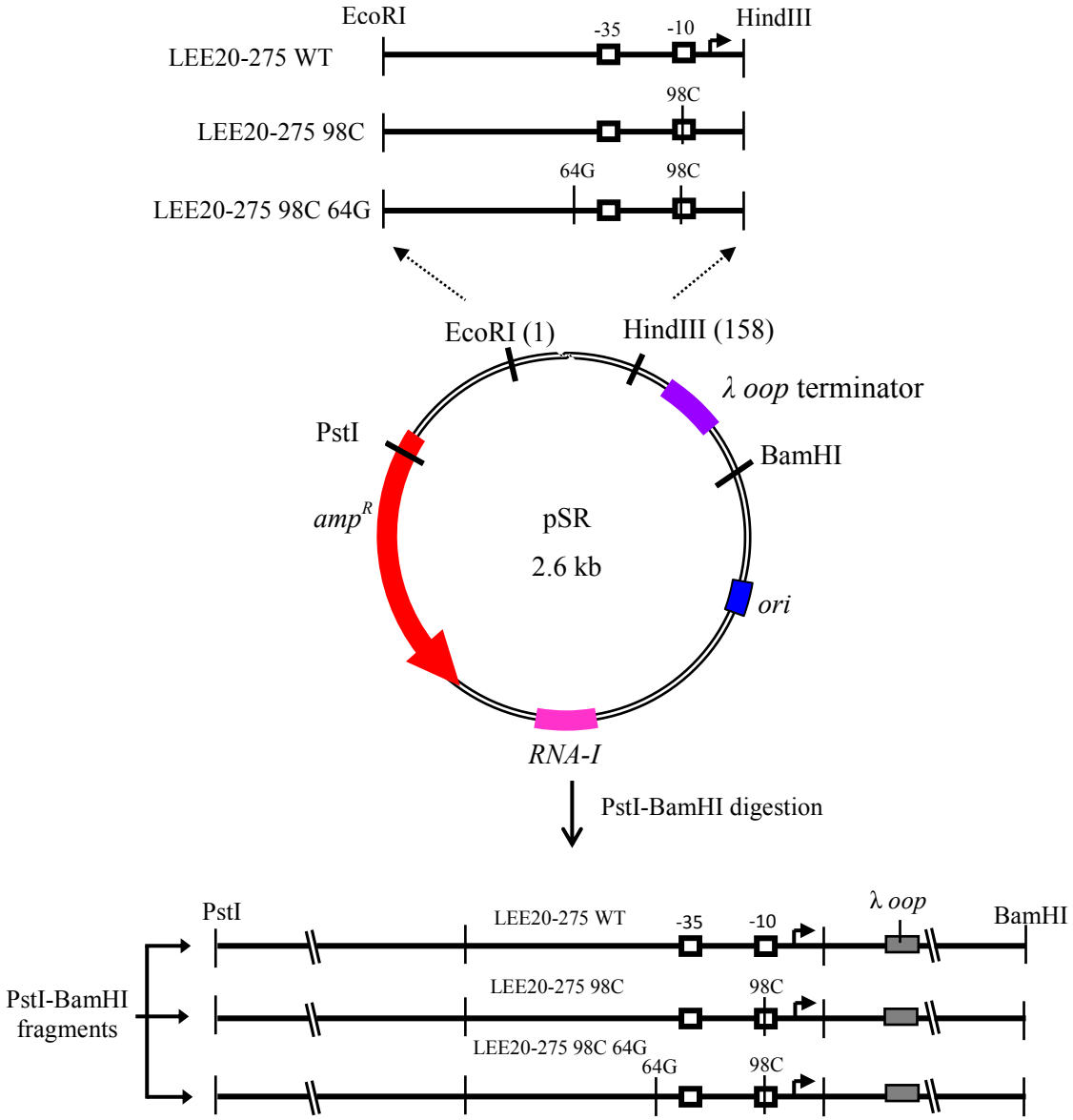


Figure 2.15. Construction of pSR derivatives and preparation of PstI-BamHI fragments.

EcoRI-HindIII promoter fragments such as LEE20-275, LEE20-275 98C and LEE20-275 98C 64G were cloned into upstream of the λ *oop* terminator site of a multicopy plasmid pSR. Plasmids containing the promoter fragments were digested with BamHI, dephosphorylated, purified and digested with PstI to get PstI-BamHI fragments. The fragments contain a λ *oop* terminator downstream of the cloned promoters. These fragments were used for *in vitro* $KMnO_4$ footprinting and transcription assays. Transcription initiates at the cloned promoter and terminates at the λ *oop* terminator, and thus produces transcripts of defined lengths.

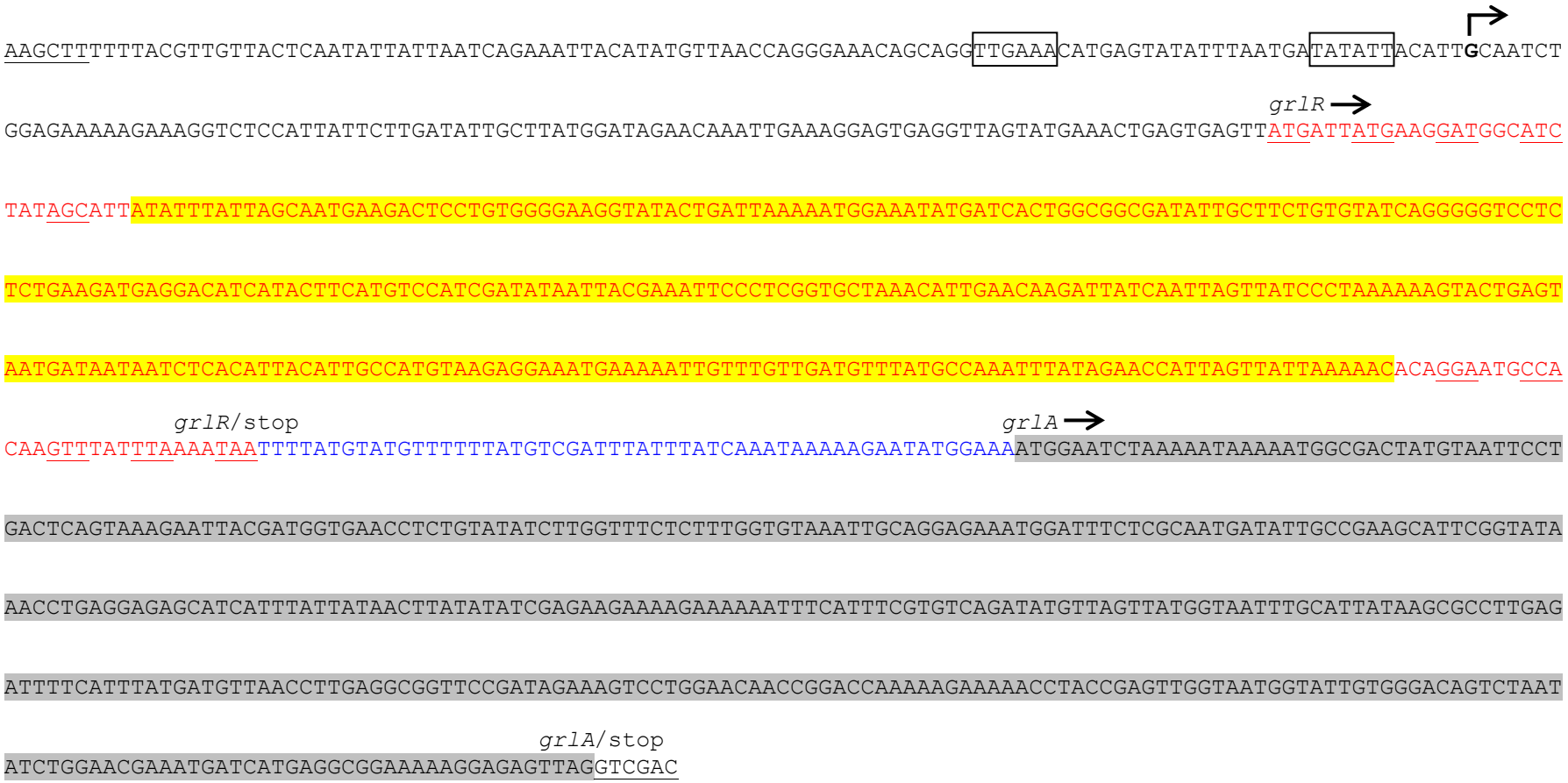


Figure 2.16. DNA sequence of the *grlRA* operon.

HindIII-SalI restriction sites are underlined. Predicted promoter -35 and -10 elements are boxed and transcription start site is indicated by a bent arrow (Mellies *et al*., 1999). Shown that *grlR* (red coloured sequence) and *grlA* (grey shaded sequence) genes are separated by a short intergenic DNA sequence (blue coloured sequence). Predicted start and stop codons of *grlR* and *grlA* genes are indicated. The *grlR* base sequence that is shaded yellow was deleted to give *grlR-A+* derivative (see Figure 2.17).

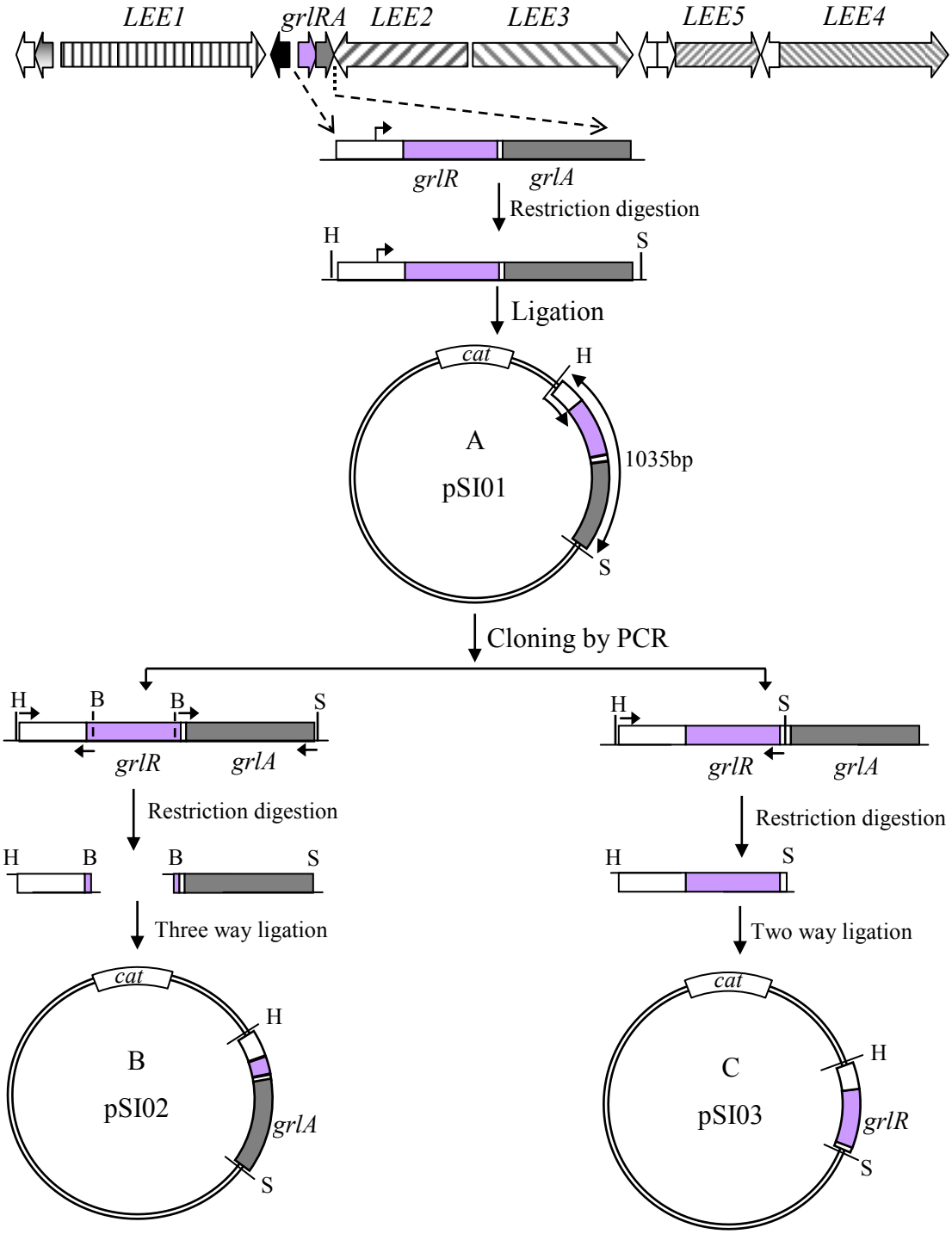


Figure 2.17. Schematic representation of construction of recombinant plasmids pSI01, pSI02 and pSI03.

(A) *grlRA* operon was amplified from *E. coli* 0157: Sakai chromosome digested and cloned into pACYC184 resulting in pSI01 (GrlR+A+).
(B) An in-phase deletion in the *grlR* gene of *grlRA* operon leads to the excision of most part of the gene (BamHI deletion of 312 bp), which results in pSI02 (GrlR-A+).
(C) Amplification of *grlR* gene with its regulatory sequence and cloning it into the pACYC184 resulting in pSI03 (GrlR+A-). H, S and B represent HindIII, SalI and BamHI, respectively.

which is a derivative of pSI01 deleted for *grlA*, primers D62895 and D63698 (Table 2.4) and pSI01 as a template were used for amplification. The resulting product was restricted with HindIII and SalI and cloned into pACYC184 to give pSI03 (Table 2.2).

2.11.4 Construction of pRW902 derivatives

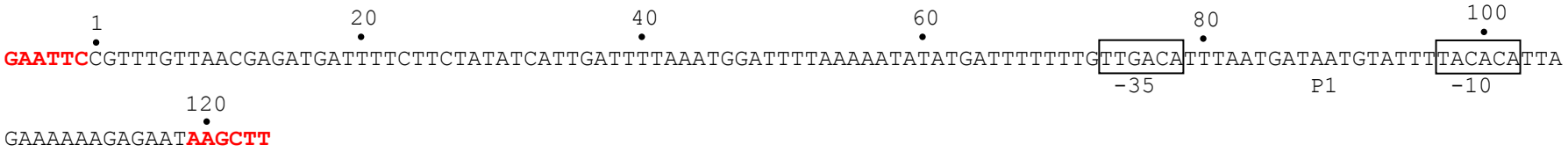
Vector DNA was prepared from pRW902 plasmid. Plasmid miniprep DNA was double digested using EcoRI and HindIII restriction enzymes, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation method (see section 2.6). PCR amplified target DNA fragments were digested with the same enzymes and extracted from 1.3% agarose gel. Insert DNA was ligated into EcoRI-HindIII digested pRW902 vector for 15 mins at room temperature. *E. coli* M182 competent cells were transformed with the ligation mix and transformants were selected for on MacConkey lactose indicator plate supplemented with tetracycline. Screening of candidates for an insert was carried out by digesting plasmid DNA with EcoRI-HindIII. Positive candidates were sequenced using primers D10520, which anneal to the plasmid DNA sequence flanking the insert site (Table 2.4).

For DNA-sampling experiments, three fragments for example, LEE20-275, LEE30-275 and LEE30-275 151T (Figure 2.18) were cloned into pRW902. LEE20-275 fragment was PCR amplified using D63949 and D64100 primer pair (Table 2.4) from LEE20-275/pRW224 plasmid DNA. LEE30-275 and LEE30-275 151T fragments were PCR amplified from LEE30-275/pRW224 and LEE30-275 151T/pRW224 plasmid DNA, respectively using D63949 and D64101 or D63949 and D53463 primer pairs, respectively (Table 2.4). The fragments were double digested with EcoRI-HindIII and cloned into the EcoRI-HindIII sites of the vector pRW902 (Figure 2.4, Table 2.2).

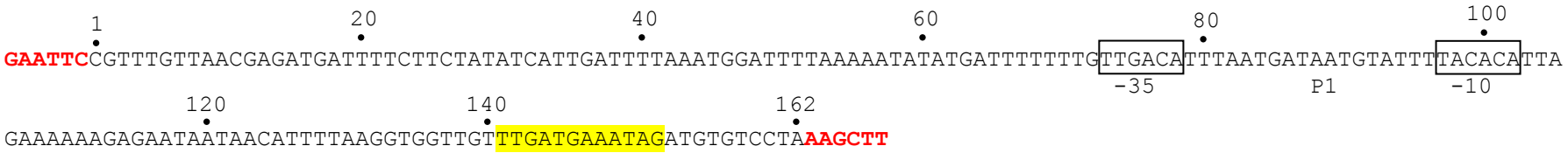
2.11.5 Construction of pJW15Δ100 derivatives

Vector DNA was prepared from pJW15Δ100 plasmid (see plasmid detail in Figure 2.5). Plasmid miniprep DNA was digested with NsiI and HindIII restriction enzymes, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation method (see section 2.6). PCR amplified target DNA fragments were digested with the same enzymes and extracted from 1.3% agarose gel. Insert DNA was ligated into NsiI-HindIII

LEE20-275



LEE30-275



LEE30-275 151T

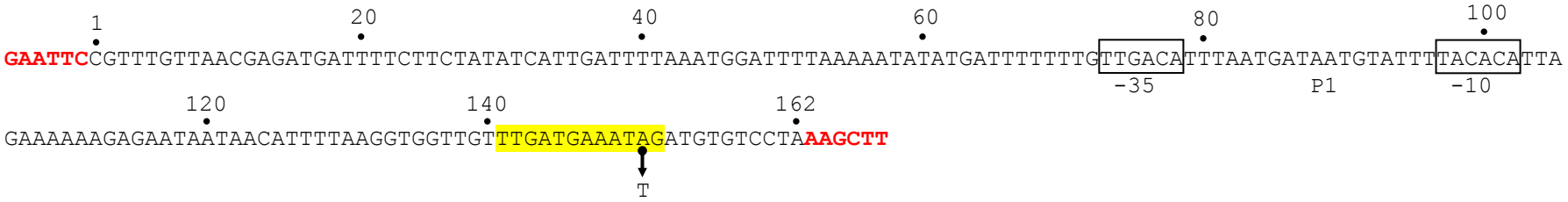


Figure 2.18. DNA sequences of LEE20-275, LEE30-275 and LEE30-275 151T promoter fragments.

EcoRI and HindIII sites are coloured red. Nucleotide sequences are numbered starting with the first cloned *LEE1* regulatory region base. Predicted promoter elements are boxed. Yellow shaded sequence corresponds to the predicted small ORF both in LEE30-275 and LEE30-275 151T fragments and a single base change at position 151 in LEE30-275 151T fragment is indicated by arrow.

digested pJW15Δ100 vector for 15 mins at room temperature. *E. coli* M182 competent cells were transformed with the ligation mix and transformants were selected for on nutrient agar plates supplemented with ampicillin. Screening of candidates for an insert was carried out by digesting plasmid DNA with NsiI and HindIII. Positive candidates were sequenced using primers D3407 and/or D4600 which anneal to the plasmid DNA sequence flanking the insert site (Table 2.4).

DNA sequences of different genes cloned into pJW15Δ1-100 are shown in Figure 2.19. To construct pSI04, the *ler* gene (390 bp) was PCR amplified using genomic DNA of EHEC Sakai as template and primers D61652 and D61223 (Table 2.4). The resulting product was digested with NsiI and HindIII cloned into pJW15Δ100 (Figure 2.20). The *pchC* gene (315 bp) was PCR amplified using genomic DNA of EHEC Sakai strain as template and primers D61658 and D61659 (Table 2.4). The resulting product was digested with NsiI and HindIII and cloned into pJW15Δ100 to give pSI05 (Figure 2.20). The *cesD* (456 bp) was amplified using genomic DNA of EHEC Sakai strain as template and primers D64149 and D64150 (Table 2.4). The PCR product was then cloned into the NsiI-HindIII sites of the vector pJW15Δ100 to give pSI06 (Figure 2.20). Plasmid pSI07 was constructed by amplifying the *cesT* (471 bp) gene using genomic DNA of EHEC Sakai strain as template and primers D64151 and D64152 (Table 2.4) and cloning the product after NsiI-HindIII restriction digestion into the pJW15Δ100 vector (Figure 2.20). To construct pSI08, *cesL* gene (354 bp) (Younis *et al*., 2010) was amplified using EHEC Sakai genomic DNA as template and primers D64149 and D64150 (Table 2.4). The resulting product was cloned into the NsiI-HindIII sites of the vector pJW15Δ100 (Figure 2.20).

2.11.6 Construction of pRW225 derivatives

Vector DNA was prepared from pRW225. Plasmid miniprep DNA was digested with EcoRI and HindIII or EcoRI and BamHI, treated with alkaline phosphatase and purified by phenol/chloroform extraction and ethanol precipitation method (see section 2.6). Insert DNA was ligated into EcoRI-HindIII or EcoRI-BamHI digested vector for 15 mins at room temperature. *E. coli* M182 competent cells were transformed with ligation mix and transformants were selected on MacConkey lactose indicator plates supplemented with tetracycline. Screening of candidates for an insert was carried out by digesting plasmid DNA

***ler*(390bp)**

Δ

ATGCATATGCGGAGATTATTTATTATGAATATGGAAAATAATTCACATACAACAAGTCCATACATTCAGCTTATAGAGCAAATTGCAGTTCTACAGCAGGAAGCAAAGCGA

CTGCGAGAGCAGGAAGTTCAAAGTGTAATTGAGTCGATTCAGAAGCAGATTACTTATTACAATATAACCTTACAAGAGCTGGGATATACTAATGTGCCTGATGATGGACTC

GCTCGCCGGAACTCATCGAAAGGTGTTTACTACCGCAATGAAGAAGGGCAGACCTGGTCGGGCGTAGGCCGACAGCCACGCTGGCTTAAAGAAGCACTGTTGAATGGAATG

AAGAAAGAAGATTTTCTTGTGAAGGACACTGAAGAAGAAATAATACCGCTGAAAAATATTTAAAAGCTT

***pchC*(315bp)**

Δ

ATGCATATGCTACATGATCACCTGGCAGAATGTCTGGAGAAAAAAGGACTGTACCGGAGAGCAGCTGAACGATGGGCAAAAGTGATGGTACAGCTAAGTGATGACCAGAAA

AGAAAAGTGGCGGCACAGAAACGAGCAGAGTGTTTGCGTAAGGCGCGCCGGACTCCGGTTTCACCGATGAACCTGACAGAAATAAAACAAGCGGTCAACAGACTACATTCT

GAGTTGGGAATGGGATTTGAAGAGCGGCGGGTATTCCGACGATATAAAGGGACAGGAGAACAGAATACGTCCGGAAACGCGCGGTCAAAAAAATGCTAAAAGCTT

***cesD*(456bp)**

Δ

ATGCATATGAGCAGGAAATTTAGCTCTCTAGAGGATATTTATGATTTCTACCAGGATGGTGGCACATTAGCGTCATTAACAAATCTGACACAACAAGATCTCAATGACCTT

CATTCTTATGCCTATACAGCATATCAATCTGGTGATGTTATAACCGCAAGAAATCTATTCCATTTGCTCACATATCTGGAACACTGGAATTATGACTACACCTTATCTCTG

GGCTTATGTCATCAGCGTTTATCAAATCATGAAGATGCACAACTGTGTTTCGCACGCTGTGCAACTTTAGTTATGCAAGATCCCAGGGCATCTTATTATTCTGGAATTAGC

TACTTACTCGTCGGAAATAAGAAAATGGCCAAGAAAGCCTTTAAGGCTTGTTTAATGTGGTGTAATGAAAAAGAAAAATACACTACATATAAAGAAAATATTAAAAAATTG

TTAGGTAATACAGAGTAAAAGCTT

Figure 2.19. DNA sequences of *ler, pchC, cesD, cesT* and *cesL* genes (continued on page 117).

***cesT*(471bp)**

NsiI Δ

ATGCATATGTCATCAAGATCTGAACTTTTATTAGAAAAATTTGCAGAAAAAATTGGTATTGGATCTATTTCATTTAATGAAAACAGATTGTGTTCTTTTGCTATTGATGAA

ATTTATTATATTTCGTTATCTGATGCCAATGACGAATATATGATGATTTATGGTGTCTGTGGGAAATTCCCGACAGATAACTCTAACTTCGCTCTTGAGATTTTGAATGCA

AACTTATGGTTTGCAGAGAATGGTGGGCCATATCTGTGCTATGAGGCTGGAGCACAATCGCTGTTGTTAGCGTTACGTTTCCCTCTCGATGATGCTACCCCTGAAAAACTC

GAGAATGAAATAGAAGTCGTTGTTAAGTCAATGGAAAACCTGTATTTGGTATTACATAATCAGGGAATAACATTAGAAAACGAACATATGAAAATAGAGGAAATCAGTTCA

AGCGACAATAAACATTATTACGCCGGAAGATAAAAGCTT

***cesL* (354bp)**

NsiI Δ

ATGCATATGCATAATCTTTTAGTTAAAAGAAACGTTGAAGAGTTTTTAAGATTATTGGGAAATGATTTTTATTTATTTGATAACCGTGTTGAAATTGATTTTAATGGGTTT

TCTTTTTTTTATTGAAATAATTGATAATAATGTTTTTGTTACGTTTGCTTTAGAGTATAACGAAAACGCATTTTTCTCTTTCTTTAGTGCCCTTGCTCCTGAGCGTACGCAG

GGAGTGATTGAACATATTTTTGTCTACGATAATAAATTGTGCTTAAGTTGTTTGTTAACCAATATCGATGTTTTTTTTCTAATGAATACTTTTCAACAGCATGTGCAGATT

ATTGAGCGCGTTCGCAGGATGACATCATGAAAGCTTAAGCTT

Figure 2.19. DNA sequences of *ler, pchC, cesD, cesT* and *cesL* genes (continued).

The figure shows the base sequences together with restriction maps of *ler, pchC, cesD*, *cesT*, and *cesL* genes. NsiI and HindIII restriction sites are coloured red. Translation start codon (ATG) of each gene, indicated by yellow colour was deleted (Δ) since NsiI restriction site contains ATG.

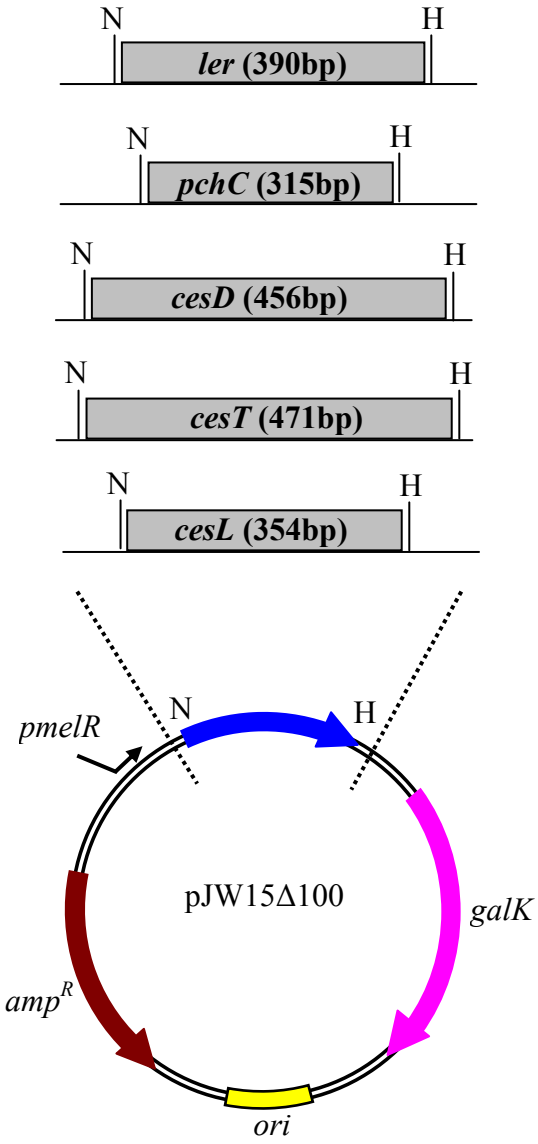


Figure 2.20. Cloning of genes into plasmid pJW15Δ100.

Promoterless *ler*, *pchC*, *esD*, *cesT and cesL* genes were amplified from *E. coli* O157: Sakai chromosome with NsiI-HindIII sites. Amplified fragments were digested and cloned into the NsiI-HindIII sites of the cloning vector pJW15Δ100 that contains *melR* promoter (see plasmid detail in Figure 2.5) resulting in plasmids pSI04, pSI05, pSI06, pSI07 and pSI08, respectively. N and H represent NsiI and HindIII restriction sites, respectively.

with EcoRI and HindIII or EcoRI and BamHI. Positive candidates were sequenced using primers D10520 or D53463 (Table 2.4), which anneal to the plasmid DNA sequence flanking the insert site.

To determine experimentally whether the upstream mini-ORF located in the downstream site of LEE30-275 fragment (Figure 2.18) is translated, the LEE30-275 fragment was cloned into the EcoRI-BamHI sites of plasmid pRW225 as translation fusion to *lacZ*. In order to incorporate the BamHI restriction site in the downstream end of the EcoRI-HindIII LEE30-275 fragment and to make sure that the predicted translation start codon in the LEE30-275 fragment could be in frame with *lacZ* gene in the plasmid, the downstream primer D65474 containing BamHI site together with the upstream primer D63949 was used to amplify the EcoRI-BamHI LEE30-275 fragment from plasmid LEE30-275/pRW224. The resulting LEE30-275 fragment was then cloned into EcoRI-BamHI sites of pRW225 to give LEE30-275/pRW225 (Figure 2.21).

To experimentally determine the *ler* functional translation start site, LEE151 (553 bp) containing the predicted translation start codon that is assigned in coliBASE (Chaudhuri *et al*., 2004) (Figure 2.22) was PCR amplified using primers D61221 and D65812 (Table 2.4) and cloned into the EcoRI-HindIII sites of pRW225, which results in the plasmid LEE151/pRW225 (Table 2.2). The regulatory region LEE150 (571 bp) containing the second predicted translation start codon located 15 bp downstream of the first one (Figure 2.22) was also PCR amplified using primers D61221 and D65811 and cloned into the EcoRI-HindIII sites of pRW225 to give LEE150/pRW225 (Table 2.2).

The predicted ORF in the LEE150 fragment (Figure 2.22) was replaced with KpnI site, which results in LEE150-1 fragment. This fragment was cloned into the EcoRI-HindIII sites pRW225 to give LEE150-1/pRW225 (Figure 2.23).

2.12 β-galactosidase assays

β-galactosidase assays (Miller, 1972) were used to measure the *lacZ* activity in cells containing plasmid- encoded promoter:: *lacZ* fusions in a range of genetic backgrounds and under different growth conditions.

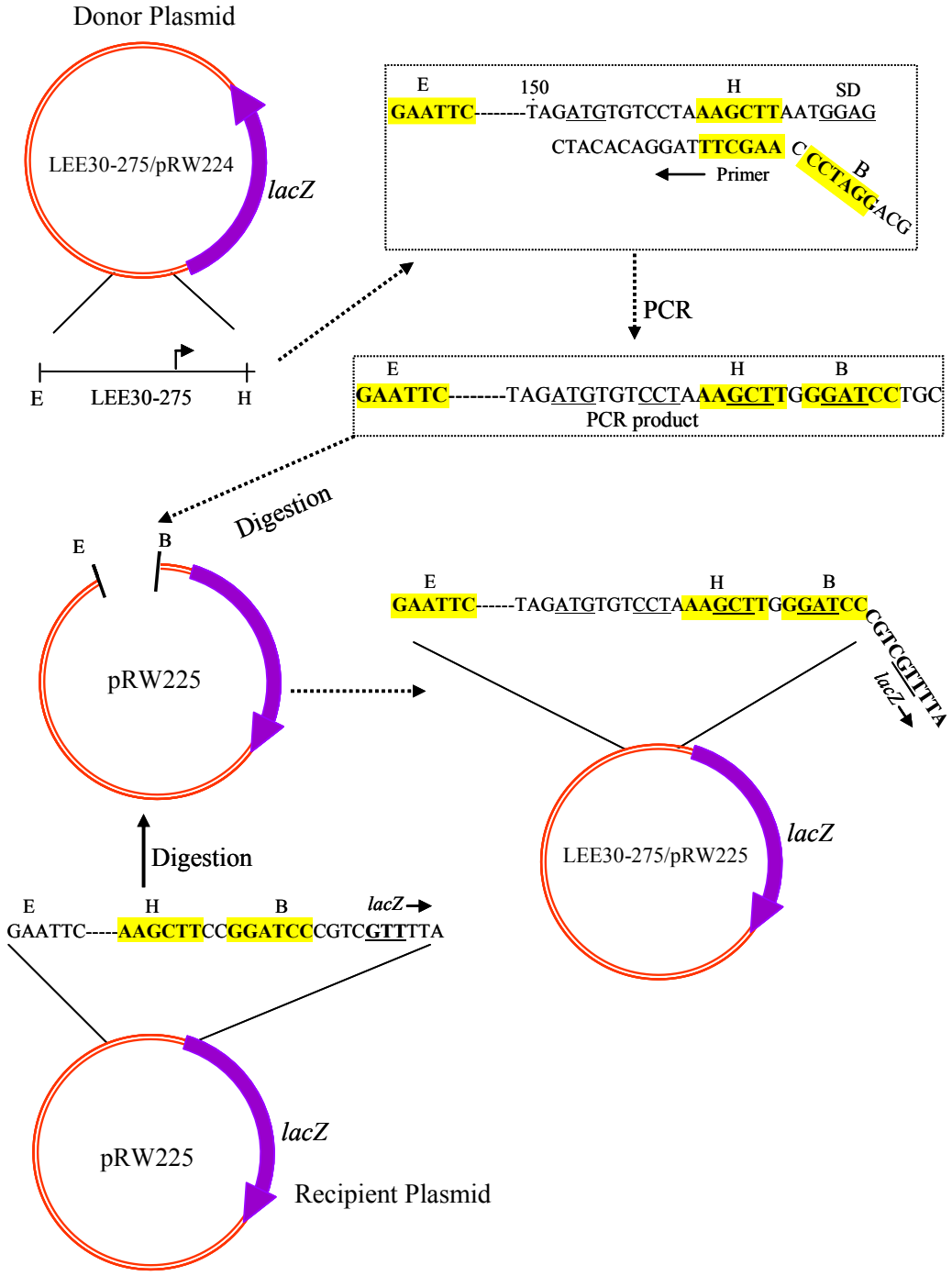


Figure 2.21. Cloning of the LEE30-275 fragment into pRW225 as translational fusion to *lacZ*.

LEE30-275 promoter fragment was amplified from LEE30-275/pRW224 using a primer pair D65474 and D63949 and cloned into EcoRI-BamHI sites of pRW225. The D65474 was designed in such a way that the start codon (ATG, position 44-46) (Figure 2.18) in the promoter fragment was in frame with the start codon of *lacZ*. Therefore, translation initiated on the cloned promoter fragment would be in-frame with the *lacZ* coding sequence in pRW225. E, H and B represent EcoR1, HindIII and BamHI restriction sites, respectively.

LEE151 (553bp)

```
     -568
GAATTCTGCACCCGTTCCAGGTTAGTGCTGGCTGTAGCTTATGTCCGGGAAACAGCTAATAGATATATATACTCGTCATACTTCAAGTTGCATGTGCTGCGACTGC

GTTCGCTTACCCCAATCACTTACTTATGTAAGCTCCTGGGGATTCACTCGCTTGCCGCCTTCCTGTAACTCGAATTAAGTAGAGTATAGTGAAACGGTTCAGCTTG

GTTTTTATTCTGTTTTATTTGTTTATGCAATGAGATCTATCTTATAAAGAGAAACGCTTAACTAAATGGAAATGCAATTATTAAAGTCGTTTGTTAACGAGATGAT

TTTCTTCTATATCATTGATTTTAAATGGATTTTAAAAATATATGATTTTTTTGTTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAATAACA
                                                          -35          P1         -10

TTTTAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAATTTGATAGATAAACGTTATCTCACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTGATAAGG
                          -19                                                      -35          P2         -10

TCGCTAATAGCTTAAAATATTAAAGCATGAAGCTT
```

LEE150 (571bp)

```
     -568
GAATTCTGCACCCGTTCCAGGTTAGTGCTGGCTGTAGCTTATGTCCGGGAAACAGCTAATAGATATATATACTCGTCATACTTCAAGTTGCATGTGCTGCGACTGC

GTTCGCTTACCCCAATCACTTACTTATGTAAGCTCCTGGGGATTCACTCGCTTGCCGCCTTCCTGTAACTCGAATTAAGTAGAGTATAGTGAAACGGTTCAGCTTG

GTTTTTATTCTGTTTTATTTGTTTATGCAATGAGATCTATCTTATAAAGAGAAACGCTTAACTAAATGGAAATGCAATTATTAAAGTCGTTTGTTAACGAGATGAT

TTTCTTCTATATCATTGATTTTAAATGGATTTTAAAAATATATGATTTTTTTGTTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAGAGAATAATAACA
                                                          -35          P1         -10

TTTTAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAATTTGATAGATAAACGTTATCTCACATAATTTATATCATTTGATTAATTGTTGGTCCTTCCTGATAAGG
                          -19                  +1                                  -35          P2         -10

TCGCTAATAGCTTAAAATATTAAAGCATGCGGAGATTATTTATTATGAAGCTT
```

Figure 2.22. DNA sequences of LEE151 and LEE150 fragments of the *LEE1* regulatory region.

EcoRI and HindIII sites are coloured red. Predicted P1 and P2 promoters (Sperandio *et al*., 2002) are boxed and underlined, respectively. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene reported by Yerushalmi *et al*. (2008). Dotted boxes both in LEE151 and LEE150 fragments correspond to a database assigned translation start point (TSP) of the *ler* gene. Closed grey shaded box in LEE150 corresponds to an additional predicted TSP that has a SD sequence (blue coloured) in front of it. Yellow colour in both fragments represents the predicted mini-ORF that has a SD sequence (blue coloured) in front of it.

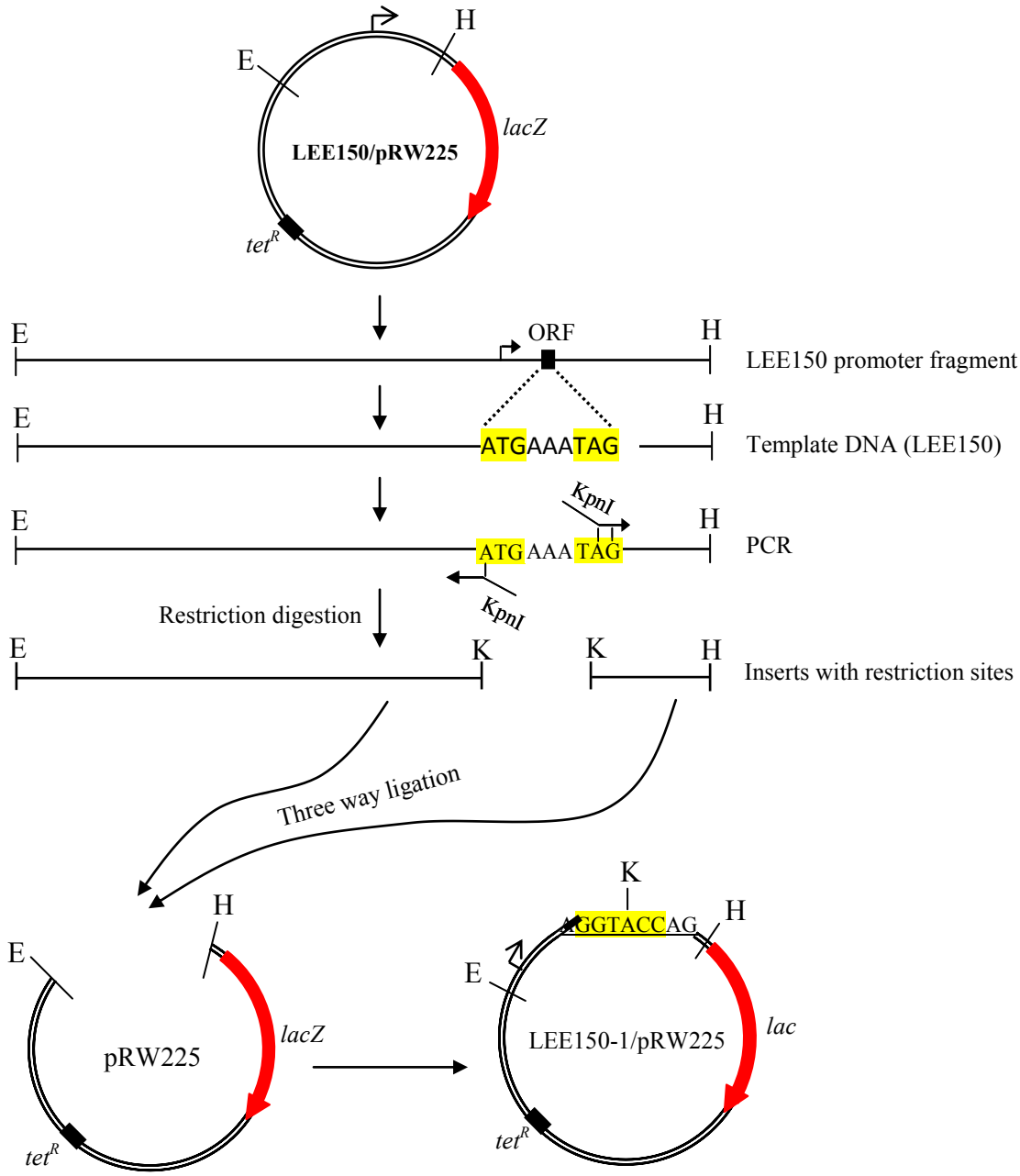


Figure 2.23. Construction of the LEE150-1/pRW225.

The predicted ORF in the LEE150 fragment was replaced with the KpnI (K) site. Plasmid LEE150/pRW225 was used as template. Primer D66033 (containing the KpnI site) together with D61221, and D66032 (containing the KpnI site) together with D65811 were used to amplify two fragments containing a KpnI site in each. Both fragment were digested, purified and cloned into pRW225 as translation fusion to *lacZ* through three way ligation, resulting in plasmid LEE150-1/pRW225. E, H and K represent EcoRI, HindIII and KpnI restriction sites, respectively.

2.12.1 β-galactosidase assays during exponential growth phase

5 ml LB supplemented with approximate antibiotic(s) was inoculated with a fresh colony of each strain carrying the plasmid-encoded promoter::*lacZ* fusion to be assayed. The strains were grown overnight at 37°C with shaking, and the following morning, 100 µl overnight culture was sub-cultured into 5 ml of the same medium for each strain. These cultures were incubated at 37°C until they reached an OD_{650} of 0.3-0.6 (exponential phase). Each culture was lysed using 2 drops each of toluene and 1% sodium deoxycholate, vortexed for 15 s, and shaken for 20 mins at 37°C with the bung removed from the flask, to allow the toluene to evaporate off. 100 µl of each lysate was assayed for β- galactosidase activity by adding 2.5 ml 13 mM 2-Nitrophenyl β-D-galactopyranoside (ONPG) solution and incubating at 37°C until a yellow colour developed. At this point, the reaction was stopped by adding 1 ml of 1M sodium carbonate and the optical density at 420 nm was read. β-galactosidase activity was calculated using the formula:

$$\text{β-galactosidase activity} = \frac{1000 \text{ X } 2.5 \text{ X } 3.6 \text{ X } OD_{420nm}}{OD_{650nm} \text{ x } 4.5 \text{ x t x v}} \text{ nmol/min/mg bacterial mass}$$

Where,

2.5 = factor for conversion of OD_{650} into bacterial mass, based on OD_{650} of 1 being equivalent to 0.4 mg/ml bacteria (dry weight).

3.6 = final assay volume (ml)

100/4.5 = factor for conversion of OD_{420} into nmol o-nitrophenyl (ONP), based on 1 nmol ml^{-1} ONP having an OD_{420} of 0.0045

t = incubation time (min)

v = volume of lysate added (ml)

Each set of conditions was assayed in triplicate with at least two separate experiments, and the mean β-galactosidase activity and standard deviation calculated under each condition. In all plasmid fusion assays, cells containing only the plasmid were used as control.

2.12.2 β-galactosidase assays during stationary growth phase

β-galactosidase assays of cultures grown to stationary phase were carried out as detailed in section 2.12.1, but the assays were carried out on the overnight culture directly. Overnight

cultures were diluted 1 in 10 in the appropriate medium, the OD_{650} calculated, and the diluted culture was lysed and assayed. Again, each strain/set of conditions was assayed in triplicate in each of at least two separate experiments.

2.13 DNA sampling

An out line of DNA sampling is shown in the Figure 2.24.

2.13.1 Sample preparation

LEE20-275, LEE30-275 and LEE30-275 151T promoter fragments (Figure 2.18) were cloned into the EcoRI-HindIII sites of pRW902 (Table 2.2). Resulting plasmids together with pACBSR were introduced into *E. coli* strain MG1655 encoding 3xFLAG-tagged LacI. A single fresh colony was inoculated in a 5 ml LB medium supplemented with appropriate antibiotics and grown overnight at 37°C. The 2.5 ml overnight cultures were inoculated in 500 ml LB (1:200 dilutions) supplemented with required antibiotics and grown with aeration at 37°C. When cells reached mid-exponential growth phase (OD650 = 0.5), 0.4% l-arabinose was added to induce the expression of the I-SceI meganuclease and the bacteriophage lambda Gam protein from plasmid pACBSR. After 20 mins, bacteria were harvested by centrifugation for 15 mins at ~11300 x g at room temperature. Cell pellets were resuspended in 20 ml of extraction buffer (see section 2.2.4) and a Roche Diagnostics protease inhibitor cocktail tablet was added. RNase A (300 μg/ml) and lysozyme (400 μg/ml) were than added and, after incubation at room temperature for 10 mins, mixture was cooled on ice and sonicated for three periods of 30 s. Samples were then centrifuged at ~40000 x g for 20 mins at 4°C to get clear lysates.

2.13.2 Confirmation of discrete promoter fragments in cells

To check the discrete promoter fragment in the cells, 7 ml of cultures before addition of arabinose and after 20 mins induction by arabinose were collected and crude DNA was extracted using Qiagen Mini prep kit and PCR purification kit and run on 7.5% polyacrylamide gel.

2.13.3 Isolation and analysis of protein-DNA complexes

The 20 ml of clear lysate was transferred to a 50 ml centrifuge tube and mixed with 25 mg of

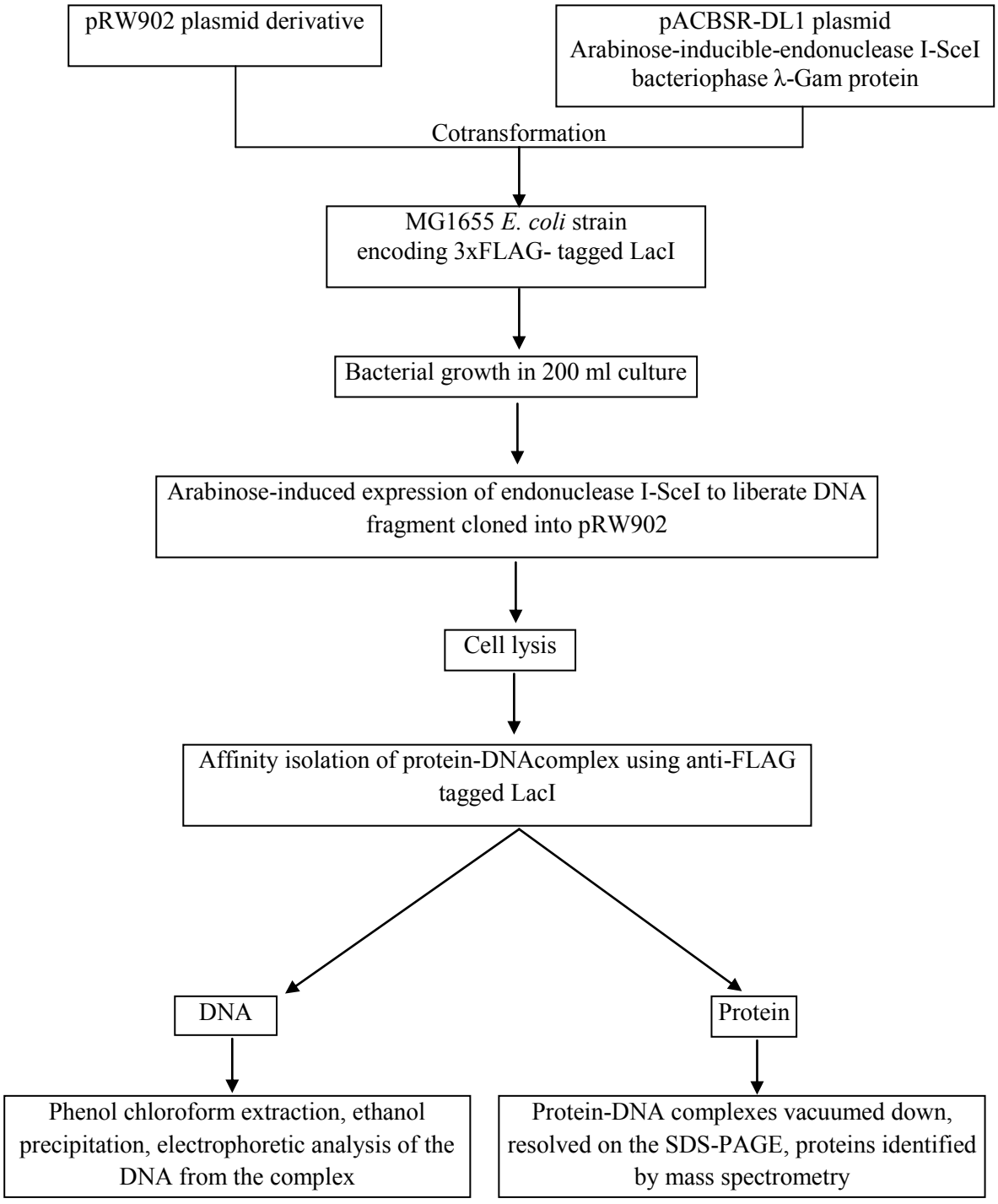


Figure 2.24. Outline of DNA sampling (Butala *et al*., 2009).

Dynabeads (M270-epoxy; Invitrogen) cross-linked to mouse anti-FLAG antibody (F3165, Sigma) and incubated on ice for 10 mins with gentle agitation. Dynabeads were collected to the side of the tube with a magnet and the lysate removed. The isolated beads were transferred to a 2 ml protein Lo-Bind tube (Eppendorf) and washed five times with 1 ml of wash buffer (See section 2.2.4). The protein-DNA complexes were eluted from the anti-FLAG antibody by addition of 500 µl of elution buffer (see section 2.2.4), and mixing on a rotating wheel for 5 mins. The supernatant was removed from the beads and transferred to a fresh tube. DNA in the supernatant was purified by phenol-chloroform extraction and ethanol precipitated and analyzed by electrophoresis in 1.3% agarose gels. Proteins in the supernatant were vacuum dried and dissolved in SDS-PAGE loading buffer containing 10 mM *tris* (2-carboxymethyl) phosphine-HCl (Sigma), which reduced any disulphide bonds. The samples were heated at 95°C for 5 mins. Samples were then alkylated by addition of 50 mM iodoacetamide (Sigma) for 30 mins at room temperature, resolved by SDS-PAGE in 4-12% gradient gels (Invitrogen), and visualized by Coomassie blue staining. Quantities and relative stoichiometries of DNA and proteins in the complexes were determined after scanning agarose gels and SDS-PAGE gels, that had been calibrated with the NEB 100-bp DNA ladder and Invitrogen SeeBlue Plus 2 protein markers, respectively.

To identify different proteins in SDS-PAGE gels, 1 mm gel slices were excised. The gel slices were placed in protein Lo-Bind tubes and destained for 1 hour, with agitation, by the addition of a 1 ml mixture of 50 % (v/v) acetonitrile and 50 mM ammonium bicarbonate. After 1 hr, the destaining solution was removed and the gel slices were dehydrated by the addition of 100 µl of acetonitrile. When the gel slices became opaque, the acetonitrile was removed, and the gel slices allowed to air dry for 5 minutes. Finally, proteins in each gel slice were digested with 8 ng of trypsin (Promega), in 100 µl of 50 mM ammonium bicarbonate solution, at 37°C for 4 hrs. Peptides from each slice were analyzed on a Thermo-Finnigan FT-ICR mass spectrometer using a NanoMate chip-based electrospray system operated by the University of Birmingham Functional Genomics and Proteomics Unit.

The experimental masses of peptides in each gel slice, as determined by mass spectrometry, were then compared to an *in silico* derived *E. coli* protein digestion database, using the Thermo-Finnigan Proteome Discoverer Software. Proteins whose expectation values were

lower than the 0.01 cut-off, and that were identified by the presence of more than 1 individual peptide, were considered to be positive identifications. Proteins whose expectation values were higher than the 0.01 cut-off, or that were identified by the presence of only one individual peptide were considered to be false-positives, and were therefore discarded from the identification list.

2.14 Pull down assays

In order to determine whether GrlA can associate with RNAP before interacting with the target site at the *LEE1* P1 promoter, pull down assays were used (Lee *et al*., 2008) in which protein A affinity tagged RNAP with associated proteins from Sakai *rpoC*::*4PrA* strain containing plasmid encoded GrlA was analyze.

Plasmids, pACYC184ΔHN (R-A-), pSI01 (R+A+) or pSI02 (R-A+) were introduced into Sakai *rpoC*::*4PrA* strain. In each case, a single fresh colony was inoculated in a 5 ml LB medium supplemented with chloramphenicol and grown overnight at 37°C. 2.5 ml overnight cultures were inoculated in 500 ml LB (1:200 dilutions) supplemented with chloramphenicol and grown with aeration at 37°C. When cells reached mid-exponential growth phase (OD_{650} = 0.7), they were harvested by centrifugation for 15 mins at ~11300 x *g* at 4°C and cell pellet was stored at -80°C. The cell pellet was then resuspended in 20 ml of extraction buffer (see section 2.2.5). DNase I (20 μg/ml), RNase A (300 μg/ml), and lysozyme (200 μg/ml) were then added. After incubation at room temperature for 10 mins, the mixture was cooled on ice and sonicated three times for 1 min. Sample was then centrifuged at ~40000 x *g* for 20 mins at 4°C to get clear lysates.

The supernatant was removed and incubated with 20 mg of Dynabeads (M270-epoxy; Dynal) coated with rabbit immunoglobulin G for 3 mins. The beads were collected with a magnet and washed with buffer (see section 2.2.5). The proteins were then eluted from the beads by incubation for 5 mins with elution buffer (see section 2.2.5). Coeluted proteins were vacuum dried and reduced and alkylated by resuspension in SDS-PAGE loading buffer containing 10 mM tris (2-carboxyethyl) phosphine-HCl (Sigma) and 50 mM iodoacetamide (Sigma). The proteins were resolved by SDS-PAGE in 4% to 12% gradient gels (Invitrogen) and visualized with Coomassie blue staining and/or silver staining following SilverQuest™ Microwave

Silver Staining Protocol (Invitrogen). Target protein bands were sliced and digested with trypsin for 8 hrs at 37°C. The peptides from each gel slice were purified and analyzed by mass spectrometry.

2.15 *In vitro* DNA footprinting analysis

2.15.1 Preparation of fragments for DNA footprinting

DNA footprinting experiments were performed using PstI-BamHI fragments, excised from pSR carrying the EcoRI-HindIII promoter fragment of interest (Figure 2.15). Approximately 200 µl pSR maxiprep DNA was digested with 16µl BamHI in a final volume of 240 µl 1 x NEBuffer 3 supplemented with 100 µg/ml BSA, then treated with 8 µl alkaline phosphatase to remove the 5' phosphate. After phenol/chloroform extraction and ethanol precipitation, the DNA was digested with 8 µl PstI in 80 µl 1 x NEBuffer 3 supplemented with 100 µg/ml BSA. PstI-BamHI fragments were purified by electroelution from a 7.5% polyacrylamide gel, extracted using phenol/chloroform and ethanol precipitation (see section 2.6), and then resuspended in 50 µl TE buffer.

2.15.2 Radio-labeling of DNA fragments

PstI-BamHI DNA fragments were end-labeled with [γ-^{32}P] on the BamHI end using T4 polynucleotide kinase (New England Biolabs), which catalyses the transfer of the γ phosphate from ATP to 5' terminal hydroxyl groups in DNA. 8-16 µl DNA was mixed with 1 µl T4 polynucleotide kinase and 1 µl [γ-^{32}P]-ATP in a final volume of 20 µl T4 polynucleotide kinase buffer, and incubated for 30 mins at 37°C. Unincorporated [γ ^{32}P]-ATP was then removed by passing the labelling mix down a Sephadex G-50 column: 400 µl 50% Sephadex -G50 suspension was loaded onto a Micro Bio-Spin column (Bio-Rad) and centrifuged for 2 mins at ~1300 x *g*. The column was then placed into a fresh collection tube, and the 20 µl labeling reaction was loaded onto the column and centrifuged for a further 2 mins at ~1300 x *g*.

2.15.3 Preparation of G+A ladder

Radio-labeled PstI-BamHI promoter fragments were treated with formic acid and piperidine to generate Maxam-Gilbert G+A sequence ladders, used to calibrate DNA footprinting gels. 3-4 µl labeled promoter fragment was made up to a final volume of 12 µl using sterile

distilled water, mixed with 50 µl formic acid, and incubated for 90 s at room temperature. The reaction was stopped by adding 200 µl 0.3 M sodium acetate (pH 7) and 700 µl ice-cold 100% ethanol, and the DNA sample ethanol precipitated as described in section 2.6. The DNA pellet was resuspended in 100 µl 1 M piperidine, and incubated for 30 mins at 90ºC. Samples were then ethanol precipitated, resuspended in 20 µl denaturing gel loading buffer, and heated to 90ºC before loading 0.5-2 µl onto footprinting gels.

2.15.4 *In vitro* potassium permanganate footprinting

Potassium permanganate footprinting was used to detect open complex formation at promoters *in vitro.* Potassium permanganate oxidizes unpaired thymine residues in single stranded regions of DNA, and subsequent treatment with piperidine results in cleavage of the DNA backbone on the 3' side of the modified nucleotide. Permanganate footprints were performed following the protocol of Savery *et al.* (1996). In each reaction, 0.1-0.5 µl PstI-BamHI promoter fragment, labeled with [γ-^{32}P] on the BamHI end, was mixed with a final concentration of 50 nM RNAP holoenzyme in a final volume of 20 µl 1 x HEPES/glutamate binding buffer containing 0.5 mg/ml BSA. After incubation for 20 mins at 37º C, reactions were treated with 1 µl 100 mM $KMnO_4$ solution for 4 mins at 37ºC, then stopped with 50 µl potassium permanganate stop solution. After phenol/chloroform extraction and ethanol precipitation, samples were resuspended in 40 µl 1 M piperidine and incubated at 90ºC for 30 mins. Samples were purified by phenol/chloroform extraction and ethanol precipitation, resuspended in denaturing gel loading buffer, and analysed by electrophoresis on a denaturing 6% polyacrylamide sequencing gel as described in the section 2.5.3.

2. 16 *In vitro* transcription assays

Multi round *in vitro* transcription analysis was performed using PstI-BamHI fragments as template excised from pSR containing promoter of interest (Figure 2.15), using purified RNAP holoenzyme, following the method as described by Browning *et al*. (2009). RNAP holoenzyme was diluted to 25 x the required final concentration in 1x transcription buffer. For each transcription reaction, 40 ng linear DNA fragment carrying the promoter of interest was mixed with 1 µl 25 x concentrated RNAP holoenzyme in a final volume of 25 µl 1 x Tris acetate transcription buffer containing 100 µg/ml BSA, 200 µM GTP, 200 µM ATP, 200 µM CTP, 10 µM UTP and 5µCi [α^{32}P]-UTP. Reactions were started by adding the RNAP,

incubated for exactly 15 mins, and stopped by adding 25 µl transcription stop solution. 6 µl each reaction was run on a denaturing 5.5% polyacrylamide sequencing gel and analyzed as described in the section 2.5.3.

2.17 Fluorescent actin staining (FAS) tests

To confirm whether the Ler protein encoded from the plasmid pSI04 plasmid is functional, Robert Shaw, Research Fellow, School of Biosciences, University of Birmingham performed a Ler complementation test.

2.17.1 HeLa cell culture

HeLa cell culture was done by Robert Shaw. Cells were maintained in bicarbonate buffered Minimum Essential Medium (MEM) (Sigma) supplemented with 10% foetal calf serum (PAA laboratories) at 37°C in a Sanyo CO_2 incubator at a ratio of 5% CO_2 in air. Cells were removed from 75 cm flasks with 1:1 trypsin/EDTA (Sigma) and seeded onto ethanol sterilized 13 mm circular glass coverslips in 24 well plates, at a split ratio of 1:20. Cells were grown for a further 48 hrs for cell attachment and spreading.

2.17.2 FAS tests

The assay was done by Robert Shaw. Adhesion of EHEC to HeLa cells was tested as described by Knutton *et al*. (1989). Tissue culture cells on 13-mm-diameter glass cover slips were placed in the wells of a multiwell tissue culture plate with 1 ml of HEPES-buffered essential medium containing 2% serum and 0.5% D-mannose. A sample (10 µl) of overnight bacterial broth culture was added to each well, and the cells were incubated for 6 hrs at 37°C. In 6 hrs assays, cells were washed and fresh medium was added after 3 hrs. After three washes to remove nonadhering bacteria, cells were fixed for 20 mins in 3% Formalin.

Formalin fixed coverslips were washed with phosphate-buffered saline (PBS) three times to remove formalin. Cells were permeabilised with 0.1% Triton X100 in (PBS) for 4 mins. After three washes in PBS, cover slips were stained with a mixture of 5 µg/ml solution of fluorescein isothiocyanate-phalloidin (Sigma) and 2 µg/ml propidium iodide (Invitrogen) in PBS for ~30 mins to specifically stain filamentous actin and the adherent bacteria. Cover slips were washed three times in PBS and mounted in citifluor mountant (Agar Scientific). The

slides were viewed on a Leica SP2 confocal microscope with spectral detection and XYZ maximum projections of full cell thickness were obtained.

2.18 Helix-turn-helix predictions for GrlA

Lewis Bingle, Research Fellow, School of Biosciences, University of Birmingham performed helix-turn-helix prediction for GrlA. The DNA-binding helix-turn-helix of GrlA, suggested by Deng *et al.* (2004), was confirmed computationally by the Dodd and Egan (1990) method, http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html, the GYM2.0 method, http://users.cis.fiu.edu/~giri/bioinf/GYM2/prog.html (Narasimhan *et al.*, 2002), and the Jpred structure prediction tool, http://www.compbio.dundee.ac.uk/www-jpred/index.html (Cole *et al*., 2008).

Chapter 3

Organisation of the *LEE1* operon regulatory region and activation by GrlA

3.1 Introduction

Ler is the key activator for expression of most of the LEE genes (Mellies *et al*., 1999; Bustamante *et al*., 2001; Umanski *et al*., 2002; Haack *et al.,* 2003). The regulatory region of the *ler* gene is the target for a range of regulatory factors including GrlR and GrlA (Mellies *et al.*, 2007; Hansen and Kaper, 2009; Tree *et al*., 2009; Kendall *et al*., 2010; Deng *et al*., 2004; Barba *et al*., 2005; Jimenez *et al*., 2010). GrlA acts as a transcription activator, whereas GrlR interacts with the GrlA and prevents its activity (Jobichen *et al*., 2007).

GrlA-mediated regulation of the *ler* gene expression is not fully understood. Both Huang and Syu (2008) and Jimenez *et al*. (2010) reported that purified GrlA fusion proteins could bind to the EHEC *LEE1* regulatory region but did not find its binding target, whilst Russell *et al.* (2007) suggested that GrlA might need another factor to modulate the *LEE1* expression. The work described in this chapter was aimed to get a better understanding of the action of GrlA at the EHEC *LEE1* regulatory region. Individual regulatory components can function in isolation (Barnard *et al*., 2004) and, thus, in this work a laboratory strain of *E. coli* K-12 was used to reproduce GrlA-dependent activation at *LEE1* and to determine the functional determinants in both GrlA and in the *LEE1* promoter.

3.2 Activity of the *LEE1* promoters and nested deletion analysis

The *LEE1* operon regulatory region is located upstream from the *ler* gene. Thus, a construct was made in which *lacZ* expression is controlled by the LEE10-568 fragment. The fragment contains the base sequence from position -568 to position -19 upstream of the functional ATG start codon of the *ler* gene (Figures 2.10, 2.11, 3.1). A series of eight nested deletions of this fragment were constructed. The upstream end of each of the nested deleted fragments is shown with respect to the reported *LEE1* P1 and P2 promoters (Sperandio *et al*., 2002). Fragments were cloned into the vector, pRW224, and the resulting recombinants were transformed into both M182, and EHEC strains (both Sakai and EDL933). Effects of the deletions on the *LEE1* promoter activity were assessed by the level of the measured β-galactosidase expression (Figure 3.1B). The patterns of activity from different promoter fragments seem to be invariable in all three strains backgrounds and in both LB medium (Figure 3.1) and DMEM medium (Figure 3.2). Removal of upstream sequences causes ~2-4

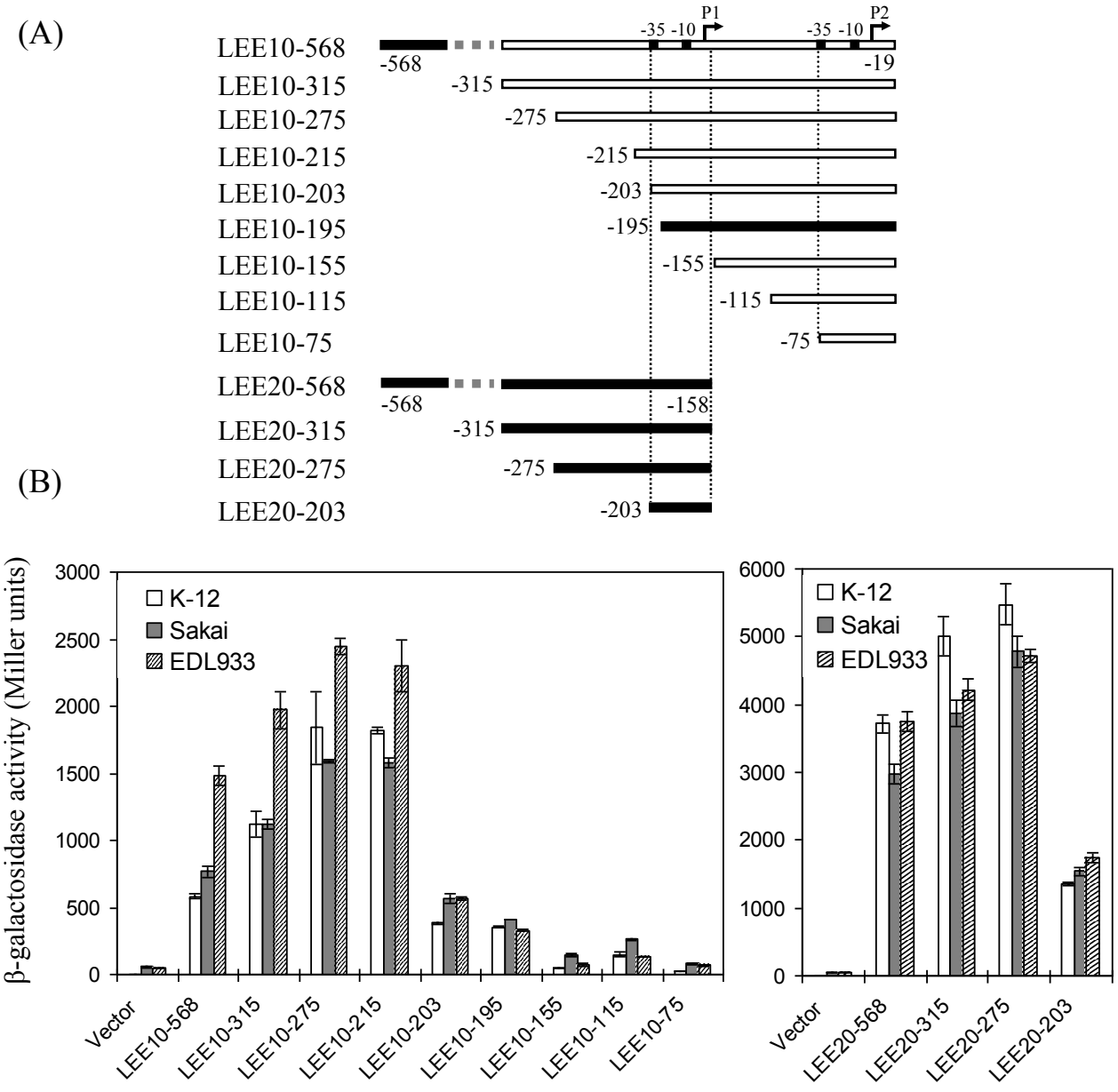


Figure 3.1. Nested deletions of the *LEE1* regulatory region.

A. Schematic representation of a set of EcoRI-HindIII DNA fragments covering the *LEE1* regulatory region. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene reported by Yerushalmi *et al*. (2008). The P1 and P2 promoters are indicated by bent arrows and the shaded black boxes represent the cognate -10 and -35 hexamer elements.
B. Bar chart to illustrate measured β-galactosidase activities in *E. coli* K-12, EHEC Sakai and EHEC EDL933 cells carrying pRW224 with each of the different fragments. Vector refers to empty pRW224 with a short linker between the EcoRI and HindIII sites. Activities were measured after growing the cells in LB medium to an optical density at 650 nm of ~0.5 at 37°C. The values are the average of three independent assays and standard deviations are shown with bars.

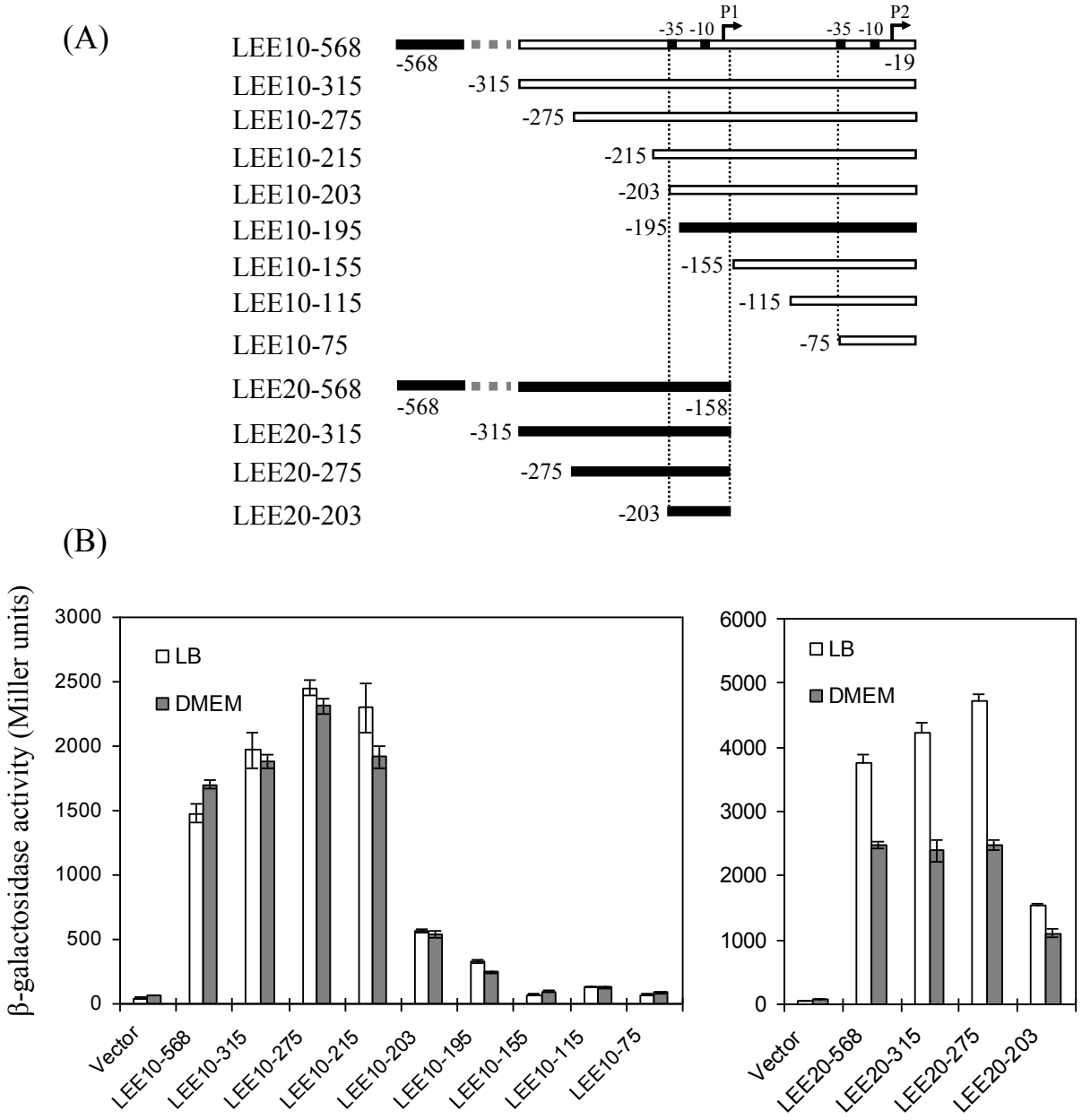


Figure 3.2. Nested deletion analysis of the *LEE1* regulatory region.

A. Schematic representation of a series of nested deleted derivatives of *LEE1* operon regulatory region. The coordinates of upstream and downstream end of each derivative refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene.
B. Bar chart to illustrate the β-galactosidase activities in EHEC EDL933 carrying pRW224 with each of the different fragments. Vector refers to empty pRW224 with a short linker between the EcoRI and HindIII sites. Activities were measured after growing the cells either in LB (open bars) or DMEM (shaded bars) media at 37ºC to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

fold increase in *lac* expression and the expression was found to be optimal with the LE10-275 and LEE10-215 fragments, which carry DNA sequence up to positions -275 and -215 upstream from the *ler* start codon, respectively. Deletions to positions -203, -195 and -155 result in a large decrease in expression. The region from positions -215 to -155 contains the proposed sequence elements for the P1 promoter (Sperandio *et al*., 2002). Hence, I conclude that, at least in my experimental conditions, P1 must be the major functional promoter for the *LEE1* operon. Absence of the P1 promoter elements in both LEE10-155 and LEE10-115 fragments reduces the expression to 10-15%. This expression is most likely due to the P2 promoter because deletion to the LEE10-275 fragment, which carries P2 only, reduces the expression to basal levels. In a complementary experiment, the effect of the deletions of the downstream sequences containing the P2 promoter elements was checked. To do this, the LEE10-568, LEE10-315, LEE10-275 and LEE10-203 fragments were shortened to move the downstream HindIII site to position -158, rather than position -19, thereby deleting the downstream P2 promoter, and generating the LEE20-568, LEE20-315, LEE20-275 and LEE20-203 fragments (Figure 3.1A). Fragments were cloned into pRW224 and *lacZ* expression was measured. The observed patterns of β-galactosidase expression in both the O157:H7 and M182 strains were similar to with the LEE10 fragment series (Figure 3.1B). The data are consistent with the conclusion that the P2 promoter makes but a minor contribution to the activity measured here. However, the increased expression seen with the LEE20 fragments is probably due to the shorter untranslated leader sequence upstream of the reporter *lacZ* gene.

3.3 Identification of functional elements at the *LEE1* regulatory region P1 promoter

The experiment illustrated in Figure 3.1 shows that the shortest fragment, which shows maximum promoter activity is the LEE20-275 fragment. This fragment carries the DNA sequence from positions -275 to -158 upstream of the *ler* start codon and contains the P1 promoter. The base sequence of this fragment is shown in Figure 3.3, where the sequence has been renumbered 1-120, starting immediately downstream of the EcoRI linker. This numbering system was adopted to describe the ensuing mutational analysis of the LEE20-275 fragment and the shorter LEE20-203 fragment.

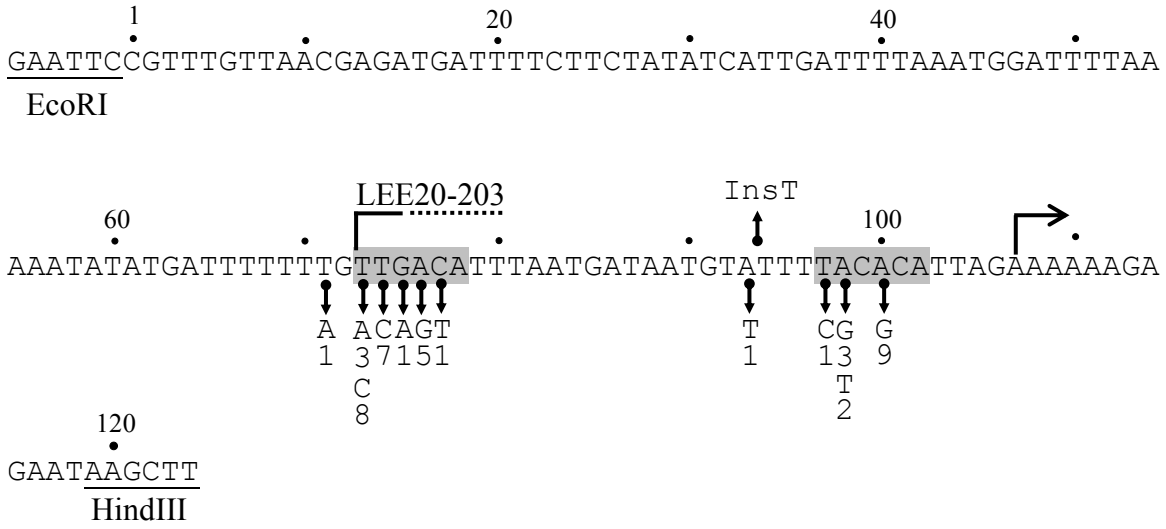


Figure 3.3. Mutational analysis of the LEE20-275 fragment.

The figure shows the nucleotide sequence of the upper strand of LEE20-275 fragment. The base sequences are numbered 1-120, starting with the first cloned *LEE1* regulatory region base (that is 275 base pairs upstream from the *ler* translation start codon). The positions of randomly generated single mutations that reduced expression from this fragment are illustrated by showing the substituted base and, in each case, the adjacent number records the number of times that the particular substitution was obtained. The locations of the P1 promoter -35 and -10 hexamer elements, deduced from this study, are shown by grey shading and the transcript start, at position 107A, determined from data in Figure 3.4B is indicated. The upstream end of the smaller LEE20-203 fragment is indicated by a bent solid/dotted line.

In order to define elements essential for promoter activity, error-prone PCR was performed, which generated ten independent preparations of the LEE20-275 fragment carrying random point mutations. The fragments were then cloned into pRW224, recombinant plasmids were transformed into *E. coli* strain M182, and transformants were grown on MacConkey lactose indicator plates. As expected, the majority of colonies were found as Lac^+ (red). However, after screening over 100,000 transformants, ~100 Lac^- (pale pink) colonies were identified. Sequence analysis showed that 43 of these carried single mutations in the LEE20-275 fragment cloned in pRW224. The locations of the different single point mutations that reduced *lac* expression from the LEE20-275 promoter fragment are illustrated in Figure 3.3.

Surprisingly, 25 of the 43 point mutations fall in the TTGACA motif at positions 73-77. The motif matches perfectly the consensus hexamer -35 element for *E. coli* promoters (Rosenberg and Court, 1979; McClure, 1985) and it was predicted to be the P1 promoter -35 element by Sperandio *et al.* (2002). The effects of the different substitutions in this element were quantified (Table 3.1). The results are consistent with this being the functional -35 element controlling expression from the LEE20-275 fragment. Another of the point mutations (71A) falls just upstream of the -35 element, and its effects are consistent with the lower promoter activity of the LEE20-203 fragment compared to the LEE20-275 fragment (Figure 3.1B).

A further 15 of the 43 point mutations form a second cluster that appears to fall in the promoter -10 hexamer (Figure 3.3). It is well understood that base pairs at positions 1 and 2 of promoter -10 hexamers in *E. coli* are most crucial for promoter activity (Rosenberg and Court, 1979; McClure, 1985). This suggests that the motif TACACA at positions 97-102 is likely to be the functional -10 hexamer element of the P1 promoter. Since the consensus -10 element for *E. coli* promoters is TATAAT, I performed site-directed mutagenesis to generate complementary mutations to check this suggestion. Effects of different mutations are presented in Table 3.1. The base substitutions 97C, 98C, 98G, 98T and 100G cause >90% reductions in expression, whilst the 99A, 99G, 101G, 102C and 102G substitutions have lesser effects. These results support my prediction and confirm the TACACA hexamer as the -10 element. However, this suggests that the spacer between the proposed -35 and -10 hexamer elements is 18 base pairs rather than the optimal 17 base pairs. To check the effect of the spacer length on the promoter activity, a LEE20-275 fragment derivative in which a T in

Table 3.1: Mutational analysis of the *LEE1* P1 promoter

Fragment	Promoter sequence from positions 71-102	β-galactosidase activity (Miller units ± sd)
Starting fragment		
LEE20-275	5′-TGTTGACATTTAATGATAATGTATTTTACACA-3′	5477 ± 300
	71 80 90 100	
Point mutation upstream of the -35 element		
LEE20-275 71A	5′-**A**GTTGACATTTAATGATAATGTATTTTACACA-3′	1034 ± 123 (18.9)
Point mutations in the -35 element (TTGACA)		
LEE20-275 73A	5′-TG**A**TGACATTTAATGATAATGTATTTTACACA-3′	1103 ± 20 (20.1)
LEE20-275 73C	5′-TG**C**TGACATTTAATGATAATGTATTTTACACA-3′	1128 ± 28 (20.6)
LEE20-275 74C	5′-TGT**C**GACATTTAATGATAATGTATTTTACACA-3′	991 ± 90 (18.1)
LEE20-275 75A	5′-TGTT**A**ACATTTAATGATAATGTATTTTACACA-3′	690 ± 27 (12.6)
LEE20-275 75C*	5′-TGTT**C**ACATTTAATGATAATGTATTTTACACA-3′	962 ± 141 (17.6)
LEE20-275 76G	5′-TGTTG**G**CATTTAATGATAATGTATTTTACACA-3′	1478 ± 120 (27.0)
LEE20-275 77A*	5′-TGTTGA**A**ATTTAATGATAATGTATTTTACACA-3′	1532 ± 40 (28.0)
LEE20-275 77G*	5′-TGTTGA**G**ATTTAATGATAATGTATTTTACACA-3′	1136 ± 80 (20.7)
LEE20-275 77T	5′-TGTTGA**T**ATTTAATGATAATGTATTTTACACA-3′	1185 ± 60 (21.6)
Point mutations in the spacer region		
LEE20-275 93T	5′-TGTTGACATTTAATGATAATGT**T**TTTTACACA-3′	812 ± 86 (14.8)
LEE20-275 InsT (93-94)	5′-TGTTGACATTTAATGATAATGTA**T**TTTTACACA-3′	205 ± 26 (3.7)
LEE20-275 Δ94T*	5′-TGTTGACATTTAATGATAATGTA—TTTACACA-3′	11622 ± 291 (212.2)
Point mutations in the -10 element (TACACA)		
LEE20-275 97C	5′-TGTTGACATTTAATGATAATGTATTT**C**ACACA-3′	318 ± 12 (5.8)
LEE20-275 98C*	5′-TGTTGACATTTAATGATAATGTATTTT**C**CACA-3′	233 ± 4 (4.3)
LEE20-275 98G	5′-TGTTGACATTTAATGATAATGTATTTT**G**CACA-3′	202 ± 15 (3.7)
LEE20-275 98T	5′-TGTTGACATTTAATGATAATGTATTTT**T**CACA-3′	234 ± 9 (4.3)
LEE20-275 99A*	5′-TGTTGACATTTAATGATAATGTATTTTA**A**ACA-3′	10169 ± 298 (185.7)
LEE20-275 99G*	5′-TGTTGACATTTAATGATAATGTATTTTA**G**ACA-3′	2721 ± 16 (50.0)
LEE20-275 100G	5′-TGTTGACATTTAATGATAATGTATTTTAC**G**CA-3′	305 ± 9 (5.6)
LEE20-275 101G*	5′-TGTTGACATTTAATGATAATGTATTTTACA**G**A-3′	2493 ± 132 (45.5)
LEE20-275 102C*	5′-TGTTGACATTTAATGATAATGTATTTTACAC**C**-3′	4375 ± 105 (80.0)
LEE20-275 102G*	5′-TGTTGACATTTAATGATAATGTATTTTACAC**G**-3′	1843 ± 72 (33.6)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-275 fragment and different mutations. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). Activities expressed as a percentage of activity with the starting LEE20-275 fragment are shown in parentheses. The central part of the table shows the fragment base sequence from position 71 to position 102, with the P1 promoter -10 and -35 hexamer elements underlined. Base substitutions and insertions in the different fragments are highlighted in boldface type and coloured red whilst the Δ94 deletion is shown by a dash. Mutations made by site directed mutagenesis are indicated by asterisks, whilst the other mutations came from the random PCR mutagenesis experiment illustrated in Figure 3.3.

the spacer was deleted (Δ94T), thus reducing the spacer length from 18 to the optimal 17 base pairs was constructed. The mutation Δ94T causes ~2 fold increase in the promoter activity (Table 3.1). Note that one of the Lac$^-$ mutants carrying the LEE20-275 fragment with an extra T between position 93 and 94 (Figure 3.3) was found during random mutagenesis of the LEE20-275 fragment. This insertion increased the spacer length from 18 to 19 base pairs, and caused a sharp reduction in *lac* expression (Table 3.1). These results suggest that an increase in the spacer length further away from the optimal 17 base pairs causes a reduction in the promoter activity.

3.4. Biochemical analyses

In a complementary experiment, complexes between purified DNA fragments and purified *E. coli* RNAP holoenzyme were studied to confirm the location of the functional promoter in the LEE20-275 fragment. In these experiments, the starting 'wild-type' DNA fragment and a corresponding fragment carrying the 98C substitution at position 2 of the -10 hexamer element were compared. To monitor the effect of the 98C base substitution on the open complex formation, an *in vitro* potassium permanganate foot-printing analysis (Savery *et al.* 1996) was performed. In this analysis, interaction between purified RNAP and purified DNA fragment results in single stranded 'bubble' produced after local unwinding of promoter DNA around the transcription start site. Potassium permanganate modifies single stranded thymine residues and subsequent treatment with the piperidine of the modified DNA causes cleavage of the DNA backbone on the 3' side of the modified nucleotide. Results of the potassium permanganate foot-printing analysis shown in Figure 3.4A reveal an extensive unwinding just downstream of the promoter -10 hexamer and, crucially, this unwinding is greatly reduced with the 98C mutant.

In a complementary experiment, multi round *in vitro* transcription assays following the protocol used by Browning *et al*. (2009) were performed in which RNAP runs to a terminator in the DNA fragment synthesising labelled RNA molecules that are detected by gel electrophoresis. The labelled transcripts that formed after labelled nucleoside triphosphates were added to the binary RNAP-DNA complexes were analyzed (Figure 3.4B). The major transcript starting in the LEE20-275 fragment is ~102 bases and this corresponds to a

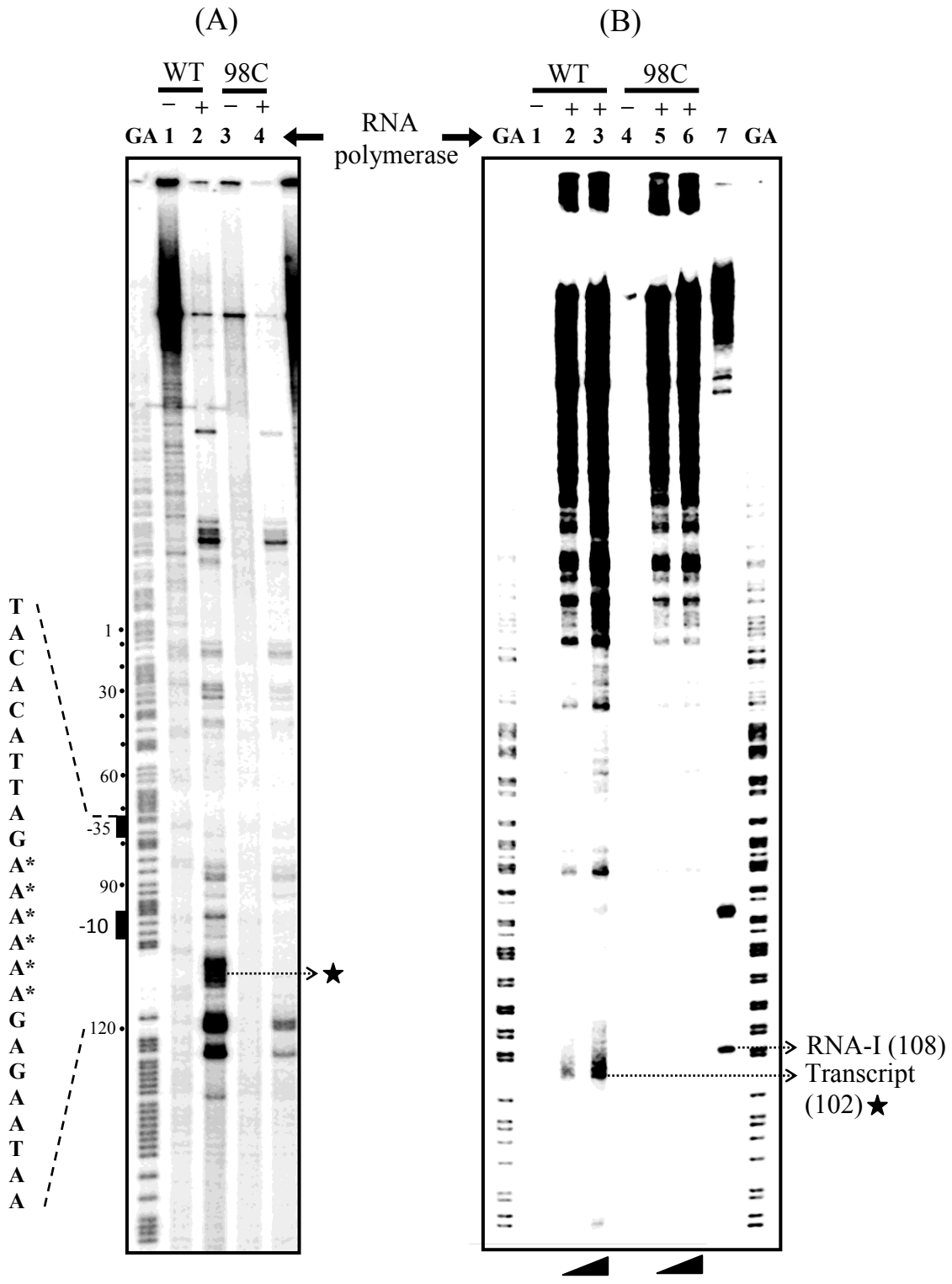


Figure 3.4. ***In vitro*** **permanganate footprinting and transcription assays (continued on page 142).**

Figure 3.4. *In vitro* permanganate footprinting and transcription assays (continued).

A. The figure shows an autoradiogram that identifies the potassium permanganate sensitive sites in complexes of holo RNA polymerase with a DNA fragment that includes the LEE20-275 sequence (WT) or a derivative carrying the 98C mutation (98C). Lanes 1 and 3 show the result after control incubations without RNA polymerase, whilst lanes 2 and 4 show the analysis of samples with 50 nM RNA polymerase. The gel was calibrated using a Maxam-Gilbert sequence reaction (GA) and relevant positions are indicated. The location of the *LEE1* P1 promoter -10 and -35 elements are indicated by black boxes and the asterisks identify 6 consecutive residues just downstream of the -10 element that display RNA polymerase-dependent reactivity to permanganate.

B. Autoradiogram of an analysis by gel electrophoresis of ^{32}P-labelled RNA transcripts made by RNA polymerase holoenzyme from PstI-BamHI fragments carrying wild type LEE20-275 sequences (lanes 1-3) and the 98C derivative (lanes 4-6). The RNA polymerase concentration was: lanes 1 and 4, no enzyme; lanes 2 and 5, 200 nM; lane 3 and 6, 400 nM. The gel was calibrated with the pSR plasmid-encoded 108 base RNA-I transcript (lane 7) and a Maxam -Gilbert sequence reaction (GA). The proposed LEE1 P1 transcript is indicated by an asterisk.

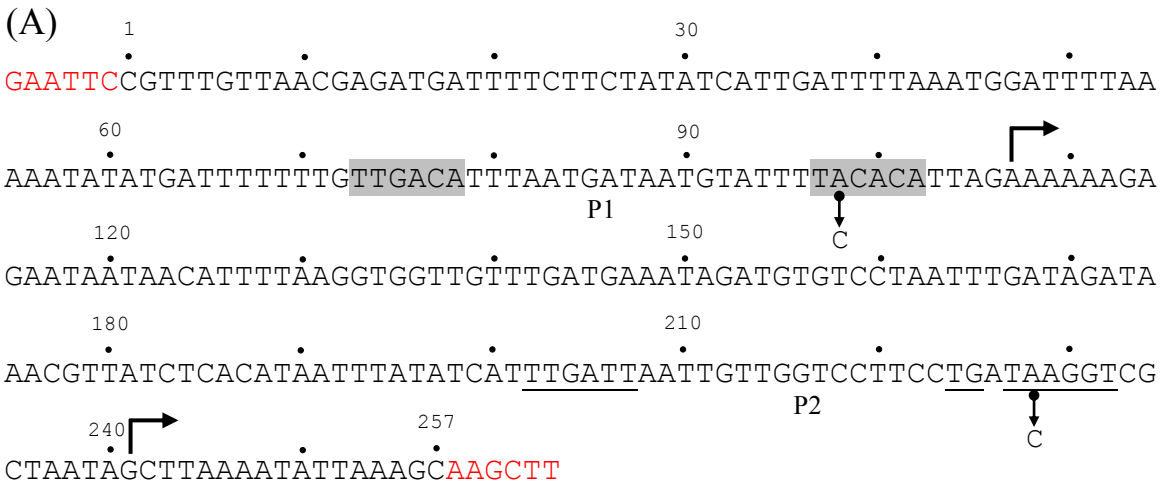
transcript starting at position 107A, 5 bases downstream from the proposed -10 element (Figure 3.3). This RNA species is completely absent when the DNA fragment carrying the 98C substitution is used (Figure 3.4B).

3.5 Contribution of P1 and P2 promoters to the expression of the *LEE1* operon

Sperandio *et al* (2002) showed that the *LEE1* P2 promoter plays a major role in the expression of the *LEE1* operon. The P2 promoter has an extended -10 element with a TG motif. The P2 promoter carries TTGATT and TAAGGT as the -35 and -10 elements, respectively, with a 19 base pairs promoter spacer (Sperandio *et al*., 2002). Transcription initiates from this promoter at a site located 35 base pairs upstream of the translation start site of *ler* gene. The nested deletion analysis presented (Figure 3.1) shows that the P1 promoter is the principal promoter, whilst the P2 promoter plays a minor role. In a complementary experiment, mutational analysis of the LEE10-275 promoter fragment that contains both P1 and P2 promoters was carried out to confirm further the contribution of two promoters in *lac* expression. The nucleotide sequence of this fragment is numbered 1-257 following the numbering system of LEE20-275 fragment (Figure 3.3). Site directed mutagenesis using megaprimer PCR was used to introduce a base substitution either at the second position of the P1 -10 element (98C) or at the second position of the P2 -10 element (228C) and at both positions (98C 228C) in the LEE10-275 fragment. Derivatives were then cloned into pRW224 and *lac* expression was quantified (Figure 3.5). Mutation of the P1 promoter (LEE10-275 98C) reduced the promoter activity sharply (~95%) whilst mutation of the P2 promoter (LEE10-275 228C) caused a ~40% decrease in *lac* expression. However, mutations in both P1 and P2 (LEE10-275 98C 228C) reduced the promoter activity to a low level similar to that found with the mutation in P1 alone (LEE10-275 98C). These results support the conclusion, based on nested deletion analysis that the *LEE1* P1 promoter plays the major role in *LEE1* expression.

3.6 Activation by GrlA

In order to carry out a thorough investigation of the role of GrlA at the *LEE1* operon regulatory region, a series of pACYC184 derivatives carrying the *grlRA* operon (pSI01), *grlA* alone (pSI02) and *grlR* alone (pSI03) were constructed (Figure 2.17). Each of recombinant



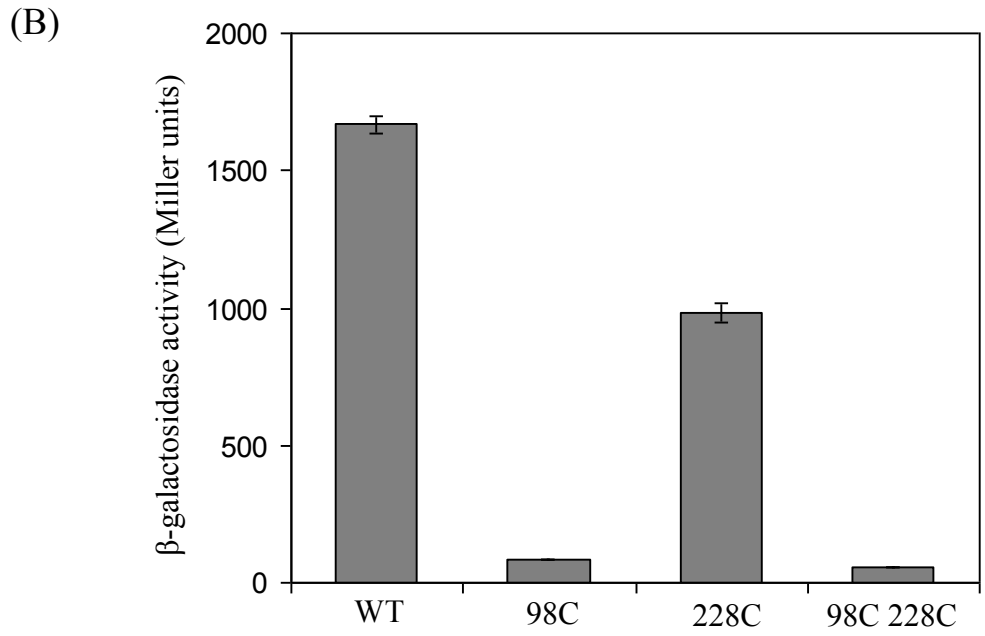


Figure 3.5. Mutational analysis of the LEE10-275 fragment.

A. The figure shows the nucleotide sequence of the upper strand of LEE10-275 fragment. The base sequences are numbered 1-257 following numbering system at Figure 3.3. P1 and P2 promoter elements are shaded grey and underlined, respectively. Transcription initiation site from both promoters are indicated by bent arrows. Vertical arrows indicate the location of the mutations. EcoRI and HindIII sites are coloured red.

B. The bar chart illustrates the measured β-galactosidase activity in M182 cells containing *lac* expression vector pRW224 with LEE10-275 fragment (WT) or mutant derivatives. Measurements were made after growing the cells in LB at 37°C to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

plasmids or empty vector was transformed into the *E. coli* K-12 M182 *Δlac* strain carrying pRW224 derivatives with different *LEE1* regulatory region fragments (Figure 3.1). Initially, M182 was co-transformed with pRW224 carrying the LEE10-568 fragment and the different pACYC184 derivatives, and *lac* expression was quantified. The results illustrated in Figure 3.6 show that GrlA causes ~6-fold activation of expression from the promoter fragment and as expected, this activation was found to be suppressed by the presence of GrlR (Jobichen *et al*., 2007).

In order to map a minimal region of *LEE1* operon regulatory region necessary for GrlA-dependent activation, the GrlA-dependent activation was measured at the *LEE1* regulatory region carried by fragments with the two series of nested deletions described earlier (Figure 3.1). Interestingly, activation is observed with the LEE20 series of fragments that lack the P2 promoter (Figure 3.7). Moreover, the activation is lost with the LEE10-195, LEE10-155, LEE10-115 and LEE10-75 fragments in which segments of the P1 promoter are deleted (Figure 3.7). These results suggest that the *LEE1* P1 promoter is the target for activation by GrlA, and this can be measured even with the short LEE20-203 fragment that carries only 52 base pairs P1 sequence.

GrlA-dependent activation was also measured in the Sakai EHEC serotype O157:H7 strain using the *LEE1* P1 promoter carried by the LEE20-203 fragment. Results illustrated in Figure 3.8 show that activation level in Sakai is similar to that seen in M182 (Figure 3.8). The expression from the LEE20-203 promoter fragment cloned into pRW224 in *E. coli* Sakai wild type background was also measured in the absence of plasmid encoded GrlA and that activity was compared with those measured in either *grlA* or *grlR* mutant strains. Results illustrated in Figure 3.8 show that no significant differences appear in the β-galactosidase activity between different strains. However, the promoter activity was found to be decreased slightly in *grlA* mutant strain and increased in *grlR* mutant strain compared with the activity observed in wild-type Sakai background. The reduced trend of the promoter activity in *grlA* mutant background suggests that the GrlA might play a positive role in the expression from the *LEE1* P1 promoter. The slight increase in the level of promoter activity in the *grlR* mutant background compared with that in the wild-type Sakai background again suggests that the free GrlA might positively regulate expression from the *LEE1* P1 promoter. Moreover, this supports the fact

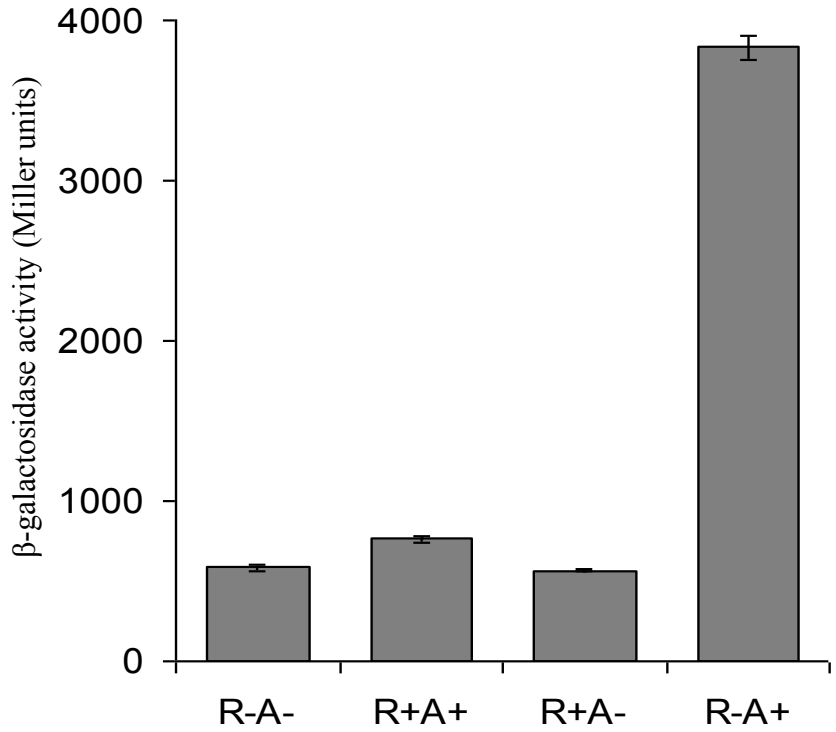
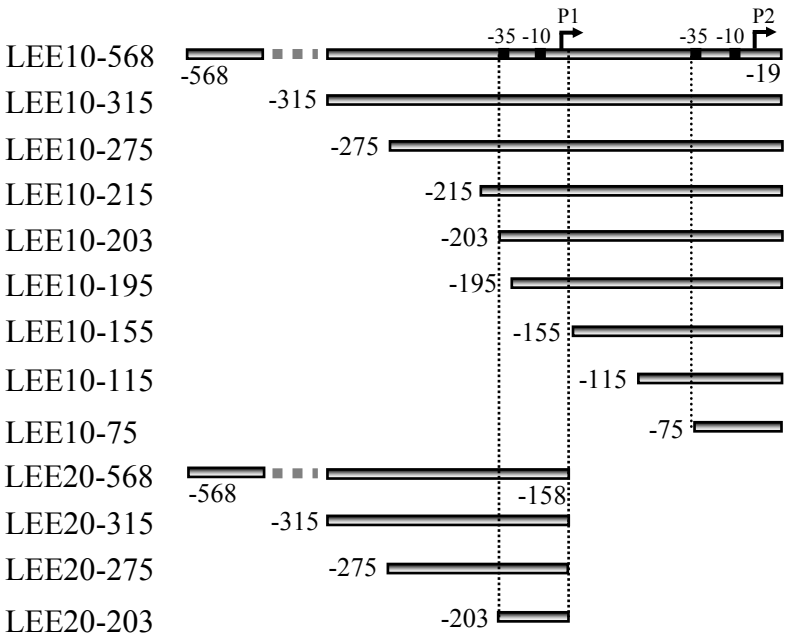


Figure 3.6. Activation of expression from the *LEE1* regulatory region by GrlA.

The bar chart illustrates measured β-galactosidase expression in M182 cells containing pRW224 with the LEE10-568 promoter fragment together with plasmid pACYC184ΔHN (R-A-), pSI01 (R+A+), pSI03 (R+A-) or pSI02 (R-A+). Measurements were made after growing the cells in LB medium at 37ºC to an optical density of ~0.5 at 650 nm. Each bar represents the mean of three independent experiments together with the standard deviation.



Expression from different promoters		
β-galactosidase activity (Miller Units ± sd)		
-GrlA	+GrlA	Fold activation
584 ± 21	3831 ± 72	6.6
1120±93	5325±81	4.8
1838±273	6178±162	3.4
1821±28	5848±231	3.2
381±5	2278±260	6
354±5	356±3	1.0
49±1	49±1	1.0
146±17	153±17	1.0
26±0	24±1	1.0
3721±130	7645±67	2.1
5012±297	10772±616	2.1
5477±300	9125±237	1.7
1353±16	4448 ± 169	3.3

Figure 3.7. GrlA-dependent activation at different *LEE1* regulatory region fragments.

The left part of the figure illustrates different fragments covering the *LEE1* regulatory region that were cloned into pRW224, using the same conventions as in Figure 3.1. The right part of the figure presents β-galactosidase activities measured in M182 cells carrying pRW224 containing each of the fragments. Cells also contained either pSI02 (+GrlA) or empty vector, pACYC184ΔHN (-GrlA). Measurements were made after growth in LB at 37ºC to an optical density of ~0.5 at 650 nm. Standard deviations (sd) were obtained from at least three independent experiments. The fold activation by GrlA for each fragment is shown in third column.

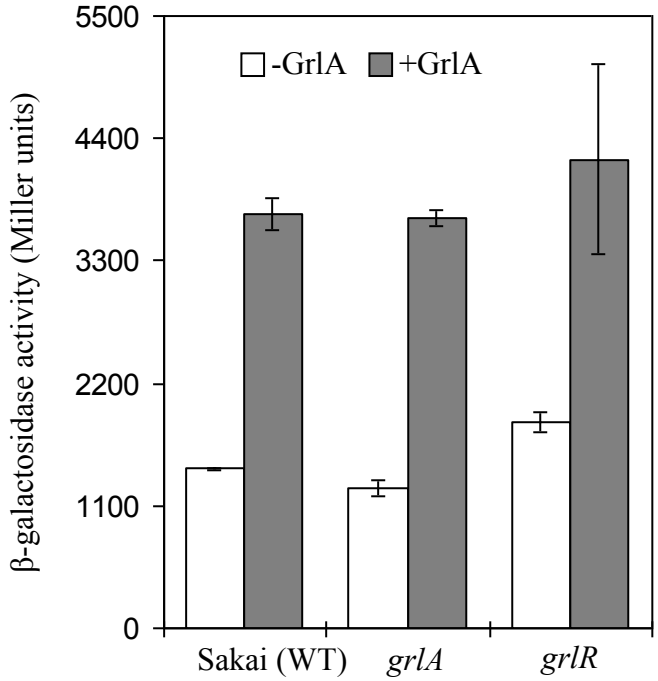


Figure 3.8. GrlA-dependent activation in an EHEC strain.

The figure illustrates β-galactosidase activities measured in *E. coli* Sakai 813 strain (Wild type: WT), and *grlA* and *grlR* mutant derivatives carrying pRW224 with the LEE20-203 promoter fragment. Cells also contained either empty vector, pACYC184ΔHN (-GrlA: open bars), or pSI02 (+GrlA: shaded bars). Measurements were made after growth in LB at 37ºC to an optical density of ~0.5 at 650nm. Standard deviations were obtained from three independent experiments.

that GrlR interacts with the GrlA and prevents the GrlA from activating the *LEE1* P1 promoter (Jobichen *et al*., 2007).

Both Barba *et al.* (2005) and Jimenez *et al.* (2010) reported that GrlA activates *LEE1* promoter mainly by counteracting H-NS-mediated repression. They also reported GrlA -dependent activation of *LEE1* in the absence of H-NS. To address this issue further, *lac* expression from four different nested derivatives of the *LEE1* regulatory region was measured in *E. coli* K-12 strain MG1655 and its isogenic Δ*hns* derivative. Data shown in Figure 3.9 illustrate that activity from the full-length LEE10-568 is clearly repressed by H-NS, but expression from the short LEE20-203 fragment is not. It also shows that both upstream (LEE20-568) and downstream (LEE10-203) sequences of the P1 promoter are needed for repression by H-NS, which might support the DNA bridging property of H-NS (Dame *et al*., 2005; Stoebel *et al*., 2008; Browning *et al*., 2010). GrlA-dependent activation was also measured in MG1655 and its isogenic Δ*hns* derivative (Figure 3.10). With the full-length LEE10-568 fragment, repression of the *LEE1* promoter by H-NS is overcome by GrlA, which activates to higher levels in both the hns^+ and Δ*hns* backgrounds. With the short LEE20-203 fragment, the *LEE1* promoter activity that is not repressed by H-NS is activated similarly by GrlA in both genetic backgrounds. These results suggest that GrlA can activate the *LEE1* P1 promoter independently of H-NS as well as counteract repression by H-NS.

3.7 Evidence that GrlA binds to the spacer region at the *LEE1* P1 promoter

GrlA activates expression from both LEE10-203 and LEE20-203 promoter sequences, which begin immediately upstream of the -35 TTGACA hexamer element (Figure 3.11A). These data suggest that GrlA interacts with the promoter DNA sequence located downstream of its -35 element. To identify essential promoter determinants for activation, a series of derivatives of the LEE20-203 fragment containing different point mutations throughout the fragment were constructed and GrlA-dependent activation was measured. The effects of 23 point mutations throughout this fragment, which carries only 46 base pairs of the *LEE1* regulatory region sequence, were measured. Results illustrated in Figure 3.11 show that GrlA-dependent activation is completely suppressed by the 89G, 90C, 91C, 92A and 98C substitutions. Mutation 98C corresponds to the base change at the second position of the promoter -10

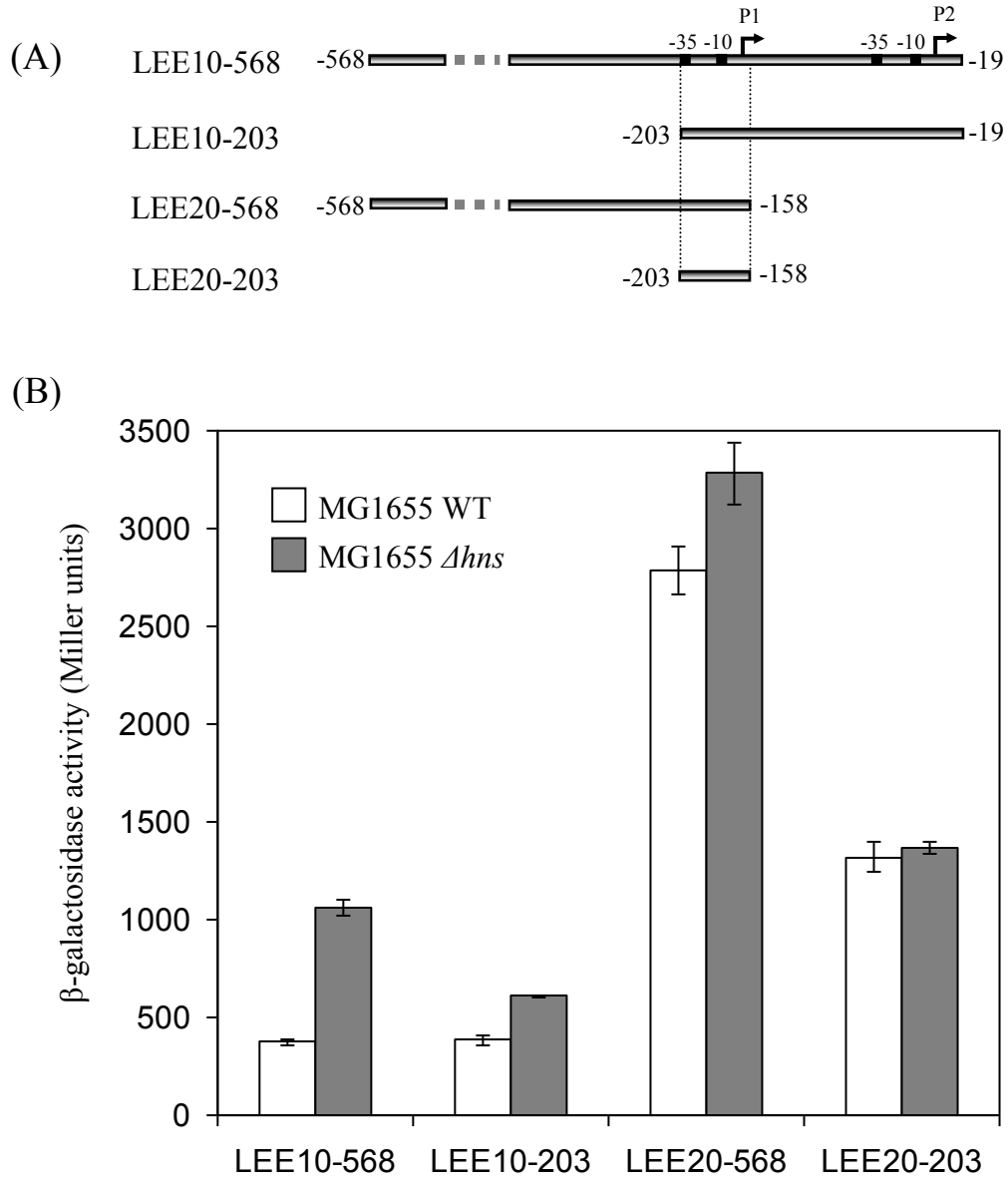


Figure 3.9. Repression by H-NS at the *LEE1* regulatory region.

A. Schematic representation of four nested deleted derivatives of *LEE1* operon regulatory region.

B. The bar chart illustrates measured β-galactosidase activities in *E. coli* K-12 strain MG1655 (WT) and a Δ*hns* derivative carrying pRW224 with LEE10-568, LEE10-203, LEE20-568 or LEE20-203 fragments. Measurements were made after growth of the cells in LB at 37ºC to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

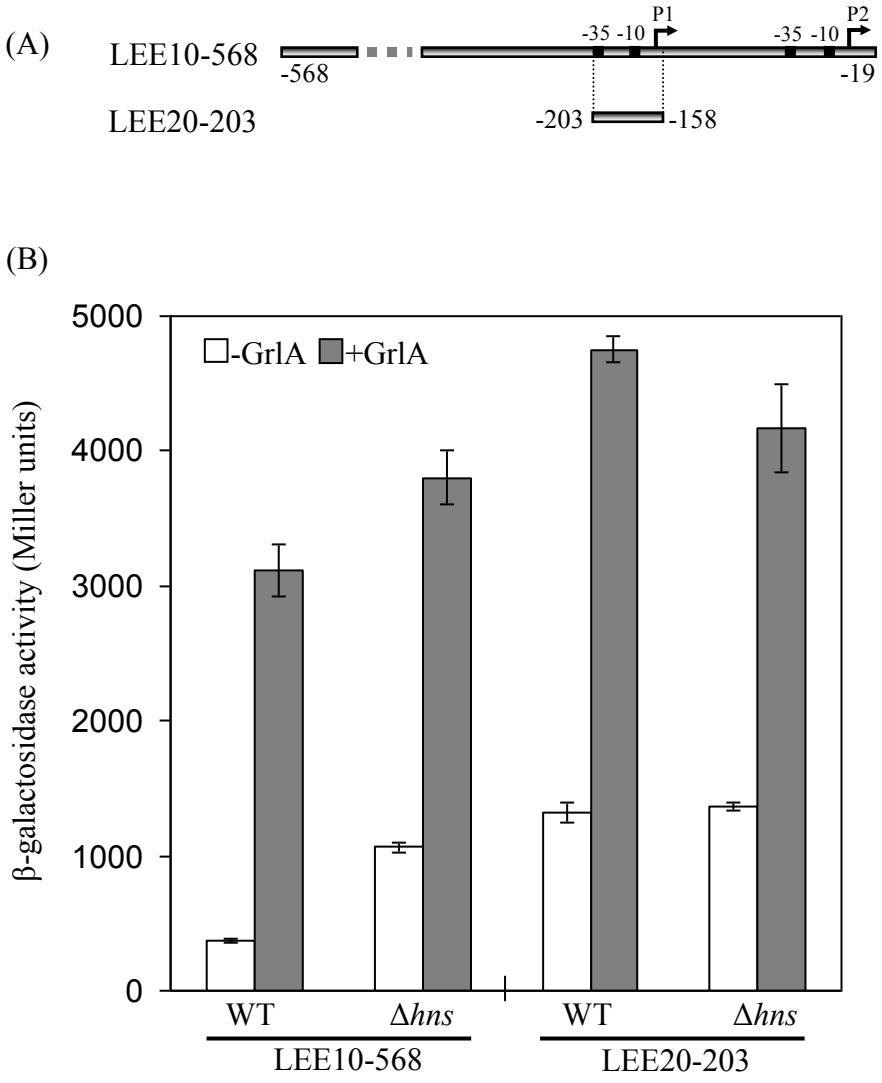


Figure 3.10. Repression by H-NS and activation by GrlA at the *LEE1* regulatory region.

A. Schematic representation of the LEE10-568 and LEE20-203 fragments of *LEE1* operon regulatory region.

B. The bar chart illustrates measured β-galactosidase activities in *E. coli* K-12 strain MG1655 (WT) and a Δ*hns* derivative carrying pRW224 with either the LEE10-568 or LEE20-203 fragments. Cells also contain either empty vector pACYC184ΔHN (-GrlA: open bars) or pSI02 (+GrlA: shaded bars). Measurements were made after growth of the cells in LB at 37ºC to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

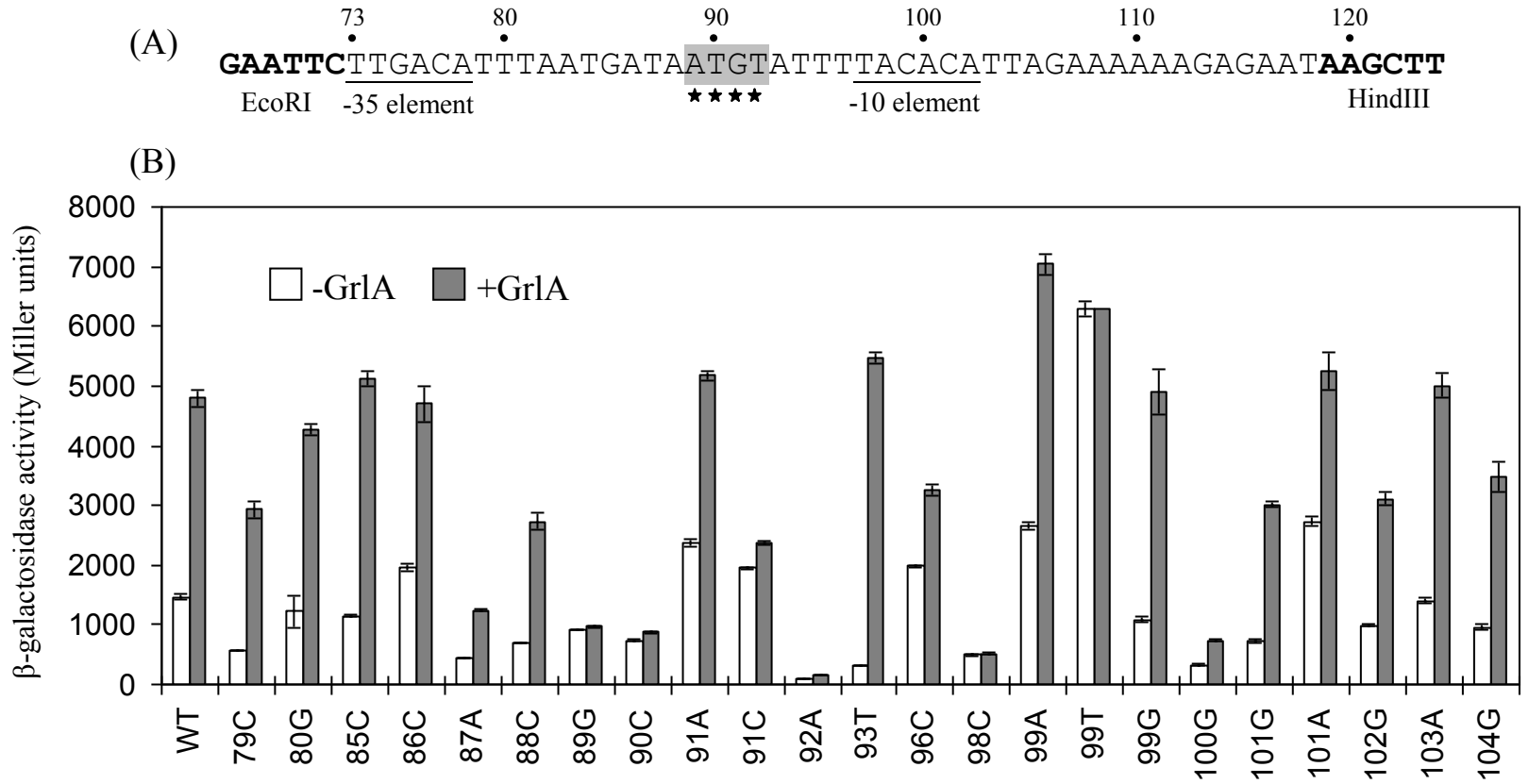


Figure 3.11. Mutational analysis of the LEE20-203 fragment.

A. Base sequence of the LEE20-203 fragment numbered and annotated as in Figure 3.3. The asterisks and shading indicate bases where substitutions suppress GrlA-dependent.

B. The figure illustrates β-galactosidase activities measured in M182 cells carrying pRW224 containing the starting LEE20-203 fragment (WT) or derivatives with different single base substitutions, indicated on the x-axis. Measurements were made in cells containing either pSI02 (+GrlA: shaded bars) or empty vector, pACYC184ΔHN (-GrlA: open bars). Standard deviations were obtained from three independent experiments.

element and as described earlier, it completely inactivates the P1 (Figure 3.4). In contrast, positions 89-92 fall within the 18 base pair promoter spacer region. The effects of changes at these positions could be explained by a simple way that they fall within the binding site for GrlA and hence, suppress GrlA binding and prevent activation. Since the *LEE1* P1 promoter carries 18 base pairs non-optimal spacer sequence between its -10 and -35 elements (Figure 3.3) and GrlA binds the spacer sequence (Figure 3.11), it was speculated that the spacer length might play a role in the GrlA-dependent activation of this promoter. To check this, a number of derivatives of the LEE20-203 fragment containing insertions or deletions in the spacer region were constructed. In order to ensure that either base deletions or insertions are not interfering with the GrlA binding site, changes were made either upstream or downstream of, but not within, positions 89-92. The effects of altering the length of the spacer length on GrlA-dependent activation at the *LEE1* P1 promoter were then measured and recorded in Table 3.2. The results show that basal promoter activity is increased by single base deletions at positions 79 or 94, but decreased by two-base deletions at positions 79 and 80, or by single base insertions between positions 78 and 79 or between positions 93 and 94. Therefore, as expected, the optimal *LEE1* P1 activity is found with a 17 base pair spacer. On the other hand, GrlA-dependent activation of the *LEE1* P1 promoter is found only with the starting 18 base pair non-optimal spacer. Interestingly, the activation is suppressed by all of the deletions and insertions that I tested. These results unambiguously suggest that the spacer length between the P1 promoter -35 and -10 elements is crucial for the GrlA-dependent activation of the promoter.

A complementary experiment was performed to check that results concerning the GrlA target in the P1 spacer obtained from experiments with the short LEE20-203 fragment are applicable for the full-length LEE10-568 fragment. Hence, LEE10-568 derivatives containing few different altered bases at P1 were constructed by megaprimer PCR and cloned to pRW224. The effects of the changes on GrlA-dependent activation were measured. Like at the short LEE20-203 fragment, base alterations in the P1 promoter spacer affect GrlA binding at full -length LEE10-568 regulatory region (Figure 3.12).

3.8. Biochemical and genetic analyses of GrlA binding to its target

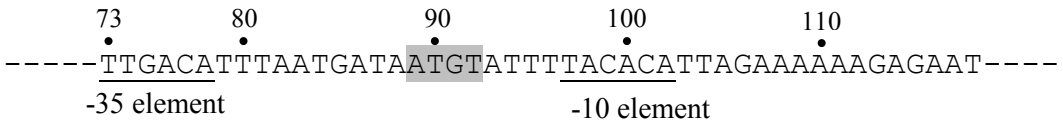
As part of this research, efforts to show direct interaction between purified GrlA and the

Table 3.2: Effects of spacer length on activation by GrlA at the *LEE1* P1 promoter

Derivative	Promoter sequence from positions 73-102	Spacer length (bp)	β-galactosidase activity (Miller units ± sd)		
			- GrlA	+GrlA	Fold activation
LEE20-203 (WT)	5'-TTGACATTTAATGATAATGTATTTTACACA-3'	18	1479 ± 36	4797 ± 140	3.2
Δ79T	5'-TTGACA—TTAATGATAATGTATTTTACACA-3'	17	6255 ± 87	7213 ± 149	1.2
Δ94T	5'-TTGACATTTAATGATAATGTA—TTTACACA-3'	17	3918 ± 79	4097 ± 55	1.0
Δ79/80T	5'-TTGACA——TAATGATAATGTATTTTACACA-3'	16	473 ± 21	491 ± 29	1.0
InsT (78-79)	5'-TTGACA**T**TTTAATGATAATGTATTTTACACA-3'	19	1054 ± 123	947 ± 12	1.0
InsT (93-94)	5'-TTGACATTTAATGATAATGTA**T**TTTTACACA-3'	19	588 ± 6	621 ± 13	1.0

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-203 fragment and different derivatives with insertions or deletions in the P1 promoter spacer. Cells also contained either plasmid pACYC184ΔHN (-GrlA) or pSI02 (+GrlA). Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). The second column of the table shows the fragment base sequence from position 73 to position 102, with the P1 promoter -10 and -35 hexamer elements underlined. Base deletions in the different fragments are indicated by dashes whilst insertions are shown in boldface type and coloured red. The fold activation by GrlA for each fragment is shown in sixth column.

(A)



(B)

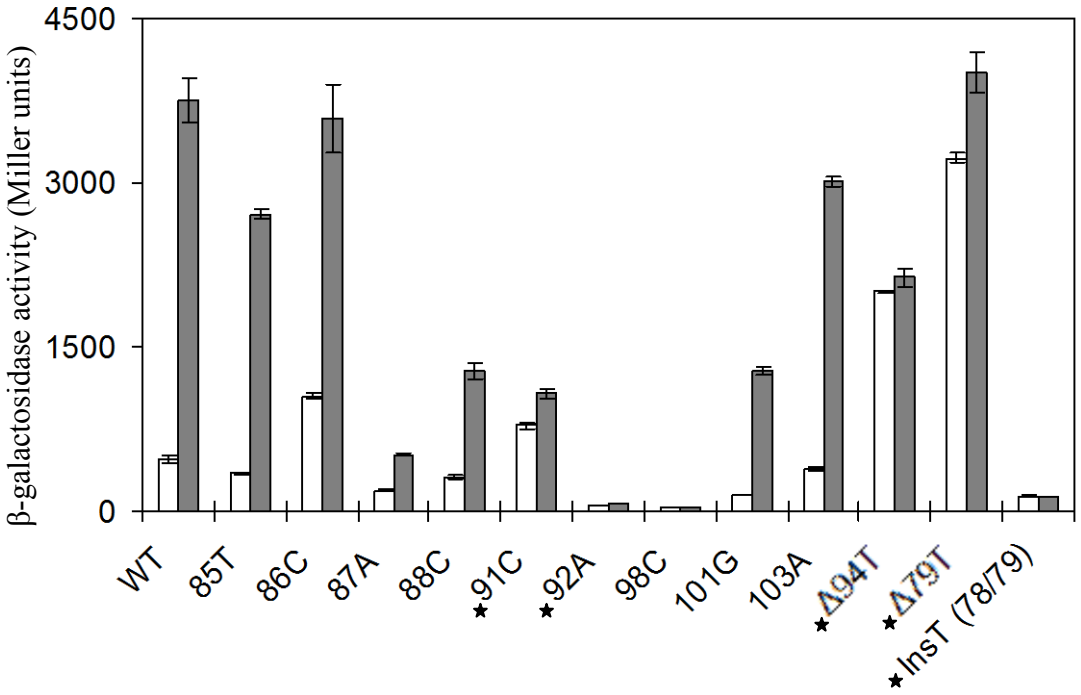


Figure 3.12. Mutational analysis of the LEE10-568 fragment.

A. Base sequence from positions 73 to 118 of the LEE10-568 fragment numbered and annotated as in Figure 3.5. The shading indicates bases where substitutions suppress GrlA-dependent activation.

B. The figure illustrates β-galactosidase activities measured in M182 cells carrying pRW224 containing the starting LEE10-568 fragment (WT) or derivatives with different single base alterations, indicated on the x-axis. Measurements were made in cells containing either pSI02 (+GrlA: shaded bars) or empty vector, pACYC184ΔHN (-GrlA: open bars). Base alterations that suppress GrlA-dependent activation are indicated by asterisks.

LEE1 P1 promoter as well as to identify the GrlA target at the promoter spacer using the well-known *in vitro* band shift and footprinting techniques, respectively (Minchin and Busby, 2009) were made. However, all efforts were unsuccessful as it proved impossible to purify soluble functional GrlA. To solve the problem, different genetic approaches were used. Firstly, suppression genetic analysis was considered. In this experiment, advange was taken of having the LEE20-203 derivative containing 92A mutation that reduces the expression from the fragment in either the presence or absence of wild type GrlA (Figure 3.11). Using suppression genetic approach, it was attempted to show whether any mutation in the GrlA could improve the expression level from the derivative of LEE20-203 containing 92A mutation. To do this, error-prone PCR was used to create 10 independent preparations of random mutations in the *grlA* coding sequence within pSI02. Mutated plasmids were then transformed into *E. coli* strain M182 containing pRW224 carrying the LEE20-203 fragment with the 92A substitution. Note that, reduced expression due to the 92A mutation results in Lac^- colonies on MacConkey lactose indicator plates. However, after screening over 75,000 transformants, 11 Lac^+ colonies containing pSI02 with single-base substitutions leading to the RK53, IT59 or IV59 substitutions in GrlA were found. Interestingly, all of these these substitutions fall in the predicted DNA recognition helix of the helix-turn-helix DNA binding motif of GrlA (Figure 3.13A). Results illustrated in Figure 3.13B show that the RK53, IT59 and IV59 substitutions partially restore GrlA-dependent activation with the LEE20-203 fragment carrying the 92A mutation.

In order to check whether the restoration of GrlA-dependent activation is specific for 92A base substitution, the effects of RK53, IT59 and IV59 substitutions on the expression from the LEE20-203 derivatives carrying either 89G, 90C or 91C substitutions were measured. Results illustrated in Figure 3.13B show that none of the substitutions (RK53, IT59 or IV59) improves the expression from LEE20-203 derivatives containing either 89G, 90C or 91C mutations. This suggests that the GrlA-dependent activation can only be restored by the RK53, IT59 or IV59 substitutions if the LEE20-203 derivative carries 92A mutation. A possible explanation of these results is that when bound at the *LEE1* P1 promoter, residues 53 and 59 may contact with the base pair at position 92. This suggests that the predicted helix-turn-helix in GrlA is functional in binding the target site at the *LEE1* P1.

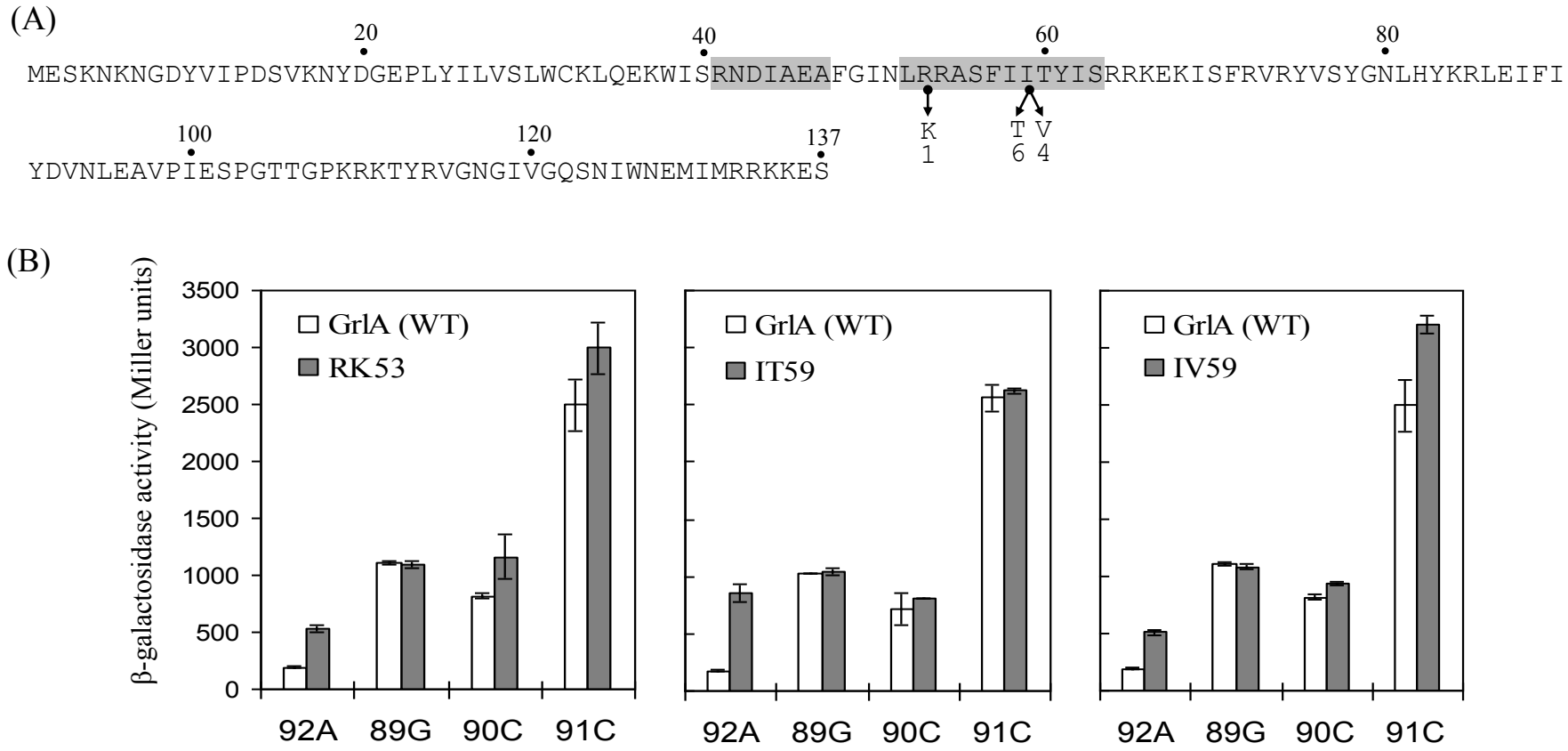


Figure 3.13. Analysis of GrlA-dependent activation using suppression genetics.

A. The figure shows the GrlA amino acid sequence. The helix-turn-helix DNA binding motif predicted by Deng *et al.* (2004), and confirmed by the Dodd & Egan, GYM2.0 and Jpred prediction tools (see section 2.18) is indicated by grey shading. Substitutions at R53 and I59, identified after random mutagenesis and selection for suppressors of the 92A mutation in the *LEE1* P1 promoter, are indicated. The number adjacent to each substitution is number of times that it was isolated.

B. The figure shows the effect of GrlA substitutions on the activity of *LEE1* P1 promoter mutants. The bar charts illustrate β-galactosidase activities in M182 cells carrying pRW224 containing the LEE20-203 fragment with single base substitutions, indicated on the x-axis. Measurements were made in cells containing either pSI02 encoding wild type GrlA (WT, open bars) or GrlA with the RK53, IT59 or IV59 substitutions, as indicated (grey bars). Standard deviations were obtained from three independent experiments.

In a complementary experiment, a set of pSI02 derivatives encoding GrlA with alanine substitutions at sequential residues from L52 to S63 were constructed and the GrlA-dependent activation of expression from the starting LEE20-203 fragment was measured. Results illustrated in Figure 3.14A show that RA54 substitution reduces the activation most (Figure 3.14A). These results together with the suppression genetic data suggest that the side chain of residue R54 in GrlA makes a contact with the GrlA target at any of four positions 89-92 in the LEE20-203 fragment and that the RA54 substitution causes a loss of the side chain and thus, ensuing the activation and that the RK53 substitution in the neighbouring side chain makes a contact that compensates for the 92A substitution.

In order to determine whether any of the alanine substitutions could restore the activity lost due to specific base substitutions in the putative GrlA operator target in the LEE20-203 promoter fragment, the expression from the LEE203 promoter derivatives containing either 89G, 90C or 91C was measured in the presence of alanine substitutions or the wild type GrlA. Among different alanine substitutions, the FA57 substitution partially restores the effects of the 89G and 90C substitutions (Figure 3.14B and C). The results suggest a direct interaction between the substitution FA57 and base pairs at positions 89-90 in the mutated LEE20-203 89G or 90C fragments. In other words, F57 collides with the base pairs at positions 89-90 in the mutated LEE20-203 89G or 90C fragments, probably because when bound, residue F57 comes closer to positions 89 and 90. Likewise, the residue R53 contacts with the position 91(Figure 3.14D).

As an alternative approach to see the direct binding of GrlA protein to the *LEE1* P1 promoter, DNA-sampling protocol was used, originally reported by Butala *et al*., (2009). It is a method of rapid isolation of a specific segment of DNA and accompanying proteins from *E. coli* K-12 strain. The DNA fragment to be sampled is produced as a discrete fragment within cells by the yeast I-SceI meganuclease, and is purified using FLAG-tagged LacI repressor and beads carrying anti-FLAG antibody.

LEE20-275 promoter fragment (Figure 3.3) was cloned into pRW902 EcoRI-HindIII sites adjacent to the 5 LacI operators with two flanking 18 base pairs target sites for the yeast meganuclease I-SceI (see plasmid detail in Figure 2.4). The resulting plasmids were

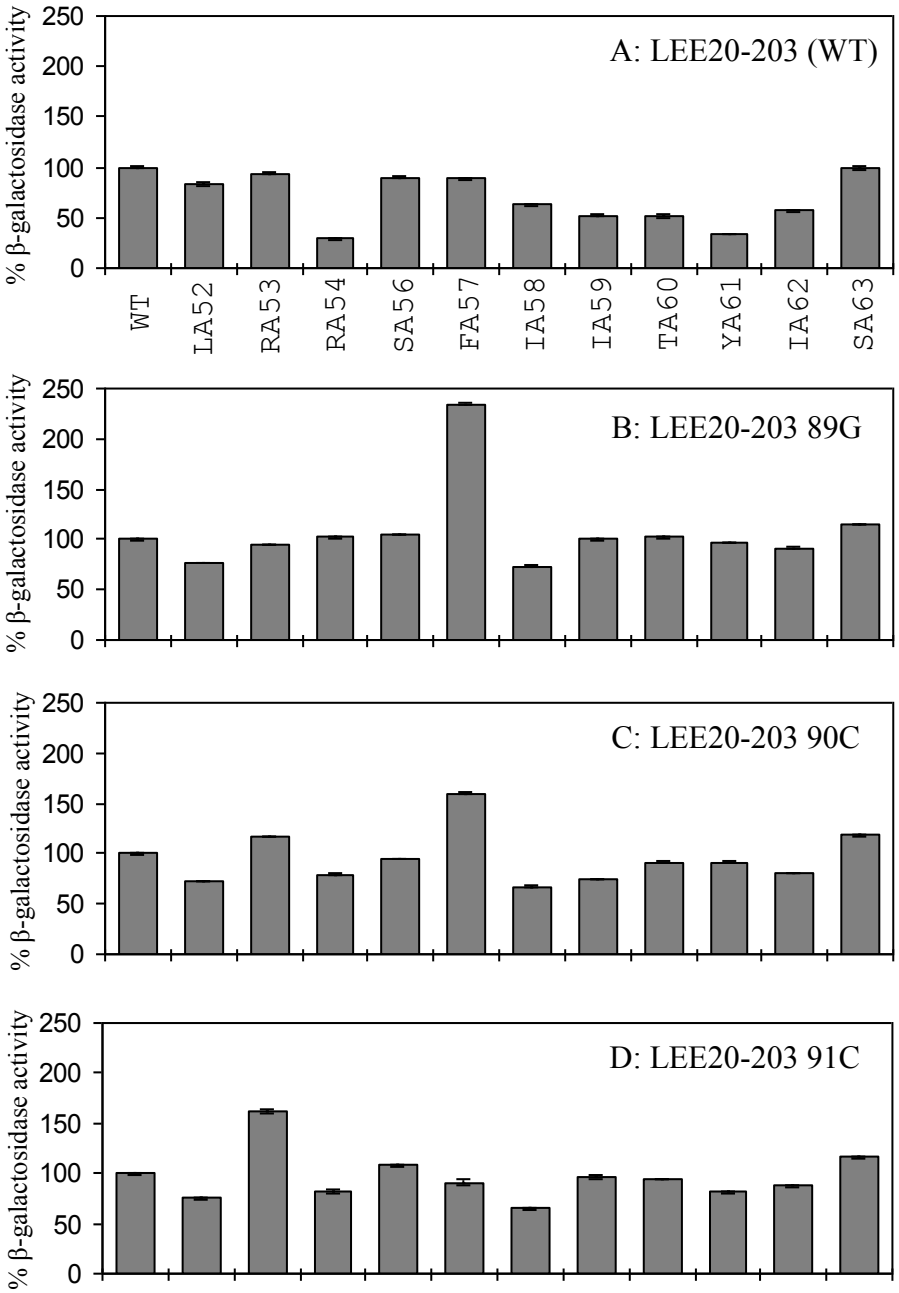


Figure 3.14. Epistasis analysis of GrlA interactions.

The figure shows bar charts that illustrate measured β-galactosidase activities in M182 cells carrying pRW224 containing the LEE20-203 fragment with the starting *LEE1* P1 promoter sequence (panel A), and with the 89G (panel B), 90C (panel C) or 91C (panel D) mutations. Measurements were made as in Figure 3.6 with cells carrying pSI02 encoding wild type GrlA (WT) or the different alanine substitutions indicated on the x-axis. For each promoter, the data are expressed as a percentage of the activity with wild type, the values are the average of three independent assays, and standard deviations are shown with bars.

transformed into *E. coli* K-12 MG1655 3xFLAG-tagged *lacI* (Table 2.1). Plasmid pBAD-SG (Figure 2.9) that encodes SceI and the bacteriophage lambda Gam protein, under the control of an arabinose inducible promoter was co-transformed. The strain was also co-transformed either with pACYC184ΔHN (GrlR-A-), pSI01 (GrlR+A+) or pSI02 (GrlR-A+). After induction by arabinose, a discrete DNA fragment carrying the promoter of interest with five LacI repressor targets was generated and the fragment was stable due to inhibition of RecBCD by the Gam protein. Immunoaffinity methods using magnetic beads carrying anti-FLAG antibodies were then used to isolate the DNA fragment together with accompanying proteins, which were finally identified using gel electrophoresis and mass spectroscopy. Proteins that were accompanied with the LEE20-275 fragment are shown in Figure 3.15. Proteins that were confirmed by the mass spectroscopy are indicated by asterisks in the Figure 3.15. These are 50S ribosomal protein L15, RNAP ω-subunit, Hfq and HU. Other protein bands are labelled according to their sizes (Lee *et al*., 2008; Butala *et al*., 2009). On basis of sizes, the analysis reveals a major band corresponding to the FLAG-tagged LacI and bands for RNAP subunits in the complex. However, no GrlA protein (~16 kDa) was detected in the lane 4. The intensity of the bands of the RNAP subunits in the lane 4 seems to be high compared with those in land 2 and 3 suggesting that the level of RNAP is high in the lane 4 compared with those in lane 2 or 3. This suggests that the presence of GrlA in the cell possibly causes increased recruitment of the RNAP to the *LEE1* P1 promoter.

In order to determine whether GrlA can bind RNAP prior to interacting with its target at the spacer between P1 promoter -10 and -35 elements, a pull down experiment was conducted essentially following Lee *et al*. (2008). The experiment was designed to analyze protein A affinity tagged RNAP (β':4PrA) with associated proteins from *E. coli* Sakai cells. In this experiment, plasmids pACYC184ΔHN (GrlR-A-), pSI01 (GrlR+A+) or pSI02 (GrlR-A+) were transformed in an *E. coli* O157:H7 Sakai strain containing chromosomal fusion of protein A affinity tag with the RNAP β'-subunit (see detail in section 2.14). RNAP was affinity isolated from the cells grown to mid-exponential growth phase in LB and the protein complexes were analyzed by SDS-PAGE (Figure 3.16). However, the experiment fails to detect GrlA protein (~16 kDa) on the gel. As expected β':4PrA fusion protein (RpoC) and the other three components of core RNA polymerase β (RpoB), α (RpoA), and ω (RpoZ) were found. These protein bands are labelled according to their sizes (Lee *et al*., 2008).

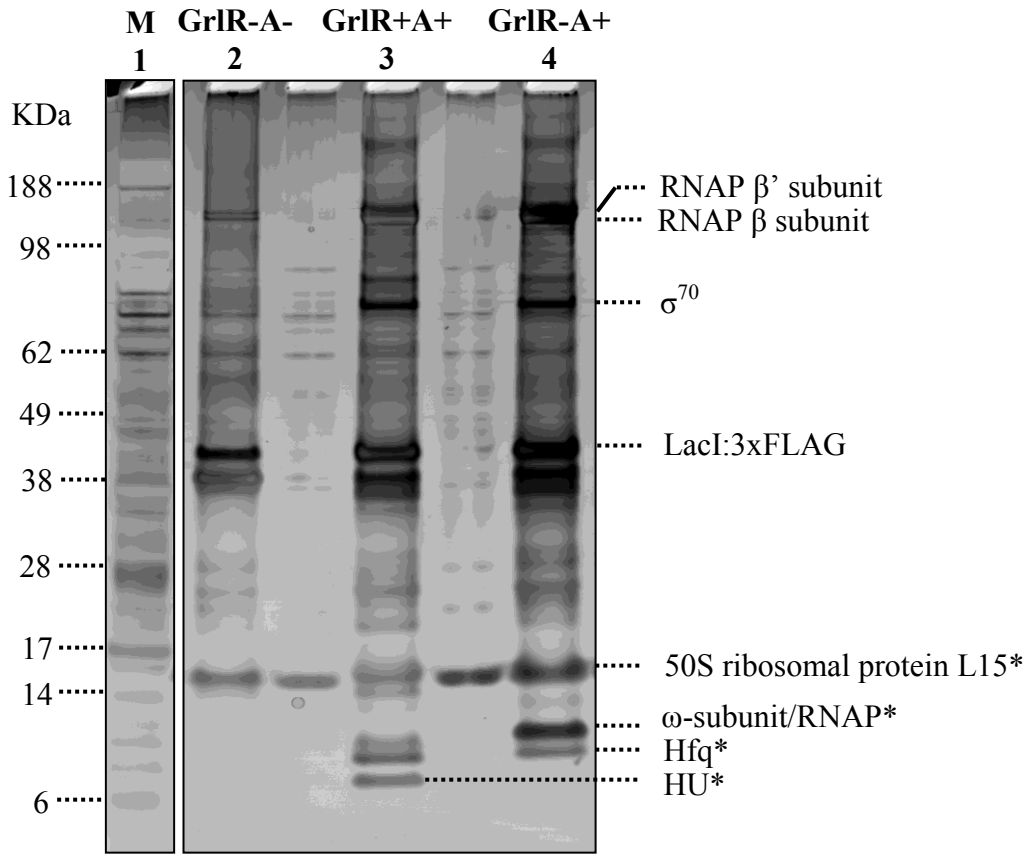


Figure 3.15. SDS-PAGE analysis of proteins bound on the LEE20-275 promoter fragment.

The figure shows the affinity-isolated proteins from *E. coli* K12 strain MG1655 grown in LB medium. Lane1 represents the calibration of the gel using Invitrogen SeeBlue Plus 2 protein markers (M). The lane 2 represents proteins from cells containing plasmid pACYC184ΔHN (GrlR-A-, as control). Lane 3 shows proteins from the cells containing plasmid pSI01 (GrlR+A+). Lane 4 shows proteins from the cells carrying plasmid pSI02 (GrlR-A+). Proteins that were confirmed by mass spectrometry are indicated by asterisks, whilst other proteins are labelled according to their sizes (Lee *et al*., 2008; Butala *et al*., 2009).

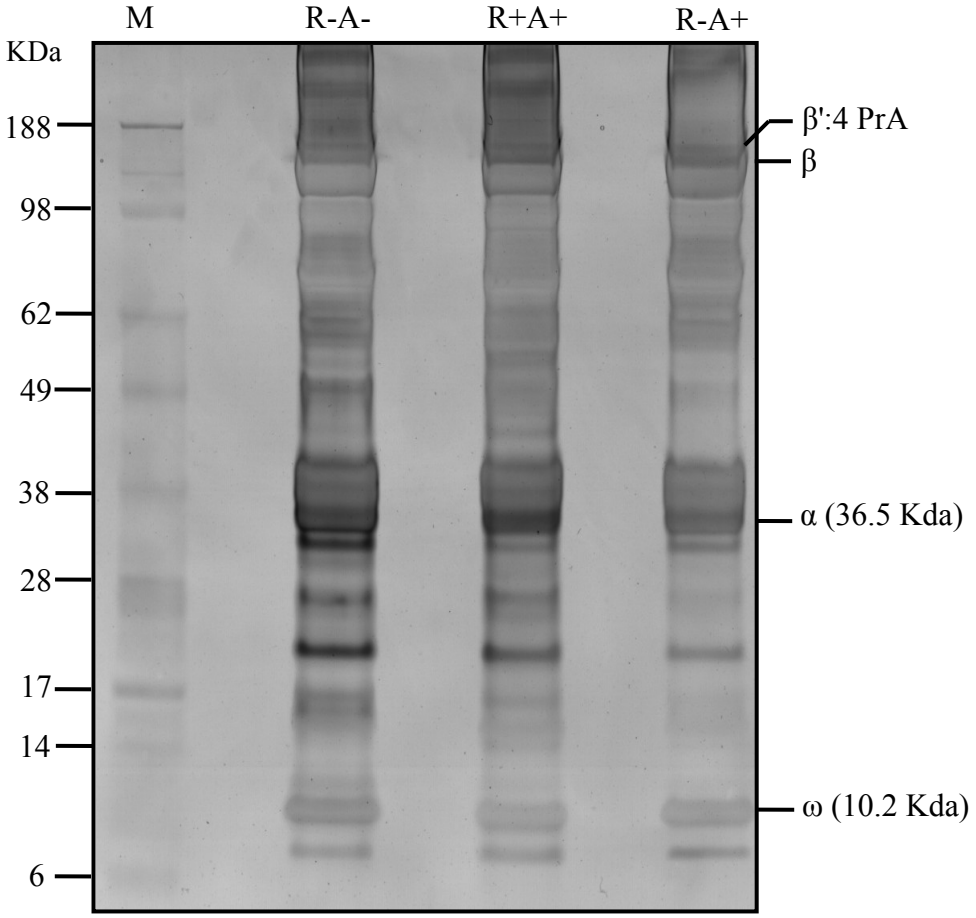


Figure 3.16. Affinity isolation of RNA polymerase binding proteins.

SDS-PAGE analysis of pulled out RNA polymerase associated proteins from *E. coli* O157:H7 Sakai carrying chromosomal *rpoC:PrA* fusion. Cells also contain plasmids pACYC184ΔHN (R-A-), pSI01 (R+A+) or pSI02 (R-A+). Analyses were made after growth to mid-exponential growth phase in LB medium. The gel was stained with Coomassie blue and calibrated with Invitrogen SeeBlue Plus 2 protein markers (M). Bands are labelled according to their sizes (Lee *et al*., 2008; Butala *et al*., 2009).

3.9 Discussion

The LEE pathogenicity island is one of the major determinants playing a leading role during infection in pathogenic *E. coli* and other related bacteria. The expression of the different LEE genes must be carefully controlled in time and space. Different regulatory inputs explore the *LEE1* operon regulatory region to control overall expression from the LEE. Here, I have investigated GrlA-dependent activation from the *LEE1* operon regulatory region using a simplified system where non-pathogenic *E. coli* K-12 strain was used as a 'test tube'.

Firstly, the relative contribution of the two promoters, P1 and P2 was determined, in the expression from the *LEE1* operon regulatory region and it was concluded that P1 is the major promoter in the *LEE1* expression. This conclusion is in agreement with Porter *et al*. (2005). In contrast, Sperandio *et al*., (2002) as well as Sharp and Sperandio (2007) reported a bigger role for P2. However, possible reasons for the differences could be the precise fusions, strains and conditions used for the experiments. Experimental conditions of this work allowed for the detailed mutational analysis of the key P1 promoter (Figure 3.3 and Table 3.1) and TTGACA and TACACA were identified as the functional -35 and -10 hexamer elements, respectively. The -35 element, which matches perfectly the consensus for *E. coli* promoters (Rosenberg and Court, 1979; McClure, 1985) was predicted by Sperandio *et al*. (2002) and other groups. In the case of the -10 element, Sperandio *et al*. (2002) originally assigned TACACA as -10 element whilst other groups marked a hexamer sequence located 2 base pairs upstream, TTTACA as -10 element (Porter *et al*., 2005; Russell *et al*., 2007; Sharp and Sperandio, 2007). Differences in tracing the -10 element are probably due to the uncertainty in the location of the P1 transcript start. It is likely that both -35 and -10 elements were identified on the basis of their relative distance from the transcript start site and by comparison to *E. coli* promoter consensus -10 and -35 hexamers. However, mutational analysis presented here argues unambiguously that the TACACA is the -10 element of the P1 promoter. This means that the P1 promoter has a non-optimal 18 base pairs spacer sequence, which is just one above the optimal 17 base pairs for promoter activity (McClure, 1985). It is important to mention that the location of the 5' end of transcripts with respect to -10 hexamer regions at bacterial promoters is not fixed (Darst, 2009). Moreover, highly AT-rich sequence downstream of the -10 element might cause a greater than usual unwinding (>20 base pairs) seen in the open complex probed by permanganate (Figure 3.4A; compare with Browning *et al*., 2009). This

may be the reason for the greater flexibility at the *LEE1* P1 promoter. Therefore, the different transcription start points reported by different groups in different experiments could all be correct. Moreover, Sclavi (2009) reported that in many cases, the upstream end of the -10 hexamer provides the major anchoring point for RNAP and the location of the first templated base is not fixed. For this reason, instead of following the customary convention of numbering base sequences with respect to the transcript start point, an arbitrary numbering system was used considering the first cloned base as a start point to describe the experiments (Figure 3.3).

Though most summaries of transcriptional regulation in the LEE (e.g. Kendall *et al*., 2010) show GrlA as an activator for expression from the *LEE1* regulatory region, the current literature is not completely unambiguous. The K-12 system reconstitutes a significant activation at *LEE1* by GrlA (Figure 3.6) and shows that the activation was suppressed by GrlR. This suggests that the experimental conditions presented here provide physiologically relevant effects. Intriguingly, it was found that GrlA could activate expression from a very short promoter fragment (LEE20-203) that contains only 52 base pairs sequence and lacks upstream sequence of promoter -35 element. More strikingly, interacting site for GrlA was found at the spacer between promoter -35 and -10 elements. The observation that GrlA exhibits dual functions in regulation of the *LEE1* regulatory region is in line with Jimenez *et al*. (2010). In one hand, GrlA counteracts H-NS-mediated repression of the expression from the *LEE1* promoter regulatory region and on the other hand, it activates the promoter in H-NS independent fashion (Figure 3.10). However, expression from the short 52 base pairs LEE20-203 fragment is not repressed by H-NS but activated by GrlA. This observation unambiguously suggests that GrlA can function as a true transcription activator.

To date, no one is able to define GrlA target at the *LEE1* promoter *in vitro*. This is most probably due to the difficulty in obtaining purified functional GrlA. Attempts were made to reproduce the *in vitro* binding of purified GrlA, reported by Huang and Syu (2008) and Jimenez *et al.* (2010) but they were unsuccessful. However, different genetic approaches were used to show that the predicted DNA-binding helix-turn-helix motif in GrlA is functional, and to define its target in the promoter. Most bacterial transcription activators bind sites upstream of or overlapping the promoter -35 element (Browning and Busby, 2004). Surprisingly, it was found that GrlA binds the spacer region between the *LEE1* P1 promoter -35 and -10 elements. It is important to recall that many members of the MerR family transcription regulators bind

to sites located between the -10 and -35 hexamers at target promoters (reviewed by Brown *et al*., 2003). Moreover, they act on their target promoters where the length of the spacer sequence between the -10 and -35 hexamers is higher than the optimum 17 base pairs. Models for activation by MerR family members show that their tight binding to the spacer between target promoters -35 and -10 elements causes physical bending or twisting the DNA resulting in reorientation of the promoter -35 and -10 elements. Thus, the rearranged promoter -35 and -10 elements could bind the RNAP effectively to initiate transcription (Ansari *et al*, 1995; Brown *et al*., 2003). The observation that shortening of the *LEE1* promoter spacer from 18 to 17 base pairs increases the promoter activity and ceases GrlA-dependent activation suggests that activation mechanism of GrlA is similar to many of the MerR family activators (Figure 3.17), even though there is no significant sequence similarity between GrlA and any of the member of the MerR family. However, association of free GrlA with holo RNAP prior to binding at target promoter cannot be ruled out. But a preliminary study using the materials and pull-down protocols described by Lee *et al*. (2008) did not find any evidence for direct association of GrlA with free RNAP (Figure 3.16).

GrlA increases the *ler* transcription but in combination with GrlR, it does not have the effect on *ler* expression suggesting that GrlR modulates the GrlA-dependent activation of *ler* gene expression. However, to make sure GrlA-dependent activation of *LEE1* promoter, bacterial cells might sense a signal(s) that may modulate anti-GrlA action of GrlR. In order to address whether any factor(s) is involved to overcome the GrlR mediated repression in GrlRA action, expression from *LEE1* promoter regulatory region were quantified in the presence of plasmid encoded GrlRA together with *in trans* supply of different small proteins (PchC, CesD, CesT, CesL or Ler) in *E. coli* K12 cells or the activity was measured in the presence of different low-molecular-weight ligands or signaling molecules such as sodium bicarbonate, small fatty acids etc. None of the substances that were tried was found to be involved in releasing the GrlR-mediated repression of GrlRA (data not shown).

To sum up, GrlA activates the *LEE1* promoter by a novel mechanism that underscores the complex regulatory system at the *LEE1* operon regulatory region. However, to get precise knowledge on the action of GrlA, other targets of GrlA need to be analysed. It has been reported that GrlA positively regulates the expression of the Ehx enterohaemolysin (Saitoh *et al.,* 2008) and repress genes responsible for the formation of flagella (Iyoda *et al.*, 2006).

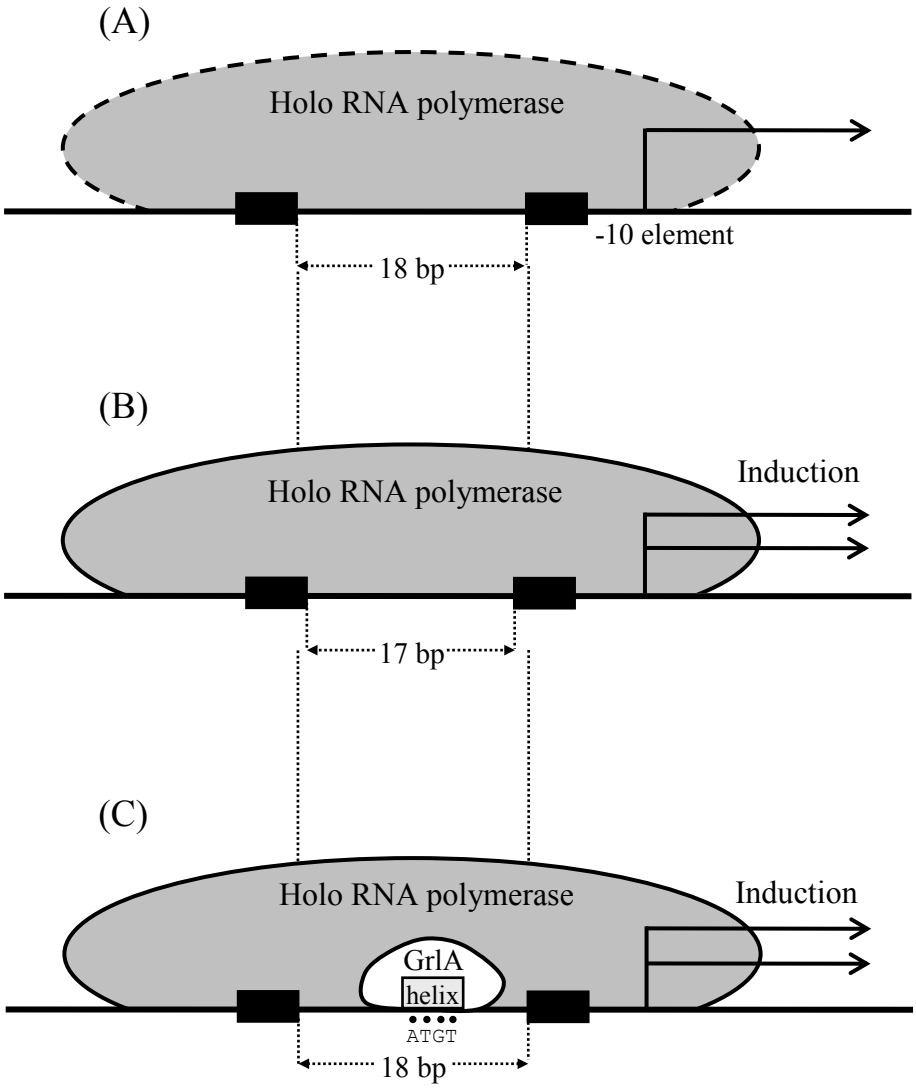


Figure 3.17. Model for activation of *LEE1* P1 promoter by GrlA.

A. Weak recognition of the promoter by RNA polymerase due to sub-optimal spacer length.
B. Efficient recognition of the promoter by RNA polymerase due to optimised spacer length.
C. GrlA interacts with the spacer sequence and facilitates RNA polymerase activity.

Chapter 4

A cryptic promoter in the *LEE1* regulatory region: promoter specificity in AT-rich gene regulatory regions

4.1 Introduction

Bacterial gene regulatory regions contain promoters that have evolved to set the level of expression of downstream genes, and the different promoter elements that are recognised by the multi-subunit bacterial RNAP are now well understood. Exhaustive analyses of *E. coli* promoters have shown that the principal promoter elements are the -35 and -10 hexamer elements, and consensus sequences of TTGACA and TATAAT respectively have been derived. Thus, to a first approximation, the strength of an *E. coli* promoter will depend on the degree to which the base sequence of the functional -10 and -35 elements correspond to the consensus (reviewed in Busby *et al*., 2009).

Computational analysis of *E. coli* sequences has revealed the existence of potential -10 and -35 elements at higher than expected frequencies in intergenic regulatory regions (Huerta and Collado-Vides, 2003; Huerta *et al*., 2006). Whilst this is probably a by-product of the evolution of gene regulatory regions, it raises the issue of the specificity of promoter recognition by RNAP, and the role played by cryptic promoter elements. This is especially relevant for regulatory regions within horizontally acquired pathogenicity islands that are generally characterised by a high proportion of AT base pairs (Hacker *et al*., 1997).

Experimental evidence that supports the existence and function of promoter-like sequences or cryptic promoters within the regions upstream of genes is limited. Woody *et al* (1993) found that mutation at the P_R promoter in the O_R region of bacteriophage λ unmasks the P_α cryptic promoter. P_α competes with both P_R and P_{RM} promoters since it is only active when P_R is mutated and can be suppressed by mutations that increase P_{RM} activity.

The *E. coli* lactose operon regulatory region has a principal promoter, *lac* P1 that is overlapped by at least five promoter-like elements or cryptic promoters (Reznikoff, 1992). One well-studied cryptic promoter is *lac* P2. Transcription initiates from this promoter at a site that is located 22 base pairs upstream of the P1 promoter transcript start. P2 promoter interferes with RNAP binding at the P1 promoter (Malan and McClure, 1984). In the *E. coli* galactose operon regulatory region, there are three overlapping promoters P1, P2 and P3 that specify transcription initiation from sites S1, S2 and S3, respectively. Musso *et al.* (1977) first found that mutational analysis of *gal* P1 promoter abolished transcription initiation from S1

site but that initiation could occur from the S2 site of P2 promoter located 5 base pairs upstream of S1. In contrast to *lac* P1 and P2 promoters, *gal* P1 and P2 promoters are distinct each other since mutation in P1 eliminates transcription initiation from S1 site without affecting the transcription initiation from S2 site of P2 promoter and vice-versa (Busby *et al*., 1982; Bingham *et al*., 1986). Ponnambalam *et al*. (1987) further reported that a point mutation that simultaneously stops transcription initiation from sites S1 and S2, unmasks a third promoter, P3 that initiates transcription from the S3 site, which is located 14 base pairs upstream of S1.

Transcription initiation from overlapping promoters both in *lac* and *gal* operons are co-ordinately regulated by the cAMP and its receptor protein, CRP (Musso *et al*., 1977; Aiba *et al*., 1981; Malan and McClure, 1984; Kuhnke *et al*., 1986). In both operons, cAMP-CRP causes shifting of transcription initiation from the upstream promoter P2 to the downstream promoter P1. Generally, CRP-independent transcription comes from the P2 promoter. The presence of CRP plus cAMP causes repression of transcription from P2 promoter and initiates transcription from P1 promoter. Thus, CRP-cAMP can exert both positive and negative control at overlapping promoters.

In Chapter 3, mutational analysis of the LEE20-275 promoter fragment was performed to identify functional elements of the *LEE1* P1 promoter. During this work, a number of double mutations provided interesting observation. Subsequent biochemical- and mutational analyses of the mutated fragments unveiled a cryptic promoter designated, P1A, which overlaps with the principal P1 promoter and initiates transcription from a site located 10 base pairs upstream of the transcription initiation site of the P1 promoter. Surprisingly, mutations in the *LEE1* P1 promoter -10 and -35 hexamer elements unmask the cryptic promoter to different extents. This chapter describes experiments aimed to investigate the cryptic promoter.

4.2 Unexpected promoter activity unmasked by double mutations in the LEE20-275 fragment

In Chapter 3, error-prone PCR was exploited to generate a library of random mutations in the LEE20-275 fragment cloned in pRW224. MacConkey lactose indicator plates were then used to screen the Lac phenotype of colonies of M182 cells carrying the pRW224 library and

single base substitutions that reduced promoter activity were selected. Strikingly, over 90% of the selected base changes fell in two hexamer elements, TTGACA and TACACA, that were identified as the P1 promoter -35 and -10 hexamers respectively (Figure 3.3). Assays with the different mutated LEE20-275 fragments cloned in pRW224 showed that single mutations in the P1 -10 element reduced measured *lac* expression to 3-6% of the level with the starting fragment, whilst mutations in the -35 element reduced expression to 12-27% of the starting level.

Since random mutagenesis is an inefficient way to produce a full repertoire of promoter 'down' mutations, site-directed mutagenesis was used to introduce some mutations at specific positions in the P1 promoter. During the screening of Lac^- M182 colonies (white or pale pink colonies) that carried mutations at position 2 of the -10 hexamer (98C or 98G or 98T) of the P1 promoter in the LEE20-275 fragment cloned in pRW224, a number of anomalous Lac^+ colonies (Pink or Red colonies) were isolated. Sequence analysis showed that, in these colonies, the LEE20-275 fragments carrying mutation at position 98, contained an additional base alteration that resulted in LEE20-275 98C 64G, LEE 98C 35C, LEE20-275 98T 75C or LEE20-275 98G 76G derivatives (Figure 4.1).

In addition, to determine functional importance of different base species substations at position 96 (one base upstream of the P1 -10 hexamer sequence) in P1 promoter activity, the nucleotide T in 96 position was replaced with G or C in LEE20-275 fragment and cloned into pRW224. M182 cells containing recombinant plasmids were then screened for their Lac phenotype in MacConkey lactose indicator plate. Most of the colonies screened were Lac^+ (red colonies) and some were Lac^- (white or pink colonies). Sequence analysis confirmed that C for T change at position 96 (LEE20-275 96C) caused a red phenotype, G for T (LEE20-275 96G) caused a white phenotype. Amongst the colonies, I isolated G for T at position 96 together with G for A at position 100 (LEE20-275 96G 100G) that gave a pink phenotype.

The *lacZ* expression from each of the derivatives containing double mutations was compared with that of wild type LEE20-275 fragment or the derivatives with single mutations. Expression from the derivatives appears to be interesting as β-galactosidase activities from these derivatives significantly differ from those containing single mutations (Figure 4.1).

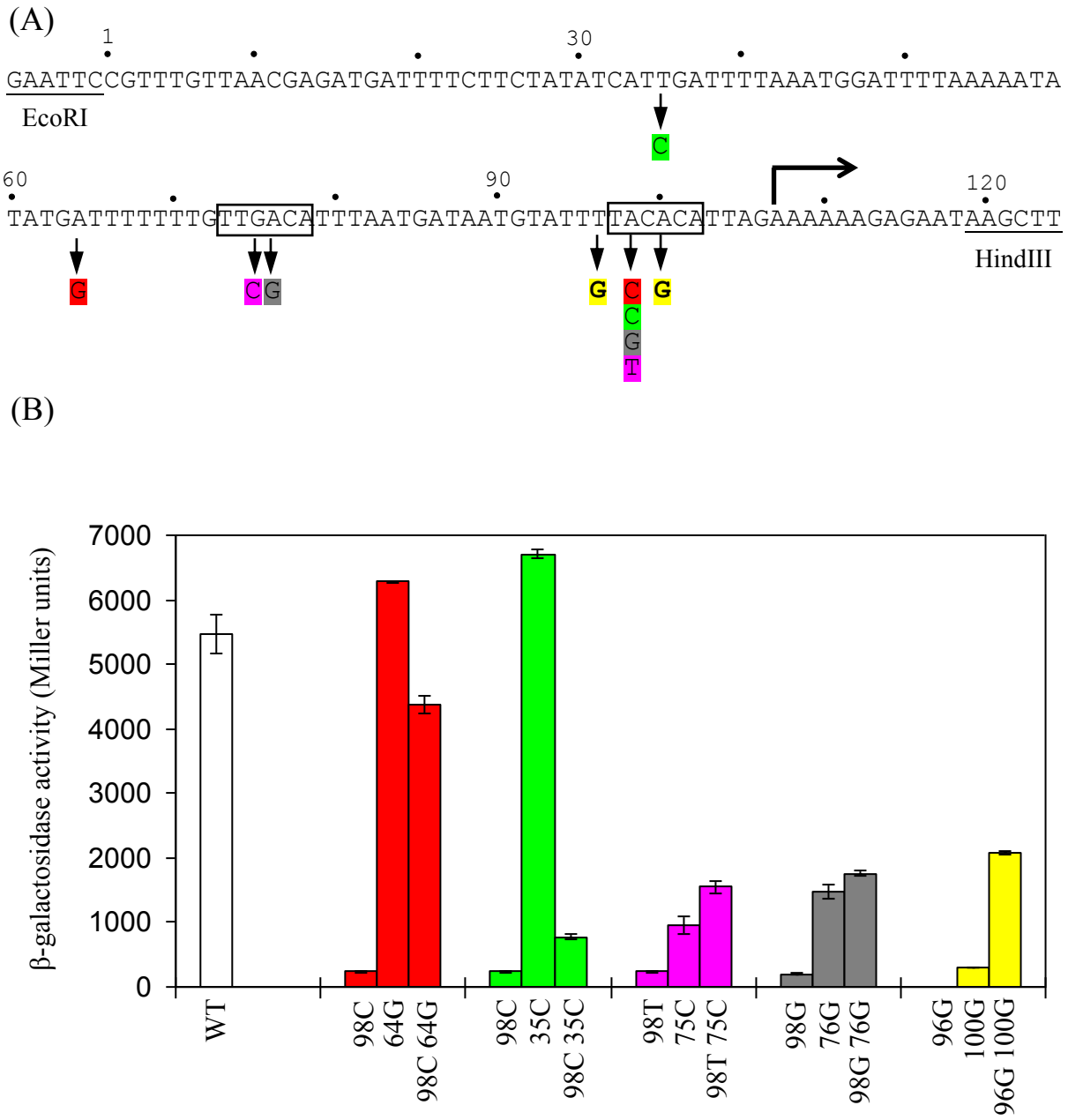


Figure 4.1. Effects of different double mutations in the expression from the LEE20-275 promoter fragment.

(A) Nucleotide sequence of LEE20-275 promoter fragment numbered and annotated as in Figure 3.3. EcoRI and HindIII restriction sites are underlined. Positions of the mutations are indicated by arrows and double base substitutions are indicated by the same colour.
(B) The figure illustrates β-galactosidase activities measured in M182 cells carrying pRW224 containing the starting LEE20-275 fragment (WT) or derivatives with different base substitutions. Measurements were made after growth of the cells in LB at 37ºC to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

Most interestingly, the base substitution 98C (LEE20-275 98C) at the second position of the P1 promoter -10 element almost inactivates the promoter as expected, but this mutation together with 64G (LEE20-275 98C 64G), located 34 base pairs upstream of the 98C base change, restores the wild type promoter activity. Table 4.1 shows measurements of the promoter activity of this fragment, together with controls. As expected, the 98C substitution reduces expression to 4% of the starting level, whilst the measured promoter activity of the LEE20-275 fragment with both the 98C and 64G substitutions is 20-fold higher, which is nearly 80% of the starting level. Results in Table 4.1 also show that, in isolation, the 64G mutation causes but a modest 14% increase in promoter activity.

The effect of each of the other double mutations is not as large as that of 98C 64G but they also gave interesting observations (Figure 4.1). Double base substitutions 98C 35C caused ~3 fold increase of the promoter activity compared with that of single base substitution 98C, whereas in isolation, the 35C causes 20% increase in promoter activity. The reason for the increased promoter activity due to 98C 35C mutations compared with 98C is not clear. Also given that base substitutions at the second position of -10 element (98T or 98G) reduced the promoter activity to ~3-4% and changes at third or fourth positions of the consensus -35 element (75C or 76G, respectively) reduces the promoter activity to 18-27%., as expected. Unexpectedly, base changes at -10 element together with changes at -35 element (98T 75C or 98G 76G) increased the promoter activity 6-8 fold compared with that of single mutations at position 98 (98T or 98G) and the activity covers 28-32% of the starting level. Results illustrated in Figure 4.1 also show that base substitution G for T at position 96 reduced the promoter activity drastically. The reason for the reduction of the promoter activity due to 96G mutation is not known. The change G for A at position 100 reduced the promoter activity to ~5%. The reduced level of promoter activity due to 100G is expected since the base change at the fourth position of the -10 element results in a weak -10 element (consensus ^{1}TATAAT^{6}). Unexpectedly, the double base substitutions 96G 100G increased the promoter activity dramatically compared with the single mutations (96G or 100G) and the activity is nearly 40% of the starting level. However, since 98C 64G double mutation restored the promoter activity largely compared with other double mutations, the promoter derivative containing 98C 64G double mutations was considered mainly for further study, aimed to address the reason for the restoration of the promoter activity.

Table 4.1: Mutational analysis of LEE20-275 promoter fragment

Promoter fragments	Promoter sequences from positions 62 to 102	β-galactosidase activity (Miller units ± sd)
	62 70 80 90 102	
LEE20-275 (WT)	5'-TGATTTTTTTGTTGACATTTAATGATAATGTATTTTACACA-3'	5513 ± 261
LEE20-275 98C	5'-TGATTTTTTTGTTGACATTTAATGATAATGTATTTT**C**CACA-3'	233 ± 4 (4.2)
LEE20-275 98C 64G	5'-TG**G**TTTTTTTGTTGACATTTAATGATAATGTATTTT**C**CACA-3'	4375 ± 133 (79.4)
LEE20-275 64G	5'-TG**G**TTTTTTTGTTGACATTTAATGATAATGTATTTT**A**CACA-3'	6276 ± 20 (114)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-275 (WT) fragment and different derivatives. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). Activities expressed as a percentage of activity with the starting LEE20-275 fragment are shown in parentheses. The central part of the table shows the fragment base sequence from position 62 to position 102, with the P1 promoter -10 and -35 hexamer elements shaded grey. Base substitutions in the different fragments are underlined, highlighted in boldface type and coloured red.

4.3 Characterisation of the cryptic *LEE1* P1A promoter

Since it is known that many *E. coli* promoters are completely inactivated by base changes at position 2 of the -10 hexamer (Rosenberg and Court, 1979; Miroslavova and Busby, 2006), the simplest explanation for the effect of the 64G mutation is that it unmasks a cryptic promoter in the LEE20-275 fragment. To investigate this, transcripts formed, after labelled nucleoside triphosphates were added to binary complexes of purified RNAP and the LEE20-275 fragment, were analysed. In this experiment, RNAP runs to a downstream transcription terminator, and the RNA transcripts are labelled by using ^{32}P-labelled UTP and analysed by gel electrophoresis (Figure 4.2). With the starting LEE20-275 fragment, the major transcript is ~102 bases, which corresponds to the *LEE1* P1 transcript that starts at position 107A. As expected, this transcript is completely absent when the DNA fragment carries the 98C substitution. However, with the LEE20-275 fragment carrying both the 98C and 64G substitutions, a new ~112 base transcript is observed. This corresponds to a transcript starting at position 97T, suggesting that the 64G substitution has unmasked a cryptic promoter (Figure 4.2). To confirm this, I used potassium permanganate footprinting to compare regions of DNA duplex unwinding in binary complexes of purified RNAP and the LEE20-275 fragment either without or with the 98C and 64G substitutions. Recall that potassium permanganate modifies T residues in the single stranded 'bubble' produced after local unwinding of promoter DNA around the transcription start at promoters (Savery *et al*., 1996). Results in Figure 4.3 show that, without the substitutions, there is extensive unwinding that starts just downstream of the P1 promoter -10 hexamer, and that this unwinding is suppressed by the 98C substitution. In contrast, with the fragment carrying both the 98C and 64G substitutions, clear duplex opening is seen at positions 86, 88, 89 and 93, just upstream from the transcript start at 97T. This argues that 98C and 64G substitutions cause RNAP to recognise a new promoter (P1A), with a transcript start that is 10 base pairs upstream from the P1 promoter start (Figure 4.2).

In order to identify sequence elements essential for the P1A promoter, error prone PCR was used to generate 3 independent preparations of the LEE20-275 fragment with the 98C and 64G substitutions. They were cloned into pRW224 and the mixture of resulting recombinant plasmids was transformed into *E. coli* strain M182, and transformants were grown on MacConkey lactose indicator plates. As expected, the majority of colonies scored as Lac$^+$,

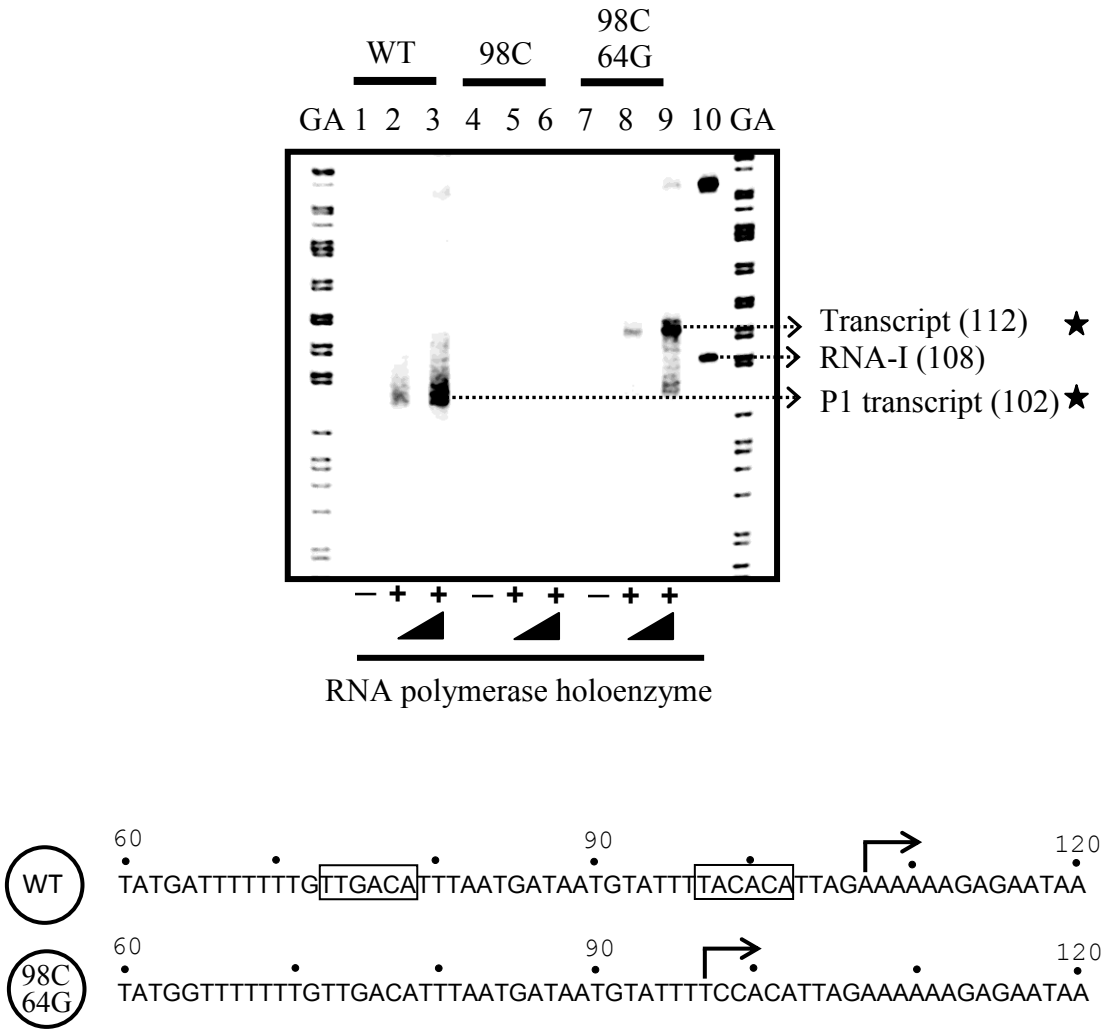


Figure 4.2. *In vitro* run off transcription assay.

The figure shows an autoradiogram of an analysis by gel electrophoresis of ^{32}P-labelled RNA transcripts made by RNA polymerase holoenzyme from PstI-BamHI fragments carrying LEE20-275 sequences (lanes 1-3) and derivatives carrying either 98C (lanes 4-6) or 98C 64G (lanes 7-9) substitutions. The RNA polymerase concentration was: lanes 1, 4 and 7, no enzyme; lanes 2, 5 and 8, 200 nM; lanes 3, 6 and 9, 400 nM. The gel was calibrated with the pSR plasmid-encoded 108 nucleotides RNA-I transcript (lane 10) and Maxam-Gilbert sequence reaction (GA). Proposed transcripts are indicated by asterisks and transcription initiation sites are shown by bent arrows in partial nucleotide sequences of LEE20-275 (WT) and LEE20-275 98C 64G fragments.

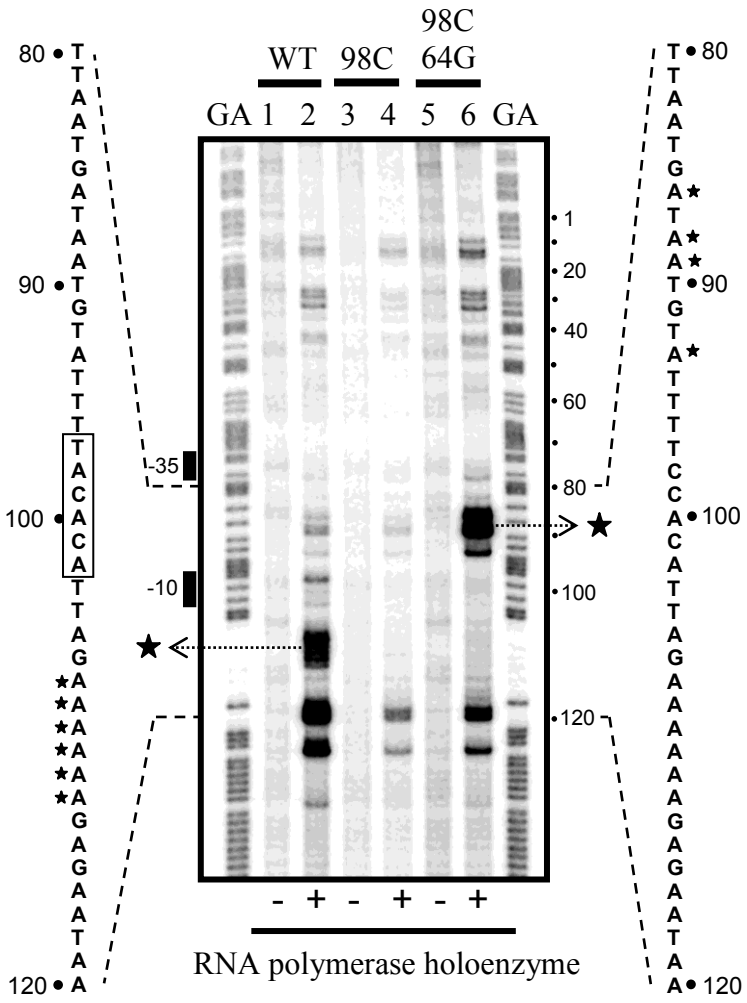


Figure 4.3. *In vitro* potassium permanganate footprinting analysis.

An autoradiogram that identifies the potassium permanganate sensitive sites in complexes of holo RNA polymerase with a DNA fragment that includes the LEE20-275 sequence (WT) or derivatives containing either 98C mutation (98C) or 98C 64G mutations (98C 64G). Lanes 1, 3 and 5 show the results after control incubations without RNA polymerase, while lanes 2, 4 and 6 show the analysis of samples with 50 nM RNA polymerase. The gel was calibrated using Maxam-Gilbert sequence reaction (GA) and relevant positions are indicated. Asterisks identify the residues that display RNA polymerase-dependent reactivity to permanganate. Partial sequences of LEE20-275 and LEE20-275 98C 64G with reactive sites (asterisks) are shown in left and right sites of the gel, respectively.

but, after screening over 15,000 transformants, I identified 4 Lac$^-$ colonies that each carried a supplementary base change. These were an A for C change at position 88 (88C), A for G substitution at position 89 (89G), a T for C substitution at position 90 (90C) and the insertion of a single T between positions 71 and 72 (Figure 4.4).

The effects of the different substitutions were quantified and the data are shown in Table 4.2. The results show that the 88C substitution reduces promoter activity by over 20 fold, whilst the other three substitutions have smaller effects. These data, together with the *in vitro* results in Figures 4.2 and 4.3, identify the hexamer, TAATGT, from position 87 to position 92 as the P1A -10 hexamer, upstream of the transcript start at position 97. Hence, the 88C mutation at position 2 of the hexamer has the biggest effects on P1A activity and is within the region of duplex unwinding following RNA. I tentatively assign TGGTTT from position 62 to position 67 as the P1A -35 hexamer (Figure 4.4). Although this corresponds to the consensus TTGACA at only 2 out of 6 positions, the concordant G at position 64 is the G that was created by the 64G substitution that unmasked the P1A promoter. Thus, the high promoter activity of the LEE20-275 fragment carrying the 98C and 64G substitutions is likely due to improvement of the -35 element from TGATTT to TGGTTT. This data also suggest that the spacer length between -10 and -35 elements is 19 base pairs. Note that one of the Lac$^-$ mutants selected after random mutagenesis carried LEE20-275 98C 64G fragment with an extra T between positions 71 and 72 (Table 4.2). This extra base changes the promoter spacer length to 20 base pairs and that reduced the promoter activity sharply. In order to determine the effect of shorter spacer length on the promoter activity, derivative LEE20-275 98C 64G with Δ71T deletion was constructed by site directed mutagenesis and *lac* expression from the derivative was measured. Data presented in the Table 4.2 shows that deletion of the base at position 71, that brings spacer length 19 to 18, closure to optimal spacer (17 base pairs) caused a significant increase in the *lac* expression.

4.4 Activity of *LEE1* P1A without the 64G substitution

The chance isolation of the 64G substitution in the LEE20-275 fragment, in combination with the 98C substitution, unmasked the P1A promoter. To address the issue of whether P1A is functional without the 64G substitution, I have exploited the 88C substitution that I had found to inactivate P1A activity (Table 4.2). Results in Table 4.3 show that the 88C substitution has

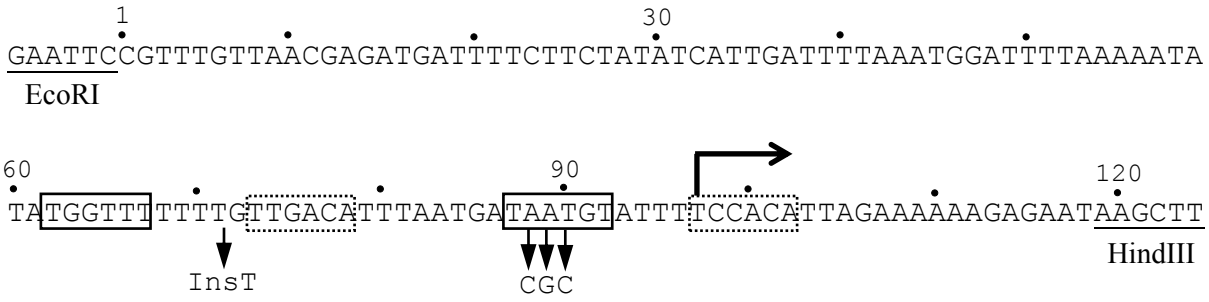


Figure 4.4. Mutational analysis of the LEE20-275 98C 64G fragment.

The figure shows the nucleotide sequence of the upper strand of the LEE20-275 98C 64G fragment. The base sequences are numbered 1-120 as Figure 4.1. The positions of randomly generated single mutations that reduced expression from this fragment are illustrated by showing the substituted base. The P1A promoter -35 and -10 hexamer elements, deduced from this study, are solid boxed and the transcript start, at position 97T, determined from the results in Figure 4.2 is indicated by a bent arrow. Dotted boxes represent the location of mutated P1 promoter elements.

Table 4.2: Identification of the functional elements of the P1A promoter

Promoter fragments	Promoter sequences from positions 62 to 102	β-galactosidase activity (Miller units ± sd)
	62 70 80 90 102	
LEE20-275 98C 64G	5'-TGGTTTTTTTGTTGACATTTAATGATAATGTATTTTCCACA-3'	4375 ± 133
LEE20-275 98C 64G 88C	5'-TGGTTTTTTTGTTGACATTTAATGATCATGTATTTTCCACA-3'	190 ± 45 (4.3)
LEE20-275 98C 64G 89G	5'-TGGTTTTTTTGTTGACATTTAATGATAGTGTATTTTCCACA-3'	517 ± 28 (11.8)
LEE20-275 98C 64G 90C	5'-TGGTTTTTTTGTTGACATTTAATGATAACGTATTTTCCACA-3'	263 ± 15 (6.0)
LEE20-275 98C 64G InsT (71-72)	5'-TGGTTTTTTTTGTTGACATTTAATGATAACGTATTTTCCACA-3'	939 ± 8 (21.5)
LEE20-275 98C 64G Δ71T*	5'-TGGTTTTTT–GTTGACATTTAATGATAACGTATTTTCCACA-3'	7109 ±107 (162.5)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-275 98C 64G fragment and different mutations. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). Activities expressed as a percentage of activity with the starting LEE20-275 98C 64G fragment are shown in parentheses. The central part of the table shows the fragment base sequence from position 62 to position 102, with the location of P1 promoter -10 and -35 hexamer elements shaded grey and P1A promoter -10 and -35 elements are underlined. Base substitutions and insertion in the different fragments are highlighted in boldface type and coloured red whilst Δ71T is indicated by a dash. Mutation made by site directed mutagenesis is indicated by asterisk, whilst the other mutations came from the random PCR mutagenesis experiment illustrated in Figure 4.4.

very little effect on the measured promoter activity of the starting LEE20-275 fragment. In contrast, the activity of the fragment carrying the 98C substitution is reduced by over 95% by the 88C substitution. Recall that the 98C substitution completely prevents *LEE1* P1 promoter activity (Figure 4.2, 4.3). Thus, I deduce that reduced expression from the LEE20-275 fragment carrying the 98C substitution is due to P1A activity, and, hence, is suppressed by the 88C change. Similarly, expression from the LEE20-275 fragment carrying the 76G substitution, in the P1 -35 element, is also greatly reduced by the 88C substitution. From these results, I deduce that P1A activity increases as P1 becomes less active, due to substitutions in either the -10 or -35 element.

Table 4.3 also shows the results from the LEE20-275 derivative carrying a substitution in the P1 -35 hexamer (76G) combined with a substitution in the P1 -10 hexamer (98G). The introduction of 76G substantially increased the activity of the LEE20-275 fragment carrying the 98G substitution, which was also described in Figure 4.1. This increased expression is completely suppressed by the 88C substitution, and, hence, must be due to the P1A promoter. From this, I conclude that, even though the P1 promoter is completely inactivated by the 98G substitution, the activity of P1A remains low because RNAP still makes abortive contacts with the P1 consensus TTGACA -35 hexamer element. However, P1A activity increases as these interactions are weakened, due to substitutions such as 76G. These observations also suggest that the activity of the promoter fragments containing either 75C or 75C 98T mutations described in the Figure 4.1 could possibly come from the P1A promoter.

In a related experiment, I strengthened P1A by deletion of one of the 7 consecutive T residues immediately upstream of the P1 promoter -35 hexamer element (Δ71T) which alters the length of the spacer between the P1A promoter -10 and -35 hexamer closer to the optimal 17 base pairs (Rosenberg and Court, 1979). Results in Table 4.3 show that the promoter activity of the resulting LEE20-275 fragment carrying the Δ71T change is reduced ~3 fold by the 88C substitution.

Results described in Table 4.3 also show that Δ71 caused ~45% reduction of the overall promoter activity. These results suggest that shortening the P1A promoter spacer length from19 to 18 base pairs strengthens the P1A (Table 4.2) and this might increase the

Table 4.3: Effect of mutations in unmasking P1A promoter

Promoter fragments	Promoter sequences from positions 62 to 102	β-galactosidase activity (Miller units ± sd)
	62 70 80 90 102	
LEE20-275 (WT)	5'-TGATTTTTTTGTTGACATTTAATGATAATGTATTTTACACA-3'	5513 ± 261
LEE20-275 88C	5'-TGATTTTTTTGTTGACATTTAATGATCATGTATTTTACACA-3'	4596 ± 30 (83.4)
LEE20-275 98C	5'-TGATTTTTTTGTTGACATTTAATGATAATGTATTTTCCACA-3'	233 ± 4
LEE20-275 98C 88C	5'-TGATTTTTTTGTTGACATTTAATGATCATGTATTTTCCACA-3'	10 ± 1 (4.3)
LEE20-275 76G	5'-TGATTTTTTTGTTGGCATTTAATGATAATGTATTTTACACA-3'	1568 ± 24
LEE20-275 76G 88C	5'-TGATTTTTTTGTTGGCATTTAATGATCATGTATTTTACACA-3'	220 ± 13 (14.0)
LEE20-275 98G	5'-TGATTTTTTTGTTGACATTTAATGATAATGTATTTTGCACA-3'	202 ± 15
LEE20-275 98G 76G	5'-TGATTTTTTTGTTGGCATTTAATGATAATGTATTTTGCACA-3'	1289 ± 3
LEE20-275 98G 76G 88C	5'-TGATTTTTTTGTTGGCATTTAATGATCATGTATTTTGCACA-3'	11 ± 1 (1.0)
LEE20-275 Δ71	5'-TGATTTTTT—GTTGACATTTAATGATAATGTATTTTACACA-3'	3107 ± 115
LEE20-275 Δ71 88C	5'-TGATTTTTT—GTTGACATTTAATGATCATGTATTTTACACA-3'	1103 ± 24 (35.5)

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 containing the LEE20-275 fragment and different mutations. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). Activities with the fragments containing P1A -10 mutants expressed as a percentage of activities with the starting fragments are shown in parentheses. The central part of the table shows the fragment base sequence from position 62 to position 102, with the P1 promoter -10 and -35 hexamer elements shaded grey and P1A promoter -10 and -35 elements are underlined. Base substitutions in the different fragments are highlighted in boldface type and red coloured whereas Δ71 deletions are shown by dashes.

competition between two promoters for RNAP, and result in low level of activity. However, an insertion of T between 71-72 positions in the LEE20-275 fragment increased the *lacZ* expression from 5513 Miller units to 7898 Miller units. This suggests that increasing the P1A spacer length from 19 to 20 base pairs weakens the P1A promoter as seen in Table 4.2 and this might reduce the competition between two promoters for RNAP and result in high level of activity.

4.5 Discussion

Many bacterial gene regulatory regions, especially those that are AT-rich, contain multiple promoters, and it is generally assumed that they are a by-product of evolution. Here I have identified a cryptic promoter, P1A, which overlaps the *LEE1* P1 promoter of EHEC serotype O157:H7. *In vivo* reporter gene assays and subsequent *in vitro* footprinting and transcription analyses revealed that base substitutions in the key elements of the P1 promoter unmask P1A promoter that specifies transcription initiation from a site that is located 10 base pairs upstream of the P1 transcription start site (Figure 4.2).

Mutational analysis revealed that DNA sequence upstream of the P1A promoter transcription initiation site contains a -10 region showing 3 of 6 base pairs homology with the consensus -10 element sequence (TAATGT vs consensus TATAAT) (McClure, 1985). There is a very weak -35 element having only 1 of 6 base pairs homology with the consensus sequence for the other promoters (TGATTT vs TTGACA) (McClure, 1985). These two key elements are separated by a non-optimal 19 base pairs spacer sequence. Moreover, there is an extended -10 element or a TG motif located one base pair upstream of the -10 element and also short runs of T residues in the spacer. These results are in agreement with previous studies supporting the fact that extended -10 element or TG motif-dependent promoters have longer spacer sequence, contain short runs of T residues in the spacer region and render the -35 hexamer sequences dispensable compared with those of non-TG promoters (Burr *et al*., 2000; Mitchell *et al*., 2003).

The P1A promoter was discovered following the chance isolation of the 64G substitution that improves its -35 element. Here, however, I have presented evidence that the P1A promoter can function even without the 64G substitution. The P1A promoter overlaps with P1 and the

two promoters compete and are mutually exclusive. In the present condition, with the starting LEE20-275 fragment, the P1 promoter clearly wins the competition. However, as P1 is weakened, or P1A is strengthened, the scale tips towards P1A (Figure 4.5). A common phenomenon is found in the O_R region of bacteriophage λ. Inactivation of P_R promoter in that region uncovered P_α cryptic promoter (Woody *et al*., 1993). Likewise, mutation(s) that weakens the principal *lac* P1 promoter favours unmasking a weak upstream P2 promoter in *lac* regulatory region (Malan and McClure, 1984).

Most important observation of this work is that mutations in P1 promoter -10 and -35 elements unmask the P1A promoter in different extents. Interestingly, even when P1 is completely inactivated by a substitution at position 2 of its -10 hexamer, measured P1A activity remains low due to the P1 -35 element, which corresponds exactly to the consensus. Thus the P1 promoter, even when inactive, can retain the ability to sequester RNAP, thereby blocking access to the P1A promoter. Specific contact of RNAP with inactivated *gal* P2 promoter has been reported (Johnston *et al*., 1987). Mutation in extended -10 element of *gal* P2 promoter reduced the promoter activity but RNAP could still make a specific contact with the inactivated promoter. In contrast with a mutation in the *gal* P2 -10 hexamer sequence, RNAP could not make this contact.

The presence of the P1A promoter may result from high evolutionary rate in the *LEE1* regulatory region that may alter the functional promoter's activity significantly or to some an extent as mutation(s) occurs nearby the functional promoter. Alignment of the base sequence of the *LEE1* regulatory region of EHEC O157:H7 Sakai and EPEC E2348/69 finds a single base deletion (ΔT) located one base upstream of the EHEC *LEE1* P1 promoter -35 element (Porter *et al*., 2005). Insertion of a T base in that position in EHEC *LEE1* sequence reduces the P1A promoter activity sharply (Table 4.2). This suggests that the influence of P1A promoter in the expression from EHEC *LEE1* regulatory region is high compared with that of EPEC. Moreover, addition of an extra T that makes the EHEC *LEE1* P1 promoter sequences more likely to EPEC *LEE1* promoter significantly increases the promoter activity. This suggests that EHEC *LEE1* P1A promoter might compete with the main P1 promoter for RNAP binding and thus reduces the overall promoter activity whereas competition between the two promoters in EPEC *LEE1* is less since P1A is weak, which might favor increased

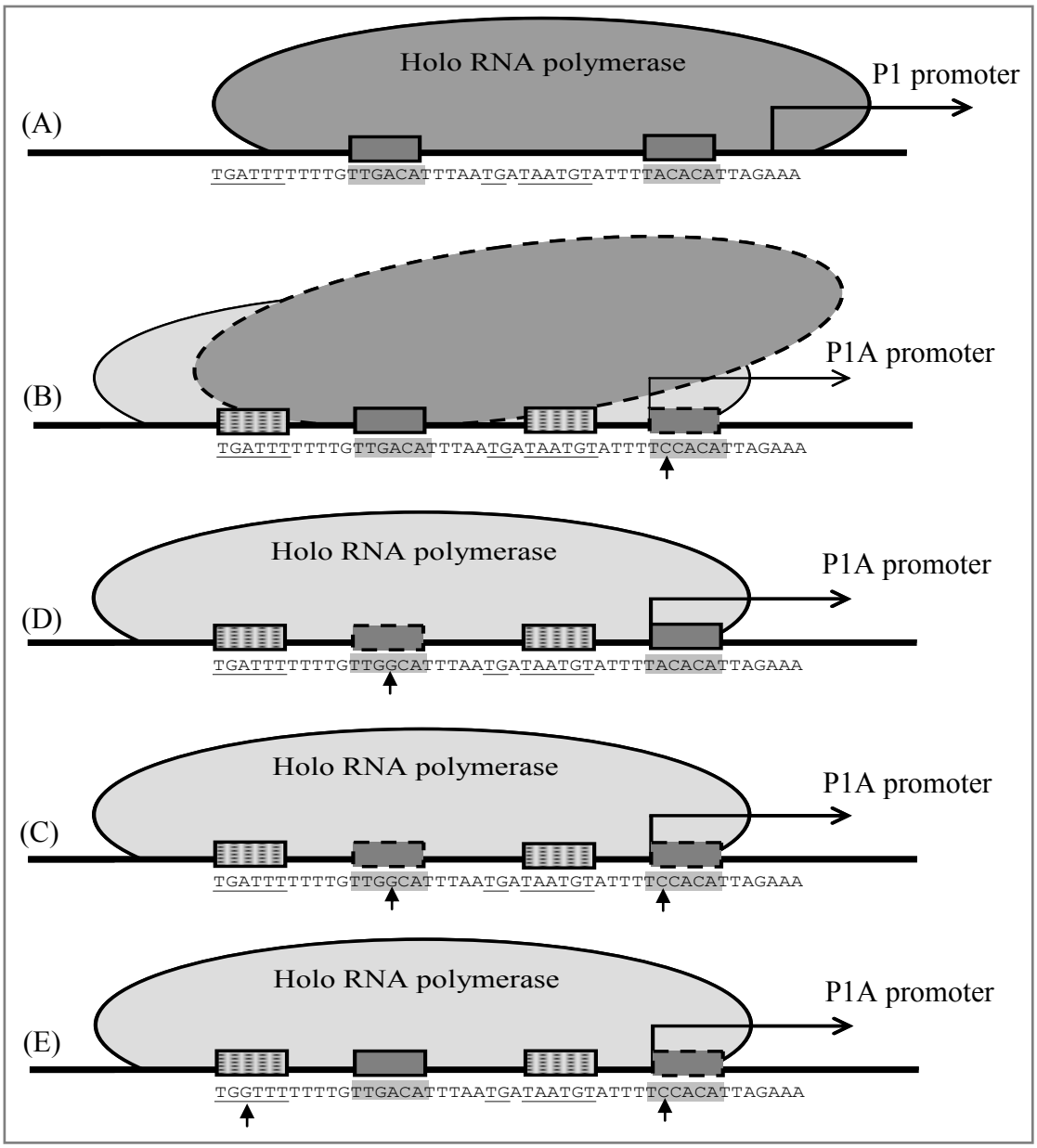


Figure 4.5. Schematic representation of the unmasking P1A promoter.

(A) RNA polymerase holoenzyme recognizes P1 promoter -35 and -10 elements (shaded grey) and initiates transcription. (B) Mutation in the second position of the -10 element causes poor -10 element that results in sharp reduction of the transcription initiation from P1 promoter. RNA polymerase holoenzyme moves towards P1A promoter -10 and -35 elements (underlined). Consensus -35 element of the inactivated P1 promoter sequestered RNA polymerase and blocked recruitment by P1A that results in very weak transcription initiation from P1A promoter. (C) Mutation in the P1 -35 hexamer causes sharp reduction of transcription initiation from the P1 promoter that facilitates initiation from the P1A promoter. (D) Mutations both in P1 promoter -35 and -10 elements cause complete destruction of the P1 promoter, which favors efficient transcription initiation from the P1A promoter. (E) Improvement of P1A -35 hexamer sequence together with weakening P1 -10 element cause complete shifting of the RNA polymerase from P1 to the P1A promoter, ensuring efficient transcription initiation from P1A. Arrows below the nucleotide sequences indicate mutations and bent arrows show the transcription start sites of the promoters.

expression from EPEC *LEE1* promoter.

Among the double mutations, described in Figure 4.1, 98C 64G or 98T 75C or 98G 76G double mutation unmasks the P1A. Though other two double mutations e.g., 98C 35C or 96G 100G were not investigated, I assume that the increased *lac* expression from the promoter derivatives containing either 98C 35C or 96G 100G double mutation could be the consequence of the presence of other promoter-like signals in the *LEE1* promoter regulatory region.

Potentially, clusters of potential promoter elements in regulatory regions can have many different consequences. For example, they may quite simply give rise to multiple transcription starts and recent high-resolution analysis of the *E. coli* transcriptome has identified many cases of this (Mendoza-Vargas *et al*., 2009). This can lead to complex patterns of regulation. Alternatively, these elements could be the remains of now defunct regulatory systems, or be awaiting future adoption. My observation that the P1 -35 element retains function even when P1 is inactive, together with data with the *E. coli gal* operon regulatory region (Johnston *et al*., 1987) underscores that residual promoter elements at non-functional promoters can affect the distribution of RNAP. At present, the role of the *LEE1* P1A is unclear. However, I note that consensus -35 hexamer elements are very rare at *E. coli* promoters (Mitchell *et al.*, 2003). I speculate that the rationale for retaining TTGACA as the *LEE1* P1 -35 element is to focus RNAP at P1 and to reduce the use of neighbouring promoters such as P1A.

Chapter 5

A small translated open reading frame in the leader sequence of *LEE1* regulatory region: role in the expression of downstream genes

5.1 Introduction

The leader sequence of a messenger RNA (mRNA) can modulate the translation process of a gene. The leader sequence is the sequence at the 5' end of a mRNA molecule that is not translated into protein. It starts at the +1 position (transcription start site) of the gene and ends just before the start codon (AUG) of the coding region. This region often contains a SD sequence, which is a 4-6 base sequence that is complementary to a sequence at the 3' end of one of the ribosomal RNAs (Shine and Dalgarno, 1975; Steitz and Jakes, 1975).

Most of the 5' end sequences of *E. coli* messenger RNAs have 40-80 untranslated bases, but some are much longer (Shultzaberger *et al*., 2007). Genes that contain unusual long leader sequences are often subjected to complex regulation mechanisms. Long leader region may contain regulatory sequence, including binding sites for proteins that can affect the stability of the mRNA. Upstream ORF found in a leader region can couple with the downstream gene and can regulate its expression. This is the case for the *tap* (translation activator peptide) and *repA* (encoding replicase) genes in the regulation of plasmid R1 replication. The leader peptide *tap* is located in the region between the *copA* and *repA* genes. The CopA regulatory RNA binds to the RBS of *tap* and this results in translational inhibition, transmitted to the *repA* gene by translational coupling (Blomberg *et al*., 1992).

In some cases, 5' untranslated long leader regions encode short peptides, which influence expression of neighbouring downstream genes (Vitreschak *et al*., 2004). Some of the bacterial operons involved in biosynthesis of amino acids including threonine, leucine, histidine, tryptophan and phenylalanine are subject to this kind of post-transcriptional regulation (Gardner, 1979; Gemmill *et al*., 1979; Johnston *et. al*., 1980; Yanofsky, 1981; Gavini and Pulakat, 1991; Gurvich *et al*., 2010). In many cases, riboswitches in 5' untranslated regions of mRNA molecules can serve as receptors for specific metabolites. These metabolites bind to the targets and affect the downstream gene expression positively or negatively (Mironov *et al*., 2002; Nahvi *et al*., 2002; Winkler *et al*., 2002a,b; Sudarsan *et al*., 2003; Mandal and Breaker., 2004; Nudler and Mironov, 2004; Lemay and Lafontaine, 2007). Additionally, recently Artsimovitch (2010) reported that *cis*-acting RNA elements, termed as EAR, located between the *epsB* and *epsC* genes, increase the expression of exopolysaccharide genes in *Bacillus*

subtilis at the level of transcription termination. The EAR shares the complexity of riboswitches but uniquely function as a processive, long-range antiterminator.

During work with the ~170 nucleotides-long *LEE1* leader sequence, which contains mostly adenines and uracils, a mini-gene was identified. The translation of this mini-gene is necessary for optimal expression of the downstream *ler* gene.

5.2 Identification of the functional Ler translation start site

The 5' untranslated leader sequence of a gene is defined by a section of an mRNA molecule that starts at the +1 (where transcription begins) and ends one nucleotide before the translation start codon. Thus, the start point of this work was to determine the translation start site of the *ler* gene. According to coliBASE (ECs4588) (Chaudhuri *et al*., 2004) and other databases, *ler* translation starts from ATG located one base pair upstream of a potential SD sequence (Figure 2.22). However, there is a potential translation start codon (ATG), located 10 base pairs downstream of the SD site. Generally, translation initiation codons of genes are located 5-9 nucleotides downstream of the SD sequence (Gualerzi and Pon, 1990). Therefore, the database information about *ler* translation start site is ambiguous. In order to experimentally determine the *ler* translation start site, the LEE151 fragment containing the assigned translation start codon in the database (550 base pairs), or the LEE150 fragment carrying the second predicted translation initiation codon (571 base pairs) (Figure 2.22) were cloned into pRW225 as translation fusions to *lacZ*. The resulting plasmids were transformed into M182 cells and measured *lacZ* expression was taken as a measurement of translation activity. Results illustrated in Figure 5.1 shows no expression from the LEE151 fragment cloned as translation fusion to *lacZ*, whilst the level of *lac* expression from the LEE150 is 3597 Miller units. These results suggest that the *ler* translation starts at a site (ATG), located 10 base pairs downstream of the ribosomal binding site. However, this would mean that the *ler* regulatory region contains a ~170 nucleotides long leader sequence, which corresponds to a LEE1 mRNA section located between the site of *LEE1* P1 promoter transcription start and the nucleotide before the *ler* translation start site.

5.3 Downstream elements affect expression from the *LEE1* P1 promoter

In an attempt to understand the role of the long leader sequence in the EHEC *LEE1* regulatory

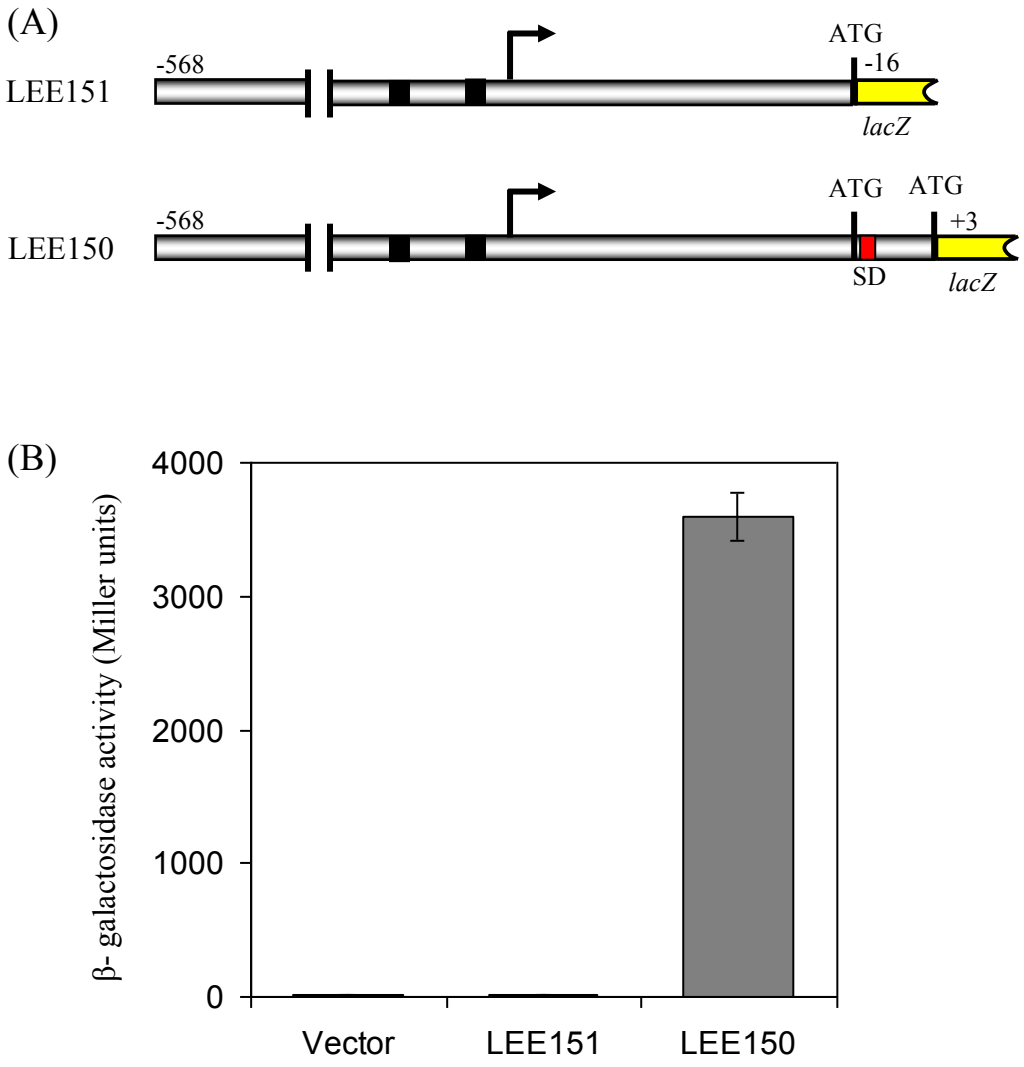


Figure 5.1. Translational activities from LEE151 and LEE150 fragments.

(A) Schematic representation of the LEE151 and LEE150 fragments showing the differences between them (nucleotide sequence details in Figure 2.22). LEE151 fragment carries the predicted translation start site of the *ler* assigned in coliBASE whereas LEE150 fragments carries second predicted translation start site assigned by Yerushalmi *et al*., (2008) that contains a ribosomal binding sequence (SD) in front of it. Bent arrows indicate the *LEE1* P1 promoter. Both fragments were cloned as translation fusions to *lacZ* in pRW225.
(B) Bar graph shows the measured β-galactosidase activities in *E. coli* K-12 strain M182 containing plasmid pRW225 with LEE151 or LEE150 fragments. Vector represents the plasmid pRW225 that contains promoterless 122 base pairs "stuffer" sequence. Measurements were made after growth of the cells in LB medium at 37°C to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

region, a nested deleted derivative, LEE30-275 was constructed. This contains extra 44 base pairs downstream *LEE1* leader sequence compared with the sequence of the LEE20-275 derivative (Figure 2.10). After cloning the derivative into pRW224, *lacZ* expression was measured in M182 cells. Figure 5.2 shows the level of *lacZ* expression from LEE30-275 fragment and compares the activity with those from LEE10-275 and LEE20-275 promoter fragments. As seen in Figure 3.1, the highest level of *lacZ* expression was quantified from LEE20-275 fragment whilst the level of *lacZ* expression from LEE10-275 was found to be ~2.5-fold lower than that from LEE20-275. Decreased expression seen from the LEE10-275 fragment compared with that from LEE20-275 fragment is probably due to the longer untranslated leader sequence upstream of the reporter *lacZ* gene. Surprisingly, inclusion of 44 base pairs downstream leader sequence caused a sharp reduction in expression from LEE30-275 promoter fragment compared with that from LEE20-275 promoter fragment. This results suggest that *cis*- and/or *trans*- acting elements in untranslated long leader sequence upstream of the reporter gene may play a role in the differential level of expressions from these three derivatives.

The complete sequence of the LEE30-275 fragment that showed decreased *lac* expression compared with that of LEE20-275 is shown in Figure 5.3, where the sequence has been numbered 1-162 following the numbering system adopted for the LEE20-275 fragment in Figure 3.3. Sequence analysis suggests that the DNA sequence of the LEE30-275 fragment has a potential SD sequence located downstream, between positions 129-137, and a potential translation start (ATG) and stop sites (TAG) located at the positions 144 and 150, respectively (Figure 5.4). Moreover, the program "RBS calculator version 1.0" (Salis *et al.*, 2009) predicted two adjacent translation start codons located at positions 141 (TTG) and 144 (ATG), respectively and the predicted translation initiation rate was 5-fold higher at ATG compared with TTG. Therefore, bioinformatics data suggest that the LEE30-275 promoter fragment might contain a small translated ORF, which may interfere with the translation of downstream reporter gene, hence cause a low level of expression from the fragment.

5.4 Analysis of up-mutants of the LEE30-275 fragment

In order to identify the determinant(s) responsible for the low level of *lac* expression from the LEE30-275, an experiment with error-prone PCR was conducted to generate 6 independent

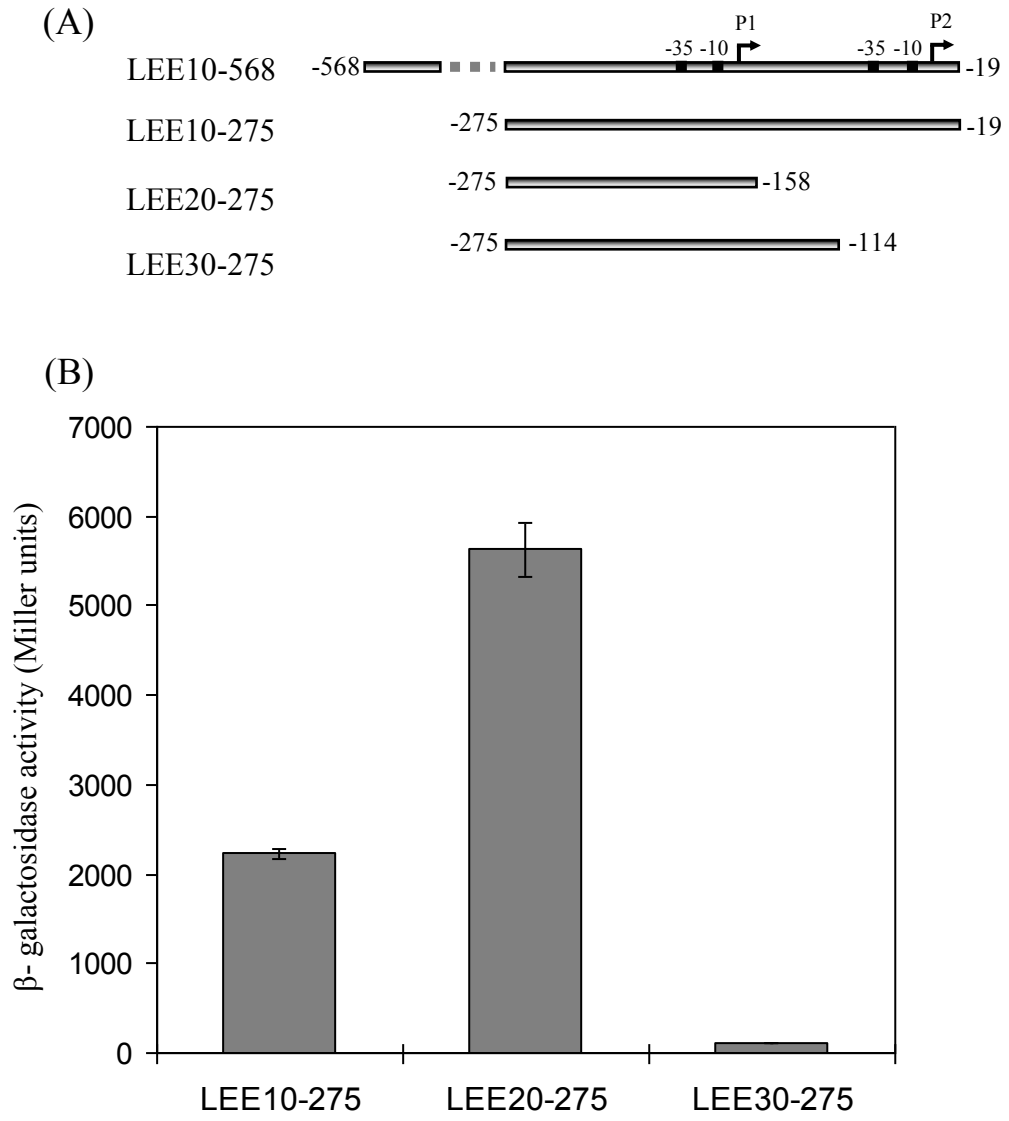


Figure 5.2. Expression from different *LEE1* promoter fragments.

(A) Schematic representation of three nested deleted derivatives of LEE10-568 promoter fragments. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene. The P1 and P2 promoters are indicated by bent arrows and the shaded black boxes represent the cognate -10 and -35 hexamer elements.
(B) The bar graph shows the measured β-galactosidase activities in *E. coli* K-12 strain M182 carrying the plasmid pRW224 with LEE10-275 and two derivatives LEE20-275 and LEE30 -275 as transcriptional fusion to *lacZ*. Measurements were made after growth of the cells in LB medium at 37°C to an optical density at 650 nm of ~0.5. Values are the average of at least three independent assays, and standard deviations are shown with bars.

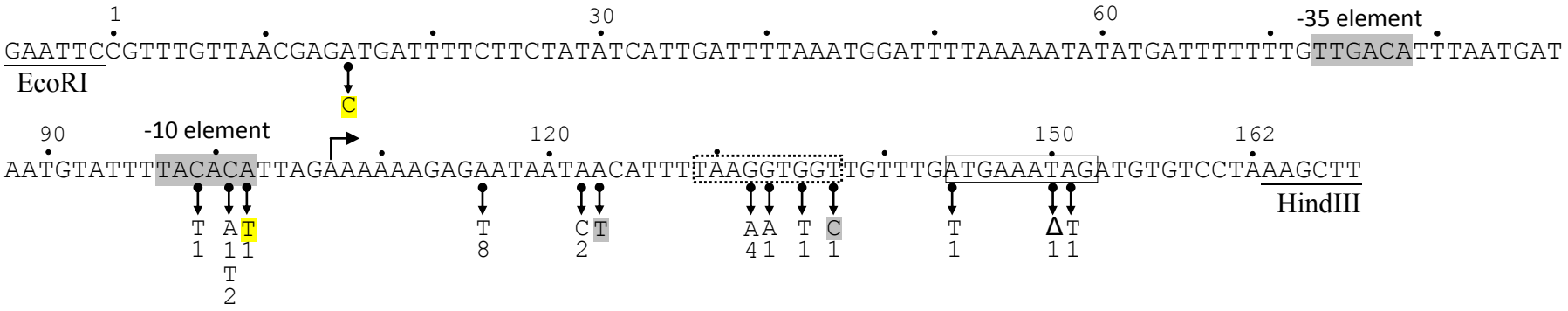


Figure 5.3. Mutational analysis of the LEE30-275 fragment.

The figure shows the nucleotide sequence of the upper strand of LEE30-275 fragment with restriction sites EcoRI and HindIII. The base sequences are numbered 1-162, starting with the first cloned *LEE1* regulatory region base following the numbering system in Figure 3.3. The positions of randomly generated single and double mutations that increased expression from this fragment are illustrated by showing the substituted base(s) and, in each case, the adjacent number records the number of times that the particular substitution was obtained. Single mutations are indicated by black letters whilst double base substitutions are indicated by the yellow or grey colours. The locations of the P1 promoter -35 and -10 hexamer elements are shaded grey and the transcript start is indicated by bent arrow. Predicted small ORF and associated SD sequence are indicated by solid and dotted boxes, respectively.

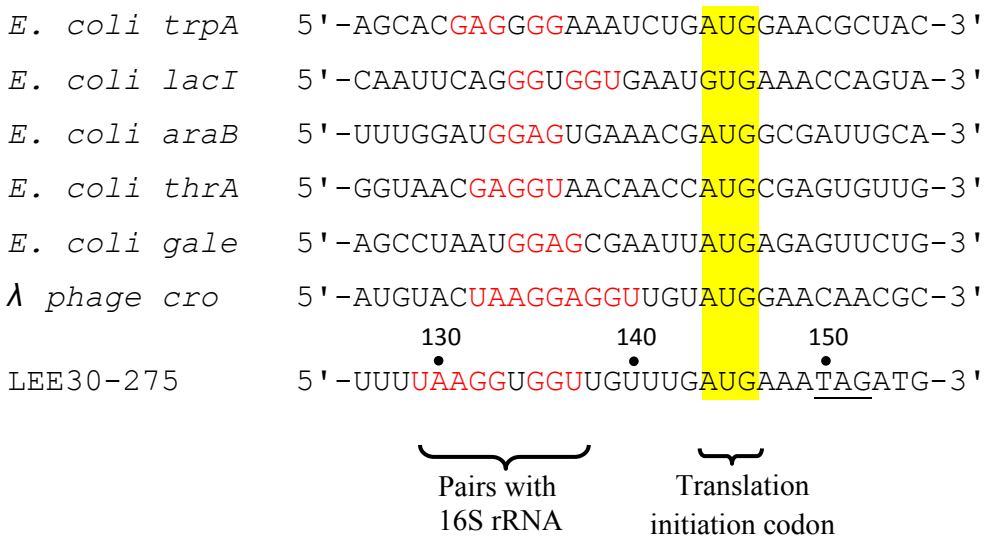


Figure 5.4. Sequence similarity between the predicted translational initiation region in LEE30-275 and some bacterial and viral mRNAs.

The figure shows the alignment of translational initiation region of some bacteria, viral and LEE30-275 mRNA molecules. Sequence of LEE30-275 corresponds to the partial sequence of LEE30-275 promoter fragment (positions 126 to 155 in Figure 5.3). Potential SD sequence marked by red colour, and translational initiation codons are shaded yellow. Translation stop codon in LEE30-275 is underlined.

preparations of the LEE30-275 fragment carrying random mutations. The fragments were then cloned into the *lac* reporter plasmid pRW224, the resulting recombinant plasmids were transformed into *E. coli* strain M182, and transformants were grown on MacConkey lactose indicator plates. As expected, the majority of colonies scored as Lac^- (white) colonies, but after screening over 10000 transformants, 36 Lac^+ (pink) colonies were identified. Sequence analysis showed that 12 of these carried single mutations whilst 5 carried double mutations. Most of the double mutations were not informative and thus not included in the further study. Four of the single mutations were found more than once (up to 8 times for LEE30-275 116T).

The location of mutations that increased *lac* expression from the LEE30-275 promoter fragment is illustrated in Figure 5.3. Of the 12 single mutations, 3 fall in the P1 promoter -10 element. These were a T for C substitution at position 99 (99T) and an A or a T for C at position 101 (101A or 101T). Also included a double base substitution that was a T for A at position 102 together with a change at position 15 (15C 102T). The effects of the base substitutions were quantified and data are recorded in the Table 5.1. Mutations at the -10 element increased the promoter activity 3-8 fold. The high promoter activity of theLEE30-275 fragment carrying the 99T, 101A, 101T or 15C 102T substitutions is due to the improvement of the P1 promoter -10 element from TACACA to a close proximity of the consensus TATAAT for *E. coli* promoters (McClure, 1985).

Other 3 point mutations were found at the predicted ribosomal binding site. Theses are an A or a T for G at position 132, an A for G at position 133 and an A for G at position 135. They also included a double mutation that was a C for A at position 137 together with a C for A at position 123. Presumably, these reduce SD activity to promote translation.

Three other single mutations fall in the coding region of predicted small ORF. These are a T for A at position 144 that changed the predicted translation start point from ATG to TTG, a base deletion at position 150 (Δ150T) or a T for A at position 151. These latter two changes disrupted the predicted translation stop site of the small ORF. Results illustrated in Table 5.1 show that base changes in the predicted SD sequence or in small ORF increased the *lac* expression 3-10 fold compared with that of starting LEE30-275 fragment. The reason for this remains unclear. My speculation is that translation of the mini-ORF might interfere with the

Table 5.1: Mutational analysis of LEE30-275 promoter fragment

Promoter fragments	β-galactosidase activity (Miller units ± sd)	Fold increase
Starting fragment		
LEE30-275	124 ± 9	-
Mutations at P1 promoter -10 element		
LEE30-275 99T	554 ± 18	4.5
LEE30-275 101A	565 ± 18	4.6
LEE30-275 101T	412 ± 9	3.3
LEE30-275 102T 15C	928 ± 31	7.5
Mutations at predicted SD sequence		
LEE30-275 132A	551 ± 28	4.5
LEE30-275 132T	498 ± 130	4.0
LEE30-275 133A	460 ± 9	3.7
LEE30-275 135T	401 ± 33	3.2
LEE30-275 123T 137C	751 ± 33	6.0
Mutations at predicted mini-ORF		
LEE30-275 144T	288 ± 11	2.3
LEE30-275 Δ150	357 ± 13	2.9
LEE30-275 151T	1270 ± 109	10.2
Mutations at other positions		
LEE30-275 116T	765 ± 53	6.2
LEE30-275 122C	948 ± 126	7.6
LEE20-275 fragment (wild type)	5247 ± 242	42

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 with the LEE30-275 fragment and different mutant derivatives. Measurements were made after growing the cells in LB medium at 37°C to an optical density of ~0.5 at 650 nm and are listed in the central column. Activities were measured in triplicate, giving a mean and standard deviation (sd). Activities expressed as fold increase compared to the starting LEE30-275 fragment are given in the third column.

translation of the reporter *lac* gene and thus reduced the *lac* expression from the LEE30-275 promoter. Base change either in SD or in mini-ORF might reduce the interference and thus increase *lac* expression to some an extent.

Increased activity from the LEE30-275 containing a change A to C at position 122 might be due to the similar reasons. A single base change A to T at position 116 increased the *lac* expression 6 fold. The increased transcriptional activity might be due to the creation of a new strong promoter because the base change makes a consensus -10 element around the position. However, though a number of up-mutations generated in the LEE30-275 fragment through error-prone PCR, none of the mutations brought the level of *lac* expression from the LEE30-275 fragment closure to that from LEE20-275 promoter fragment (Table 5.1). This might be due to the residual effect of strong ribosomal binding site of mini-ORF on the translation of reporter *lac*.

5.5 RNA secondary structure predictions

To investigate whether secondary structure in the LEE message could affect the transcriptional activity of the LEE30-275 promoter, the RNA structure for LEE30-275 fragment was predicted using the mfold program (version 3.2) (Zuker, 2003). To see whether all the mutations generated during random mutagenesis of the LEE30-275 could fit with the predicted secondary structure, the base substitutions were plotted on the structure at respective positions (Figure 5.5). Most of the up-mutations such as 132A/T, 144T, Δ150T or 151T fall outside of the hairpin like structure. Only two changes such as 116T and 133A could affect the hairpin like structure and thus the increased LEE30-275 promoter activity due to these two base substitutions could be explained by the secondary structure. But increased transcriptional activity due to the base substitution at positions 122 and 135 or other up-mutations that fall outside of the hairpin like structure cannot be explained by the secondary structure of RNA. From this analysis, I conclude that the low level of *lac* expression from the LEE30-275 fragment is difficult to explain by the secondary RNA structure.

5.6 DNA sampling

To identify possible negative regulatory proteins that might play a role in lowering *lac* expression from the LEE30-275 promoter fragment, the DNA sampling method (Butala *et. al*.,

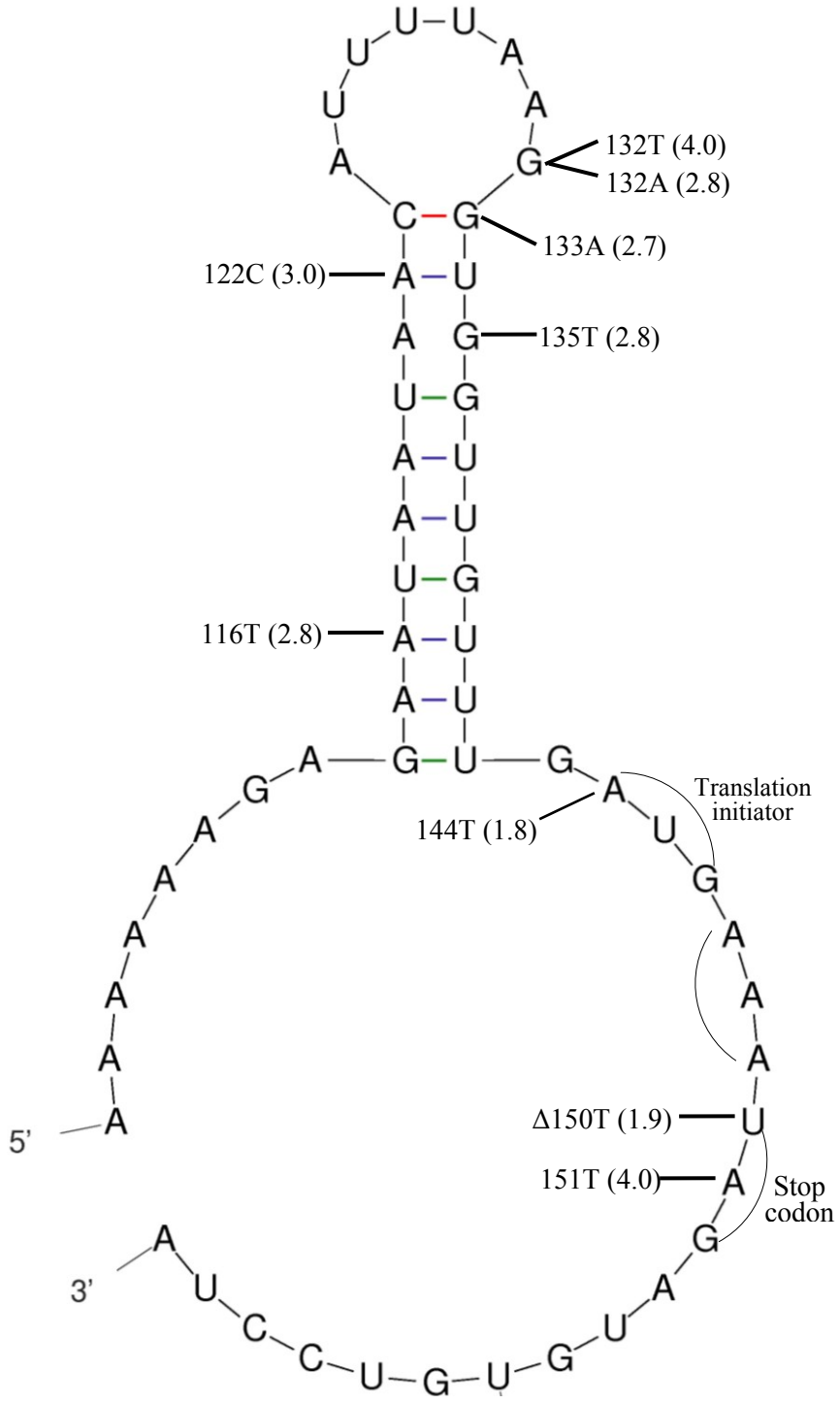


Figure 5.5. Secondary RNA structure of mRNA molecule of LEE30-275 fragment.

The figure shows the predicted secondary RNA structure of LEE30-275 fragment produced by using a program mfold (version 3.2) (Zuker, 2003). Up-mutations generated during random mutagenesis of the fragment, were plotted in the respective sites. Fold-increases in *lacZ* expressions compared to the starting LEE30-275 fragment were shown in the parentheses.

2009) was applied. The method allows rapid isolation of specific segments of DNA, together with bound proteins, from *E. coli* K-12. The DNA fragment to be sampled is produced as a discrete fragment within cells by the yeast I-SceI meganuclease, and is purified using FLAG-tagged LacI repressor and beads carrying anti-FLAG antibody.

Three DNA fragments, LEE20-275, LEE30-275 and LEE30- 275 151T were taken for DNA -sampling aiming to obtain the protein(s) specifically bound with LEE30-275 fragment. Fragments were cloned in pRW902 EcoRI-HindIII sites adjacent to the 5 LacI operators with two flanking 18 base pairs target sites for the yeast meganuclease I-SceI (plasmid detail in Figure 2.4). The resulting plasmids were transformed into *E. coli* K-12 strain MG1655 in which the *lacI* gene had been modified to encode a 3× FLAG-tagged LacI repressor. Plasmid pACBSR-DL1 that encodes SceI and the bacteriophage lambda Gam protein, under the control of an arabinose inducible promoter was co-transformed. After induction by arabinose, a discrete DNA fragment carrying the promoter of interest with five LacI repressor targets was generated and the fragment was stable due to inhibition of RecBCD by the Gam protein. Immunoaffinity methods using magnetic beads carrying anti-FLAG antibodies were then used to isolate the DNA fragment together with accompanying proteins, which were finally identified using gel electrophoresis and mass spectroscopy. Proteins bound on the LEE20-275, LEE30-275 and LEE30-275 151T fragments are shown in Figure 5.6. As expected, protein complexes contain a major band corresponding to the FLAG-tagged LacI, which is labelled by its molecular size (Butala *et al*., 2009). The intensity of these bands reflects the relative level of each protein bound on the different promoter fragments. Two protein bands such as 50S ribosomal protein L15 and HU were confirmed by mass spectrometry. The bands for RNAP β' and β sub units are also detected in the gel and labelled by their respective sizes (Lee *et al*., 2008; Butala *et al*., 2009). However, no protein bound specifically on the LEE30-275 promoter fragment was detected.

5.7 Study of a small translated ORF in LEE30-275 fragment

Random mutation analysis of the LEE30-275 fragments suggested that the LEE30-275 fragment contains a small ORF. To determine experimentally whether the predicted small ORF is translated; starting LEE30-275 fragment and a derivative containing a mutation at predicted stop codon in the small ORF (151T) were cloned into pRW225 as translation fusion

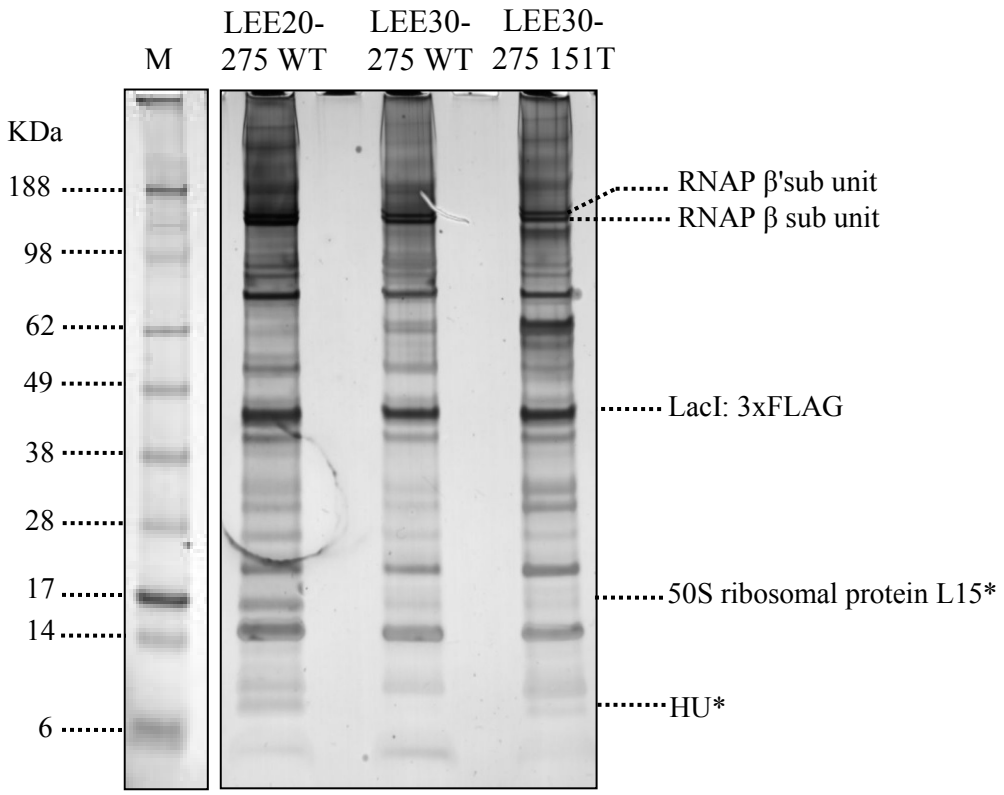


Figure 5.6. SDS-PAGE analysis of proteins that were bound on LEE20-275, LEE30-275 and LEE30-275 151T promoter fragments.

The figure shows the affinity-isolated proteins from *E. coli* K12 strain MG1655 grown in LB medium. Proteins that were identified by mass spectrometry analysis are indicated by asterisks, whilst other proteins are labelled according to their sizes (Lee *et al*., 2008; Butala *et al*., 2009). The gel was calibrated with Invitrogen SeeBlue Plus 2 protein markers (M).

to *lacZ* in a manner so that predicted translation start point of the small ORF could be in-frame with the *lacZ* initiation codon (Figure 2.21). Results illustrated in Figure 5.7 show no *lac* expression from the starting LEE30-275 fragment whereas the mutation (151T) at the predicted stop codon in small ORF results in a high level of *lac* expression. This result suggests that the LEE30-275 leader sequence contains a translated small ORF and its translation stop codon, TAG, is located at position 150 (Figure 5.3).

Some of the LEE30-275 derivatives containing mutations at the stop codon in the small ORF were generated spontaneously. M182 cells carrying pRW225 with the LEE30-275 fragment produced white colonies on MacConkey lactose indicator plates after overnight incubation at 37°C. However, when plates were incubated for two more days at 37°C, a number of small red colonies on the white colonies appeared. Sequence analysis confirmed spontaneous base substitutions at the TAG codon. These were a C (151C) or G (151G) or T (151T) for A at position 151 (Table 5.2). *lacZ* expression from these derivatives was quantified and reported in the Table 5.2. In each case, mutation at the stop codon caused a high level of *lac* expression and the expression is codon specific. The highest level of *lac* expression was found when the third nucleotide of the stop codon, G at position 152 changed to T. From this result, it appears that possibly ribosome starts the translation of the ORF from ATG at position 144 and stops at TAG codon at position 150. Single base(s) changes either at 151 or at 152 positions disrupt the stop codon that helps pass the message to the reporter gene.

In order to confirm the translation start site in the mini-gene, a number of derivatives of the LEE30-275 152T fragment were constructed. These derivatives contain base substitutions at two predicted translation start sites of the ORF. The changes were a C for G at position 143 (143C), a T for A at position 144 (144T), or G for T at position 145 (145G) and a double base change at positions 143 and 145 (143C 145G). The derivatives were cloned in pRW225 *lac* reporter plasmid and expressions were recorded (Table 5.2). Data presented in Table 5.2 show that introduction of 143C mutation that changes the second predicted start codon from TTG to TTC did not have any effect on the *lac* expression from LEE30-375 152 fragment. But mutation 144T that changed the first predicted start codon from ATG to TTG caused 4.5-fold reduction of the *lac* expression from LEE30-275 152T and 145G mutation that changed ATG to AGG caused 9-fold decrease in the expression. Recall that TTG is used as an alternative

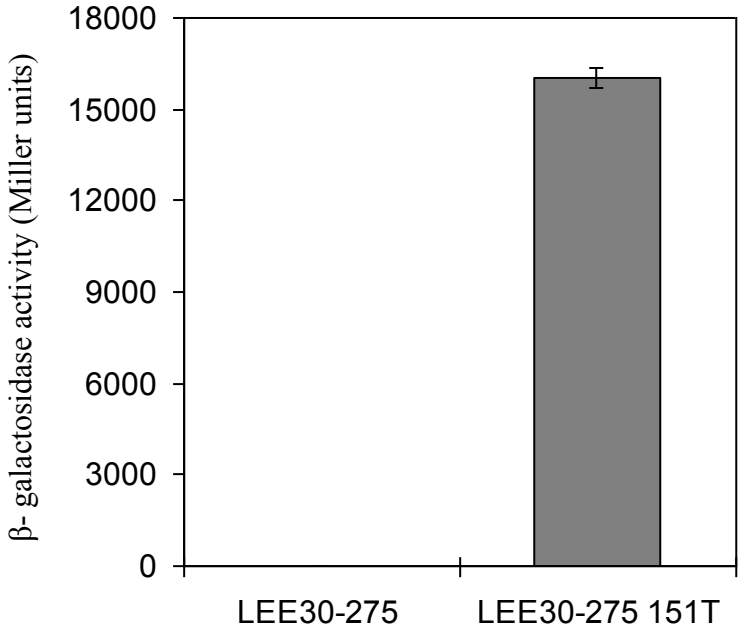


Figure 5. 7. Translation activity in the LEE30-275 fragment.

Measured β-galactosidase activity in M182 cells carrying plasmid pRW225 with starting LEE30-275 fragment or derivative cloned in pRW225 as translation fusion to *lacZ*. Measurements were made after growth of the cells in LB at 37°C to an optical density of ~0.5 at 650 nM. Values are the average of at least three independent assays, and standard deviations are shown with bars.

Table 5.2: Effects of mutations on the translation of mini-ORF (continued on page 203)

Derivatives	Sequences of cloned LEE30-275 fragment from 129 to 162::*lacZ*	β-galactosidase activity (Miller units ±sd)
Starting LEE30-275 fragment		
	129 140 150 162 *lacZ*→ TAAGGTGGTTGTTTGATGAAATAGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	16 ± 0
Mutations in the translation stop codon		
151C	TAAGGTGGTTGTTTGATGAAATCGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	8281 ± 168
151G	TAAGGTGGTTGTTTGATGAAATGGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	3751 ± 205
151T	TAAGGTGGTTGTTTGATGAAATTGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	16035 ± 353
152T	TAAGGTGGTTGTTTGATGAAATATATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	22995 ± 561
Mutations in the translation star sites		
144T 152T	TAAGGTGGTTGTTTGTTGAAATATATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	5083 ± 178
145G 152T	TAAGGTGGTTGTTTGAGGAAATATATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	2503 ± 17
143C 152T	TAAGGTGGTTGTTTCATGAAATATATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	23375 ± 801
143C 145G 152T	TAAGGTGGTTGTTTCAGGAAATATATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	52 ± 2

Table 5.2: Effects of mutations on the translation of mini-ORF (continued)

Derivatives	Sequences of cloned LEE30-275 fragment from 129 to 162::*lacZ*	β-galactosidase activity (Miller units ±sd)
Replacement of mini-ORF by KpnI site		
	129 • 140 • 150 • 162 • *lacZ*→	
LEE30-275-1	TAAGGTGGTTGTTTGA**GGTACC**AGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	542 ± 7
LEE30-275-1 143C	TAAGGTGGTTGTTT**C**A**GGTACC**AGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	22 ± 2
Mutations in the SD sequence		
132T 152T	TAA**T**GTGGTTGTTTGATGAAATA**T**ATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	8582 ± 188
132T 150C	TAA**T**GTGGTTGTTTGATGAAA**C**AGATGTGTCCTAAAGCTTGGGATCCCGTCGTTTTA	3083 ± 7

The table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW225 containing the LEE30-275 fragment and different mutant derivatives. Cultures were grown aerobically at 37°C in LB medium to an optical density of ~0.5 at 650 nm. Activities were measured in triplicate, giving a mean and standard deviation (sd). The central part of the table shows the base sequence from position 129 to position 162 of LEE30-275 fragment cloned in pRW225 as translation fusion to *lacZ* in a manner so that predicted translation start site of the mini-ORF is in-frame with the *lacZ* gene. Shine-Dalgarno sequence and mini-ORF deduced from mutational analysis are grey- and yellow respectively. In LEE30-275-1, mini-ORF was replaced with a KpnI site. Base substitutions are coloured red and underlined.

translation start codon in *E. coli* (Blattner *et al*., 1997; O'Donnell and Janssen, 2001). Therefore, a change from ATG to AGG affects the translation of the mini-gene severely compared with the change from ATG to TTG. Double mutations (143C 145G) that alter both start codons reduced the *lac* expression from the LEE30-275 152T derivative to the basal level. This argues that the mini-gene contains two adjacent alternative translation start points, TTG and ATG, located at positions 141 and 144 respectively and ribosome preferentially used ATG as the major translation start point during translation of the mini-gene.

In a complementary experiment, a new derivative LEE30-275-1, where the coding sequence of the mini-gene was altered by the introduction of a KpnI restriction site was constructed. This alteration leads to the disruption of the major translation start point (ATG to AGG), sense codon (AAA to TAC) and stop codon (TAG to CAG) of the mini-gene (Figure 2.23, Table 5.2). The LEE30-275-1 derivative was cloned into pRW225 and the *lac* expression from the derivative was quantified. Data recorded in Table 5.2 show that alteration of the mini-gene by KpnI site reduced the *lac* expression to ~2%, compared with the expression from the LEE30-275 152T fragment. Data presented in the Table 5.2 also show that *lac* expression from the LEE30-275-1 is 5-fold lower that that from LEE30-275 145G 152T. This might be due to the codon-specific effect on the *lac* expression as reported for *E. coli galE* gene by Dreyfus *et al*. (1985). Moreover, derivative of LEE30-275-1 containing a base substitution at the second translation start site (143C) reduced the *lac* expression to a barely detectable level. These results also confirms that *ler* leader sequence contain a translated mini-gene, which has two alternative translation start sites and a single sense codon.

Mutation(s) in the SD sequence can affect the translation efficiency of a gene. Recall that a number of up-mutations clustering in the predicted SD sequence were obtained during random mutagenesis of LEE30-275 cloned in pRW224 as transcription fusions to *lacZ* (Figure 5.3). In a complementary experiment, to determine the effect of the mutation at the SD sequence on the translation efficiency of the small ORF, derivatives of LEE30-275 containing mutations both at the stop codon and the SD sequence (132T 150C or 132T 152T) were constructed and cloned in pRW225 as translation fusions. The *Lac* expressions from the derivatives were quantified. Data presented in the Table 5.2 show that single base

substitutions at the SD sequence caused a sharp reduction in the translation of the mini-gene. The reduction might be due to a reduced mRNA-ribosome pairing efficiency.

5.8 Role of the mini-ORF in the context of the downstream gene expression

The LEE150 fragment that ends with *ler* translation start codon ATG, cloned in pRW225 as a translation fusion to *lacZ* (Figure 5.1), was used to investigate downstream effects of the small LEE1 leader ORF. Mutations that appeared to alter the mini-gene translation (Table 5.2) were introduced in the LEE150 fragment using megaprimer PCR. The name of each of the derivatives of LEE150 was based on the position of base substitution and base species substitution(s) adopted from LEE30-275 derivatives (Table 5.2). All the derivatives including LEE150 145G, LEE150 143C, LEE150 143C 145G, LEE 150-1 or LEE150-1 143C were cloned in pRW225 as translation fusions to the *lacZ* reporter gene. *Lac* expression from each of the derivatives was measured in M182 cells after growth in LB (Figure 5.8A). Results illustrated in Figure 5.8 show that mutation in the second alternative start codon (143C) of the mini-gene has no effect; mutation in the first start codon (145G) reduced the *lac* expression ~30% whilst mutations at both start sites caused ~55% reduction in *lac* expression compared with the starting LEE150 fragment. This suggests that translation of the upstream mini-gene has positive effect on the expression of the downstream *ler* gene. Similarly, the alteration of the mini-gene by KpnI site that results in the derivative LEE150-1 (Figure 2.23) and an additional change in the second translation start site that results in LEE150-1 143C reduced the expression of the downstream *ler* gene.

Many studies have showed that environmental signals can alter the LEE gene expression and this differs from strain to strain (Russell *et al*., 2007). In order to check whether the effect of the upstream mini-gene on *ler* gene expression varies with bacterial strains and environment, the effects were measured in EHEC Sakai cells after growth in both LB and DMEM. Results illustrated in Figure 5.8C show that translation of the upstream mini-gene reduced the downstream *ler* gene expression in EHEC Sakai cells in agreement with the results obtained in M182 cells, and the effect does not vary with different culture conditions.

The mini-gene located in the leader sequence of *LEE1* operon is conserved among other attaching and effacing bacteria including EPEC and *C. rodentium* (Figure 5.9). The sequence

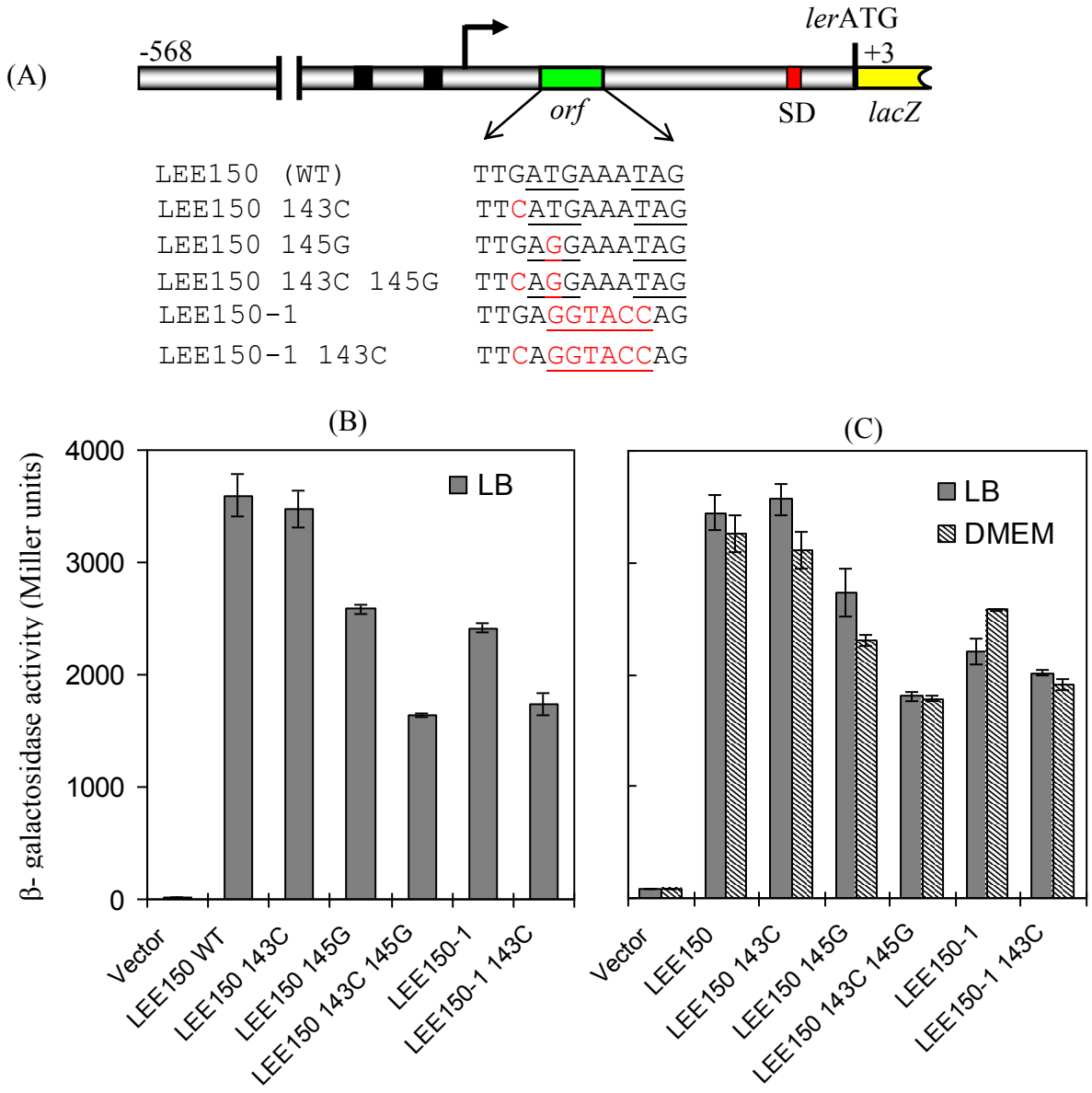


Figure 5.8. Effects of the upstream mini-ORF on the expression of downstream gene.

(A) Schematic representation of the *ler* regulatory region of LEE150, and derivatives carrying mutations in the mini-ORF. LEE150 145G carries a point mutation in the first translation start site of the mini-ORF. LEE150 143C carries a point mutation in the second predicted translation start site. LEE150 143 145G carries mutations that altered both translation start sites. In LEE150-1, the predicted mini-ORF was replaced with a KpnI site. LEE150-1 143C carries a mutation in the second predicted translation start site. Derivatives were cloned as translation fusions to *lacZ* in pRW225.

(B) Measured β-galactosidase activities in *E. coli* K-12 strain M182 containing pRW225 with LEE150 or derivatives. Measurements were made after growing the cells in LB at 37°C to an optical density of ~0.5 at 650 nM.

(C) Measurements were also made in EHEC Sakai cells after growth both in LB (grey shaded bars) and DMEM media (dashed bars). Values are the average of at least three independent assays, and standard deviations are shown with bars.

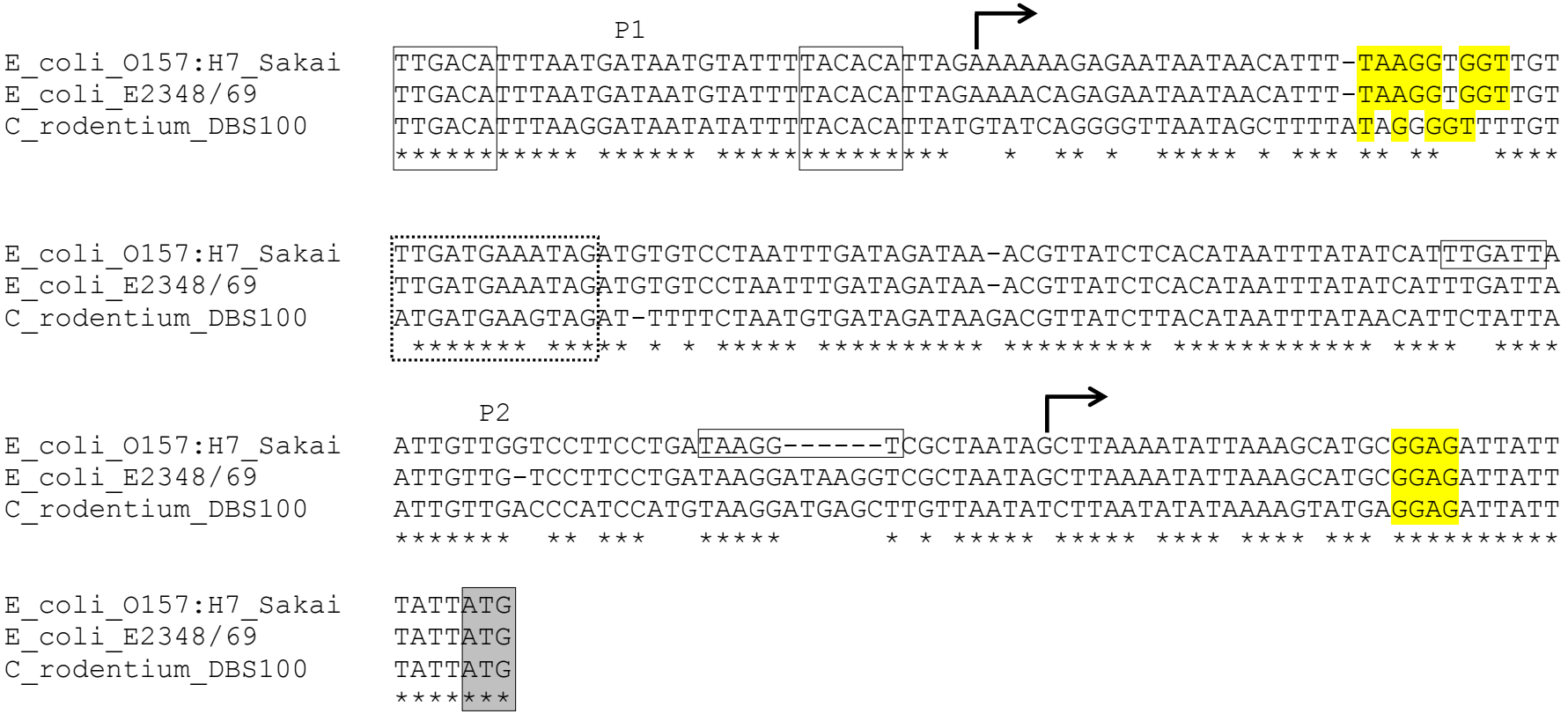


Figure 5.9. Alignment of the DNA sequences of the *LEE1* operon regulatory region of EHEC, EPEC and *C. rodentium*.

Solid white boxes correspond to two promoters: P1, which is present in EHEC (Sakai), EPEC (E2348/69) and *C. rodentium* (DBS100) and P2 promoter that is specific to EHEC (Sakai). Bent arrows indicate the predicted transcription start sites. Dotted box corresponds to the mini-ORF. Yellow regions correspond to the SD sequences. The grey shaded box represents the translation start site of the *ler* gene.

covering the SD sequence and coding sequence of the mini-gene in both EHEC and EPEC is identical. In each case, the mini-gene has a good SD sequence, two adjacent translation start sites (TTG and ATG), only one sense codon specific for lysine (AAA) and translation stop codon (TAG). In the case of *C. rodentium*, the mini-gene has a relatively weak SD sequence, two adjacent start sites (two ATG codons), one sense codon specific for lysine (AAG) and translation stop codon (TAG). However, in *C. rodentium,* both translation start sites are strong (two ATG codons compared with TTG and ATG in both EPEC and EHEC) and this perhaps compensates the weak SD sequence compared with EHEC and EPEC. However, in all three strains, the mini-gene has a lysine specific last sense codon.

In an attempt to determine whether lysine can modulate expression from the LEE30-275 promoter fragment, EHEC EDL933 cells containing pRW224 with the LEE30-275 fragment or derivatives carrying base substitutions at translation start sites (143C 145G) or a base substitution (147T) that changed the lysine specific sense-codon (AAA) to non-sense codon (TAA), were grown in minimal medium with or without L-lysine supplementation at 37°C. The *Lac* expression was measured after growing the cells to an optical density of ~0.4 at 650 nm. Results illustrated in Figure 5.10A show that lysine does not affect the *lac* expression from each of the derivatives. In another experiment, EHEC EDL933 cells containing pRW225 *lac* expression plasmid with the LEE150 or derivatives where translation of mini-gene was altered by base substitutions at translation start sites (LEE150 143C 145G) were grown in minimal medium with or without supplementation of L-lysine and *lac* expressions were quantified. Results illustrated in Figure 5.10B show that lysine does not influence the expression from the LEE150 or derivatives.

The small ORF located in the leader sequence of the LEE1 is translated and it encodes a dipeptide (Met-Lys) (Table 5.2). In order to determine whether *in-trans* supply of mini-gene encoded dipeptide can modulate the expression of downstream gene, the LEE30-275 fragment containing the mini-gene was cloned in a multicopy plasmid pSR, and the resulting plasmid was transformed into EHEC EDL933 cells containing plasmid pRW225 with the LEE150 or derivatives where mini-genes were knocked out. Results in Figure 5.11 show that, as expected, knocking out the mini-gene reduced the downstream gene expression but there is no effect of the plasmid-encoded dipeptide on the expression of the downstream gene.

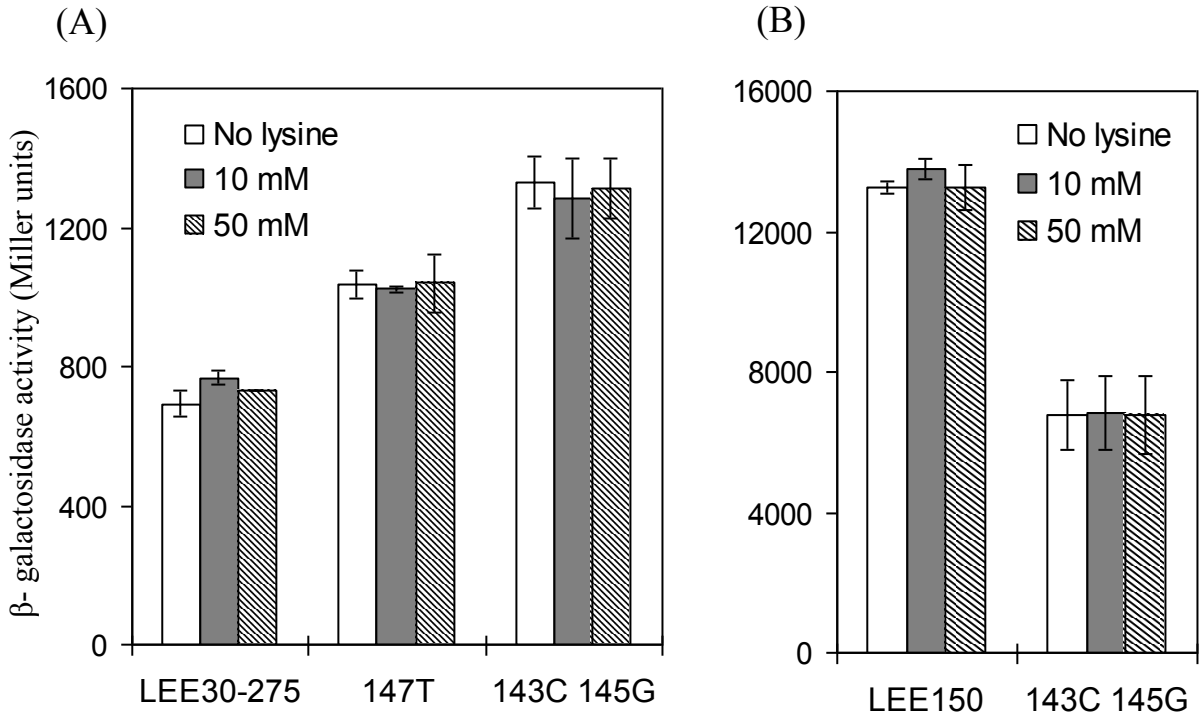


Figure 5.10. Effects of lysine on the expression from derivatives of both LEE30-275 and LEE150 fragments.

(A) Measured β-galactosidase activity in EHEC EDL933 cells containing pRW224 with LEE30-275 or derivatives cloned as transcription fusion to *lacZ*. Measurements were made after growing the cells to an optical density at 650 nM of 0.3-0.4 in minimal medium with or without supplementation of L-lysine at 37°C. Values are the average of at least three independent assays, and standard errors are shown with error bars.
(B) Measured β-galactosidase activity in EHEC EDL933 cells containing pRW225 with LEE150 or derivatives cloned as translation fusion to *lacZ*. Measurements were made after growing the cells to an optical density at 650 nM of 0.3-0.4 in minimal medium with or without supplementation of L-lysine at 37°C. Values are the average of at least three independent assays, and standard deviations are shown with bars.

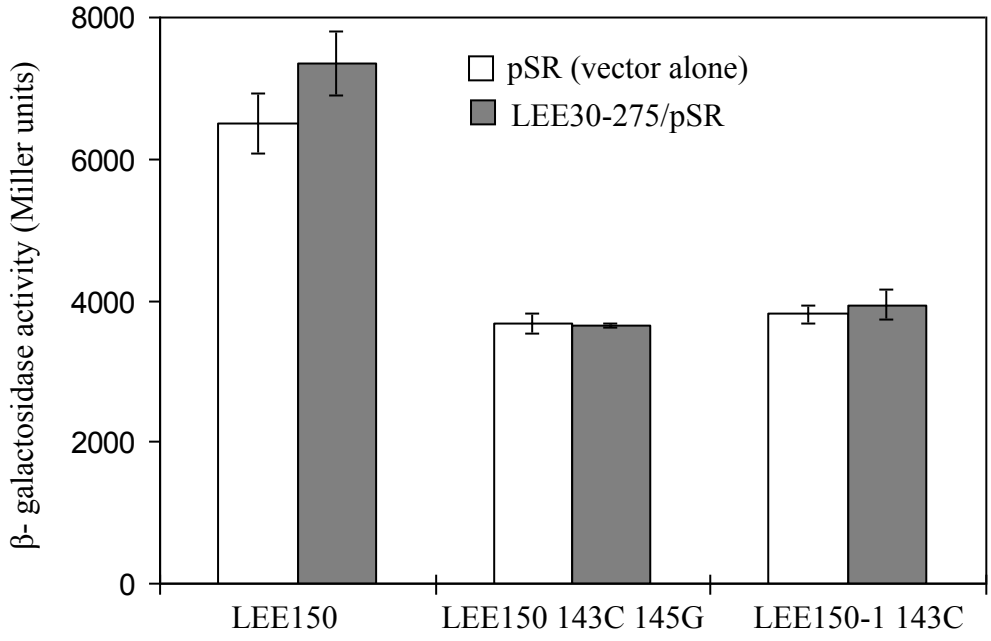


Figure 5.11. *In-trans* effect of mini-ORF on the expression from LEE150 derivatives.

The bar chart shows the measured β-galactosidase activity in EHEC EDL933 cells containing pRW225 with LEE150 or derivatives cloned as translation fusions to *lacZ*. Cells also contain multicopy plasmid pSR (vector alone: open bars) or pSR with the LEE30-275 fragment (shaded bars). Measurements were made after growing the cells in LB at 37° to an optical density of ~0.5 at 650 nm. Values are the average of at least three independent assays, and standard deviations are shown with bars.

5.9 Discussion

The *cis*-acting elements in untranslated leader sequences such as small ORFs can take part in gene expression regulation (Yanofsky, 1981; Vitreschak *et al*., 2004). Reports on the identification of short translated upstream ORFs are limited and the possibility of alternative initiation codons complicate the recognition of associated ORFs (Morris and Geballe, 2000). Mutational analysis and subsequent reporter gene assays suggested that the *LEE1* leader sequence contains a potential ribosomal binding site followed by a small ORF, which has two adjacent alternative translation initiation codons, a sense codon specific for lysine and a stop codon located just six nucleotides downstream of the first predicted start codon (Figure 5.12A). Base change at the first predicted translation initiation codon (ATG) reduced the translational activity ~9 fold but did not completely abolish the activity. Base change at a second possible initiator codon (TTG) had no effect on the translational activity. However, base changes at both initiators reduced the translational activity to a barely detectable level. This suggests that, though the predicted ORF contains two adjacent initiator codons, translation initiation mostly takes place at the first predicted initiation codon (ATG) with a trace from the second (TTG). Similar minigenes were reported to be present at the *bar* loci in the bacteriophage λ chromosome. Expression of mini-genes containing rare codons at the end results in accumulation of peptidyl-tRNAs, which inhibit protein syntheis and thus cause lethal effects in bacteria (Ontiveros *et al*., 1997; Oviedo *et al*., 2004). However, role of the *LEE1* operon leader mini-gene seems to be different.

An interesting finding from my work concerned the role of the upstream mini-gene in the expression of downstream genes. Knocking out the mini-gene reduces the expression of the down stream *ler* gene significantly, suggesting that efficient translation of down stream gene(s) requires prior translation of upstream mini-gene (Figure 5.12B). How the mini-gene increases downstream gene expression is unclear. There is a wide range of sequence variations in the *LEE1* regulatory region among EHEC, EPEC and *C. rodentium*. However, the mini-gene region is fairly conserved and in each case, it possesses a triplicate (AAA or AAG) that is specific for lysine. However, my work did not find involvement of lysine in the mini-gene-dependent regulation of the downstream gene. One possible mode of action would be that the upstream mini-gene might function as processive, long-range antiterminator for the *LEE1* operon. The mRNA transcript of the mini-gene may directly, or associated with other factors,

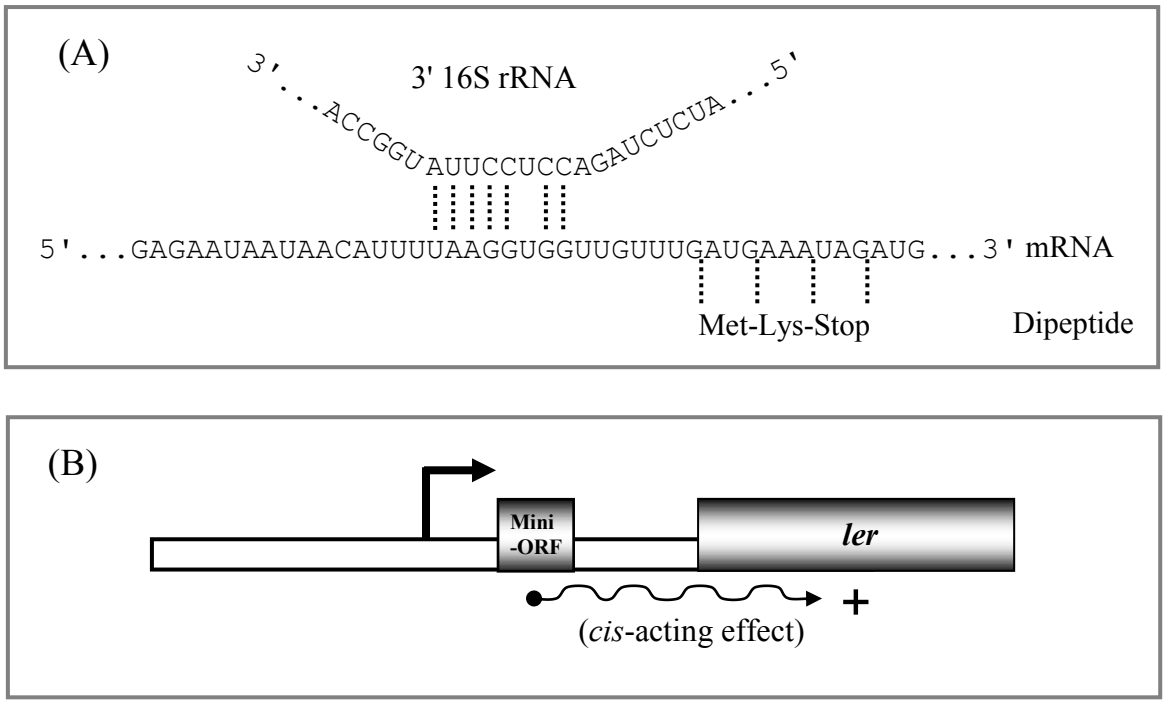


Figure 5.12. Schematic representation of the role of the mini-gene located in the *LEE1* operon leader region.

(A) Partial sequence of mRNA molecules representing leader sequence of *LEE1* operon. Ribosome recognizes the SD sequence, located upstream of the mini-ORF and initiate translation at the immediate downstream ATG start site and a dipeptide is synthesized.
(B) Translation of upstream mini-ORF increases the expression of the downstream *ler* gene.

interact with the RNA and that may allow RNAP to bypass termination signals of the genes in the *LEE1* operon. Recently Irnov and Winkler (2010) reported that a *cis*-acting RNA element, coined 'EAR', located between the second and third genes of the *eps* (exopolysaccharide) operon associates with RNAP to readthrough of distally located termination signals. The observation that knocking out the mini-gene reduced ~55% of the expression of the first cistron (*ler*) suggests that possibly the mRNA transcript acts as an EAR sequence and expression of the distal cistrons of the *LEE1* operon might be more dependent on the mini-gene transcript. To address this possibility, I made an isogenic derivative of EHEC EDL933 using the gene gorging method (Herring *et al.*, 2003). In this strain, mini-gene has been replaced with the KpnI site and also the second translation start point TTG has been changed to TTC (data not shown). Quantification of mRNA molecules for different genes in *LEE1* operon both from the wild type strain and the derivative will give an idea whether the LEE1 leader mini-gene transcript functions as a processive, long range antiterminator.

Chapter 6

Regulation of the *LEE1* and *LEE2* operons by Ler

6.1 Introduction

The *LEE1* operon is considered the master operon in the LEE region because the first cistron of this operon encodes the Ler protein, which positively regulates most of the LEE genes. Nucleoid associated protein, H-NS, plays a suppressive role in LEE gene expression. Ler is supposed to negate the H-NS mediated repression ensuring the LEE gene expression (Mellies *et al*., 1999; Elliott *et al*., 2000; Sperandio *et al*., 2000; Bustamante *et al*., 2001; Sanchez-SanMartin *et al*., 2001; Umanski *et al*., 2002; Haack *et al*., 2003). However, the role of Ler in the expression of genes from the *LEE1* operon is open to question. Most of the previous studies reported that Ler does not play any role in the expression of the *LEE1* operon (Mellies *et al*., 1999; Elliott *et al*., 2000; Haack *et al*., 2003). However, Berdichevsky *et al*. (2005) and Yerushalmi *et al*. (2008) reported that Ler negatively regulates *LEE1* transcription in EPEC.

Ler positively regulates the expression of the *LEE2* operon. In an effort to know whether Ler can activate expression from the *LEE2* promoter in EPEC, Sperandio *et al*. (2000) performed deletion analysis of the *LEE2* operon regulatory region and found that the expression from the *LEE2* regulatory region covering nucleotide sequence from -247 to +149 is activated by Ler, whereas expression from a fragment of -126 to +149 sequence is not activated. Moreover, DNA- binding assays showed that Ler bound to a DNA fragment containing -300 to +1 region of *LEE2* and DNase I footprinting revealed that Ler protected a region of 121 base pairs upstream of *LEE2*. This extended region protected by Ler suggests that Ler binds in a multimerized form. Bustamante *et al.* (2001) suggested that Ler activates the *LEE2* operon by counteracting H-NS mediated repression. To date, the role of Ler in expression of *LEE1* and *LEE2* operons has been investigated in EPEC. This study was aimed to determine the effects of Ler on expression from both *LEE1* and *LEE2* operon regulatory regions in EHEC.

6.2 The role of Ler in the expression from the *LEE1* operon regulatory region

In an attempt to study the role of Ler in expression of *LEE1*, the *ler* gene was cloned into a multicopy plasmid pJW15Δ100 under the control of the *melR* promoter to give plasmid pSI04. (Figure 2.20). Functional Ler encoded from the plasmid was confirmed by the pathogenicity test commonly known as FAS test (Knutton *et al*., 1989, 1991).

EHEC produce a type three secretion mediated characteristic A/E lesion when they adhere to intestinal mucosa *in vivo* and to tissue culture cells *in vitro*. The A/E lesion is associated with the dense concentration of microfilamentous actin that accumulates in the apical cytoplasm beneath attached bacteria. This actin can be detected by fluorescence staining in the FAS test. To test this, HeLa cells were infected with wild type EHEC EDL933 cells and infected cells were fluorescence stained with fluorescein isothyocyana-phalloidin, a phallotoxin that binds specifically to polymerized actin that results in fluorescence. Since Ler plays a positive role in the regulation of the LEE genes that are involved in the formation of A/E lesion, an EHEC EDL933 *ler* mutant bacterium should not be adherent to HeLa cells and thus should be the negative control for the FAS test as seen by Friedberg *et al*. (1999) and Elliott *et al*. (2000). However, if the Ler expressed from pSI04 plasmid is functional, EDL933 *ler* mutant cells transformed with pSI04 should be capable of adhering to HeLa cells and give a positive FAS test. Figure 6.1 clearly illustrates that deletion of *ler* gene from the EHEC Sakai strain altered the wild type adherence-phenotype (loss of actin formation) (D-F), complementation of the Δ*ler* mutation with the plasmid encoded Ler (pSI04) restored the wild type adherence-phenotype (A-C), and positive FAS staining (G-I). These results suggested that pSI04 encodes Ler, which is functional.

To investigate the action of Ler at the *LEE1* operon regulatory region, plasmids pSI04 containing *ler* or empty vector pJW15Δ100 were introduced in *E. coli* K-12 strain M182 *Δlac* carrying pRW224 derivatives with different *LEE1* regulatory region fragments (Figure 2.11). In a preliminary experiment, M182 was co-transformed with pRW224 carrying the LEE10-568 fragment and the plasmid pSI04 or empty vector pJW15Δ100, and β-galactosidase expression was measured. The results show ~30% repression of the promoter activity in the presence of Ler (Figure 6.2).

Ler-mediated repression at the *LEE1* operon regulatory region carried by fragments with the two series of nested deletions described earlier (Figure 3.1) was measured. Moreover, Ler -mediated repression was also quantified for one new nested deleted derivatives namely LEE10-235, which was cloned into pRW224. With the LEE10 series of nested fragments, repression is observed with the LEE10-568, LEE10-275, LEE10-235, LEE10-215,

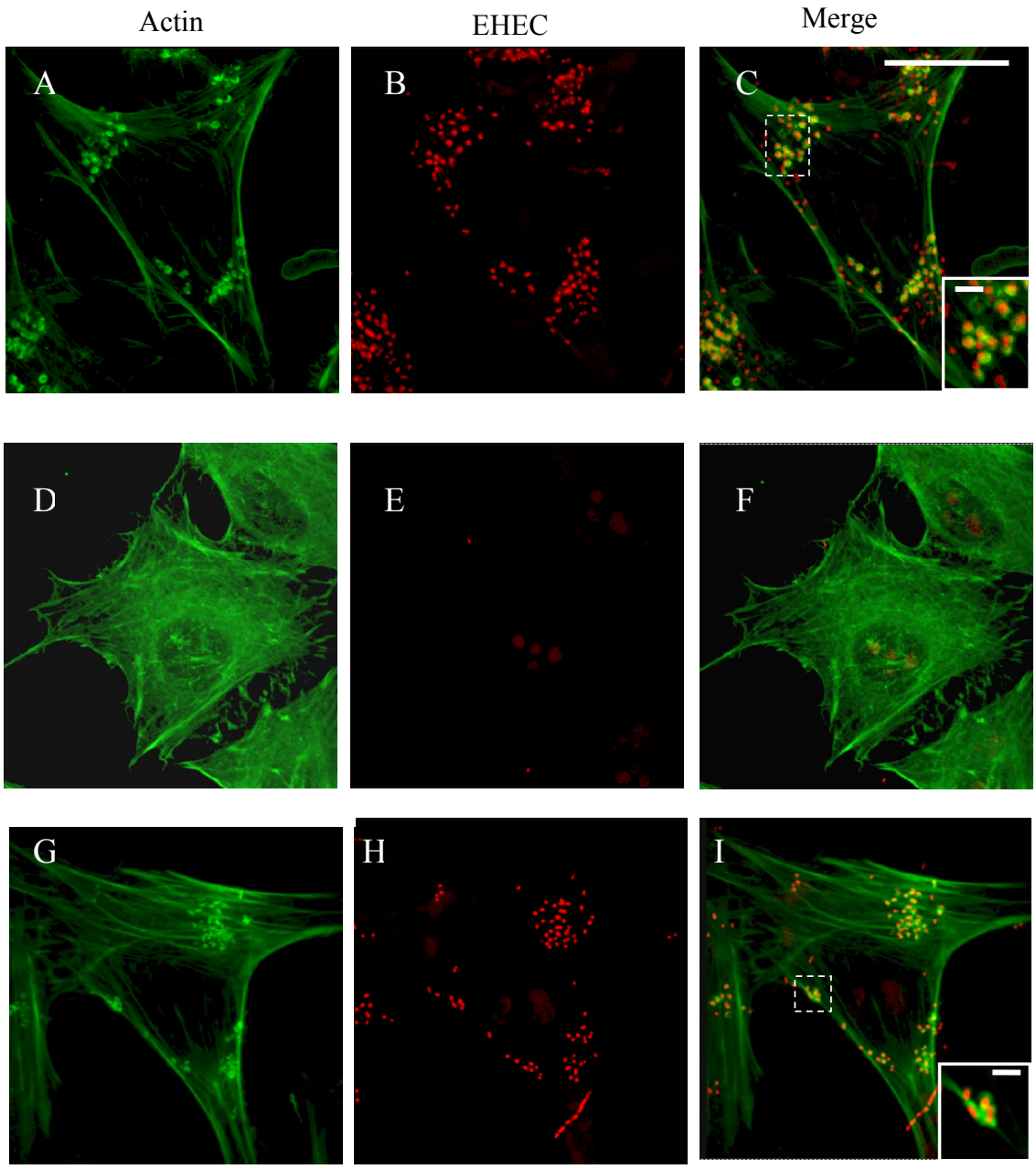


Figure 6.1. Ler complementation assay.

The figure illustrates the actin fluorescence (A, D, and G), corresponding bacterial fluorescence (B, E, and H) and merge of fluorescence and phage-contrast (C, F, and I) micrographs showing EHEC EDL933 infected HELA cells. Micrographs A-C: wild type EDL933 cells, D-F: EDL933Δ*ler* with pJW15Δ100 (no *ler*) and G-I: EDL933Δ*ler* with pJW15 containing *ler*. Bar = 20µm; 0.5µm (inset). The assay was done by Robert Shaw.

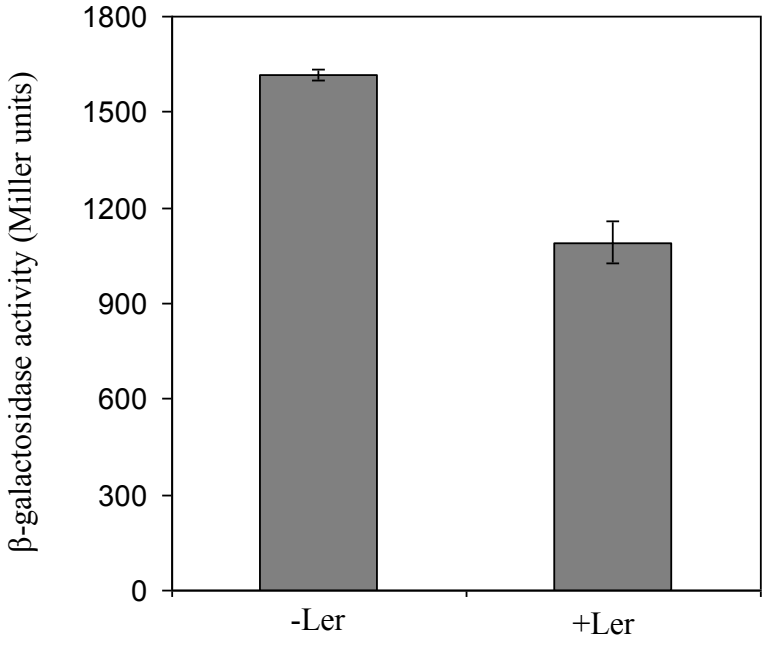


Figure 6.2. Repression of expression from the *LEE1* regulatory region by Ler.

The bar chart illustrates measured β-galactosidase expression in M182 cells containing pRW224 with the LEE10-568 promoter fragment together with plasmid pJW15Δ100 (-Ler), or pSI04 (+Ler). Measurements were made after growing the cells in LB medium at 37ºC to stationary growth phase. Each bar represents the mean of three independent experiments together with the standard deviations.

and LEE10-203 fragments containing both the P1 and P2 promoters but lost with the LEE10-155, LEE10-115 and LEE10-75 fragments, which contain only the P2 promoter (Figure 6.3), suggesting that Ler might target the *LEE1* promoter. Interestingly, the level of repression was found to be higher with LEE10-235 and LEE10-215 fragments and decreased with LEE10 -203 and LEE10-195. These suggest that possibly Ler nucleates immediately upstream of the P1 promoter and overlaps with the promoter -35 element.

6.3 Regulation of expression from the *LEE2* promoter by Ler

The regulatory region of the *LEE2* operon is located upstream of *espZ* gene. Therefore, the start point of this work is to create a construct in which *lacZ* expression is controlled by the fragment LEE2-220, which carries base sequence from position -220 to +109 with respect to the transcription start point reported by Mellies *et al*. (1999) (Figure 2.12, 2.13). Figure 6.4A illustrates three nested deletion derivatives of this fragment showing the location of the upstream end of each fragment with respect to the proposed *LEE2* promoter. Each fragment was cloned into the *lac* expression vector, pRW224, and the resulting recombinants were transformed into the *Δlac E. coli* K-12 laboratory strain, M182. Measurements of β-galactosidase expression at 37°C and 25°C illustrated in Figure 6.4 show the effects of the deletions on the activity of the *LEE2* promoters. Expression increases as upstream sequences are removed and increased expression is found both with the LEE2-125 and LEE2-100 fragments that carry DNA upstream sequence up to positions -125 and -100, respectively. These results suggest that sequence upstream of position -125 might contain negative elements that could reduce the *LEE2* promoter activity.

Nucleoid associated protein H-NS plays a negative role in the expression of the A/T rich LEE genes and it represses the expression better at lower temperature (Umanski *et al*., 2002). In order to determine whether the lower level of expression from the LEE2-220 fragment compared with two other fragments is associated with H-NS, *lac* expression from all these three fragments were measured at 25°C. Results illustrated in Figure 6.4 show that the removal of upstream sequence caused a ~3.5 fold increase in expression at 37°C whilst at 25°C, the expression increased ~6.5-fold suggesting that H-NS might be associated with the low level of activity of the LEE2-220 fragment.

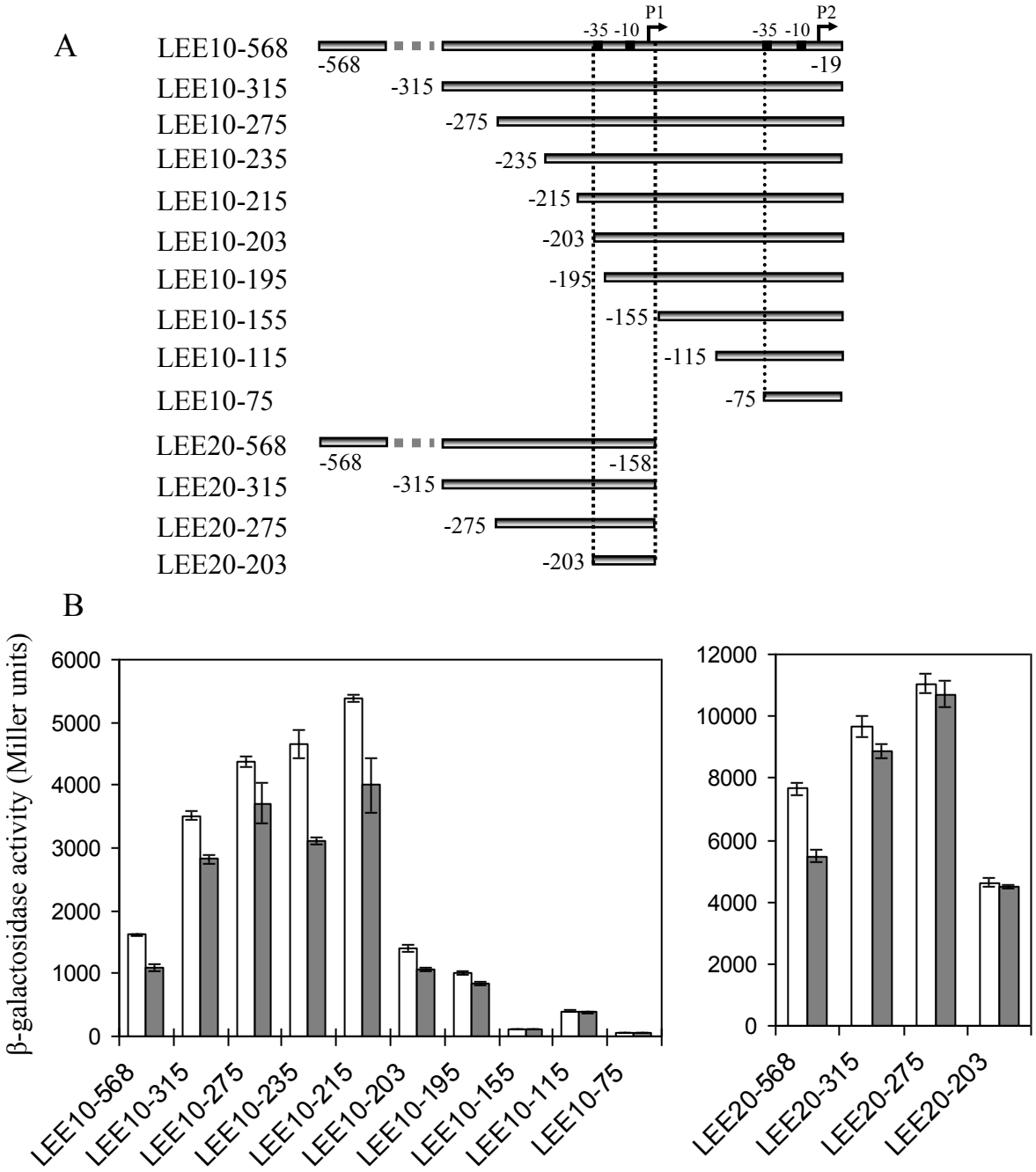


Figure 6.3. Ler-mediated repression at different *LEE1* regulatory region fragments.

A. Schematic representation of a series of nested deleted derivatives of *LEE1* operon regulatory region. The coordinates of the upstream and downstream end of each derivative refer to the number of base pairs upstream from the functional ATG start codon of the *ler* gene.

B. Bar charts illustrate the measured β-galactosidase activities in M182 carrying pRW224 with each of the different fragments together with plasmid pJW15Δ100 (-Ler: open bars) or pSI04 (+Ler: shaded bars). Measurements were made after growing the cells in LB at 37ºC to stationary growth phase.Standard deviations were obtained from three independent experiments.

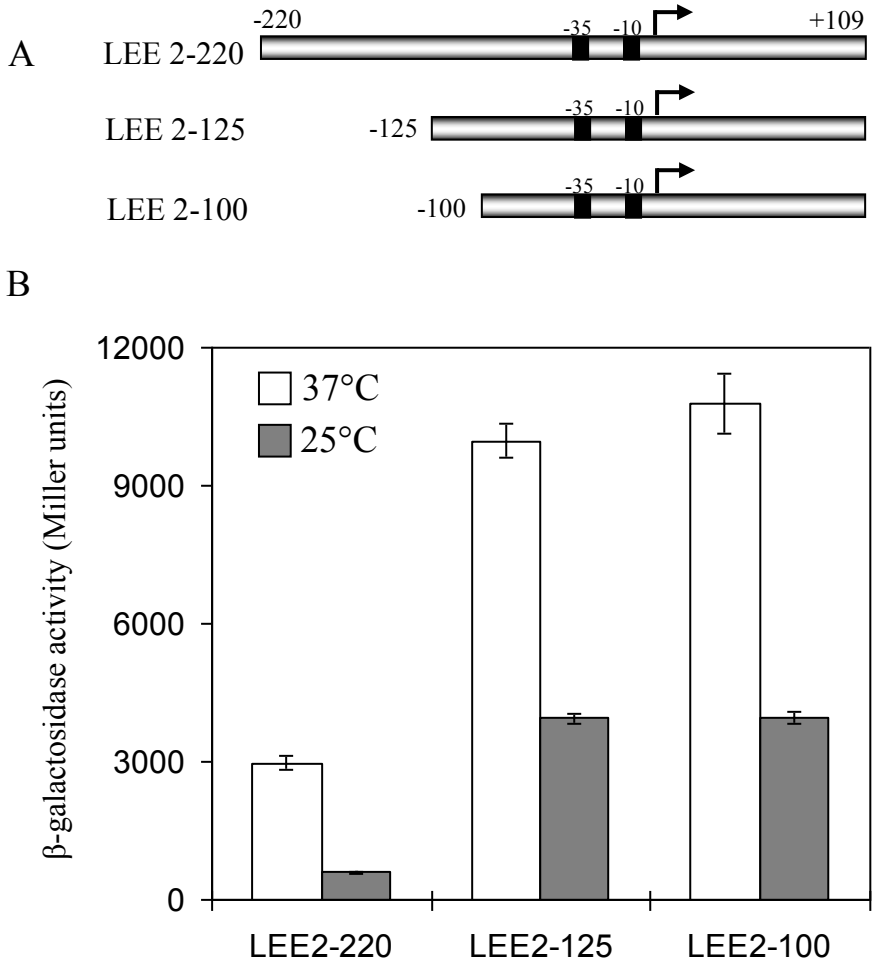


Figure 6.4. Nested deletions of the *LEE2* regulatory region.

A. Schematic representation of three EcoRI-HindIII DNA fragments of *LEE2* regulatory region. The coordinates of the upstream and downstream end of each fragment refer to the number of base pairs upstream and down stream from the transcription start point of *espZ* gene reported by Mellies *et al*. (1999). The transcription start site is indicated by bent arrow and the shaded black boxes represent the cognate -10 and -35 hexamer elements.
B. The bar chart to illustrate measured β-galactosidase activities in *E. coli* K-12 cells carrying pRW224 with each of the different fragments. Activities were measured after growing the cells in LB medium to stationary growth phase at 37°C (open bars) or at 25°C (shaded bars). The values are the average of three independent assays. Standard deviations are shown with bars.

In order to investigate the action of Ler at the *LEE2* operon regulatory region, plasmid pSI04 (Ler) or empty vector were introduced with the *E. coli* K-12 M182 *Δlac* strain carrying pRW224 derivatives with three different *LEE2* regulatory region fragments (Figure 6.4A). In a preliminary experiment, M182 was co-transformed with pRW224 carrying the LEE2-220 fragment and the plasmid pSI04 (Ler) or empty vector, and β-galactosidase expression was measured both at 37°C and 25°C. Results illustrated in Figure 6.5 show that Ler activates the *lac* expression from the LEE2-220 promoter fragment and the activation was higher at 25°C (~4 fold) than that at 37°C (~2 fold). Ler-dependent activation was also checked for the nested derivatives, LEE2-125 and LEE2-100 after growing the cells at 25°C. Results illustrated in Figure 6.6 show that there is no influence of Ler in the expression from both LEE2-125 and LEE2-100. This suggests that *LEE2* promoter no longer needs Ler for transcription activation if the negative elements located between positions -125 and -220 are removed.

6.4 Discussion

Whilst most of the previous studies reported that Ler does not have any influence in expression of its own operon, the conclusion of my work that over-expression of the Ler represses expression from the *LEE1* promoter is interesting and is in agreement with the findings of Berdichevsky *et al*. (2005) and Yerushalmi *et al*. (2008). In *E. coli*, many central regulators are negatively autoregulated. This mechanism enables rapid promoter activation upon stimulation while maintaining a relatively low steady-state level of the regulator (Rosenfeld *et al*., 2002). In addition, cell-to-cell variability in the levels of promoter activity is reduced by an autorepression mechanism (Becskei and Serrano, 2000). This study made a detailed nested deletion analysis of Ler regulatory region and checked the effects of Ler on the expression from each of the promoter fragment. Deletion analysis revealed that Ler does not have any effect on the *LEE1* P2 promoter but it represses expression from the *LEE1* P1 promoter. Ler-mediated repression was found with upstream (LEE20-568) and down stream sequences (LEE10-203) but not with short fragment LEE20-203 of the *LEE1* promoter. The action of Ler in the *LEE1* regulatory region resembles the action of H-NS, as H-NS needs both upstream and downstream sequences for repression of the expression from the *LEE1* promoter regulatory region (Figure 3.9).

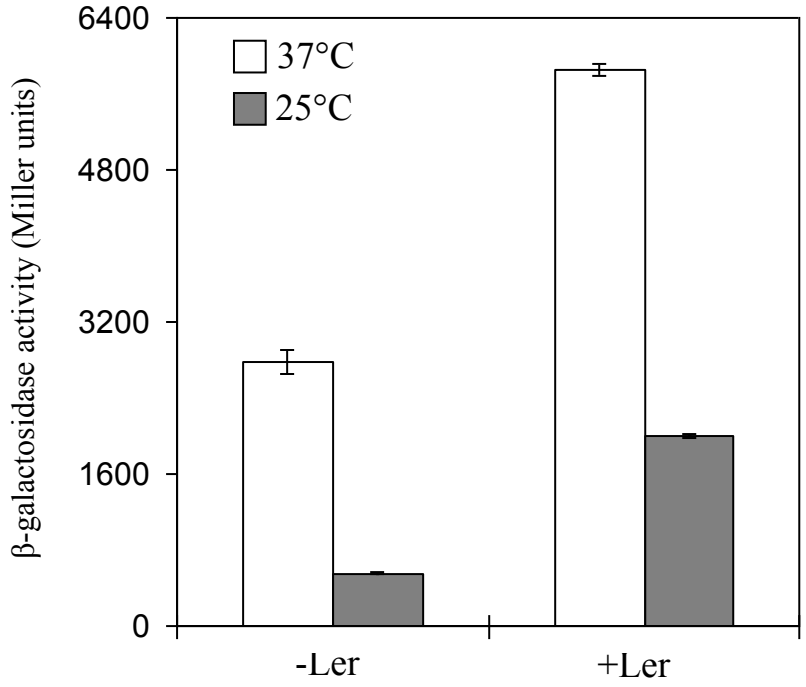
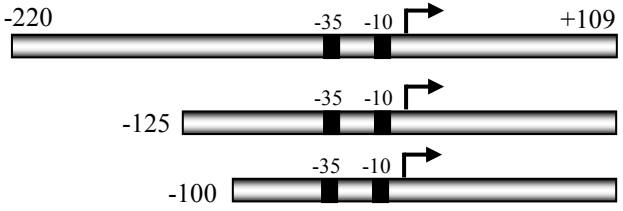


Figure 6.5. Activation of expression from the *LEE2* operon regulatory sequence by Ler.

The bar chart illustrates measured β-galactosidase activity in M182 cells containing pRW224 with the LEE2-220 promoter fragment together with plasmid pJW15Δ100 (-Ler), and pSI04 (+Ler). Measurements were made after growth of the cells in LB medium to stationary growth phase at 37°C (open bars) or at 25°C (shaded bars). The values are the average of three independent assays. Standard deviations are shown with bars.



Derivatives	Expression from different promoters		
	β-galactosidase activity (Miller Units ± sd)		
	-Ler	+Ler	Fold activation
LEE2-220	552 ± 11	1993 ± 20	3.6
LEE2-125	3960 ± 295	3972 ± 219	1.0
LEE2-100	3537 ± 82	3521 ± 50	1.0

Figure 6.6. Ler-dependent activation at different *LEE2* regulatory region fragments.

The left part of the figure illustrates three different fragments covering the *LEE2* regulatory region. The fragments were cloned into pRW224, using the same conventions as in Figure 6.4. The right part of the figure presents β-galactosidase activities measured in M182 cells carrying pRW224 containing each of the fragments. Cells also contained either empty vector pJW15Δ100 (-Ler) or pSI04 (+Ler). Measurements were made after growth in LB at 25ºC to stationary growth phase. Standard deviations (sd) were obtained from at least three independent experiments. The fold activation by Ler is shown in fourth column.

Ler plays an opposite role in the case of expression from the *LEE2* promoter. It activates the *LEE2* promoter. Expression from the nested deleted derivatives of the *LEE2* promoter suggests that upstream region of the promoter located between -125 and -220 contains a negative element that represses the expression from the promoter. Nucleoid associated protein H-NS plays major negative role in LEE gene expression. The sequences to which it binds are usually A+T-rich and are often associated with regions of intrinsic curvature (Stoebel *et al*., 2008; Browning *et al.,* 2010). Increase in temperature reduces the degree of curvature and thus H-NS cannot function effectively. A negative element located between positions -125 and -220 of the *LEE2* regulatory region represses the promoter activity higher at 25°C than at 37°C. This suggests that possibly H-NS binds to the upstream sequence and conveys the negative effect to the promoter. Deletion of the upstream sequence important for H-NS mediated repression thus results in increase in promoter activity. The conclusion that H-NS is the player that represses the expression from the *LEE2* promoter is in line with Sperandio *et al.* (2000) and Bustamante *et al*. (2001).

The role of Ler in the expression from the *LEE2* seems to be interesting. Ler only increases the expression from the fragment that is repressed by H-NS. This suggests that the role of Ler is to displace H-NS (Stoebel *et al*., 2008). However, Ler increases the activity from the repressed fragment but the activation did not reach the level of activity of the fragments that lack the negative elements. This implies that Ler partially antagonises the H-NS repression. In other words, Ler is not needed for maximal expression from the *LEE2* promoter.

In conclusion, this preliminary work finds that overexpression of Ler represses the *LEE1* promoter activity by interacting with an extended region of P1 promoter and activates the *LEE2* promoter possibly by displacing H-NS from a region located between positions -125 and -220 with respect to the transcription start site.

Chapter 7

Final discussion

In EHEC or related pathogenic bacteria, horizontally acquired LEE genes play vital roles during infection. Upon the completion of full genome sequencing of pathogenic bacteria, researchers are investing their efforts to understand the functional mechanisms of the expression of the LEE genes. In over a decade and a half, there has been remarkable progress in understanding the regulation of LEE genes. Genetic and biochemical studies revealed that both LEE and non-LEE-encoded regulatory proteins regulate LEE gene expression. Among them, Ler, encoded by the first gene of *LEE1* operon, functions as the central receiver of regulatory inputs. Most of the both negative and positive regulatory proteins and signals associated with the LEE genes convey their regulatory role to the LEE genes via modulating *ler* gene expression (Figure 1.9). Knowledge on regulation of the *ler* gene is therefore, crucial for understanding of the regulation of other LEE genes. Hence, to upgrade the present understanding of the LEE gene regulation, I put my efforts to determine the organization of the *LEE1* operon regulatory region and to know how LEE-encoded GrlA exploits the complex structure of *LEE1* operon regulatory region to activate *ler* expression. Study of the *LEE1* operon regulatory region has enabled me to identify the P1A cryptic promoter that overlaps the main promoter, and a mini-gene in the leader region, further suggesting that *ler* is subject to very complex regulations. Moreover, it was found that overexpression of Ler regulates its own transcription and activates the expression from the *LEE2* promoter by negating H-NS repression.

Until recently, there is a big debate whether there are one or two functional promoters in the EHEC *LEE1* operon regulatory sequence (Sperandio *et al*., 1999, 2002; Porter *et al*., 2005). To advance understanding of the issue, I performed a deletion analysis of the *LEE1* regulatory region. The deletion analysis described in Chapter 3 finds several clues concerning *LEE1* promoter structure: (i) P1 is the major functional promoter, (ii) P2 plays but a minor role, (iii) the immediate upstream sequence from the P1 promoter contains up-elements, (iv) further upstream as well as downstream sequences from the P1 promoter contain negative elements, which are thought to be the sites for interacting sites for H-NS that needs both upstream and downstream sequences for repression. To date, the *LEE1* promoter functional -10 and -35 elements had been predicted based on their relative positions from the transcription start site and sequence homology with the consensus *E. coli* σ^{70}-dependent promoter. This results in different sets of -10 and -35 elements suggested by different groups (Mellies *et al*., 1999;

Sperandio *et al*., 2002; Deng *et al*., 2004; Russell *et al*., 2007; Sharp and Sperandio, 2007; Kendall *et al*., 2010). Using both genetic and biochemical assays, the precise location of the P1 promoter -35 and -10 hexamers was determined. These are TTGACA and TACACA respectively and they are separated from each other by a non-optimal 18 base pairs spacer sequence.

The major aim of my project was to investigate GrlA-dependent activation of the *LEE1* promoter. This is one of the big issues among the researchers dealing with LEE gene regulation since the LEE-encoded GrlA was reported to play a positive role in LEE gene expression (Deng *et al*., 2004). Many of the reports published to date, concerning the action of GrlA at the *LEE1* promoter are confusing and contradictory. Whilst some groups (Barba *et al*. 2005; Huang and Syu, 2008) suggested that GrlA activates expression from the *LEE1* promoter independently, Russell *et al* (2007) assumed that GrlA might need other factor(s) to activate expression of the *ler*. In view of these reports, and to obtain advance understanding of the action of GrlA, a systematic study of the *LEE1* promoter and its activation by GrlA in both EHEC and K-12 backgrounds was performed.

GrlA activates the *LEE1* promoter in both *E. coli* K12 and EHEC backgrounds suggesting that GrlA can activate the *LEE1* promoter independently. GrlA-dependent activation from a series of nested deleted *LEE1* regulatory region fragments located the GrlA targets at the *LEE1* P1 promoter. Mutational analysis of a 52 base pair-long P1 promoter fragment, containing the functional promoter elements, activated by GrlA defined that GrlA activates *LEE1* promoter by targeting the non-optimal spacer sequence between P1 -35 and -10 elements. It is important to recall that both Barba *et al* (2005) and Jimenez *et al*. (2010) found GrlA-dependent *LEE1* activation from a promoter fragment that lacks upstream sequence of the promoter -35 element but, amazingly, they never tested a fragment that lacks both upstream and downstream sequences.

Concerning the point, raised by Jimenez *et al*. (2010), that GrlA mainly acts to counteract repression by H-NS, my work revealed that both upstream and downstream sequences are necessary for H-NS-mediated repression of the P1 promoter, and interestingly, the short 52 base pairs promoter fragment that is activated by GrlA, is not repressed by H-NS (Figure 3.9).

This strongly argues that GrlA is a *bona fide* transcription activator, likely acting directly.

Attempts were made to perform DNA electrophoresis mobility-shifting assay for providing evidence that GrlA-dependent activation occurs via a direct binding at the promoter region, and to perform footprinting assay to map GrlA binding site. To do these, GrlA protein fused either to His tag, a maltose binding protein (MBP) tag, or a glutathione-S-transferase (GST) tag, were used to purify the protein, but no functional GrlA protein was obtained. Purified functional GrlA protein is difficult to obtain (Barba *et al*., 2005). However, Huang and Syu (2008) reported GrlA binding to the *LEE1* promoter using purified GrlA-GST but their results are unconvincing since there were no controls to establish specificity. Recently, Jimenez *et al*, (2010) reported convincing GrlA band shifts, together with controls and showed that it bound to a target downstream of position -54 at the *LEE1* P1. However, they did not present footprints to show a GrlA binding site. In view of these problems, a genetic suppression approach was used to solve the problem. This classical approach revealed that GrlA interacts with the *LEE1* P1 spacer region and provided the first experimental evidence that the proposed GrlA helix-turn-helix is functional.

Based on the mutational analysis described in chapter 3, I argue that GrlA compensates the abnormal spacing of the *LEE1* P1 promoter. If the spacer, which is 18 bp between the -10 and -35 region is changed to the optimum 17 bp, GrlA regulation is no longer necessary. This suggests that GrlA activates transcription in a manner similar to members of the MerR family regulators, which activate transcription by binding to the spacer of target promoters at which the distance between the -10 and-35 elements is greater than normal (O'Halloran *et al*., 1989; Parkhill and Brown, 1990). Models for activation mechanism of MerR family members suggest that the binding of MerR family members to their target promoter spacers causes deformation, which is defined by a local unwinding of the spacer DNA by 33°, causing realignment of the -10 and -35 elements and thus facilitating RNAP binding (O'Halloran *et al*., 1989; Frantz and O'Halloran, 1990; Ansari *et al*., 1992). Evidence that GrlA targets the *LEE1* P1 promoter spacer and compensates abnormal spacer length suggests that, like MerR family members, GrlA possibly activate transcription by provoking a change in the conformation of the target DNA that makes the promoter more attractive to RNAP. However, an alternative mechanism cannot be ruled out. One possibility is that GrlA directly interacts

with RNAP before interacting with its DNA target. However, pull down assays in which proteins bound to RNAP are resolved failed to find evidence for any "off-DNA" direct interaction between GrlA and RNAP.

Although further investigations are necessary, the available evidence suggests that the function of the GrlA is unusual. Whilst previous studies suggest that GrlA activates the *LEE1* promoter mainly by negating H-NS mediated repression, my work shows that GrlA can directly interact with a target promoter and its mode of action is unusual. Instead of following the usual role of an activator, by recruiting RNAP by binding upstream or overlapping the promoter -35 element, GrlA activates the *LEE1* promoter by interacting with its non-optimal spacer sequence, possibly following the role of many of the MerR family activators. However, it is worth mentioning that there is no sequence similarity between GrlA and any of the MerR family activators. Whole genome analysis to identify GrlA targets and investigation of those regulons might provide more precise knowledge on the mode of action of GrlA.

The *LEE1* promoter regulatory region is very complex. This is the target for many regulators that switch on and off LEE gene expression via *ler* gene expression. Identification of the cryptic promoter P1A overlapped with the *LEE1* P1 described in the Chapter 4 underscores its complexity. Weakening or inactivating the functional P1 promoter unmasks the P1A promoter suggesting that two promoters compete each other for RNAP and the P1 promoter wins in competition over P1A since it does contain a consensus -35 element. Though the regulatory role of the P1A promoter in *ler* gene expression is still unknown, investigation of this promoter gives some observations of general interest, presumably applicable for other promoters. My most important observation is that sequestration of RNAP by inactivated P1 promoter consensus -35 element prevents the cryptic promoter from functioning. This suggests that a consensus -35 elements make the P1 promoter more attractive to RNAP compared with the P1A ensuring the dominancy of the P1 over the P1A. Residual effects of an inactivated promoter have been reported for the *gal* P2 promoter (Johnston *et al*., 1987), where an inactivated *gal* P2 promoter created by an alteration of a base at the extended -10 element can make contact with RNAP.

A regulatory role for the P1A promoter cannot be ruled out. It may sense different environmental signal(s). Note that both the *E. coli gal* and *lac* promoters contain upstream overlapping promoters. Expression from the *gal* or *lac* P1 promoters is up-regulated by CRP (Musso *et al*., 1977; Malan and McClure, 1984). In each case, a weak overlapping promoter overlaps the target site for CRP and the cryptic promoter activity is suppressed by the activator. In the case of *LEE*, the P1A cryptic promoter -10 element region overlaps the recognition site for GrlA. However, GrlA does not have any influence in the expression from the cryptic promoter (data not shown). Possibly other transcription factors particularly EHEC specific transcription factor(s) act on this promoter in response to environmental signals.

Another important finding of my work is the identification of a short ORF in the ~170 bases long leader sequence of the *LEE1* operon described in Chapter 5. Interestingly, translation of this mini-gene is important for the optimal expression of the downstream *ler* gene. However, the precise functional mechanism needs to be addressed. The positive role of the mini-gene on the expression of the *ler* gene implies that translation of the mini-gene could stabilize the mRNA. It is well known that the bound ribosome can affect the mRNA stability by functioning as a barrier for mRNA decay (Cole and Nomura, 1986; Nilsson *et al*., 1987; Petersen, 1992; Rauhut and Klug, 1999; Arraiano *et al*., 2010). Alternatively, transcription of the mini-gene could help RNAP to bypass termination sites ensuring full length transcript as described by Artsimovitch (2010).

It is thought that Ler acts as an anti-silencer by negating H-NS mediated repression of most of the *LEE* operons. It overcomes H-NS repression as it has higher affinity for DNA than H-NS and functions in multimerized form (Yerushalmi *et al*., 2008; Mellies *et al*., 2011). The repressive role of overexpressed Ler in the expression from *LEE1* operon regulatory described in Chapter 6 is unusual. However, autorepression of the Ler may maintain a steady-state concentration of the Ler, which is sufficient to activate other LEE genes. In contrast, Ler functions as an activator in *LEE2* expression and it can only function indirectly by negating repressing elements since it can not activate the expression from a *LEE2* promoter fragment that contains an upstream sequence up to position -125 with respect to the transcription start site. In order to know the role of Ler in the expression from both *LEE1* and *LEE2* operon regulatory region effectively, more study is necessary. Moreover, study of its other targets

need to be considered. Genome wide analysis found that Ler regulates a number of non-LEE genes. Some of these are are associated with attaching and effacing phenotype whereas many are found in the non-pathogenic *E. coli* K-12 strain (Abe *et al*., 2008).

In summary, the studies communicated here, indicate that the expression of *ler* is subject to complex regulatory systems. The presence of an overlapping cryptic promoter and a translated mini-ORF in the leader region indicate that the *ler* regulatory system is much more complex than was thought before. GrlA is part of this regulatory system and plays an important role in the induction of the expression from the *LEE1* promoter by an unusual mechanism. Moreover, over-expression of Ler autorepresses its own transcription. On the other hand, H-NS mediated repression of the *LEE2* promoter is overcome partially by Ler (Figure 7.1). However, though my laboratory-based *in vitro* system has produced several interesting observations concerning the LEE gene regulation, much more research considering the consequences of *in vivo* conditions needs to be done to get precise knowledge on these aspects.

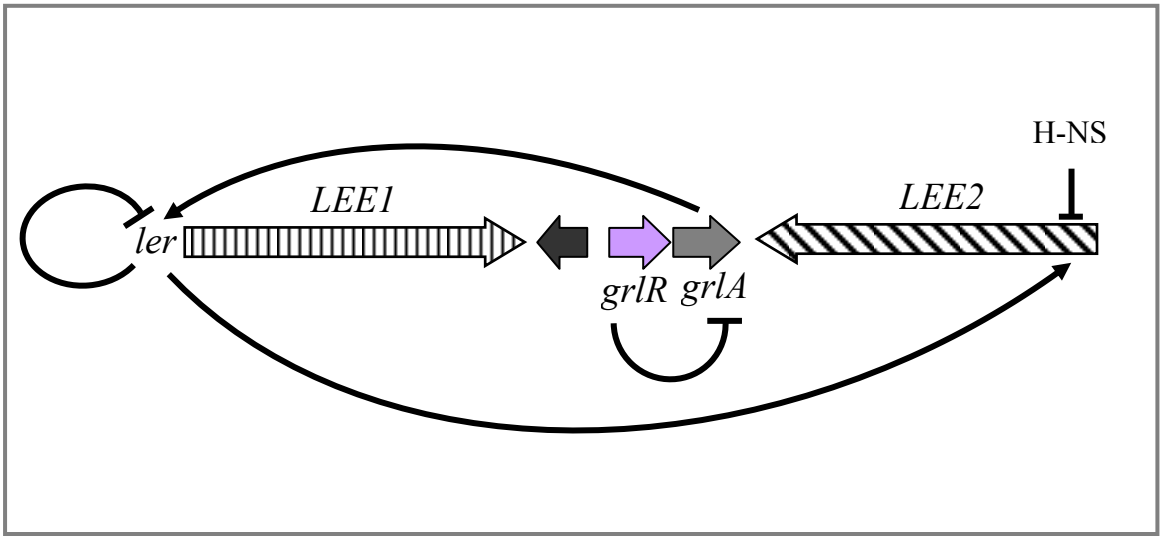


Figure 7.1. Model for the regulation of the *LEE1* and *LEE2* operons by Ler and GrlR/GrlA.

GrlA activates expression from the *LEE1* operon regulatory region ensuring induction of the *ler* expression. Ler then increases the expression from the *LEE2* promoter by negating the H-NS mediated repression whilst overexpressed Ler autorepresses its own expression. GrlR interacts with GrlA and negates the capability of the GrlA to activate expression from the *LEE1* promoter.

References

Abe, A., De Grado, M., Pfuetzner, R.A., Sánchez-SanMartín, C., DeVinney, R., Puente, J.L., Strynadka, N.C., and Finlay, B.B. (1999) Enteropathogenic *Escherichia coli* translocated intimin receptor, Tir, requires a specific chaperone for stable secretion. *Mol Microbiol* **33**: 1162-1175.

Abe, H., Tatsuno, I., Tobe, T., Okutani, A., and Sasakawa, C. (2002) Bicarbonate ion stimulates the expression of locus of enterocyte effacement-encoded genes in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **70**: 3500-3509.

Abe, H., Miyahara, A., Oshima, T., Tashiro, K., Ogura, Y., Kuhara, S., Ogasawara, N., Hayashi, T., and Tobe, T. (2008) Global regulation by horizontally transferred regulators establishes the pathogenicity of *Escherichia coli*. *DNA Res* **15**: 25-38.

Adaikkalam, V., and Swarup, S. (2002) Molecular characterization of an operon, *cueAR*, encoding a putative P1-type ATPase and a MerR-type regulatory protein involved in copper homeostasis in *Pseudomonas putida*. *Microbiology* **148**: 2857-2867.

Ahmed, M., Lyass, L., Markham, P.N., Taylor, S.S., Vazquezlaslop, N., and Neyfakh, A.A. (1995) Two highly similar multidrug transporters of *Bacillus subtilis* whose expression is differentially regulated. *J Bacteriol* **177**: 3904-3910.

Aiba, H., Adhya, S., and de Crombrugghe, B. (1981) Evidence for two functional *gal* promoters in intact *Escherichia coli* cells. *J Biol Chem* **256**: 11905-11910.

Aizawa, S.I. (2001) Bacterial flagella and type III secretion systems. *FEMS Mirobiol Lett* **202**: 157-164.

Akashi, S., Joh, K., Tsuji, A., Ito, H., Hoshi, H., Hayakawa, T., Ihara, J., Abe, T., Hatori, M., Mori, T., *et al*. (1994) A severe outbreak of haemorrhagic colitis and haemolytic uraemic syndrome associated with *Escherichia coli* O157:H7 in Japan. *Eur J Pediatr* **153**: 650-655.

Ansari, A.Z., Chael, M.L., and O'Halloran, T.V. (1992) Allosteric underwinding of DNA is a critical step in positive control of transcription by Hg-MerR. *Nature* **355**: 87-89.

Ansari, A.Z., Bradner, J.E., and O'Halloran, T.V. (1995) DNA-bend modulation in a repressor-to- activator switching mechanism. *Nature* **374**: 371-375.

Arold, S.T., Leonard, P.G., Parkinson, G.N., and Ladbury, J.E. (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proc Natl Acad Sci USA* **107**: 15728-15732.

Arraiano, C.M., Andrade, J.M., Domingues, S., Guinote, I.B., Malecki, M., Matos, R.G., Moreira, R.N., Pobre, V., Reis, F.P., Saramago, M., Silva, I.J., and Viegas, S.C. (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**: 883-923.

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

Artsimovitch, I. (2010) A processive riboantiterminator seeks a switch to make biofilms. *Mol Microbiol* **76**: 535-539.

Azam, T.A., and Ishihama, A. (1999) Twelve species of the nucleoid associated protein from *Escherichia coli*. *J Biol Chem* **274**: 33105-33113.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006.0008.

Bahrani-Mougeot, F.K., Buckles, E.L., Lockatell, C.V., Hebel, J.R., Johnson, D.E., Tang, C.M., and Donnenberg, M.S. (2002) Type 1 fimbriae and extracellular polysaccharides are preeminent uropathogenic *Escherichia coli* virulence determinants in the murine urinary tract. *Mol Microbiol* **45**: 1079-1093.

Barba, J., Bustamante, V.H., Flores-Valdez, M.A., Deng, W., Finlay, B.B., and Puente, J.L. (2005) A positive regulatory loop controls expression of the locus of enterocyte effacement-encoded regulators Ler and GrlA. *J Bacteriol* **187**: 7918-7930.

Barrineau, P., Gilbert, P., Jackson, W.J., Jones, C.S., Summers, A.O., and Wisdom, S. (1985) The structure of the *mer* operon. *Basic Life Sci* **30**: 707-718.

Barnard, A., Wolfe, A., and Busby, S. (2004) Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. *Curr Opin Microbiol* **7**: 102-108.

Barne, K.A., Bown, J.A., Busby, S.J.W., and Minchin, S.D. (1997) Region 2.5 of the *Escherichia coli* RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. *EMBO J* **16**: 4034-4040.

Becskei, A., and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* **405**: 590-593.

Bell, B.P., Goldoft, M., Griffin, P.M., Davis, M.A., Gordon, D.C., Tarr, P.I., Bartleson, C.A., Lewis, J.H., Barrett, T.J., Wells, J.G., Baron, R., and Kobayashi, J.A. (1994) Multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers: the Washington experience. *JAMA* **272**: 1349-1353.

Berdichevsky, T., Friedberg, D., Nadler, C., Rokney, A., Oppenheim, A., and Rosenshine, I. (2005) Ler is a negative autoregulator of the *LEE1* operon in enteropathogenic *Escherichia coli*. *J Bacteriol* **187**: 349-357.

Bertin, P., Benhabiles, N., Krin, E., Laurent-Winter, C., Tendeng, C., Turlin, E., Thomas, A., Danchin, A., and Brasseur, R. (1999) The structural and functional organization of H-NS-like proteins is evolutionarily conserved in gram-negative bacteria. *Mol Microbiol* **31**: 319-329.

Bettelheim, K.A. (2003) Non-O157 verotoxin-producing *Escherichia coli*: a problem, paradox, and paradigm. *Exp Biol Med* **228**: 333-344.

Beutin, L., Geier, D., Steinrück, H., Zimmermann, S., and Scheutz, F. (1993) Prevalence and some properties of verotoxin (Shiga-like toxin)-producing *Escherichia coli* in seven different species of healthy domestic animals. *J Clin Microbiol* **31**: 2483-2488.

Bhatt, S., Romeo, T., and Kalman, D. (2011) Honing the message: post-transcriptional and post-translational control in attaching and effacing pathogens. *Trends Microbiol* **19**: 217-224.

Bhavsar, A.P., Guttman, J.A., and Finlay, B.B. (2007) Manipulation of host-cell pathways by bacterial pathogens. *Nature* **449**: 827-834.

Bingham, A.H.A., Ponnambalam, S., Chan, B., and Busby, S. (1986) Mutations that reduce expression from the *P*2 promoter of the *Escherichia coli* galactose operon. *Gene* **41**: 67-74.

Blatter, E.E., Ross, W., Tang, H., Gourse, R.L., and Ebright, R.H. (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78**: 889-896.

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

Bleves, S., Viarre, V., Salacha, R., Michel, G.P., Filloux, A., and Voulhoux, R. (2010) Protein secretion systems in *Pseudomonas aeruginosa*: a wealth of pathogenic weapons. *Int J Med Microbiol* **300**: 534-543.

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

Blocker, A., Komoriya, K., and Aizawa, S. (2003) Type III secretion systems and bacterial flagella: insights into their function from structural similarities. *Proc Natl Acad Sci USA* **100**: 3027-3030.

Blomberg, P., Nordstrom, K., and Wagner, E.G. (1992) Replication control of plasmid R1: RepA synthesis is regulated by CopA RNA through inhibition of leader peptide translation. *EMBO J* **11**: 2675-2683.

Bodey, G.P., Bolivar, R., Fainstein, V., and Jadeja, L. (1983) Infections caused by *Pseudomonas aeruginosa*. *Rev Infect Dis* **5**: 279-313.

Borremans, B., Hobman, J.L., Provoost, A., Brown, N.L., and Van der Lelie, D. (2001) Cloning and functional analysis of the *pbr* lead resistance determinant of *Ralstonia metallidurans* CH34. *J Bacteriol* **183**: 5651-5658.

Borukhov, S., and Nudler, E. (2008) RNA polymerase: the vehicle of transcription. *Trends Microbiol* **16**: 126-134.

Breaker, R.R. (2010) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **2**: a003566.

Brennan, R.G., and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**: 125-133.

Brown, L., and Elliott, T. (1997) Mutations that increase expression of the *rpoS* gene and decrease its dependence on *hfq* function in *Salmonella typhimurium*. *J Bacteriol* **179**: 656-662.

Brown, N.L., Ford, S.J., Pridmore, R.D., and Fritzinger, D.C. (1983) Nucleotide sequence of a gene from the *Pseudomonas transposon* Tn*501* encoding mercuric reductase. *Biochemistry* **22**: 4089-4095.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., and Hobman, J.L. (2003) The MerR family of transcriptional regulators. *FEMS Microbiol Rev* **27**: 145-163.

Browning, D.F., and Busby, S.J. (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2**: 57-65.

Browning, D.F, Savery, N., Kolb, A., and Busby, S. (2009) Assays for transcription factor activity. In *Methods in Molecular Biology: DNA-Protein interactions.* 3rd Edition. Moss, T., and Leblanc, B. (eds). Heidelberg: Springer Science, volume 543, pp 369-387.

Browning, D.F., Grainger, D.C., and Busby, S.J.W. (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol* **13**: 773-780.

Brutinel, E.D., and Yahr, T.L. (2008) Control of gene expression by type III secretory activity. *Curr Opin Microbiol* **11**: 128-133.

Buchet, A., Nasser, W., Eichler, K., and Mandrand-Berthelot, M.A. (1999) Positive co-regulation of the *Escherichia coli* carnitine pathway *cai* and *fix* operons by CRP and the CaiF activator. *Mol Microbiol* **34**: 562-575.

Burns, S.M., and Hull, S.I. (1999) Loss of resistance to ingestion and phagocytic killing by O(-) and K(-) mutants of a uropathogenic *Escherichia coli* O75:K5 strain. *Infect Immun* **67**: 3757-3762.

Burr, S.E., Stuber, K., and Frey, J. (2003) The ADP-ribosylating toxin, AexT, from *Aeromonas salmonicida* subsp. *salmonicida* is translocated via a type III secretion pathway. *J Bacteriol* **185**: 6583-6591.

Burr, T., Mitchell, J., Kolb, A., Minchin, S., and Busby, S. (2000) DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Res* **28**: 1864-1870.

Busby, S., Aiba, H., and de Crombrugghe, B. (1982) Mutations in the *Escherichia coli* operon that define two promoters and the binding site of the cyclic AMP receptor protein. *J Mol Biol* **154**: 211-227.

Busby, S., and Ebright, R.H. (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79**: 743-746.

Busby, S., and Ebright, R.H. (1999) Transcription activation by catabolite activator protein (CAP). *J Mol Biol* **293**: 199-213.

Busby, S., Kolb, A., and Buc, H. (2009) Where it all begins: an overview of promoter recognition and open complex formation. In *RNA Polymerases as Molecular Motors* (Buc, H. & Strick, T., eds), RSC Publishing, Cambridge, chapter 1, pp 13-17.

Bustamante, V.H., Santana, F.J., Calva, E., and Puente, J.L. (2001) Transcriptional regulation of type III secretion genes in enteropathogenic *Escherichia coli*: Ler antagonizes H-NS-dependent repression. *Mol Microbiol* **39**: 664-678.

Butala, M., Busby, S.J., and D.J. (2009) DNA sampling: a method for probing protein binding at specific loci on bacterial chromosomes. *Nucleic Acids Res* **37**: e37.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell* **9**: 527-539.

Campellone, K.G., Robbins, D., and Leong, J.M. (2004) EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Dev Cell* **7**: 217-228.

Canny, G.O., and McCormick, B.A. (2008) Bacteria in the intestine, helpful residents or enemies from within? *Infect Immune* **76**: 3360-3373.

Carter, A.O., Borczyk, A.A., Carlson, J.A., Harvey, B., Hockin, J.C., Karmali, M.A., Krishnan, C., Korn, D.A., and Lior, H. (1987) A severe outbreak of *Escherichia coli* O157:H7-associated hemorrhagic colitis in a nursing home. *N Engl J Med* **317**: 1496-500.

Casadaban, M.J., and Cohen, S.N. (1980) Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol* **138**: 179-207.

Caslake, L.F., Ashraf, S.I., and Summers, A.O. (1997) Mutations in the alpha and sigma-70 subunits of RNA polymerase affect expression of the *mer* operon. *J Bacteriol* **179**: 1787-1795.

Castillo, A., Eguiarte, L.E., and Souza, V. (2005) A genomic population genetics analysis of the pathogenic enterocyte effacement island in *Escherichia coli*: the search for the unit of selection. *Proc Natl Acad Sci USA* **102**: 1542-1547.

Chang, A.C., and Cohen, S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol* **134**: 1141-1156.

Chalut, C., and Egly, J.M. (1995) AUC is used as a start codon in *Escherichia coli*. *Gene* **156**: 43-45.

Chatterji, D., and Ojha, A.K. (2001) Revisiting the stringent response, ppGpp and starvation signaling. *Curr Opin Microbiol* **4**: 160-165.

Chaudhuri, R.R., Khan, A.M., and Pallen, M.J. (2004) *coli*BASE: an online database for *Escherichia coli*, *Shigella* and *Salmonella* comparative genomics. *Nucleic Acids Res* **32**: D296-D299.

Chaudhuri, R.R., Sebaihia, M., Hobman, J.L., Webber, M.A., Leyton, D.L., Goldberg, M.D., Cunningham, A.F., Scott-Tucker, A., Ferguson, P.R., Thomas, C.M., Frankel, G., Tang, C.M., Dudley, E.G., Roberts, I.S., Rasko, D.A., Pallen, M.J., Parkhill, J., Nataro, J.P., Thomson, N.R., and Henderson, I.R. (2010) Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. *PLoS One* **5**: e8801.

Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res* **22**: 4953-4957.

Chen, H.D., and Frankel, G. (2005) Enteropathogenic *Escherichia coli*: unravelling pathogenesis. *FEMS Microbiol Rev* **29**: 83-98.

Chismon, D.L., Browning, D.F., Farrant, G.K., and Busby, S.J. (2010) Unusual organisation, complexity and redundancy at the *Escherichia coli hcp-hcr* operon promoter. *Biochem J* **430**: 61-68.

Choy, H., and Adhya, S. (1996) Negative control. In *Escherichia coli* and *Salmonella*. Neidhardt, F.C. (ed.). Washington DC: American Society for Microbiology Press, pp. 1287-1299.

Clarke, S.C., Haigh, R.D., Freestone, P.P.E., and Williams, P.H. (2003) Virulence of enteropathogenic *Escherichia coli*, a global pathogen. *Clin Microbiol Rev* **16**: 365-378.

Clarke, M.B., and Sperandio, V. (2005) Transcriptional regulation of *flhDC* by QseBC and sigma28 (FliA) in EHEC. *Mol Microbiol* **57**: 1734-1749.

Cody, S.H., Glynn, M.K., Farrar, J.A., Cairns, K.L., Griffin, P.M., Kobayashi, J., Fyfe, M., Hoffman, R., King, A.S., Lewis, J.H., Swaminathan, B., Bryant, R.G., and Vugia, D.J. (1999) An outbreak of *Escherichia coli* O157:H7 infection from unpasteurized commercial apple juice. *Ann Intern Med* **130**: 202-209.

Cole, C., Barber, J.D., and Barton, G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res*. **36**: W197-W201.

Cole, J.R., and Nomura, M. (1986) Changes in the half-life of ribosomal protein messenger RNA caused by translational repression. *J Mol Biol* **188**: 383-392.

Cornelis, G.R. (2002) The Yersinia Ysc-Yop 'type III' weaponry. *Nat Rev Mol Cell Biol* **3**: 742-752.

Corrigan, J.J., and Boineau, F.G. (2001) Hemolytic-uremic syndrome. *Pediatr Rev* **22**: 365-369.

Creasey, E.A., Delahay, R.M., Daniell, S.J., and Frankel, G. (2003) Yeast two-hybrid system survey of interactions between LEE-encoded proteins of enteropathogenic *Escherichia coli*. *Microbiology* **149**: 2093-2106.

Croxen, M.A., and Finlay, B.B. (2010) Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* **8**: 26-38.

Crump, J.A., Sulka, A.C., Langer, A.J., Schaben, C., Crielly, A.S., Gage, R., Baysinger, M., Moll, M., Withers, G., Toney, D.M., Hunter, S.B., Hoekstra, R.M., Wong, S.K., Griffin, P.M., and Van Gilder, T.J. (2002) An outbreak of *Escherichia coli* O157:H7 infections among visitors to a dairy farm. *N Engl J Med* **347**: 555-560.

Dahan, S., Wiles, S., La Ragione, R.M., Best, A., Woodward, M.J., Stevens, M.P., Shaw, R.K., Chong, Y., Knutton, S., Phillips, A., and Frankel, G. (2005) EspJ is a prophage-carried type III effector protein of attaching and effacing pathogens that modulates infection dynamics. *Infect Immun* **73**: 679-686.

Dame, R.T., Luijsterburg, M.S., Krin, E., Bertin, P.N., Wagner, R., and Wuite, G.J. (2005) DNA bridging: a property shared among H-NS-like proteins. *J Bacteriol* **187**: 1845-1848.

Daniell, S.J., Takahashi, N., Wilson, R., Friedberg, D., Rosenshine, I., Booy, F.P., Shaw, R.K., Knutton, S., Frankel, G., and Aizawa, S. (2001) The filamentous type III secretion translocon of enteropathogenic *Escherichia coli*. *Cell Microbiol* **3**: 865-871.

Daniell, S.J., Kocsis, E., Morris, E., Knutton, S., Booy, F.P., and Frankel, G. (2003) 3D structure of EspA filaments from enteropathogenic *Escherichia coli*. *Mol Microbiol* **49**: 301-308.

Darnton, N.C., Turner, L., Rojevsky, S., and Berg, H.C. (2007) On torque and tumbling in swimming *Escherichia coli*. *J Bacteriol* **189**: 1756-64.

Darst, S. (2009) A structural atlas of RNAP. In *RNA polymerases as molecular motors*. (Buc, H. and Strick, T., eds) Royal Society of Chemistry Publishing, Cambridge, UK, pp. 1-10.

Darwin, K.H., and Miller, V.L. (2001) Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. *EMBO J* **20**: 1850-1862.

Dean, P., Scott, J.A., Knox, A.A., Quitard, S., Watkins, N.J., and Kenny, B. (2010) The enteropathogenic *E. coli* effector EspF targets and disrupts the nucleolus by a process regulated by mitochondrial dysfunction. *PLoS Pathog* **6**: e1000961.

Deane, J.E., Abrusci, P., Johnson, S., and Lea, S.M. (2010) Timing is everything: the regulation of type III secretion. *Cell Mol Life Sci* **67**: 1065-1075.

deHaseth, P.L., and Helmann, J.D. (1995) Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. *Mol Microbiol* **16**: 817-824.

Deng, W., Li, Y., Vallance, B.A., and Finlay, B.B. (2001) Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. *Infect Immun* **69**: 6323-6335.

Deng, W., Puente, J.L., Gruenheid, S., Li, Y., Vallance, B.A., Vazquez, A., Barba, J., Ibarra, J.A., O'Donnell, P., Metalnikov, P., Ashman, K., Lee, S., Goode, D., Pawson, T., and Finlay, B.B. (2004) Dissecting virulence: systematic and functional analyses of a pathogenicity island. *Proc Natl Acad Sci USA* **101**: 3597-3602.

Deng, W., Li, Y., Hardwidge, P.R., Frey, E.A., Pfuetzner, R.A., Lee, S., Gruenheid, S., Strynakda, N.C., Puente, J.L., and Finlay, B.B. (2005) Regulation of type III secretion hierarchy of translocators and effectors in attaching and effacing bacterial pathogens. *Infect Immun* **73**: 2135-2146.

de Smit, M.H., and Van Duin, J. (1994a) Translational initiation on structured messengers. Another role for the Shine-Dalgarno interaction. *J Mol Biol* **235**: 173-184.

de Smit, M.H., and van Duin, J. (1994b) Control of translation by mRNA secondary structure in *Escherichia coli*. A quantitative analysis of literature data. *J Mol Biol* **244**: 144-150.

Deutscher, M.P. (1993) Promiscuous exoribonucleases of *Escherichia coli*. *J Bacteriol* **175**: 4577-4583.

Dodd, I.B., and Egan, J.B. (1990) Improved detection of helix-turn-helix DNA-binding motifs in protein sequences. *Nucleic Acids Res* **18**: 5019-5026.

Dombroski, A.J., Walter, W.A., and Gross, C.A. (1993) Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* **7**: 2446-2455.

Dorman, C.J., Hinton, J C.D., and Free, A. (1999) Domain organization and oligomerization among H-NS-like nucleoid-associated proteins in bacteria. *Trends Microbiol* **7**: 124-128.

Dorman, C.J. (2004) H-NS: a universal regulator for a dynamic genome. *Nat Rev Microbiol* **2**: 391-400.

Dorman, C.J. (2007) H-NS, the genome sentinel. *Nat Rev Microbiol* **5**: 157-161.

Dorman, C.J., and Kane, K.A. (2009) DNA bridging and antibridging: a role for bacterial nucleoid-associated proteins in regulating the expression of laterally acquired genes. *FEMS Microbiol Rev* **33**: 587-592.

Dove, S.L., Darst, S.A., and Hochschild, A. (2003) Region 4 of sigma as a target for transcription regulation. *Mol Microbiol* **48**: 863-874.

Draper, D.E., Gluick, T.C., and Schlax, P.J. (1998) Pseudoknots, RNA folding and translational regulation. In *RNA Structure and Function.* Simons, R.W., and Grunberg-Manago, M. (eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 415-436.

Dreyfus, M., Kotlarz, D., and Busby, S. (1985) Point mutations that affect translation initiation in the *Escherichia coli gal*E gene. *J Mol Biol* **182**: 411-417.

Dundas, S., Todd, W.T., Stewart, A.I., Murdoch, P.S., Chaudhuri, A.K., and Hutchinson, S.J. (2001) The central scotland *Escherichia coli* O157:H7 outbreak: risk factors for the hemolytic uremic syndrome and death among hospitalized patients. *Clin Infect Dis* **33**: 923-931.

Duran, N., and Menck, C.F. (2001) *Chromobacterium violaceum*: a review of pharmacological and industrial perspectives. *Crit Rev Microbiol* **27**: 201-222.

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

Ebright, R.H., and Busby, S. (1995) The *Escherichia coli* RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**: 197-203.

Ebright, R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* **304**: 687-698.

Eckert, K.A., and Kunkel, T.A. (1991) DNA polymerase fidelity and the polymerase chain reaction. *Genome Res* **1**: 17-24.

Edqvist, P.J., Olsson, J., Lavander, M., Sundberg, L., Forsberg, A., Wolf-Watz, H., and Lloyd, S.A. (2003) YscP and YscU regulate substrate specificity of the *Yersinia* type III secretion system. *J Bacteriol* **185**: 2259-2266.

Edwards, P.R., and Ewing, W.H. (1972) *Identification of Enterobacteriaceae*, 3rd ed. Burgess Publishing Co., Minneapolis, MN.

Eichler, K., Buchet, A., Lemke, R., Kleber, H.P., and Mandrand-Berthelot, M.A. (1996**)** Identification and characterization of the *caiF* gene encoding a potential transcriptional activator of carnitine metabolism in *Escherichia coli*. *J Bacteriol* **178**: 1248-1257.

Elliott, S.J., Wainwright, L.A., McDaniel, T.K., Jarvis, K.G., Deng, Y.K., Lai, L.C., McNamara, B.P., Donnenberg, M.S., and Kaper, J.B. (1998) The complete sequence of the locus of enterocyte effacement (LEE) from enteropathogenic *Escherichia coli* E2348/69. *Mol Microbiol* **28**: 1-4.

Elliott, S.J., Hutcheson, S.W., Dubois, M.S., Mellies, J.L., Wainwright, L.A., Batchelor, M., Frankel, G., Knutton, S., and Kaper, J.B. (1999) Identification of CesT, a chaperone

for the type III secretion of Tir in enteropathogenic *Escherichia coli*. *Mol Microbiol* **33**: 1176-1189.

Elliott, S.J., Sperandio, V., Girón, J.A., Shin, S., Mellies, J.L., Wainwright, L.A., Hutcheson, S.W., McDaniel, T.K., and Kaper, J.B. (2000) The locus of enterocyte effacement (LEE)-encoded regulator controls expression of both LEE- and non-LEE-encoded virulence factors in enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect Immun* **68**: 6115-6126.

Ellison, D.W., and Miller, V.L. (2006) Regulation of virulence by members of the MarR/SlyA family. *Curr Opin Microbiol* **9**: 153-159.

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

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998) Identification of an UP element consensus sequence for bacterial promoters. *Proc Natl Acad Sci USA* **95**: 9761-9766.

Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W., Niu, W., Ebright, R.H., and Gourse, R.L. (1999) Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes Dev* **13**: 2134-2147.

Feklistov, A., and Darst, S.A. (2009) Promoter recognition by bacterial alternative sigma factors: the price of high selectivity? *Genes Dev* **23**: 2371-2375.

Fenton, M.S., Lee, S.J., and Gralla, J.D. (2000) *Escherichia coli* promoter opening and -10 recognition: mutational analysis of σ^{70}. *EMBO J* **19**: 1130-1137.

Ferens, W.A., and Hovde, C.J. (2011) *Escherichia coli* O157:H7: Animal reservoir and sources of human infection. *Foodborne Pathog Dis* **8**: 465-487.

Forsman, M., Sandström, G., and Jaurin, B. (1990) Identification of *Francisella* species and discrimination of type A and type B strains of *F. tularensis* by 16S rRNA analysis. *Appl Environ Microbiol* **56**: 949-955.

Francis, M.S., Lloyd, S.A., and Wolf-Watz, H. (2001) The type III secretion chaperone LcrH co-operates with YopD to establish a negative, regulatory loop for control of Yop synthesis in *Yersinia pseudotuberculosis*. *Mol Microbiol* **42**: 1075-1093.

Frankel, G., Phillips, A.D., Rosenshine, I., Dougan, G., Kaper, J.B., and Knutton, S. (1998) Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol Microbiol* **5**: 911-921.

Frantz, B., and O'Halloran, T.V. (1990) DNA distortion accompanies transcriptional activation by the metal-responsive gene-regulatory protein MerR. *Biochemistry* **29**: 4747-4751.

Friedberg, D., Umanski, T., Fang, Y., and Rosenshine I. (1999) Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. *Mol Microbiol* **34**: 941-952.

Friedmann, H.C. (2004) From "butyribacterium" to "*E. coli*": an essay on unity in biochemistry. *Perspect Biol Med* **47**: 47-66.

Friedmann, H.C. (2006) Escherich and *Escherichia*. *Adv Appl Microbiol* **60**: 133-196.

Gaal, T., Bartlett, M.S., Ross, W., Turnbough, C.L. Jr, and Gourse, R.L. (1997) Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**: 2092-2097.

Gardner, J.F. (1979) Regulation of the threonine operon: tandem threonine and isoleucine codons in the control region and translational control of transcription termination. *Proc Natl Acad Sci USA* **76**: 1706-1710.

Garmendia, J., Phillips, A.D., Carlier, M.F., Chong, Y., Schüller, S., Marches, O., Dahan, S., Oswald, E., Shaw, R.K., Knutton, S., and Frankel, G. (2004) TccP is an enterohaemorrhagic *Escherichia coli* O157:H7 type III effector protein that couples Tir to the actin-cytoskeleton. *Cell Microbiol* **6**: 1167-1183.

Garmendia, J., Frankel, G., and Crepin, V.F. (2005) Enteropathogenic and enterohemorrhagic *Escherichia coli* infections: translocation, translocation, translocation. *Infect Immun* **73**: 2573-2585.

Gasser, C., Gautier, E., Steck, A., Siebenmann, R.E., and Oechslin, R. (1955) Hemolytic -uremic syndrome: bilateral necrosis of the renal cortex in acute acquired hemolytic anemia (in German). *Schweiz Med Wochenschr* **85**: 905-909.

Gauthier, A., Puente, J.L., and Finlay, B.B. (2003) Secretin of the enteropathogenic *Escherichia coli* type III secretion system requires components of the type III apparatus for assembly and localization. *Infect Immun* **71**: 3310-3319.

Gavini, N., and Pulakat, L. (1991) Role of translation of the *pheA* leader peptide coding region in attenuation regulation of the *Escherichia coli pheA* gene. *J Bacteriol* **173**: 4904-4907.

Geissmann, T.A., and Touati, D. (2004) Hfq, a new chaperoning role: binding to messenger RNA determines access for small RNA regulator. *EMBO J* **23**: 396-405.

Gemmill, R.M., Wessler, S.R., Keller, E.B., and Calvo, J.M. (1979) *leu* operon of *Salmonella typhimurium* is controlled by an attenuation mechanism. *Proc Natl Acad Sci USA* ***76***: 4941-4945.

Ghosh, P., Ishihama, A., and Chatterji, D. (2001) *Escherichia coli* RNA polymerase subunit ω and its N-terminal domain bind full-length ß' to facilitate incorporation into the α2β subassembly. *Eur J Biochem* **268**: 4621-4627.

Ghosh, P., Ramakrishnan, C., and Chatterji, D. (2003) Inter-subunit recognition and manifestation of segmental mobility in *Escherichia coli* RNA polymerase: a case study with ω-β′ interaction. *Biophys chem* **103**: 223-237.

Godessart, N., Munoa, F.J., Regue, M., and Juarez, A. (1988) Chromosomal mutations that increase the production of a plasmid-encoded haemolysin in *Escherichia coli*. *J Gen Microbiol* **134**: 2779-2787.

Goldberg, M.D., Johnson, M., Hinton, J.C., and Williams, P.H. (2001) Role of the nucleoid -associated protein Fis in the regulation of virulence properties of enteropathogenic *Escherichia coli*. *Mol Microbiol* **41**: 549-559.

Gourse, R.L., Ross, W., and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol* **37**: 687-695.

Grainger, D.C., Hurd, D., Goldberg, M.D., and Busby, S.J. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res* **34**: 4642-4652.

Grainger, D.C., Goldberg, M.D., Lee, D.J., and Busby, S.J. (2008) Selective repression by Fis and H-NS at the *Escherichia coli dps* promoter. *Mol Microbiol* **68**: 1366-1377.

Gruenheid, S., Sekirov, I., Thomas, N.A., Deng, W., O'Donnell, P., Goode, D., Li, Y., Frey, E.A., Brown, N.F., and other authors (2004) Identification and characterization of NleA, a non-LEE-encoded type III translocated virulence factor of enterohaemorrhagic *Escherichia coli* O157 : H7. *Mol Microbiol* **51**: 1233-1249.

Gualerzi, C.O., and Pon, C.L. (1990) Initiation of mRNA translation in prokaryotes. *Biochemistry* **29**: 5881-5889.

Gurvich, O.L., Näsvall, S.J., Baranov, P.V., Björk, G. R., and Atkins, J.F. (2010) Two groups of phenylalanine biosynthetic operon leader peptides genes: a high level of apparently incidental frameshifting in decoding *Escherichia coli pheL*. *Nucleic Acids Res* **39**: 3079-3092.

Guzman, L.M., Belin, D., Carson, M.J., and Beckwith, J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J Bacteriol* **177**: 4121-4130.

Haack, K.R., Robinson, C.L., Miller, K.J., Fowlkes, J.W., and Mellies, J.L. (2003) Interaction of Ler at the *LEE5* (*tir*) operon of enteropathogenic *Escherichia coli*. *Infect Immun* **71**: 384-392.

Habdas, B.J., Smart, J., Kaper, J.B., and Sperandio, V. (2010) The LysR-type transcriptional regulator QseD alters type three secretion in enterohemorrhagic *Escherichia coli* and motility in K-12 *Escherichia coli*. *J bacterial* **192**: 3699-3712.

Hacker, J., Blum-Oehler, G., Mühldorfer, I., and Tschäpe, H. (1997) Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol Microbiol* **23**: 1089-1097.

Hacker, J., Kaper, J.B. (2000) Pathogenicity islands and the evolution of microbes. *Annu Rev Microbiol* **54**: 641-679.

Hansen, A.M., and Kaper, J.B. (2009) Hfq affects the expression of the LEE pathogenicity island in enterohaemorrhagic *Escherichia coli*. *Mol Microbiol* **73**: 446-465.

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**: 1069-1082.

Helmann, J.D., and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem* **57**: 839-872.

Hérias, M.V., Midtvedt, T., Hanson, L.A., and Wold, A.E. (1997) *Escherichia coli* K5 capsule expression enhances colonization of the large intestine in the gnotobiotic rat. *Infect Immun* **65**: 531-536.

Herring, C.D., Glasner, J.D., and Blattner, F.R. (2003) Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene* **311**: 153-163.

Hicks, S., Frankel, G., Kaper, J.B., Dougan, G., and Phillips, A.D. (1998) Role of intimin and bundle-forming pili in enteropathogenic *Escherichia coli* adhesion to pediatric intestinal tissue *in vitro*. *Infect Immun* **66**: 1570-1578.

Hidalgo, E., and Demple, B. (1997) Spacing of promoter elements regulates the basal expression of the soxS gene and converts SoxR from a transcriptional activator into a repressor. *EMBO J* **16**: 1056-1065.

Higami, S., Nishimoto, K., Kawamura, T., Tsuruhara, T., Isshiki, G., and Ookita, A. (1998) Retrospective analysis of the relationship between HUS incidence and antibiotics among patients with *Escherichia coli* O157 enterocolitis in the Sakai outbreak. *Kansenshogaku Zasshi* **72**: 266-272.

Higgins, C.F., Peltz, S.W., and Jacobson, A. (1992) Turnover of mRNA in prokaryotes and lower eukaryotes. *Curr Opin Genet Dev* **2**: 739-747.

Hilborn, E.D., Mshar, P.A., Fiorentino, T.R., Dembek, Z.F., Barrett, T.J., Howard, R.T., and Cartter, M.L. (2000) An outbreak of *Escherichia coli* O157:H7 infections and haemolytic uraemic syndrome associated with consumption of unpasteurized apple cider. *Epidemiol Infect* **124**: 31-36.

Hobman, J.L. (2007) MerR family transcription activators: similar designs, different specificities. *Mol Microbiol* **63**: 1275-1278.

Hobman, J.L., Penn, C.W., and Pallen, M.J. (2007) Laboratory strains of *Escherichia coli*: model citizens or deceitful delinquents growing old disgracefully? *Mol Microbiol* **64**: 881-885.

Hockin, J.C., and Lior, H. (1987) Haemorrhagic colitis and hemolytic uremic syndrome caused by *Escherichia coli* 0157:H7 in Canada. *Can Dis Weekly Rep* **13**: 203-204.

Holmes, A., Mühlen, S., Roe, A.J., and Dean, P. (2010) The EspF effector, a bacterial pathogen's Swiss army knife. *Infect Immun* **78**: 4445-4453.

Huang, L.H., and Syu, W.J. (2008) GrlA of enterohemorrhagic *Escherichia coli* O157:H7 activates *LEE1* by binding to the promoter region. *J Microbiol Immunol Infect* **41**: 9-16.

Hueck, C.J. (1998) Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* **62**: 379-433.

Huerta, A.M., and Collado-Vides, J. (2003) Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *J Mol Biol* **333**: 261-278.

Huerta, A.M., Francino, M.P., Morett, E., and Collado-Vides, J. (2006) Selection for unequal densities of σ^{70} promoter-like signals in different regions of large bacterial genomes. *PLoS Genet* **2**: e185.

Hughes, K.T., Gillen, K.L., Semon, M.J., and Karlinsey, J.E. (1993) Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator. *Science* **262**: 1277-1280.

Irnov, I., and Winkler, W.C. (2010) A regulatory RNA required for antitermination of biofilm and capsular polysaccharide operons in Bacillales. *Mol Microbiol* **76**: 559-575.

Ishihama, A. (2000) Functional modulation of *Escherichia coli* RNA polymerase. *Annu Rev Microbiol* **54**: 499-518.

Islam, M.S., Bingle, L.E.H., Pallen, M.J., and Busby, S.J.W. (2011) Organization of the *LEE1* operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Mol Microbiol* **79**: 468-483.

Iyoda, S., and Watanabe, H. (2004) Positive effects of multiple *pch* genes on expression of the locus of enterocyte effacement genes and adherence of enterohaemorrhagic *Escharichia coli* O157:H7 to HEp-2 cells. *Microbiology* **150**: 2357-2371.

Iyoda, S., and Watanabe, H. (2005) ClpXP protease controls expression of the type III protein secretion system through regulation of RpoS and GrlR levels in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **187**: 4086-4094.

Iyoda, S., Koizumi, N., Satou, H., Lu, Y., Saitoh, T., Ohnishi, M., and Watanabe, H. (2006) The GrlR-GrlA regulatory system coordinately controls the expression of flagellar and LEE-encoded type III protein secretion systems in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **188**: 5682-5692.

Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* **3**: 318-356.

Jerse, A.E., Yu, J., Tall, B.D., and Kaper, J.B. (1990) A genetic locus of enteropathogenic *Escherichia coli* necessary for the production of attaching and effacing lesions on tissue culture cells. *Proc Natl Acad Sci USA* **87**: 7839-7843.

Jimenez, R., Cruz-Migoni, S.B., Huerta-Saquero, A., Bustamante, V.H., and Puente, J.L. (2010) Molecular characterization of GrlA, a specific positive regulator of *ler* expression in enteropathogenic *Escherichia coli*. *J Bacteriol* **192**: 4627-4642.

Jobichen, C., Li, M., Yerushalmi, G, Tan, Y.W., Mok, Y.K., Rosenshine, I., Leung, K.Y., and Sivaraman, J. (2007) Structure of GrlR and the implication of its EDED motif in mediating the regulation of type III secretion system in EHEC. *PLoS Pathog* **3**: e69.

Jobichen, C., Fernandis, A.Z., Velazquez-Campoy, A., Leung, K.Y., Mok, Y.K., Wenk, M.R., and Sivaraman, J. (2009) Identification and characterization of the lipid-binding property of GrlR, a locus of enterocyte effacement regulator. *Biochem J* **420**: 191-199.

Johannes, L., and Römer, W. (2010) Shiga toxins-from cell biology to biomedical applications. *Nat Rev Microbiol* **8**: 105-116.

Johnson, R.P., Clarke, R.C., Wilson, J.B., Read, S.C., Rahn, K., Renwick, S.A., Sandhu, K.A., Alves, D., Karmali, M.A., Lior, H., McEwen, S.A., Spika, J.S., and Gyles, C.L. (1996) Growing concerns and recent outbreaks involving non-O157:H7 serotypes of verotoxigenic *Escherichia coli*. *J Food Prot* **59**: 1112-1122.

Johnston, F., Ponnambalam, S., and Busby, S. (1987) Binding of *Escherichia coli* RNA polymerase to a promoter carrying mutations that stop transcription initiation. *J Mol Biol* **195**: 745-748.

Johnston, H.M., Barnes, W.M., Chumley, F.G., Bossi, L. and Roth, J.R. (1980) Model for regulation of the histidine operon of *Salmonella*. *Proc Natl Acad Sci USA* **77**: 508-512.

Kanamaru, K., Tatsuno, I., Tobe, T., and Sasakawa, C. (2000) SdiA, an *Escherichia coli* homologue of quorum-sensing regulators, controls the expression of virulence factors in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol Microbiol* **38**: 805-816.

Kaper, J.B. (1996) Defining EPEC. *Rev Microbiol* **27** (suppl. 1): 130-133.

Kaper, J.B., Nataro, J. P., and Mobley, H.L. (2004) Pathogenic *Escherichia coli*. *Nat Rev Microbiol* **2**: 123-140.

Karmali, M.A., Steele, B.T., Petric, M., and Lim, C. (1983) Sporadic cases of haemolytic-uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* **1**: 619-620.

Karmali, M.A. (1989) Infection by verocytotoxin-producing *Escherichia coli*. *Clin Microbiol Rev* **2**: 15-38.

Karmali, M.A., Gannon, V., and Sargeant, J.M. (2010) Verocytotoxin-producing *Escherichia coli* (VTEC). *Vet Microbiol* **140**: 360-370.

Keene, W.E., Hedberg, K., Herriott, D.E., Hancock, D.D., McKay, R.W., Barrett, T.J., and Fleming, D.W. (1997) A prolonged outbreak of *Escherichia coli* O157:H7 infections caused by commercially distributed raw milk. *J Infect Dis* **176**: 815-818.

Keilty, S., and Rosenberg, M. (1987) Constitutive function of a positively regulated promoter reveals new sequences essential for activity. *J Biol Chem* **262**: 6389-6395.

Kendall, M.M., Rasko, D.A., and Sperandio, V. (2010) The LysR-type regulator QseA regulates both charatcterised and putative virulence genes in enterohaemorrhagic *Escherichia coli* O157:H7. *Mol Microbiol* **76**: 1306-1321.

Kennell, D. (2002) Processing endoribonucleases and mRNA degradation in bacteria. *J Bacteriol* **184**: 4645-4657.

Kenny, B., and Finlay, B.B. (1995) Protein secretion by enteropathogenic *Escherichia coli* is essential for transducing signals to epithelial cells. *Proc Natl Acad Sci USA* **92**: 7991-7995.

Kenny, B., Abe, A., and Stein, M., and Finlay, B.B. (1997a) Enteropathogenic *Escherichia coli* protein secretion is induced in response to conditions similar to those in the gastrointestinal tract. *Infect Immun* **65**: 2606-2612.

Kenny, B., DeVinney, R., Stein, M., Reinscheid, D.J., Frey, E.A., and Finlay, B.B. (1997b) Enteropathogenic *E. coli* (EPEC) transfers its receptor for intimate adherence into mammalian cells. *Cell* **91**: 511-520.

Kenny, B., and Jepson, M. (2000) Targeting of an enteropathogenic *Escherichia coli* (EPEC) effector protein to host mitochondria. *Cell Microbiol* **2**: 579-590.

Kenny, B., Ellis, S., Leard, A.D., Warawa, J., Mellor, H., and Jepson, M.A. (2002) Co-ordinate regulation of distinct host cell signalling pathways by multifunctional enteropathogenic *Escherichia coli* effector molecules. *Mol Microbiol* **44**: 1095-1107.

Knutton, S., Baldwin, T., Williams, P.H., and McNeish, A.S. (1989) Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. *Infect Immun* **57**: 1290-1298.

Knutton, S., Phillips, A.D., Smith, H.R., Gross, R.J., Shaw, R., Watson, P., and Price, E. (1991) Screening for enteropathogenic *Escherichia coli* in infants with diarrhea by the fluorescent-actin staining test. *Infect Immun* **59**: 365-371.

Knutton, S., Rosenshine, I., Pallen, M.J., Nisan, I., Neves, B.C., Bain, C., Wolff, C., Dougan, G., and Frankel, G. (1998) A novel EspA-associated surface organelle of enteropathogenic *Escherichia coli* involved in protein translocation into epithelial cells. *EMBO J* **17**: 2166-2176.

Kolb, A., Kotlarz, D., Kusano, S., and Ishihama, A. (1995) Selectivity of the *Escherichia coli* RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res* **23**: 819-826.

Konkel, M.E., Klena, J.D., Rivera-Amill, V., Monteville, M.R., Biswas, D., Raphael, B., and Mickelson, J. (2004) Secretion of virulence proteins from *Campylobacter jejuni* is dependent on a functional flagellar export apparatus. *J Bacteriol* **186**: 3296-3303.

Konowalchuk, J., Speirs, J.I., and Stavric, S. (1977) Vero response to a cytotoxin of *Escherichia coli*. *Infect Immun* **18**: 775-779.

Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. *Microbiol Rev* **47**: 1-455.

Kuhnke, G., Krause, A., Heibach, C., Gieske, U., Fritz, H.J., and Ehring, R. (1986) The upstream operator of the *Escherichia coli* galactose operon is sufficient for repression of transcription initiated at the cyclic AMP-stimulated promoter. *EMBO J* **5**: 167-173.

Kulesus, R.R., Diaz-Perez, K., Slechta, E.S., Eto, D.S., and Mulvey, M.A. (2008) Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. *Infect Immun* **76**: 3019-3026.

Laaberki, M.H., Janabi, N., Oswald, E., and Repoila, F. (2006) Concert of regulators to switch on LEE expression in enterohemorrhagic *Escherichia coli* O157:H7: interplay between Ler, GrlA, HNS and RpoS. *Int J Med Microbiol* **296**: 197-210.

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

Lease, R.A., and Belfort, M. (2000) Riboregulation by DsrA RNA: trans-actions for global economy. *Mol Microbiol* **38**: 667-672.

Lee, D.J., Busby, S.J., Westblade, L.F., and Chait, B.T. (2008) Affinity isolation and I-DIRT mass spectrometric analysis of the *Escherichia coli* O157:H7 Sakai RNA polymerase complex. *J Bacteriol* **190**: 1284-1289.

Lee, D.J., Bingle, L.E.H, Heurlier, K., Pallen, M.J., Penn, C.W., Busby, S.J.W., and Hobman, J.L. (2009) Gene doctoring a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. *BMC Microbiol* **9**: 252.

Lee, S.H., and Galan, J.E. (2004) *Salmonella* type III secretion-associated chaperones confer secretion pathway specificity. *Mol Microbiol* **51**: 483-495.

Lee, S.W., Glickmann, E., and Cooksey, D.A. (2001) Chromosomal locus for cadmium resistance in *Pseudomonas putida* consisting of a cadmium-transporting ATPase and a MerR family response regulator. *Appl Environ Microbiol* **67**: 1437-1444.

Lemay, J.F., and Lafontaine, D.A. (2007) Core requirements of the adenine riboswitch aptamer for ligand binding. *RNA* **13**: 339-350.

Leung, D.W., Chen, E., and Goeddel, D.V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**: 11-15.

Levine, M.M., DuPont, H.L., Formal, S.B., Hornick, R.B., Takeuchi, A., Gangarosa, E.J., Snyder, M.J., and Libonati, J.P. (1973) Pathogenesis of *Shigella dysenteriae* 1 (Shiga) dysentery. *J Infect Dis* **127**: 261-270.

Levine, M.M. (1987) *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *J Infect Dis* **155**: 377-389.

Li, X.Y., and McClure, W.R. (1998) Characterization of the closed complex intermediate formed during transcription initiation by *Escherichia coli* RNA polymerase. *J Biol Chem* **273**: 23549-23557.

Lio, J.C.W., and Syu, W.J. (2004) Identification of a negative regulator for the pathogenicity island of enterohemorrhagic *Escherichia coli* O157:H7. *J Biomed Sci* **11**: 855-863.

Lisser, S., and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Res* **7**: 1507-1516.

Lithgow, J.K., Haider, F., Roberts, I.S., and Green, J. (2007) Alternate SlyA and H-NS nucleoprotein complexes control *hlyE* expression in *Escherichia coli* K-12. *Mol Microbiol* **66**: 685-698.

Lodato, P.B., and Kaper, J.B. (2009) Post-transcriptional processing of the *LEE4* operon in enterohaemorrhagic *Escherichia coli*. *Mol Microbiol* **71**:273-290.

Lodge, J., Fear, J., Busby, S., Gunasekaran, P., and Kamini, N-R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. *FEMS Microbiol Lett* **74**: 271-276.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992) The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* **174**: 3843-3849.

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

Lukjancenko, O., Wassenaar, T.M., and Ussery, D.W. (2010) Comparison of 61 sequenced *Escherichia coli* genomes. *Microb Ecol* **60**: 708-720.

Madrid, C., Nieto, J.M., Paytubi, S., Falconi, M., Gualerzi, C.O., and Juárez, A. (2002) Temperature- and H-NS-dependent regulation of a plasmid-encoded virulence operon expressing *Escherichia coli* hemolysin. *J Bacteriol* **184**: 5058-5066.

Mahajan, A., Currie, C.G., Mackie, S., Tree, J., McAteer, S., McKendrick, I., McNeilly, T.N., Roe, A., La Ragione, R.M., Woodward, M.J., Gally, D.L., and Smith, D.G.E. (2009) An investigation of the expression and adhesin function of H7 flagella in the interaction of *Escherichia coli* O157 : H7 with bovine intestinal epithelium. *Cell Microbiol* **11**: 121-137.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T., and Gottesman, S. (1998) DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proc Natl Acad Sci USA* **95**: 12462-12467.

Malan, T.P, and McClure, W.R. (1984) Dual promoter control of the *Escherichia coli* lactose operon. *Cell* **39**: 173-180.

Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C., and Breaker, R.R. (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**: 577-586.

Mandal, M., and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**: 451-463.

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

Markham, P.N., Ahmed, M., and Neyfakh, A.A. (1996) The drug-binding activity of the multidrug-responding transcriptional regulator BmrR resides in its C-terminal domain. *J Bacteriol* **178**: 1473-1475.

Marzi, S., Myasnikov, A.G., Serganov, A., Ehresmann, C., Romby, P., Yusupov, M., and Klaholz, B.P. (2007) Structured mRNAs regulate translation initiation by binding to the platform of the ribosome. *Cell* **130**: 1019-1031.

Martin, R.G., Gillette, W.K., Martin, N.I., and Rosner, J.L. (2002) Complex formation between activator and RNA polymerase as the basis for transcriptional activation by MarA and SoxS in *Escherichia coli*. *Mol Microbiol* **43**: 355-370.

Martinez-Antonio, A., and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* **6**: 482-489.

Mathew, R., and Chatterji, D. (2006) The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol* **14**: 450-455.

Mathy, N., Bénard, L., Pellegrini, O., Daou, R., Wen, T., and Condon C. (2007) 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* **129**: 681-692.

McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W., Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K. (2001) Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* **413**: 852-856.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem* **54**: 171-204.

McDaniel, T.K., Jarvis, K.G., Donnenberg, M.S., and Kaper, J.B. (1995) A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc Natl Acad Sci USA* **92**: 1664-1668.

McNamara, B.P., Koutsouris, A., O'Connell, C.B., Nougayrede, J.P., Donnenberg, M.S., and Hecht, G. (2001) Translocated EspF protein from enteropathogenic *Escherichia coli* disrupts host intestinal barrier function. *J Clin Investig* **107**: 621-629.

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002) Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**: 599-614.

Mellies, J.L., Elliott, S.J., Sperandio, V., Donnenberg, M.S., and Kaper, J.B. **(**1999) The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE) -encoded regulator (Ler). *Mol Microbiol* **33**: 296-306.

Mellies, J.L, Barron, A.M.S., and Carmona, A.M. (2007) Enteropathogenic and enterohemorrhagic *Escherichia coli* virulence gene regulation. *Infect Immun* **75**: 4199-4210.

Mellies, J.L., Benison, G., McNitt, W., Mavor, D., Boniface, C., and Larabee, F.J. (2011) Ler of pathogenic *Escherichia coli* forms toroidal protein-DNA complexes. *Microbiology* **157**: 1123-1133.

Melton-Celsa, A.R., and O'Brien, A.D. (1998) Structure, biology and relative toxicity of Shiga toxin family members for cells and animals. In *Escherichia coli O157:H7 and other Shiga toxin producing E. coli strains*. Kaper, J.B., and O'Brien, A.D. (eds). ASM press, Washington DC, USA, pp. 121-128.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juárez, K., Contreras-Moreira, B., Huerta, A.M., Collado-Vides, J., and Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE* **4**: e7526.

Michino, H., Araki. K., Minami, S., Takaya, S., Sakai, N., Miyazaki, M., Ono, A., and Yanagawa, H. (1999) Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. *Am J Epidemiol* **150**: 787-796.

Middlebrook, J.L., and Dorland, R.B. (1984) Bacterial toxins: cellular mechanisms of action. *Microbiol Rev* **48**: 199-221.

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

Mills, E., Baruch, K., Charpentier, X., Kobi, S., and Rosenshine, I. (2008) Real-time analysis of effector translocation by the type III secretion system of enteropathogenic *Escherichia coli*. *Cell Host Microbe* **3**:104-113.

Minamino, T., and Pugsley, A.P. (2005) Measure for measure in the control of type III secretion hook and needle length. *Mol Microbiol* **56**: 303-308.

Minchin, S.D., and Busby, S.J. (2009) Analysis of mechanisms of activation and repression at bacterial promoters. *Methods* **47**: 6-12.

Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. (2002) Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**: 747-756.

Miroslavova, N.S., and Busby, S.J.W. (2006) Investigations of the modular structure of bacterial promoters. *Biochem Soc Symp* **73**: 1-10.

Mitchell, J.E., Zheng, D., Busby, S.J., and Minchin, S.D. (2003) Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res* **31**: 4689-4695.

Mitchell, J.E., Oshima, T., Piper, S.E., Webster, C.L., Westblade, L.F., Karimova, G., Ladant, D., Kolb, A., Hobman, J.L., Busby, S.J., and Lee, D.J. (2007) The *Escherichia coli* regulator of sigma 70 protein, Rsd, can up-regulate some stress-dependent promoters by sequestering sigma 70. *J Bacteriol* **189**: 3489-3495.

Mitobe, J., Morita-Ishihara, T., Ishihama, A., and Watanabe, H. (2008) Involvement of RNA-binding protein Hfq in the post-transcriptional regulation of *invE* gene expression in *Shigella sonnei*. *J Biol Chem* **283**: 5738-5747.

Moine, H., Romby, P., Springer, M., Grunberg-Manago, M., Ebel, J.P., Ehresmann, C., and Ehresmann, B. (1988) Messenger RNA structure and gene regulation at the translational level in *Escherichia coli*: the case of threonine: tRNAThr ligase. *Proc Natl Acad Sci USA* **85**: 7892-7896.

Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., and Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature* **466**: 77-81.

Morabito, S., Dell'Omo, G., Agrimi,U., Schmidt, H., Karch, H., Cheasty, T., and Caprioli, A. (2001) Detection and characterization of Shiga toxin-producing *Escherichia coli* in feral pigeons. *Vet Microbiol* **82**: 275-283.

Morris, D.R., and Geballe, A.P. (2000) Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**: 8635-8642.

Moseley, S. L., Echeverria, P., Seriwatana, J., Tirapat, C., Chaicumpa, W., Sakuldaipeara, T., and Falkow, S. (1982) Identification of enterotoxigenic *Escherichia coli* by colony hybridization using three enterotoxin gene probes. *J Infect Dis* **145**: 863-869.

Müller, D., Benz, I., Liebchen, A., Gallitz, I., Karch, H., and Schmidt, M.A. (2009) Comparative analysis of the locus of enterocyte effacement and its flanking regions. *Infect Immun* **77**: 3501-3513.

Muller-Hill, B. (1998) Some repressors of bacterial transcription. *Curr Opin Microbiol* **1**: 145-151.

Mundy, R., Petrovska, L., Smollett, K., Simpson, N., Wilson, R.K., Yu, J., Tu, X., Rosenshine, I., Clare, S., Dougan, G., and Frankel, G. (2004) Identification of a novel *Citrobacter rodentium* type III secreted protein, EspI, and roles of this and other secreted proteins in infection. *Infect Immun* **72**: 2288-2302.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002a) Structural basis of the transcription initiation: RNA polymerase holoenzyme at 4Å resolution. *Science* **296**: 1280-1284.

Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002b) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**: 1285-1290.

Murakami, K.S., and Darst, S.A. (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol* **13**: 31-39.

Murray, P.R., Rosenthal, K.S., and Pfaller, M.A. (2009) *Medicial Microbiology* (6th ed.), Philadelphia: Mosby Elsevier.

Musso, R.E., Lauro, R.D., Adhya, S., and de Crombrugghe, B. (1977) Dual control for transcription of the galactose operon by cyclic AMP and its receptor protein at two interspersed promoters. *Cell* **12**: 847-854.

Nahvi A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002) Genetic control by a metabolite binding RNA. *Chem Biol* **9**: 1043-1049.

Nakanishi, N., Abe, H., Ogura, Y., Hayashi, T., Tashiro, K., Kuhara, S., Sugimoto, N., and Tobe, T. (2006) ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. *Mol Microbiol* **61**: 194-205.

Nakao, H., Watanabe, H., Nakayama, S., and Takeda, T. (1995) *yst* gene expression in *Yersinia enterocolitica* is positively regulated by a chromosomal region that is highly homologous to *Escherichia coli* host factor 1 gene (*hfq*). *Mol Microbiol* **18**: 859-865.

Narasimhan, G., Bu, C., Gao, Y., Wang, X., Xu, N., and Mathee, K., (2002) Mining protein sequences for motifs. *J Comput Biol* **9**: 707-720.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R.H. (2000) Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**: 601-611.

Nataro, J.P. (2006) Atypical enteropathogenic *Escherichia coli*: typical pathogens? *Emerg Infect Dis* **12**: 696.

Nataro, J.P., and Kaper, J.B. (1998) Diarrheagenic *Escherichia coli*. *Clin Microbiol Rev* **11**: 142-201.

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

Naylor, S.W., Gally, D.L., and Low, J.C. (2005) Enterohaemorrhagic *E. coli* in veterinary medicine. *Int J Med Microbiol* **295**: 419-441.

Neidhardt, F.C., and Curtiss, R. (1996) *Escherichia coli and Salmonella: Cellular and Molecular Biology*. Washington, DC: American Society for Microbiology Press.

Newberry, K.J., and Brennan, R.G. (2004) The structural mechanism for transcription activation by MerR family member multidrug transporter activation, N terminus. *J Biol Chem* **279**: 20356-20362.

Nieto, J.M., Mourino, M., Balsalobre, C., Madrid, C., Prenafeta, A., Munoa, F.J., and Juarez, A. (1997) Construction of a double *hha hns* mutant of *Escherichia coli*: effect on DNA supercoiling and alpha-haemolysin production. *FEMS Microbiol Lett* **155**: 39-44.

Nilsson, G., Belasco, J.G., Cohen, S.N., and von Gabain, A. (1987) Effect of premature termination of translation on mRNA stability depends on the site of ribosome release. *Proc Natl Acad Sci USA* **84**: 4890-48944.

Nougayrede, J.P., and Donnenberg, M.S. (2004) Enteropathogenic *Escherichia coli* EspF is targeted to mitochondria and is required to initiate the mitochondrial death pathway. *Cell Microbiol* **6**: 1097-1111.

Nudler, E., and Mironov, A.S. (2004) The riboswitch control of bacterial metabolism. *Trends Biochem Sci* **29**: 11-17.

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

O'Brien, A.D., and LaVeck, G.D., Thompson, M.R., and Formal, S.B. (1982) Production of *Shigella dysenteriae* type 1-like cytotoxin by *Escherichia coli*. *J Infect Dis* **146**: 763-769.

O'Brien, A.D., and Holmes, R.K. (1987) Shiga and Shiga-like toxins. *Microbiol Rev* **51**: 206-220.

O'Connell, C.B., Creasey, E.A., Knutton, S., Elliott, S., Crowther, L.J., Luo, W., Albert, M.J., Kaper, J.B., Frankel, G., and Donnenberg, M.S. (2004) SepL, a protein required for enteropathogenic *Escherichia coli* type III translocation, interacts with secretion component SepD. *Mol Microbiol* **52**: 1613-1625.

O'Donnell, S.M., and Janssen, G.R. (2001) The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of *c*I mRNA with or without the 5' untranslated leader. *J Bacteriol* **183**: 1277-1283.

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

Ogino, T., Ohno, R., Sekiya, K., Kuwae, A., Matsuzawa, T., Nonaka, T., Fukuda, H., Imajoh-Ohmi, S., and Abe, A. (2006) Assembly of the type III secretion apparatus of enteropathogenic *Escherichia coli*. *J Bacteriol* **188**: 2801-2811.

Ontiveros, C., Valadez J.G., Hernández, J., and Guarneros, G. (1997) Inhibition of *Escherichia coli* protein synthesis by abortive translation of phage lambda minigenes. *J Mol Biol* **6**: 167-175.

Orth, D., Grif, K., Fisher, I., Fruth, A., Tschäpe, H., Scheutz, F., Dierich, MP., and Würzner, R. (2006) Emerging Shiga toxin-producing *Escherichia coli* serotypes in Europe: O100:H- and O127:H40. *Curr Microbiol* **53**: 428-429.

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

Ou, J.T., and Anderson, T.F. (1970) Role of pili in bacterial conjugation. *J Bacteriol* **102**: 648-654.

Oviedo, A-N., Salgado, H., Collado-Vides, J., and Guarneros, G. (2004) Distribution of minigenes in the bacteriophage lambda chromosome. *Gene* **329**: 115-124.

Pallen, M.J., Beatson, S.A., and Bailey, C.M. (2005a) Bioinformatics analysis of the locus for enterocyte effacement provides novel insights into type-III secretion. *BMC Microbiol* **5**: 9.

Pallen, M.J., Beatson, S.A., and Bailey, C.M. (2005b) Bioinformatics, genomics and evolution of non-flagellar type-III secretion systems: a Darwinian perspective. *FEMS Microbiol Rev* **29**: 201-229.

Pallen, M.J., and Gophna, U. (2007) Bacterial flagella and Type III secretion: case studies in the evolution of complexity. *Genome Dyn* **3**: 30-47.

Pallen, M.J., and Wren, B.W. (2007) Bacterial pathogenomics. *Nature* **449**: 835-842.

Park, K.S., Ono, T., Rokuda, M., Jang, M.H., Okada, K., Iida, T., and Honda, T. (2004) Functional characterization of two type III secretion systems of *Vibrio parahaemolyticus*. *Infect Immun* **72**: 6659-6665.

Parkhill, J., and Brown, N.L. (1990) Site-specific insertion and deletion mutants in the *mer* promoter-operator region of Tn*501*; the nineteen base-pair spacer is essential for normal induction of the promoter by MerR. *Nucleic Acids Res* **18**: 5157-5162.

Parsot, C., Hamiaux, C., and Page, A.L. (2003) The various and varying roles of specific chaperones in type III secretion systems. *Curr Opin Microbiol* **6**: 7-14.

Parsot, C., Ageron, E., Penno, C., Mavris, M., Jamoussi, K., d'Hauteville, H., Sansonetti, P., and Demers, B. (2005) A secreted anti-activator, OspD1, and its chaperone, Spa15, are involved in the control of transcription by the type III secretion apparatus activity in *Shigella flexneri*. *Mol Microbio*l **56**: 1627-1635.

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

Paul, B.J., Berkmen, M.M., and Gourse, R.L. (2005). DksA potentiates direct activation of amino acid promoters by ppGpp. *Proc Natl Acad Sci USA* **102**: 7823-7828.

Pawar, D.M., Rossman, M.L., and Chen, J. (2005) Role of curli fimbriae in mediating the cells of enterohaemorrhagic *Escherichia coli* to attach to abiotic surfaces. *J Appl Microbiol* **99**: 418-425.

Perederina, A., Svetlov, V., Vassylyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I., and Vassylyev, D.G. (2004) Regulation through the secondary channel-structural framework for ppGpp-DksA synergism during transcription. *Cell* **118**: 297-309.

Perez-Rueda, E., and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. *Nucleic Acids Res* **28**: 1838-1847.

Perna, N.T., Mayhew, G.F., Pósfai, G., Elliott, S., Donnenberg, M.S., Kaper, J.B., and Blattner, F.R. (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **66**: 3810-3817.

Perrin, S., and Gilliland, G. (1990) Site-specific mutagenesis using asymmetric polymerase chain reaction and a single mutant primer. *Nucleic Acids Res* **18**: 7433-7438.

Petersen, C. (1992) Control of functional mRNA stability in bacteria: multiple mechanisms of nucleolytic and non-nucleolytic inactivation. *Mol Microbiol* **6**: 277-282.

Ponnambalam, S., Webster, C., Bingham, A., Busby, S. (1986) Transcription initiation at the *Escherichia coli* galactose operon promoters in the absence of the normal -35 region sequences. *J Biol Chem* **261**: 16043-16048.

Ponnambalam, S., Spassky, A., and Busby, S. (1987) Studies with the *Escherichia coli* galactose operon regulatory region carrying a point mutation that simultaneously inactivates the two overlapping promoters. Interactions with RNA polymerase and the cyclic AMP receptor protein. *FEBS Lett* **219**: 189-196.

Porter, M.E., Mitchell, P., Free, A., Smith, D.G.E., and Gally, D.L. (2005) The *LEE1* promoters from both enteropathogenic and enterohemorrhagic *Escherichia coli* can be activated by PerC-like proteins from either organism. *J Bacteriol* **187**: 458-472.

Pugsley, A.P. (1993) The complete general secretory pathway in gram-negative bacteria. *Microbiol Rev* **57**: 50-108.

Qing, G., Xia, B., and Inouye, M. (2003) Enhancement of translation initiation by A/T-rich sequences downstream of the initiation codon in *Escherichia coli*. *J Mol Microbiol Biotechnol* **6**: 133-144.

Rauhut, R., and Klug, G. (1999) mRNA degradation in bacteria. *FEMS Microbiol Rev* **23**: 353-370.

Reading, N.C., Torres, A.G., Kendall, M.M., Hughes, D.T., Yamamoto, K., and Sperandio, V. (2007). A novel two-component signaling system that activates transcription of an enterohemorrhagic *Escherichia coli* effector involved in remodeling of host actin. *J Bacteriol* **189**: 2468-2476.

Reid, G., Howard, J., and Gan, B.S. (2001) Can bacterial interference prevent infection? *Trends Microbiol* **9**: 424-428.

Reznikoff, W.S. (1992) The lactose operon-controlling elements: a complex paradigm. *Mol Microbiol* **6**: 2419-2422.

Riley, L.W., Remis, R.S., Helgerson, S.D., McGee, H.B., Wells, J.G., Davis, B.R., Hebert, R.J., Olcott, E.S., Johnson, L.M., Hargrett, N.T., Blake, P.A., and Cohen, M.L. (1983) Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N Engl J Med* **308**: 681-685.

Riley, L.W. (1987) The epidemiologic, clinical, and microbiologic features of hemorrhagic colitis. *Annu Rev Microbiol* **41**: 383-407.

Rimsky, S., Zuber, F., Buckle, M., and Buc, H. (2001) A molecular mechanism for the repression of transcription by the H-NS protein. *Mol Microbiol* **42**: 1311-1323.

Roe, A.J., Yull, H., Naylor, S.W., Woodward, M.J., Smith, D.G., and Gally, D.L. (2003) Heterogeneous surface expression of EspA translocon filaments by *Escherichia coli* O157:H7 is controlled at the posttranscriptional level. *Infect Immun* **71**: 5900-5909.

Roe, A.J., Tysall, L., Dransfield, T., Wang, D., Fraser-Pitt, D., Mahajan, A., Constandinou, C., Inglis, N., Downing, A., Talbot, R., Smith, D.G., and Gally, D.L. (2007) Analysis of the expression, regulation and export of NleA-E in *Escherichia coli* O157: H7. *Microbiology* **153**: 1350-1360.

Romero, A., and García, P. (1991) Initiation of translation at AUC, AUA and AUU codons in *Escherichia coli*. *FEMS Microbiol Lett* **68**: 325-330.

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

Rosenfeld, N., Elowitz, M.B., and Alon, U. (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323**: 785-793.

Rosenshine, I., Ruschkowski, S., and Finlay, B.B. (1996). Expression of attaching/effacing activity by enteropathogenic *Escherichia coli* depends on growth phase, temperature, and protein synthesis upon contact with epithelial cells. *Infect Immun* **64**: 966-973.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**: 1407-1413.

Russell, R.M., Sharp, F.C., Rasko, D.A., and Sperandio, V. (2007) QseA and GrlR/GrlA regulation of the locus of enterocyte effacement genes in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **189**: 5387-5392.

Russo, T.A., and Johnson, J.R. (2003) Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. *Microbes Infect* **5**: 449-456.

Saitoh, T., Iyoda, S., Yamamoto, S., Lu, Y., Shimuta, K., Ohnishi, M., Terajima, J., and Watanabe, H. (2008) Transcription of the *ehx* enterohemolysin gene is positively regulated by GrlA, a global regulator encoded within the locus of enterocyte effacement in enterohemorrhagic *Escherichia coli*. *J Bacteriol* **190**: 4822-4830.

Salis, H., Mirsky, E.A., and Voigt, C.A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**: 946-950.

Sal-Man, N., Biemans-Oldehinkel, E., and Finlay, B.B. (2009) Structural microengineers: pathogenic *Escherichia coli* redesigns the actin cytoskeleton in host cells. *Structure* **17**: 15-19.

Sanchez-SanMartin, C., Bustamante, V.H., Calva, E., and Puente, J.L. (2001) Transcriptional regulation of the *orf19* gene and the *tir-cesT-eae* operon of enteropathogenic *Escherichia coli*. *J Bacteriol* **183**: 2823-2833.

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

Savery, N.J., Belyaeva, T., and Busby, S. (1996) Introduction to protein: DNA interactions, DNase I footprinting, hydroxyl radical footprinting, permanganate footprinting and supplementary protocols. In *Essential techniques: gene transcription*. (Docherty, K., ed.) Oxford: BIOS Scientific Publishers, pp. 1-5 and 21-33.

Schauer, D.B., and Falkow, S. (1993) Attaching and effacing locus of a *Citrobacter freundii* biotype that causes transmissible murine colonic hyperplasia. *Infect Immun* **61**: 2486-2492.

Schilling, J.D., Mulvey, M.A., Vincent, C.D., Lorenz, R.G., and Hultgren, S.J. (2001) Bacterial invasion augments epithelial cytokine responses to *Escherichia coli* through a lipopolysaccharide-dependent mechanism. *J Immunol* **166**: 1148-1155.

Schmidt, M.A. (2010) LEEways: tales of EPEC, ATEC and EHEC. *Cell Microbiol* **12**: 1544 -1552.

Schneider, T.D., Stormo, G.D., and Gold, L. (1986) Information content of binding sites on nucleotide sequences. *J Mol Biol* **188**: 415-431.

Schüller, S., Heuschkel, R., Torrente, F., Kaper, J.B., and Phillips, A.D. (2007) Shiga toxin binding in normal and inflamed human intestinal mucosa. *Microbes Infect* **9**: 35-39.

Sclavi, B. (2009) Opening the DNA at the promoter: the energetic challenge. In *RNA polymerases as molecular motors*. (Buc, H. and Strick, T., eds) Royal Society of Chemistry Publishing, Cambridge, UK, pp. 38-68.

Sharma, V.K. (2002) Detection and quantitation of enterohemorrhagic *Escherichia coli* O157, O111, and O26 in beef and bovine feces by real-time polymerase chain reaction. *J Food Prot* **65**: 1371-1380.

Sharma, V.K., and Zuerner, R.L. (2004) Role of *hha* and *ler* in transcriptional regulation of the *esp* operon of enterohemorrhagic *Escherichia coli* O157:H7. *J Bacteriol* **186**: 7290-7301.

Sharp F.C., and Sperandio, V. (2007) QseA directly activates transcription of *LEE1* in enterohemorrhagic *Escherichia coli*. *Infect Immun* **75**: 2432-2440.

Shaw, R.K., Smollett, K., Cleary, J., Garmendia, J., Straatman-Iwanowska, A., Frankel, G., and Knutton, S. (2005) Enteropathogenic *Escherichia coli* type III effectors EspG and EspG2 disrupt the microtubule network of intestinal epithelial cells. *Infect Immun* **73**: 4385-4390.

Shin, M., Kang, S., Hyun, S.-J., Fujita, N., Ishihama, A., Valentin-Hansen, P., and Choy, H.E. (2001) Repression of *deo*P2 in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *EMBO J* **20**: 5392-5399.

Shin, S., Castanie-Cornet, M.P., Foster, J.W., Crawford, J.A., Brinkley, C., and Kaper, J.B. (2001) An activator of glutamate decarboxylase genes regulates the expression of enteropathogenic *Escherichia coli* virulence genes through control of the plasmid-encoded regulator, Per. *Mol Microbiol* **41**: 1133-1150.

Shine, J., and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* **71**: 1342-1346.

Shine, J., and Dalgarno, L. (1975) Determinant of cistron specificity in bacterial ribosomes. *Nature*, **254**: 34-38.

Shulman, S.T., Friedmann, H.C., and Sims, R.H. (2007) Theodor Escherich: the first pediatric infectious diseases physician? *Clin Infect Dis* **45**: 1025-1029.

Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007) Anatomy of *Escherichia coli* σ^{70} promoters. *Nucleic Acids Res* **35**: 771-788.

Siegele, D.A., Hu, J.C., Walter, W.A., and Gross, C.A. (1989) Altered promoter recognition by mutant forms of the sigma 70 subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **206**: 591-603.

Sittka, A., Pfeiffer, V., Tedin, K., and Vogel, J. (2007) The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol* **63**: 193-217.

Sittka A, Lucchini S, Papenfort K, Sharma, C.M., Rolle, K., Binnewies, T.T., Hinton, J.C., and Vogel, J. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS genet* **4**: e1000163.

Skorupski, K., and Taylor, R.K. (1997) Control of the ToxR virulence regulon in *Vibrio cholerae* by environmental stimuli. *Mol Microbiol* **25**: 1003-1009.

Sowers, E.G., Wells, J.G, and Strockbine, N.A. (1996) Evaluation of commercial latex reagents for identification of O157 and H7 antigens of *Escherichia coli*. *J Clin Microbiol* **34**: 1286-1289.

Spano, L.C., Sadovsky, A.D.I., Segui, P.N., Saick, K.W., Kitagawa, S.M.S., Pereira, F.E.L., Fagundes-Neto, U., and Scaletsky, I.C.A. (2008) Age specific prevalence of diffusely adherent *Escherichia coli* in Brazilian children with acute diarrhoea. *J Med Microbiol* **57**: 359-363.

Spears, K.J., Roe, A.J., and Gally, D.L. (2006) A comparison of enteropathogenic and enterohaemorrhagic *Escherichia coli* pathogenesis. *FEMS Microbiol Lett* **255**: 187-202.

Sperandio, V., Mellies, J.L., Nguyen, W., Shin, S., and Kaper, J.B. (1999) Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proc Natl Acad Sci USA* **96**: 15196-15201.

Sperandio, V., Mellies, J.L., Delahay, R.M., Frankel, G., Crawford, J.A., Nguyen, W., and Kaper, J.B. (2000) Activation of enteropathogenic *Escherichia coli* (EPEC) *LEE2* and *LEE3* operons by Ler. *Mol Microbiol* **38**: 781-793.

Sperandio, V., Torres, A.G., Giron, J.A., and Kaper, J.B. (2001) Quorum sensing is a global regulatory mechanism in enterohemorrhagic *Escherichia coli* O157:H7. *J Bacteriol* **183**: 5187-5197.

Sperandio, V., Li, C.C., and Kaper, J.B. (2002) Quorum-sensing *Escherichia coli* regulator A: a regulator of the LysR family involved in the regulation of the locus of enterocyte effacement pathogenicity island in enterohemorrhagic *E. coli*. *Infect Immun* **70**: 3085-3093.

Stacy-Phipps, S., Mecca, J.J., and Weiss, J.B. (1995) Multiplex PCR assay and simple preparation method for stool specimens detect enterotoxigenic *Escherichia coli* DNA during the course of infection. *J Clin Microbiol* **33**: 1054-1059.

Steitz, J.A., and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3' terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. *Proc Natl Acad Sci USA* **72**: 4734-4738.

Stephan, R., Ragettli, S., and Untermann, F. (2000) Prevalence and characteristics of verotoxin-producing *Escherichia coli* (VTEC) in stool samples from asymptomatic human carriers working in the meat processing industry in Switzerland. *J Appl Microbiol* **88**: 335-341.

Stevens, M.P., Wood, M.W., Taylor, L.A., Monaghan, P., Hawes, P., Jones, P.W., Wallis, T.S., and Galyov, E.E. (2002) An Inv/Mxi-Spa-like type III protein secretion system in *Burkholderia pseudomallei* modulates intracellular behaviour of the pathogen. *Mol Microbiol* **46**: 649-659.

Stock, J.B., Rauch, B., and Roseman, S. (1977) Periplasmic space in *Salmonella typhimurium* and *Escherichia coli*. *J Biol Chem* **252**: 7850-7861.

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

Stryer, L. (1995) *Biochemistry,* (4th ed.), New York, W.H. Freeman and Co.

Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**: 2688-2697.

Takao, T., Tanabe, T., Hong, Y.-M., Shimonishi, Y., Kurazono, H., Yutsudo, T., Sasakawa, C., Yoshikawa, M., and Takeda, Y. (1988) Identity of the molecular structure of Shiga-like toxin I (VT1) from *Escherichia coli* 0157: H7 with that of Shiga toxin. *Microb Pathog* **5**: 357-369.

Tatsuno, I., Nagano, K., Taguchi, K., Rong, L., Mori, H., and Sasakawa, C. (2003) Increased adherence to Caco-2 cells caused by disruption of the *yhiE* and *yhiF* genes in enterohemorrhagic *Escherichia coli* O157:H7. *Infect Immun* **71**: 2598-2606.

Tauschek, M., Yang, J., Hocking, D., Azzopardi, K., Tan, A., Har, E., Praszkier, J., and Robins-Browne, R.M. (2010) Transcriptional analysis of the *grlRA* virulence operon from *Citrobacter rodentium*. *J Bacteriol* **192**: 3722-3734.

Tobe, T., Ando, H., Ishikawa, H., Abe, H., Tashiro, K., Hayashi, T., Kuhara, S., and Sugimoto, N. (2005) Dual regulatory pathways integrating the RcsC–RcsD–RcsB signalling system control enterohaemorrhagic *Escherichia coli* pathogenicity. *Mol Microbiol* **58**: 320-333.

Tobe, T., Beatson, S.A., Taniguchi, H., Abe, H., Bailey, C.M., Fivian, A., Younis, R., Matthews, S., Marches, O., Frankel, G., Hayashi, T., and Pallen, M.J. (2006) An extensive repertoire of type III secretion effectors in *Escherichia coli* O157 and the role of lambdoid phages in their dissemination. *Proc Natl Acad Sci USA* **103**: 14941-14946.

Tomoyasu, T., Takaya, A., Isogai, E., and Yamamoto, T. (2003) Turnover of FlhD and FlhC, master regulator proteins for *Salmonella* flagellum biogenesis, by the ATP-dependent ClpXP protease. *Mol Microbiol* **48**: 443-452.

Toshima, H., Yoshimura, A., Arikawa, K., Hidaka, A., Ogasawara, J., Hase, A., Masaki, H., and Nishikawa, Y. (2007) Enhancement of Shiga toxin production in enterohemorrhagic *Escherichia coli* serotype O157:H7 by DNase colicins. *Appl Environ Microbiol* **73**: 7582-7588.

Tree, J.J., Wolfson, E.B., Wang, D., Roe, A.J., and Gally, D.L. (2009) Controlling injection: regulation of type III secretion in enterohaemorrhagic *Escherichia coli*. *Trends Microbiol* **17**: 361-370.

Troisfontaines, P., and Cornelis, G.R. (2005) Type III secretion: more systems than you think. *Physiology (Bethesda)* **20**: 326-339.

Tsai, N. P., Y. C. Wu, J. W. Chen, C. F. Wu, C. M. Tzeng, and W. J. Syu. (2006) Multiple functions of *l0036* in the regulation of the pathogenicity island of enterohaemorrhagic *Escherichia coli* O157:H7. *Biochem J* **393**: 591-599.

Tu, X., Nisan, I., Yona, C., Hanski, E., and Rosenshine, I. (2003) EspH, a new cytoskeleton -modulating effector of enterohaemorrhagic and enteropathogenic *Escherichia coli*. *Mol Microbiol* **47**: 595-606.

Tuttle, J., Gomez, T., Doyle, M.P., Wells, J.G., Zhao, T., Tauxe, R.V., and Griffin, P.M. (1999) Lessons from a large outbreak of *Escherichia coli* O157:H7 infections: insights

into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiol Infect* **122**: 185-192.

Umanski, T., Rosenshine, I., and Friedberg, D. (2002) Thermoregulated expression of virulence genes in enteropathogenic *Escherichia coli*. *Microbiology* **148**: 2735-2744.

Urbanowski, M.L., Lykken, G.L., and Yahr, T.L. (2005) A secreted regulatory protein couples transcription to the secretory activity of the *Pseudomonas aeruginosa* type III secretion system. *Proc Natl Acad Sci USA* **102**: 9930-9935.

Van Etten, W. J., and Janssen, G. R. (1998), An AUG initiation codon, not codon-anticodon complementarity, is required for the translation of unleadered mRNA in *Escherichia coli*. *Mol Microbiol* **27**: 987-1001.

Vial, P.A., Mathewson, J.J., DuPont, H.L., Guers, L., and Levine, M.M. (1990) Comparison of two assay methods for patterns of adherence to HEp-2 cells of *Escherichia coli* from patients with diarrhea. *J Clin Microbiol* **28**: 882-885.

Vimberg, V., Tats, A., Remm, M., and Tenson, T. (2007) Translation initiation region sequence preferences in *Escherichia coli*. *BMC Mol Biol* **8**: 100.

Vitreschak, A.G., Lyubetskaya, E.V., Shirshin, M.A., Gelfand, M.S., and Lyubetsky, V.A. (2004) Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol Lett* **234**: 357-370.

Vuthoori, S., Bowers, C.W., McCracken, A., Dombroski, A.J., and Hinton, D.M. (2001) Domain 1.1 of the σ^{70} subunit of *Escherichia coli* RNA polymerase modulates the formation of stable polymerase/promoter complexes. *J Mol Biol* **309**: 561-572.

Wade, J.T., Belyaeva, T.A., Hyde, E.I., and Busby, S.J. (2000) Repression of the *Escherichia coli melR* promoter by MelR: evidence that efficient repression requires the formation of a repression loop. *Mol Microbiol* **36**: 223-229.

Wade, J.T., Struhl, K., Busby, S.J.W., and Grainger, D.C. (2007) Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol Microbiol* **65**: 21-26.

Wagner, R. (1984) Transcription regulation in prokaryotes. Oxford University Press, Oxford, UK, 2001, 365p.

Wainwright, L.A., and Kaper, J.B. (1998) EspB and EspD require a specific chaperone for proper secretion from enteropathogenic *Escherichia coli*. *Mol Microbiol* **27**: 1247-1260.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615-628.

Wendel, A.M., Johnson, D.H., Sharapov, U., Grant, J., Archer, J.R., Monson, T., Koschmann, C., and Davis, J.P. (2009) Multistate outbreak of *Escherichia coli* O157:H7 infection associated with consumption of packaged spinach, August-September 2006: the Wisconsin investigation. *Clin Infect Dis* **48**: 1079-1086.

Wilson, J.W., Ott, C.M., Höner zu Bentrup, K., Ramamurthy, R., Quick, L., Porwollik, S., Cheng, P., McClelland, M., Tsaprailis, G., Radabaugh, T., Hunt, A., Fernandez, D., Richter, E., Shah, M., Kilcoyne, M., Joshi, L., Nelman-Gonzalez, M., Hing, S., Parra, M., Dumars, P., Norwood, K., Bober, R., Devich, J., Ruggles, A., Goulart, C., Rupert, M., Stodieck, L., Stafford, P., Catella, L., Schurr, M.J., Buchanan, K., Morici, L.,

McCracken, J., Allen, P., Baker-Coleman, C., Hammond, T., Vogel, J., Nelson, R., Pierson, D.L., Stefanyshyn-Piper, H.M., and Nickerson, C.A. (2007) Space flight alters bacterial gene expression and virulence and reveals a role for global regulator Hfq. *Proc Natl Acad Sci USA* **104**: 16299-304.

Wittmann, H.G. (1982) Components of bacterial ribosomes. *Annu Rev Biochem* **51**: 155-183.

Wittmann, H.G., Müssig, J., Piefke, J., Gewitz, H.S., Rheinberger, H.J., and Yonath, A. (1982) Crystallization of *Escherichia coli* ribosomes. *FEBS Lett* **146**: 217-220.

Wittmann, H.G. (1983) Architecture of prokaryotic ribosomes. *Annu Rev Biochem* **52**: 35-65.

Whittam, T.S., Wolfe, M.L., Wachsmuth, I.K., Orskov, F., Orskov, I., Wilson, R.A. (1993) Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun* **61**:1619-1629.

Wick, L.M., Qi, W.H., Lacher, D.W., and Whittam, T.S. (2005) Evolution of genomic content in the stepwise emergence of *Escherichia coli* O157:H7. *J Bacteriol* **187**: 1783-1791.

Williams, J., Michán, C., Webster, C., and Busby, S. (1994) Interactions between the *Escherichia coli* MelR transcription activator protein and operator sequences at the *melAB* promoter. *Biochem J* **300**: 757-763.

Winkler W., Nahvi, A., and Breaker, R.R. (2002a) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952-956.

Winkler, W.C., Cohen-Chalamish, S., and Breaker, R.R. (2002b) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci USA* **99**: 15908-15913.

Woody, S.T., Fong, R.S.C., and Gussin, G.N. (1993) A cryptic promoter in the O_R region of bacteriophage λ. *J Bacteriol* **175**: 5648-5654.

Yamada, H., Yoshida, T., Tanaka, K., Sasakawa, C., and Mizuno, T. (1991) Molecular analysis of the *Escherichia coli* H-NS gene encoding a DNA-binding protein, which preferentially recognizes curved DNA sequences. *Mol Gen Genet* **230**: 332-336.

Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. *Nature* **289**: 751-758.

Yanofsky, C. (2000) Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol* **182**: 1-8.

Yerushalmi, G., Nadler, C., Berdichevski, T., and Rosenshine, I. (2008) Mutational analysis of the locus of enterocyte effacement-encoded regulator (Ler) of enteropathogenic *Escherichia coli*. *J Bacteriol* **190**: 7808-7818.

Yip, C.K., Kimbrough, T.G., Felise, H.B., Vuckovic, M., Thomas, N.A., Pfuetzner, R.A., Frey, E.A., Finlay, B.B., Miller, S.I., and Strynadka, N.C. (2005) Structural characterization of the molecular platform for type III secretion system assembly. *Nature* **435**: 702-707.

Yona-Nadler, C., Umanski, T., Aizawa, S., Friedberg, D, and Rosenshine, I. (2003) Integration host factor (IHF) mediates repression of flagella in enteropathogenic and enterohaemorrhagic *Escherichia coli*. *Microbiology* **149**: 877-884.

Young, B.A., Gruber, T.M., and Gross, C.A. (2002) Views of transcription initiation. *Cell* **109**: 417-420.

Young, B.M., and Young, G.M. (2002) YplA is exported by the Ysc, Ysa, and flagellar type III secretion systems of *Yersinia enterocolitica*. *J Bacteriol* **184**: 1324-1334.

Younis, R., Bingle, L.E., Rollauer, S., Munera, D., Busby, S.J., Johnson, S., Deane, J.E., Lea, S.M., Frankel, G., and Pallen, M.J. (2010) SepL resembles an aberrant effector in binding to a class 1 type III secretion chaperone and carrying an N-terminal secretion signal. *J Bacteriol* **192**: 6093-6098.

Yu, R.R., and DiRita, V.J. (2002) Regulation of gene expression in *Vibrio cholerae* by ToxT involves both antirepression and RNA polymerase stimulation. *Mol Microbiol* **43**: 119-134.

Zhang, A., Wassarman, K.M., Rosenow, C., Tjaden, B.C., Storz, G., and Gottesman, S. (2003) Global analysis of small RNA and mRNA targets of Hfq. *Mol Microbio.* **50**: 1111-1124.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* **98**: 811-824.

Zhang, L., Chaudhuri, R.R., Constantinidou, C., Hobman, J.L., Patel, M.D., Jones, A.C., Sarti, D., Roe, A.J., Vlisidou, I., Shaw, R.K., Falciani, F., Stevens, M.P., Gally, D.L., Knutton, S., Frankel, G., Penn, C.W., and Pallen, M.J. (2004) Regulators encoded in the *Escherichia coli* type III secretion system 2 gene cluster influence expression of genes within the locus for enterocyte effacement in enterohemorrhagic *E. coli* O157:H7. *Infect Immun* **72**: 7282-7293.

Zhang, S., Kingsley, R.A., Santos, R.L., Andrews-Polymenis, H., Raffatellu, M., Figueiredo, J., Nunes, J., Tsolis, R.M., Adams, L.G., and Bäumler, A.J. (2003) Molecular pathogenesis of *Salmonella enterica* serotype Typhimurium-induced diarrhea. *Infect Immun* **71**: 1-12.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406-15.

Appendix I

Appendix II

Islam, M.S., Bingle, L.E.H., Pallen, M.J., and Busby, S.J. (2011) Organization of the *LEE1* operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. *Mol Microbiol* **79**: 468-483.

Islam, M.S., Pallen, M.J., and Busby, S.J. (2011) A cryptic promoter in the *LEE1* regulatory region of enterohaemorrhagic *Escherichia coli*: promoter specificity in AT-rich gene regulatory regions. *Biochem J* **436**: 681-686