Position-dependent modulation of promoter activity in *Escherichia coli*

By

Jack Alfred Bryant

A thesis submitted to the
University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

School of Biosciences
The University of Birmingham
Edgbaston
Birmingham
B15 2TT
UK
January 2013

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Synopsis

Bacterial gene expression is primarily regulated at the initial step of transcription initiation. This regulation has traditionally been viewed as due to the DNA sequence of promoters and factors that bind to them. However, the bacterial chromosome is a highly compacted and ordered structure. To date, the effects of position within the chromosome on promoter activity have barely been investigated.

Here, I report that activity of a single promoter, inserted at different loci in the *Escherichia coli* chromosome, is dependent upon position within the chromosome. Quantification of genomic DNA at a number of loci showed this was unrelated to gene dosage. Drastic changes in position-dependent hierarchies of promoter activity were observed on entry into stationary phase. However, no correlation between the orientation of a promoter, with respect to replication, and its activity was found.

Several factors are involved in this variation, including local sequence context. Increased promoter activity at origin-proximal loci was shown to be due to DNA supercoiling, maintained by the action of DNA gyrase. Transcription was shown to have a negative influence on the activity of downstream promoters. Further to this, the position of a promoter, with respect to a repressor-encoding gene, had an effect on its repression, further suggesting a role for genome organisation in the regualtion of transcription. Finally, extended protein occupancy domains were shown to repress activity of inserted promoters, indicating that these domains act to silence the genes they encode. The combined data show that the position of a promoter within the *E. coli* chromosome plays a significant part in transcriptional regulation, and that structure of the chromosome is also likely to play a role.

To Ellie and my family for their support and encouragement

Acknowledgments

This thesis is the result of 4 years of work, through which I have been supported by a number of people. I have performed all the work presented, except where mentioned. I am now pleased to thank all those who have guided and supported me throughout.

First and foremost I must thank Professor Steve Busby for giving me this opportunity and for his continued support, encouragement and faith in my abilities. It has been a pleasure and a privilege to begin my career in research with such an enthusiastic and accomplished scientist. Next I must thank Dr Kerry Hollands for her support and tutelage as I began my career in the laboratory. I also owe my gratitude to Dr Dave Lee and Dr Doug Browning for their expert help, guidance, inspiration and friendship throughout my time in the Busby lab. I am grateful to Chris Webster and Rita Godfrey for technical support and the University of Birmingham, School of Biosciences for funding. I give my thanks to students, Siân Russell, Steven Bevan and Sarah Moola for their assistance in construction of donor plasmids and strains.

I thank Busby and Cole lab members, past and present for their invaluable help and friendship, especially: Dr David Chismon; Dr Derrick Squire; Dr Mohammed Shahidul Islam; Laura Rowley and Patcharawarin Ruanto. To those closest to me during these four years, Dr Claire Vine, Dr Amanda Hopper and Ian Cadby, I am incredibly grateful for your support in the lab, the lecture theatre, the coffee room and most importantly the bar! Special mention must be given to my partner in crime, Ian Cadby, who has accompanied me on this journey from our first day in the lab and has made these past four years fly by (Club Tropicana seems just like yesterday!). I am grateful to my family for their unrelenting belief in my abilities, without them I would have never even started an undergraduate degree. Finally, I must express my deepest gratitude to my biggest inspiration and future wife, Ellie Ellis, for her encouragement, understanding, love and support throughout these four years.

Contents

List of Figures

List of Tables

List of Abbreviations

αCTD: Subunit α carboxyl-terminal-domain

αNTD: Subunit α amino-terminal-domain

AFM: Atomic force microscopy

Amp^R: Amplicillin-resistance

APS: Ammonium persulfate

ATP: Adenosine triphosphate

bp: Base Pair

BSA: Bovine serum albumin

cAMP: Cyclic adenosine monophosphate

Cat: Chloramphenicol acetyl transferase

CbpA: Curved DNA binding protein

ChIP: Chromatin immuno-precipitation

ChIP-chip: Chromatin immuno-precipitation on microarray

CIP: Calf intestinal alkaline phosphate

Cm^R: Chloramphenicol-resistance

CRP: cAMP receptor protein

dNTP: Deoxynucleoside triphosphate

DNA: Deoxyribonucleic acid

Dps: DNA protection during starvation protein

EPOD: Extended protein occupancy domain

FIS: Factor for inversion stimulation

H-NS: Histone-like nucleoid structuring protein

heEPOD: Highly expressed extended protein occupancy domain

IHF: Integration host factor

IPTG: Isopropyl β-D-1-thiogalactopyranoside

kb: kilo base pair

Mb: mega base pair

mRNA: Messenger ribonucleic acid

NAP: Nucleoid associated protein

NSL: Non-structured left

NSR: Non-structured right

RNA: Ribonucleic acid

RNAP: RNA polymerase

rRNA: Ribosomal ribonucleic acid

Tet^R: Tetracycline-resistance

tsEPOD: Transcriptionally silent extended protein occupancy domain

UV: Ultra violet

Chapter 1:

Introduction

1.1 The model bacterium *Escherichia coli*

Bacteria are unicellular, prokaryotic microorganisms, which are able to inhabit and shape any niche on the planet, including human and animal hosts. The adaptability required for bacteria to survive, and thrive, in variable environments is achieved through the ability to regulate expression of a variety of genes in response to changing environmental signals. The key distinguishing feature between prokaryotic and eukaryotic organisms is the way that chromosomal DNA is packaged within the cell. Unlike eukaryotic chromosomes, bacterial DNA is not separated from the rest of the cell by a membrane, but is instead concentrated into a region of the cell called the nucleoid. While the textbooks argue that this constitutes a less complex arrangement than in eukaryotic cells, one might argue that the lack of a membrane requires complicated organisation of the bacterial chromosome, to allow access by the gene expression machinery at the appropriate time (Campbell and Reece, 2004; Madigan and Martinko, 2006).

Bacteria usually come under public interest due to their role in health and disease, especially as the human gastrointestinal tract plays host to a wide range of bacteria, some of which can be pathogenic (Madigan and Martinko, 2006). *Escherichia coli* are members of the enteric bacteria, enterobacteriaceae, which are gram-negative, facultatively anaerobic rods containing a circular chromosome. *E. coli* inhabits the gastrointestinal tract of humans and other animals and is widely accepted as a model bacterium for the study of bacterial biochemistry, physiology and genetics (Madigan and Martinko, 2006; Murray *et al*, 2009). Most *E. coli* cells measure approximately 2 μm in length and 0.5 μm in diameter, which can make *E. coli* a difficult organism to study (Zaritsky and Woldringh, 1978). However, new microscopy techniques have now been developed to allow visualisation of the inner workings of the cell (Hohlbein *et al*., 2010).

There are many pathogenic and non-pathogenic *E. coli* strain types, which are closely related to the well known pathogenic *Salmonella* and *Shigella* species (Falkow, 1996; Fukushima *et al*., 2002). However, the lab-cultivated, non-pathogenic, *E. coli* K-12 MG1655 strain is often used as a model organism, due to the strain having lost the surface "O" antigens (lipopolysaccharide) and being considered harmless. The MG1655 strain is a derivative of the original *E. coli* K-12 strain, maintained with minimal genetic manipulation, however the strain was cured of the F plasmid and phage lambda by acridine orange and UV radiation treatment (Bachmann, 1996). Further to this, the genome of *E. coli* K-12 MG1655 was sequenced in 1997 and detailed sequence annotation is now available via the EcoCyc website. Therefore *E. coli* K-12 MG1655 is an ideal strain for the study of bacterial genetics and general molecular microbiology (Blattner *et al*., 1997; Keseler *et al*., 2011). Studies of *E. coli* cellular processes create data that is applicable to any bacteria that uses those processes. Therefore, study of the *E. coli* model bacterium can, and has, yielded transferrable information to be applied to other important or pathogenic strains of bacteria (Madigan and Martinko, 2006).

1.2 The *Escherichia coli* chromosome

1.2.1 The nucleoid

Although bacteria have been studied intensively for many years as models for systems in higher organisms, we have, until recent years, remained unaware of the complexities of their internal cellular structures. This problem was mainly due to a lack of technology sufficient to probe these minute structures, however new techniques and better technology have allowed some of these structures to be explored in greater detail (reviewed in Hohlbein *et al*., 2010). The chromosomes of bacteria are unlike those of eukaryotes, in that they are not separated

from the cytoplasm by a nuclear membrane, however they still form a tightly compacted nucleus-like body, referred to as the nucleoid (Robinow, 1956; Kleppe *et al*., 1979). The nucleoid is defined as a pseudo-compartment occupying a distinct region of the cell, characterised by the presence of highly condensed chromatin and an absence of ribosomes (reviewed in Kleppe *et al*., 1979 and Robinow and Kellenberger, 1994; Figure 1.1). The nucleoid was first isolated in its folded and supercoiled state by Stonington and Pettijohn (1971). Their method was utilised by Kavenoff and Bowen (1976) to visualise the nucleoid by electron microscopy, which revealed that the chromosomal DNA formed rosette structures of approximately 140 supercoiled DNA loops, radiating from a central core containing RNA and protein (Figure 1.1d). Previous work, analysing nucleoid superhelicity through the assessment of nucleoid sedimentation rates, demonstrated that the folded chromosome could not be entirely relaxed by a single DNase-induced nick and that many were required for full relaxation (Worcel and Burgi, 1972). This result showed that the nucleoid is organised into many independently supercoiled, topologically distinct domains, separated by barriers to supercoil diffusion. These results were the first suggestions that the bacterial nucleoid is structured and organised into separate domains.

The *E. coli* chromosome must indeed be highly structured and organised to fit into the nucleoid, which occupies only a fraction of the space within the cell. For example, the size of a single *E. coli* genome is 4,639,221 bp, resulting in an extended length of approximately 1.58 mm, based on there being a residue on each chain every 3.4 Å (Watson and Crick, 1953; Blattner *et al*., 1997). This exceptionally large molecule must be compacted into a cell that measures approximately 2 μm in length and 0.5 μm in diameter (Zaritsky and Woldringh, 1978). Several mechanisms are used to sufficiently compact the chromosome into the *E. coli* nucleoid region, which represents a condensation from free coiled DNA of approximately 1000-fold (Holmes and Cozzarelli, 2000). The mechanisms of compaction occur on different scales, much like that of eukaryotes, and include DNA supercoiling, binding of nucleoid

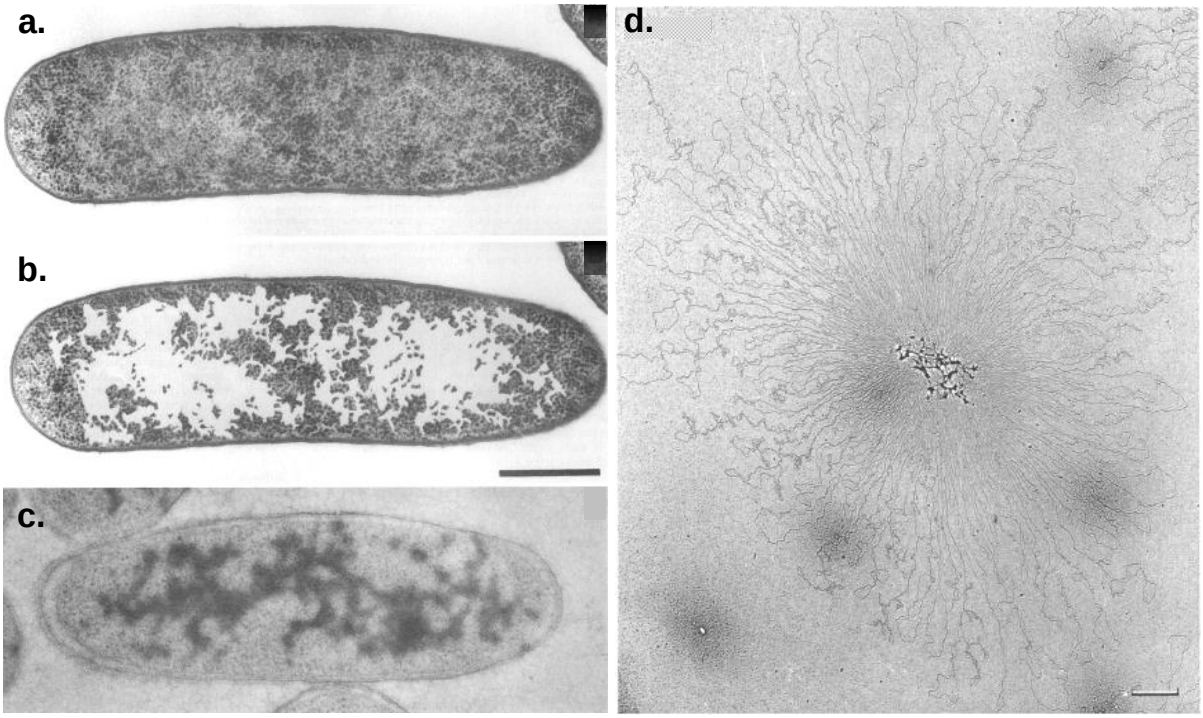

Figure 1.1: Electron micrograph images of the *E. coli* nucleoid

Electron micrograph images of the *E. coli* nucleoid. **a.** Thin section micrograph of growing *E. coli* preserved by freeze-substitution (CFS). **b.** The same section as **a.** however the ribosome-free spaces have been enhanced by hand coloring. **c.** A similar *E. coli* cell, immunostained specifically for DNA. The images in panels **a.** to **c.** demonstrate that the ribosomes of the cell are specifically segregated from the DNA-containing nucleoid (from Robinow and Kellenberger, 1994). **d.** Typical electron micrograph of an isolated and spread *E. coli* exponential-phase membrane-free chromosome. The chromosome had 141 ± 3 loops demonstrating supercoiling and radiating from a single, central, electron-dense core. (from Kavenoff and Bowen, 1976).

associated proteins (NAPs) and macromolecular crowding. Further to this small scale structuring, the *E. coli* chromosome is organised into micron-scale domains, referred to as macrodomains. While a severe compaction of the chromosome must be achieved, the nucleoid must also remain functional and accessible to allow the processes of replication, transcription, translation and recombination to be maintained. Hence the folded chromosome is dynamic and responsive to the ever-changing processes required for bacterial survival (reviewed in Thanbichler *et al*., 2005 and Dame *et al*., 2011).

1.2.2 DNA supercoiling

Sedimentation analysis by Worcel and Burgi (1972) first demonstrated that the *E. coli* chromosomal DNA is negatively supercoiled. Electron microscopy images of the *E. coli* nucleoid, produced by Kavenoff and Bowen (1976), also suggested that the chromosomal loops were inter-wound and supercoiled (Figure 1.1d). DNA supercoiling is an effective mechanism by which the volume of the DNA molecule is reduced. The relatively hydrophobic base pairs of DNA molecules avoid interactions with water by close stacking on top of each other. Due to the backbone phosphate bonds being relatively rigid, the molecule forms a right-handed double helix through twisting of the base pairs to allow optimal base stacking interactions. Therefore, these optimal interactions define the number of helical turns present in a DNA molecule of defined length, such as a bacterial chromosome, which is a closed circle of DNA (Watson and Crick, 1953; Benham and Mielke, 2005; Travers and Muskhelishvili, 2005). If the optimal number of turns were allowed then the molecule would be relaxed. However if the ends were rotated with respect to each other, before the molecule was covalently joined, then this would mean that the molecule was either overwound (positively supercoiled) or underwound (negatively supercoiled) and torsional strain would be applied to the molecule (Figure 1.2). The strain on the molecule would cause slight variations

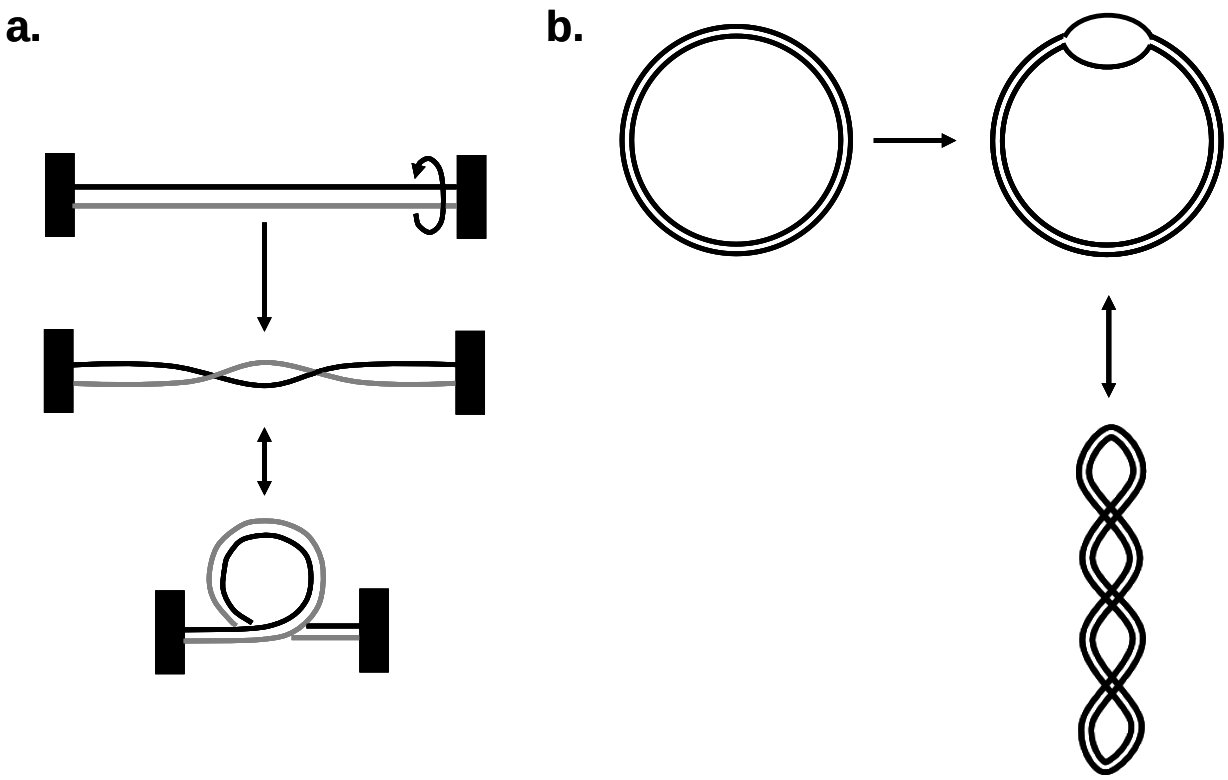

Figure 1.2: Schematic representation of DNA supercoiling

a. Representation of a double-stranded DNA molecule, with the thin black and grey lines each representing the single strands of DNA. Upon rotation by 360˚ the single strands of the molecule cross each other and become twisted. The twist is relieved by writhing of the molecule to allow the duplex to cross itself and become more compact. **b.** Representation of a circular, double-stranded DNA molecule, with each of the black lines representing a strand. Unwinding of the duplex to form a bubble increases the torsional strain in the remainder of the molecule, therefore leading to writhe of the molecule to inter-wind around itself.

in the number of helical turns, but the larger effect is to cause the DNA to writhe, which is when the double helix coils around itself to form a "coiled coil" (Benham and Mielke, 2005; Travers and Muskhelishvili, 2005; Figure 1.2).

Supercoils are introduced, and removed, by topoisomerases, which increase or decrease the number of times that the two strands of a DNA double helix are wrapped around each other. Sedimentation analysis of chromosomes, and chloroquine agarose gel electrophoresis of plasmid DNA by Pruss *et al*. (1982), demonstrated that the superhelicity of the *E. coli* chromosome is maintained by the opposing action of the DNA-relaxing topoisomerase, Topo I, and DNA gyrase. Topo I is known to relax DNA molecules, whereas DNA gyrase has been shown to introduce negative supercoils into DNA (Wang, 1971; Gellert *et al*., 1976; reviewed in Nöllmann *et al*., 2007). Transcription of the gene encoding Topo I, *topA*, is regulated through a complex homeostatic control mechanism to increase transcription of *topA* in response to increased DNA supercoiling. In contrast, transcription of the genes encoding the DNA gyrase subunits is increased in response to decreased levels of DNA supercoiling (Menzel and Gellert, 1983; Tse-Dinh and Beran, 1988).

Plasmid and chromosomal DNA in *E. coli* stationary phase cells becomes more relaxed (less super-coiled) than the negatively supercoiled state during exponential growth (Balke and Gralla, 1987). The mechanism behind this relaxation has yet to be explicitly investigated. However, DNA gyrase is the target of the *sbmC (gyrI)* gene product, which binds to and inhibits the action of DNA gyrase (Nakanishi *et al*., 1998). Expression of the gyrase inhibitor is increased upon growth transition to stationary phase and this increase of DNA gyrase inhibitor concentration is likely to be the causative agent of DNA relaxation in stationary phase cells (Baquero *et al*., 1995; Oh *et al*., 2001).

The work of Worcel and Burgi (1972) showed that the chromosome of *E. coli* is organised into topologically isolated, supercoiled loops. The presence of independently supercoiled

domains was confirmed by work on the closely related *Salmonella typhimurium*, using the supercoiling sensitive γδ resolution system (Higgins *et al*., 1996). These topologically isolated loops give rise to variation in effective superhelicity, which is a measure of the locally available superhelical density, therefore differing from the global superhelicity of the genome. Differences in effective superhelicity across the genome could have an effect on local cellular processes, such as transcription, replication and recombination. The compaction caused by supercoiling and the constraint of topologically isolated loops brings distant regions of the genome closer together and allows further compaction of the genome by the binding of NAPs (Travers and Muskhelishvili, 2005; Luijsterburg *et al*., 2008; Saier, 2008).

1.2.3 Nucleoid associated proteins

The idea of bacterial proteins with similar DNA-structuring functions to those of eukaryotic histones was first suggested by the isolation of the histone-like protein, HU, from *E. coli* extracts (Rouvière-Yaniv and Gros, 1975). The *E. coli* genome is associated with and structured by several major DNA-binding proteins, referred to as NAPs. The work of Azam *et al.* (1999) was the first to catalogue and characterise NAPs, measuring intra-cellular concentrations, binding affinities and preferred DNA binding sites of 12 NAP species. *In vitro* studies, utilising techniques such as atomic force microscopy (AFM), revealed that binding of individually purified NAPs can cause bending, wrapping, bridging and clustering of DNA, some of which can be seen in figure 1.3 (reviewed in Luijsterburg *et al*., 2006). It is by these mechanisms of DNA manipulation that NAPs contribute to the overall structuring and compaction of the *E. coli* chromosome.

The NAP, H-NS (histone-like nucleoid structuring protein), is commonly referred to as the "universal repressor" or "sentinel", due to its role in gene silencing (Browning *et al*., 2010). H-NS functions as a dimer or higher oligomers, interacting first with a single, high-affinity,

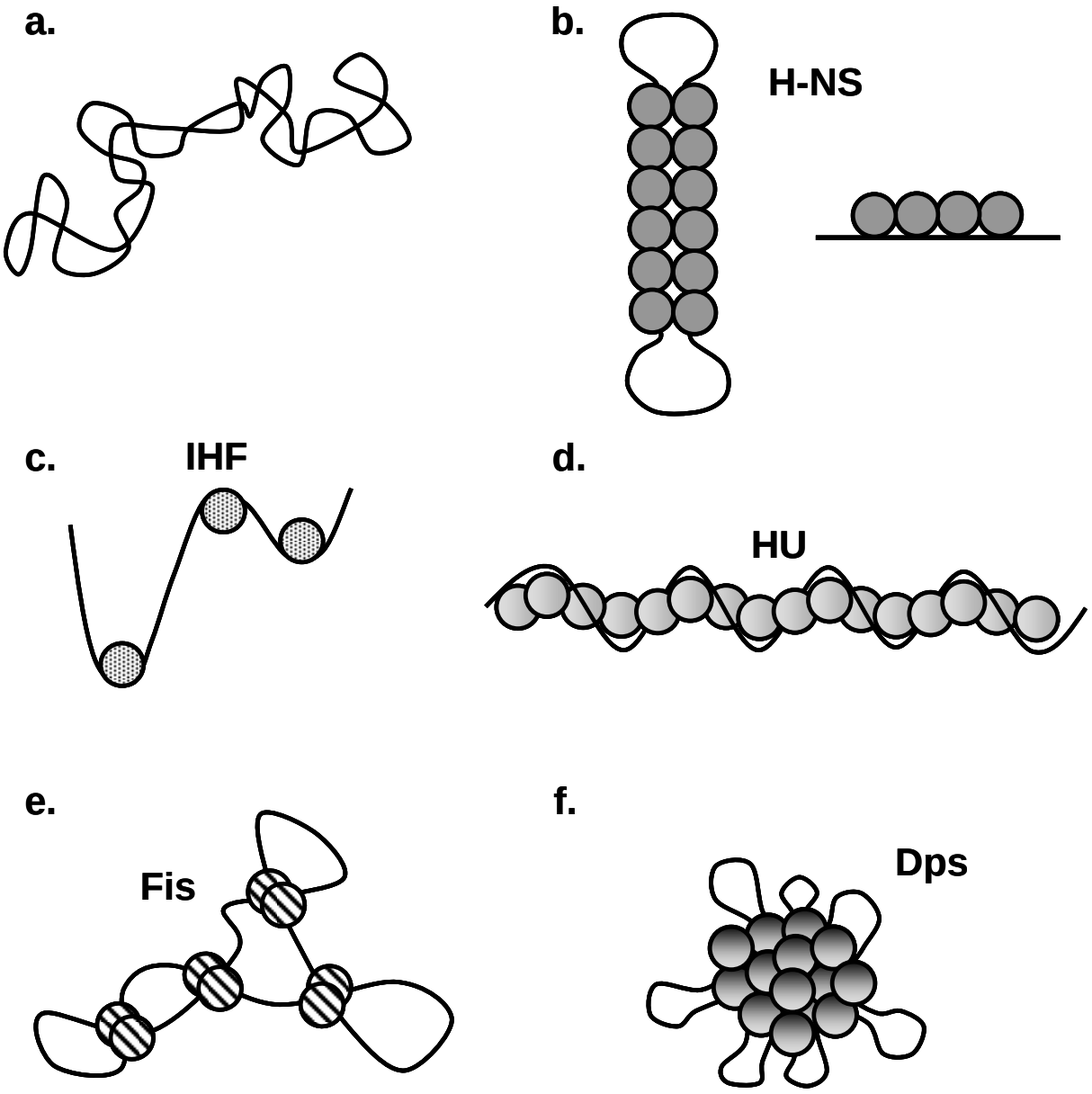

Figure 1.3: DNA structuring properties of nucleoid associated proteins

The diagram represents the DNA structuring properties exhibited by some *E. coli* nucleoid associated proteins (NAPs). Black lines represent DNA, whereas shaded spheres represent NAPs. **a.** Free supercoiled circular DNA. **b.** DNA-bridging properties of H-NS protein form a dumbbell structure out of small circular DNA, but can also form nucleo-protein filaments. **c.** Extreme bending of DNA, by 160˚, caused by binding of IHF. **d.** Formation of rigid filaments by HU binding at high concentrations. **e.** Binding of FIS at nodes where two DNA segments cross each other. **f.** Highly compacted Dps-DNA complexes caused by DNA-Dps co-crystallisation. Dps is highly abundant during stationary phase only.

DNA binding site followed by cooperative binding to multiple sites. This can lead to nucleo-protein filament formation, which was recently shown to be the mechanism behind gene silencing (Lim *et al*., 2012). H-NS can also bridge between DNA strands, therefore bringing distant regions closer together and compacting the DNA (Lang *et al*., 2007; Dame *et al*., 2000; Dame *et al*., 2005; Figure 1.3b). Recent work by Wang *et al*. (2011) utilised H-NS::fluorescent protein fusions and 3C technology (chromosome conformation capture) to show that H-NS forms two compact clusters *in vivo*, per chromosome. This work also demonstrated that H-NS sequesters operons, regulated by the protein, within these clusters, bringing broadly distributed regions of the chromosome into close proximity. Therefore this work emphasises the importance of H-NS in the structuring and organisation of the *E. coli* chromosome.

Wang *et al*. (2011) also analysed the distribution of the NAPs: FIS (factor for inversion stimulation), IHF (integration host factor), StpA (a paralogue of H-NS) and HU within the cell by tracking of fluorescently labelled proteins. However, FIS, IHF, HU and StpA were found to be dispersed throughout the *E. coli* nucleoid, despite StpA being a paralogue of the H-NS protein. Regardless of the lack of clustering, these proteins are still expected to contribute to chromosome compaction and structuring. For example, HU and its sequence-specific homologue IHF, bend DNA upon binding and, at high concentrations, HU forms nucleoprotein filaments, therefore counteracting the compaction caused by H-NS in supercoiled plasmid DNA (van Noort *et al*., 2004; Ljusterburg *et al*., 2006; Browning *et al*., 2010; Figure 1.3c and d).

Fis is another DNA bending NAP, which is the most abundant during exponential growth phase and has approximately 68,000 binding sites in the *E. coli* genome. However, the concentration of FIS protein drops to undetectable levels on entry into stationary phase (Azam *et al*., 1999). The pattern of decreasing FIS concentration upon entry into stationary phase is

the opposite to that of Dps (DNA protection during starvation) and CbpA (curved DNA binding protein) proteins, which are strongly induced during stationary phase (Azam *et al*., 1999). Dps forms sheets of dodecamers, with DNA threaded through the pores created by the structure. Therefore Dps tightly packages the chromosome to form Dps-DNA co-crystals, which protect the chromosome from damage (Ceci *et al*., 2004; Frenkiel-Krispin *et al*., 2004; Luijsterburg *et al*., 2006; Figure 1.3f). It is likely that all of the NAPs identified so far, and potentially those left undiscovered, contribute towards structuring the chromosome into higher order structures, such as topologically isolated loops and the micron-scale macrodomains.

1.2.4 Macrodomains

Work published by Boccard and co-workers, since 2004, has developed an excellent body of evidence in support of the *E. coli* chromosome being structured into micron-scale organisational structures, referred to as "macrodomains" (reviewed in Dame *et al*., 2012). The origin of this idea stems from work completed by Niki *et al*. (2000), in which fluorescence *in situ* hybridisation (FISH) was used to visualise the chromosome at approximately 230 kb intervals. The results of this study revealed that several DNA segments within a large 920 kb chromosomal section encompassing the origin of replication, *oriC*, localised to a similar position. Similar results were found for a second 920 kb chromosomal section encompassing the *dif* site, which is positioned directly opposite to *oriC* on the circular chromosome and is a target for the XerCD site-specific recombinase involved in the resolution of chromosome dimers (Louarn *et al*., 1994). These two large domains were referred to as the Ori and Ter domains. Work by Valens *et al*. (2004) confirmed the presence, and more precise boundaries, of the Ori and Ter macrodomains, by measuring the recombination efficiency of λ *att* sites scattered throughout the *E. coli* chromosome. Recombination between *att* sites is dependent

upon spatial proximity and recombination frequencies revealed that interactions were restricted to sites within the same sub-domain, with little interaction of sites in different domains. This analysis revealed two extra macrodomains, which flank the Ter macrodomain and are referred to as Left and Right, and two less structured regions (NSR and NSL) flanking the Ori macrodomain. The four macrodomains (Ori, Right, Ter and Left) are large and each contains approximately 1 Mbp of DNA, therefore forming the largest chromosomal sub-structures (Figure 1.4).

Further work from the Boccard group utilised arrays of binding sites for fluorescently labelled ParB, positioned throughout the chromosome, to track chromosomal loci throughout the cell cycle (Espeli *et al*., 2008). This work revealed that the macrodomains occupy distinct territories within the cell throughout the cell cycle with the Ori and Ter macrodomains positioned at mid-cell and the Left and Right domains positioned close to the poles in newly-divided cells. Newly replicated Ori domains migrate towards the cell poles while Ter domain segregation occurs shortly before cell division and the original positions are resumed in daughter cells (Figure 1.5). Constraints on mobility of the macrodomains were clear, however markers in the two less structured regions demonstrated a greater mobility than that of markers in the macrodomains, which could account for the greater frequency of interaction with other domains (Espeli *et al*., 2008).

Macrodomain structure and location within the cell was hypothesised to be due to organisational proteins, therefore *in silico* methods were used to search for domain-specific consensus binding motifs (Mercier *et al*., 2008). This work discovered a 13 bp motif, referred to as *matS*, which was found to be repeated 23 times specifically within the 800 Kb Ter macrodomain. Electrophoretic mobility shift assays using *E. coli* crude cell extracts helped to identify the product of the *ycbG* gene, renamed *matP*, as the factor specifically binding the *matS* site. The MatP protein was found to accumulate in the cell as a discrete focus that

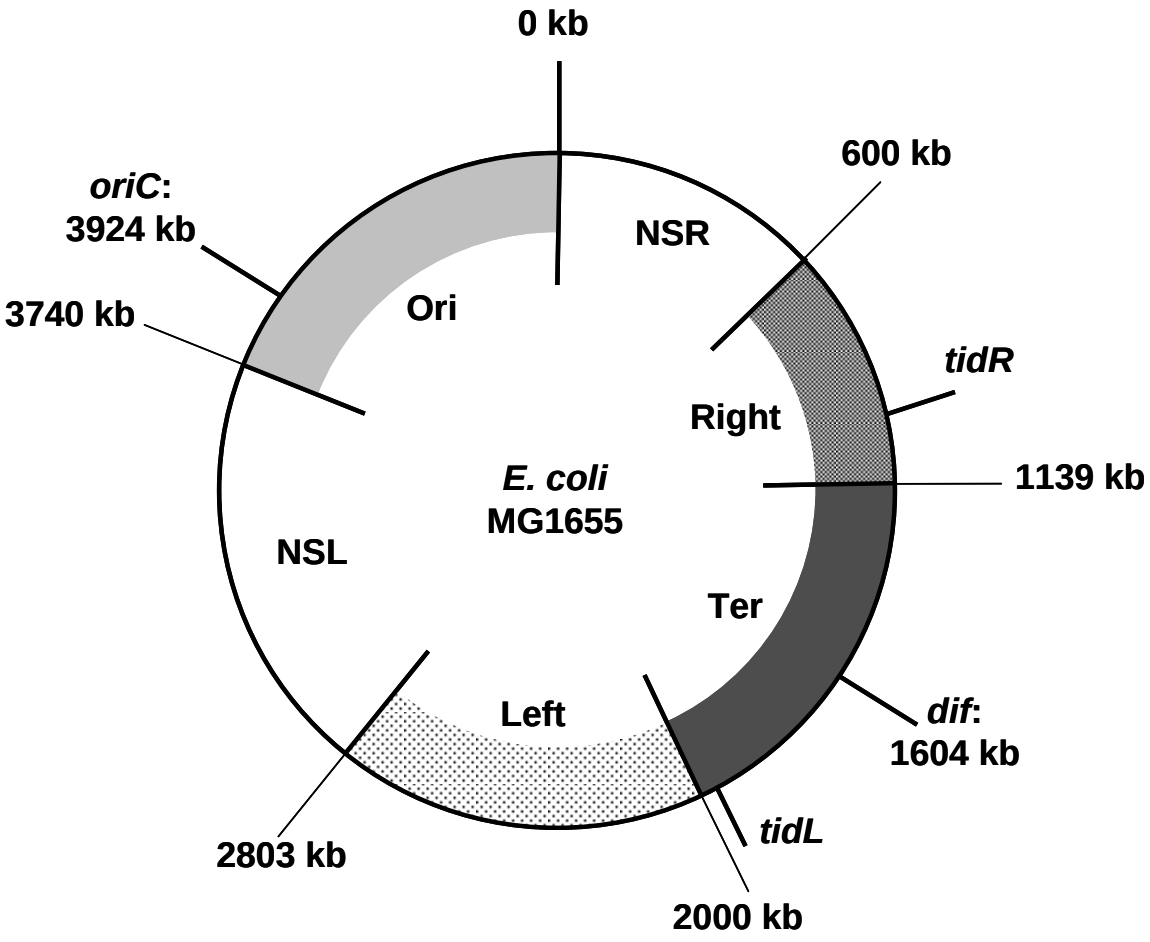

Figure 1.4: Macrodomain organisation of the *E. coli* chromosome

Circular map of the *E. coli* MG1655 chromosome, which is 4,639,221 bp in total. Shaded black arcs represent the positions of the four macrodomains (Ori, Right, Ter, and Left), while the gaps represent the less structured domains (NSL and NSR). Macrodomain boundaries are given in kb with respect to the coordinate system origin defined by Keseler *et al*. (2011) in the construction of the EcoCyc database. The position of the origin of replication (*oriC*) and *dif* site, involved in resolution of sister chromosomes, are indicated by black lines and labelled. Also, the positions of the *tidL* and *tidR* palindromes, involved in insulation of the Ter domain, are indicated. The figure is approximately to scale.

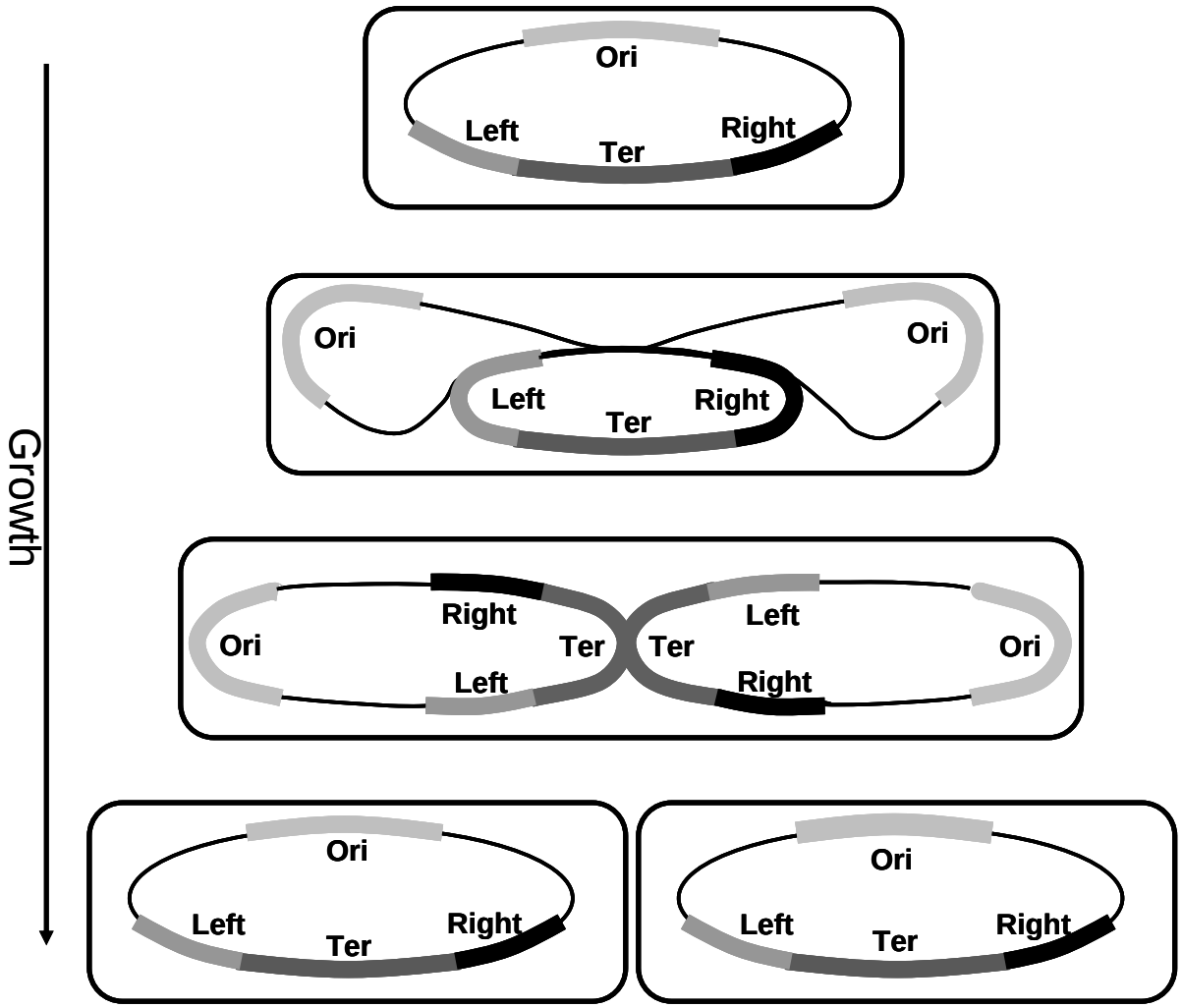

Figure 1.5: Localisation of *E. coli* chromosomal macrodomains throughout the cell cycle

Schematic representation of *E. coli* chromosomal macrodomain localisation throughout the cell cycle. Black rectangles represent the *E. coli* cell membrane, while the black oval represents the *E. coli* chromosome, with the positions of the macrodomains (Ori, Right, Left and Ter) indicated by thick shaded grey lines. The Ori and Ter macrodomains are positioned at mid-cell and the Left and Right domains positioned close to the poles in new-born cells. Newly replicated Ori domains migrate towards the cell poles while Ter domain segregation occurs shortly before cell division and the original positions are resumed in daughter cells. The Ter macrodomain is anchored to the septal ring through interaction between Ter-specific sequences, MatP protein and the divisome component ZapB. (Figure adapted from Dame *et al*., 2011).

co-localises with the Ter macrodomain. Deletion of the MatP protein causes the DNA to become less compacted, the markers within the domain to become more mobile and the Ter macrodomain to segregate earlier in the cell cycle (Mercier *et al*., 2008). Recent work from the Boccard group has solved the structure of the MatP protein. Visualisation of MatP-*matS* complexes by atomic force microscopy and electron microscopy has demonstrated that the protein flexibly links distantly located *matS* sites by formation of MatP tetramers through interaction of flexible C-terminal coiled-coil domains (Dupaigne *et al*., 2012). MatP localisation was analysed in strains containing mutations in proteins involved in formation of the division apparatus. This experiment showed that MatP localised the Ter macrodomain to mid-cell through interaction with *matS* sites and the septal ring protein, ZapB (Espeli *et al*., 2012). These results were confirmed through use of a two-hybrid assay demonstrating a strong direct interaction between ZapB and the C-terminal region of MatP. However, this interaction was only observed when ZapB was associated with the septal ring (Espeli *et al*., 2012). Interaction of the Ter macrodomain with the division machinery, via the MatP protein, causes constrained mobility of the domain. Finally, the Boccard group has identified two palindromic sequences, *tid*R and *tid*L, positioned in the Right and Left macrodomains, which are required to insulate the rest of the chromosome from the effects of Ter macrodomain interaction with the division machinery (Thiel *et al*., 2012; Figure 1.4). The *tid*R and *tid*L sequences are bound by the protein YfbV and this interaction was found to be required to maintain mobility of the less structured NS macrodomains (Thiel *et al*., 2012). This large body of work evidenced the mechanism by which the Ter macrodomain is structured, compacted, insulated and spatially linked to the cell division cycle.

While the Ter domain-specific MatP protein has been extensively characterised, organisational proteins for the other domains have yet to be characterised. However, macrodomain-specific DNA-binding properties have been shown for the SeqA and SlmA proteins (reviewed in Dame *et al*., 2012). The SeqA protein recognises pairs of

hemi-methylated GATC motifs in newly replicated DNA and prevents multiple initiation of chromosome replication from the origin. (von Freiesleben *et al*., 1994; reviewed in Waldminghaus and Skarstad, 2009). Recent ChIP-chip (ChIP – Chromatin immuno-precipitation followed by microarray) analyses of SeqA binding throughout the chromosome demonstrated that the origin of replication *oriC* is most densely populated with binding sites, however SeqA binding was detected at sites distal to *oriC* (Sánchez-Romero *et al*, 2010). Strikingly, the work of Sánchez-Romero *et al*. (2010) showed that SeqA binding is present within the Ori, Left, Right and less-structured macrodomains, however SeqA is entirely excluded from the Ter macrodomain. The work of Hiraga *et al*. (1998) utilised immunofluorescence microscopy to show that the SeqA protein forms discrete, localised, fluorescent foci, which are tethered at mid-cell until septal ring formation, after which foci migrate to the quarter cell position. It is also known that SeqA associates with the cellular membrane, therefore, due to the dynamic replication-dependent distribution of SeqA binding, SeqA may be involved in chromosome organisation during DNA replication (reviewed in Dame *et al*., 2011). The work reviewed in this section outlined how the *E. coli* chromosome is structured by DNA supercoiling, binding of NAPs and macrodomain scale organisation into the densely compacted structure, referred to as the nucleoid. While this compaction must be achieved, the chromosome must remain accessible to the transcription machinery in order for the process of gene expression to begin.

1.3 An overview of the regulation of transcription in *Escherichia coli*

The ability of bacteria to effectively and efficiently regulate cell physiology in response to environmental conditions is essential to survival. The necessary amount of the appropriate protein must be produced in response to the correct signal at the proper time. To achieve this gene expression is regulated in response to intra- and extra-cellular signals at every stage,

including rate of transcription, translation initiation, mRNA decay and the stability of the finished protein product. However, transcription initiation is the first and perhaps the most efficient stage at which the level of gene product can be modulated. The regulation of transcription involves interactions of the RNA polymerase with sigma factors, promoter sequence elements, transcription factors, small ligands, and chromosomal architecture (reviewed in Browning and Busby, 2004).

1.3.1 RNA polymerase

Eukaryotes contain multiple RNA polymerase (RNAP) enzymes, however in bacteria, a single multi-subunit RNA polymerase (RNAP) enzyme is solely responsible for all DNA-dependent RNA synthesis (Ebright, 2000). The number of RNAP molecules present in a single *E. coli* cell is dependent upon growth rate. This number has been estimated at approximately 4600 molecules per cell during growth with a doubling time of 51 minutes (Pedersen *et al*., 1978; Bremer and Dennis, 1996; Bakshi, 2012). The structure of core RNAP enzyme from *Thermus aquaticus* was determined by x-ray crystallography and correlates very well with low resolution cryoelectron microscopy images of *E. coli* core RNAP (Zhang *et al*., 1999; Finn *et al*., 2000). The core RNAP enzyme resembles a "crab-claw structure", similar to that of the yeast RNAP II, and has a molecular mass of approximately 400 kDa. The enzyme is comprised of the β and β' subunits, two identical α subunits and the ω subunit (Darst *et al*., 1998; Fu *et al.*, 1999; Zhang *et al.*, 1999; Figure 1.6). The pincers of the "crab-claw" are formed by the β and β' subunits and the active site of the enzyme is contained at the base of the cleft, which is formed in the jaw of the structure (Figure 1.6). The active site contains an essential catalytic Mg^{2+} ion chelated by the three aspartic acid residues of the conserved -NADFDGD- motif in the β' subunit (Zhang *et al*., 1999). During transcription, the channel containing the active site is occupied by the DNA template and the 3' end of the RNA transcript (Naryshkin *et al*., 2000).

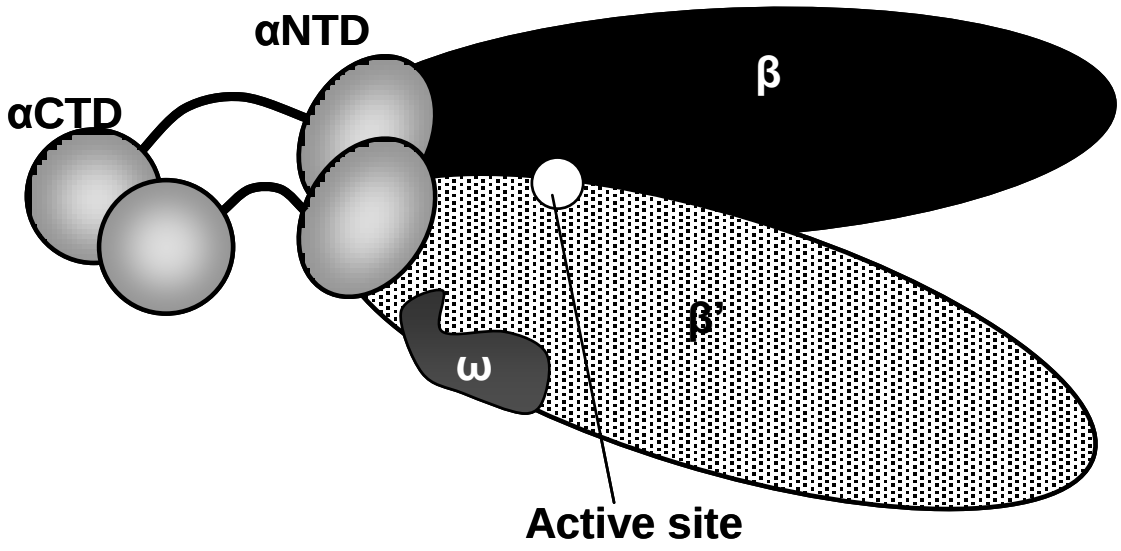

Figure 1.6: Schematic representation of RNA polymerase subunit architecture

Schematic diagram of the bacterial RNA polymerase core enzyme. The large β (black oval) and β' (dotted oval) subunits form a "crab-claw" structure with the active site (white circle) contained in the cleft formed by the jaws of the "crab-claw". The ω subunit (dark grey) is thought to be required for folding of the β' subunit to allow core enzyme formation. The two α-subunits (grey) dimerise and consist of two separate domains, the carboxy-terminal-domain (αCTD) and amino-terminal-domain (αNTD), which are connected by a flexible linker (black lines).

The two α subunits dimerise and each consists of two separate domains, the α subunit carboxyl-terminal-domain (αCTD) and the α subunit amino-terminal-domain (αNTD), the latter of which is responsible for assembly of the β and β' subunits (Blatter *et al*., 1994; Zhang *et al*., 1999). The αCTD is connected to the αNTD, and the rest of the enzyme complex, by a flexible linker and has an important role in promoter recognition during transcription initiation at certain promoters (Gaal *et al*., 1996; Murakami *et al*., 1996). The ω subunit is small and has no direct effect on transcription, however it is thought to function as a chaperone to assist in folding of the β' subunit (Mukherjee and Chatterji, 1997; Minakhin *et al*., 2001). The core enzyme is catalytically competent, however for transcription to be initiated at specific promoters, the core enzyme must interact with another subunit, the sigma subunit (σ), to form what is known as the holoenzyme (Burgess *et al*., 1969).

1.3.2 Sigma factors and promoter recognition

Most bacteria have multiple σ factors, which allow the global expression of different specialised regulons in response to different growth phases, external stimuli or stresses. The major σ factor of *E. coli*, σ^{70}, is referred to as the "housekeeping" σ factor due to the σ^{70}-containing holoenzyme being responsible for the majority of transcription during the exponential growth phase. *E. coli* has six alternative σ factors, which direct transcription of subsets of genes enabling the cell to respond and cope with different stresses (Ishihama, 2000; Gruber and Gross, 2003). For example the *E. coli* general stress sigma factor, σ^{38}, is responsible for transcription of genes required for survival during stationary phase (Lange and Hengge-Aronis, 1991; Loewen *et al*., 1998). The work of Kusano *et al*. (1996) shows that the stationary phase σ^{38}-containing holoenzyme is more efficient on relaxed DNA templates, therefore it is likely that relaxation of the chromosome during stationary phase plays a major role in transcription of the σ^{38} regulon (Kusano *et al*., 1996; Balke and Gralla, 1987). The

majority of σ factors are part of a family with homology to the *E. coli* primary σ factor, σ^{70} (Stragier *et al*., 1985; Gribskov and Burgess, 1986). Members of this family consist of up to four conserved, structurally autonomous domains, σ_1, σ_2, σ_3, σ_4, connected by flexible linkers and comprising conserved regions 1.1, 1.2–2.4, 3.0–3.1, and 4.1–4.2, respectively (reviewed in Lonetto *et al*., 1992; Campbell *et al*., 2002). The role of the σ subunit in transcription initiation is to recognise specific sequences at promoters, which will allow the holoenzyme to be positioned correctly at the target promoter, and to facilitate unwinding of double stranded DNA around the transcription start site (Gross *et al*., 1998; Murakami *et al*., 2002a; Murakami *et al*., 2002b).

Transcription initiation occurs by the recognition of specific DNA sequence elements within the upstream regulatory regions of genes, referred to as promoters, by the RNAP holoenzyme. Each *E. coli* cell encodes over 4000 genes, which are transcribed by the limited supply of RNAP molecules (Pedersen *et al*., 1978; Bremer and Dennis, 1996; Bakshi, 2012). Therefore, the sequence elements present at each promoter, and their similarity to the consensus, are a major contributor to the competition for limited RNAP holoenzyme molecules. The interactions of RNAP, carrying the major σ factor, with promoter elements have been elucidated by combining previous biochemical and genetic analyses (reviewed by Gross *et al*. (1998)) with structural data for different combinations of the RNAP, σ factor and DNA complexes (Murakami *et al*., 2002a; Murakami *et al*., 2002b; Vassylyev *et al*., 2002). The principal interactions at most promoters are between regions 2.4 and 4.2 of σ^{70} and promoter DNA sequence motifs, positioned 10 and 35 bp upstream of the transcription start site, respectively (Fenton *et al*., 2000; Campbell *et al*., 2002; Murakami *et al*., 2002b; Figure 1.7). These promoter elements have the consensus sequence 5'-TTGACA-n17-TATAAT-3', with similarity to the consensus defining the strength of promoter activity. However it should be noted that not all positions of the consensus sequence are equal in importance for promoter recognition (Rosenburg and Court, 1979; McClure, 1985; Miroslavova and Busby, 2006).

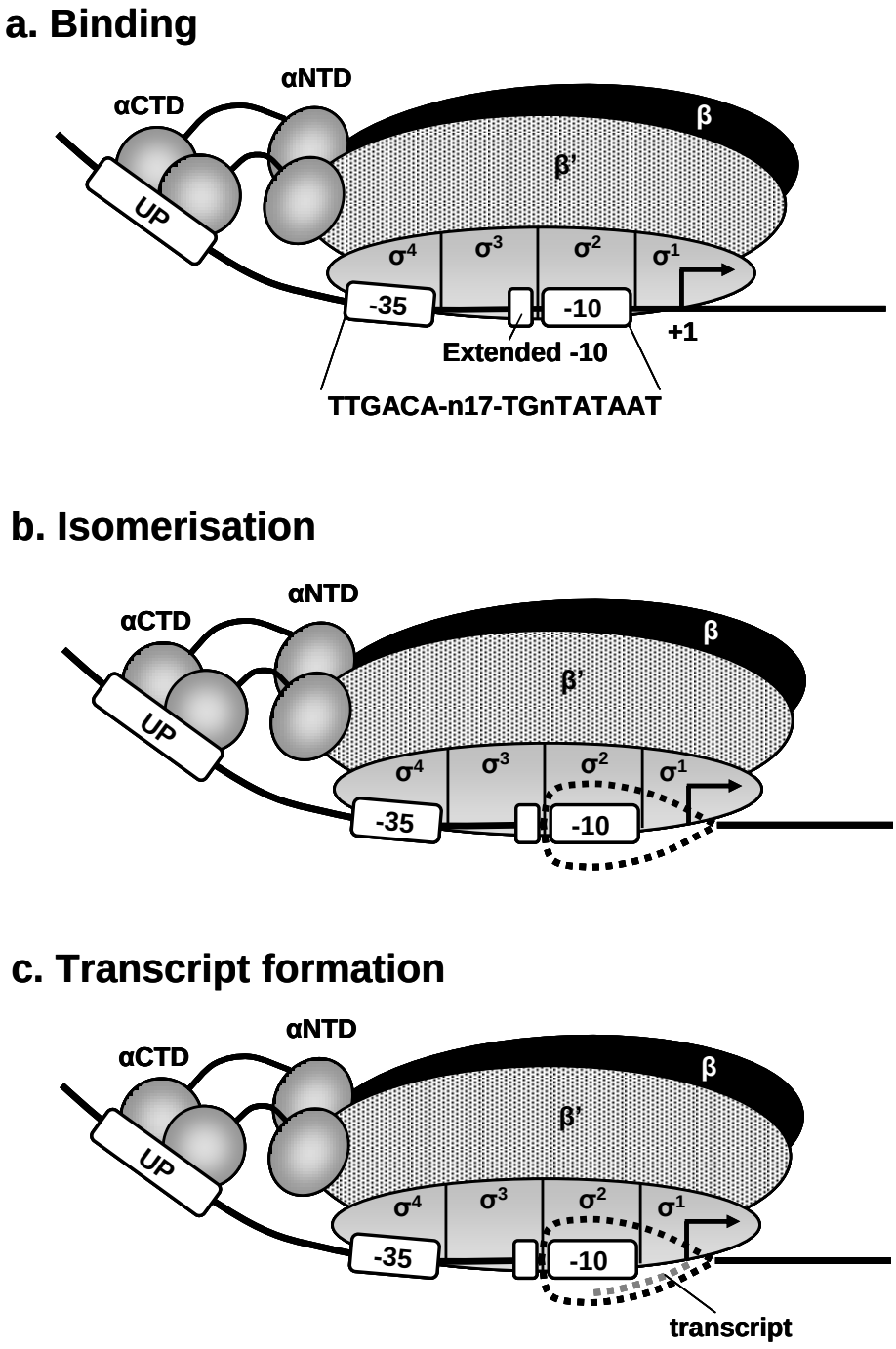

Figure 1.7: Promoter recognition and transcription initiation by RNA polymerase

Schematic representation of promoter recognition and transcription initiation by the bacterial RNA polymerase. The β (black oval) and β' (dotted oval) subunits, the active site (white circle) and two α-subunits (grey) of RNAP are represented and the promoter DNA is represented by a black line. **a.** The promoter -10, extended -10, and -35 elements (white boxes) are recognised by the σ subunit of the RNAP holoenzyme to form the promoter closed complex. The αCTD is attached to the αNTD and the rest of the holoenzyme by a flexible linker and recognises the promoter UP element. **b.** Isomerisation from the closed complex to the promoter open complex is signified by unwinding of the promoter DNA around the -10 and transcription start site (bent arrow). **c.** Transcript formation begins in the presence of nucleotide triphosphates.

After recognition of the -10 sequence by region 2.4 of the σ subunit, region 2.3 of σ initiates and stabilises DNA melting from position -11 bp to + 3 bp. The template strand is then directed into the main RNAP channel, with the transcription start site positioned at the active site ready for transcription to begin (Gross *et al*., 1998; Fenton *et al*., 2000; Campbell *et al*., 2002; Murakami *et al*., 2002b; Figure 1.7).

The sigma factor makes contact with one other promoter element at some promoters, which is the extended -10 element. The extended -10 promoter element is located immediately upstream of the -10 element and is 3-4 base pairs, with the consensus sequence TGn (Burr *et al*., 2000). This element is recognised by region 3.0 of the σ^{70} subunit, significantly increasing promoter strength, and is capable of compensating for poor -10 and -35 sequences at some promoters (Barne *et al*., 1997; Mitchell *et al*., 2003; Murakami *et al*., 2002a; Murakami *et al*., 2002b; Figure 1.7). In addition to contacts made by the σ subunit, the αCTDs of the RNAP holoenzyme interact with another promoter element at some promoters, the UP element, which is a ~ 20 bp A/T rich sequence located upstream of the -35 motif (Ross *et al*., 1993; Gourse *et al*., 2000; Figure 1.7). Many promoters are known to contain UP elements, however they are particularly associated with highly active promoters driving expression of genes encoding ribosomal and transfer RNA (Gourse *et al*., 2000).

In summary, the RNAP core enzyme must first interact with a σ factor before it is competent for promoter directed transcription. The αCTD of the RNAP enzyme can then interact with UP elements at some promoters to assist in recognition of the -10 and -35 promoter elements by region 2.4 and 4.2 of the σ^{70} subunit. Binding of the DNA double helix by RNAP holoenzyme results in what is referred to as the closed complex RP_c (reviewed in Browning and Busby, 2004; Figure 1.7). Region 2.3 of the σ^{70} subunit then promotes and stabilises unwinding of the DNA double helix around the transcription start site, allowing the template strand to be moved into the active site to form the open complex RP_o, a process referred to as

isomerisation (reviewed in Browning and Busby, 2004; Figure 1.7). The complex is then ready for RNA synthesis to begin, eventually leading to promoter escape and dissociation of the σ factor from the elongation complex.

1.3.3 Regulation of transcription initiation by transcription factors

Regulation of the *E. coli* transcriptional output is hierarchical, with sigma factors giving the upper level of hierarchical control in general response to stresses, growth phases or external stimuli. Subsets of promoters controlled by different sigma factors can be large and the next level of specificity in control of genes is achieved by interaction of transcription factors at specific subsets or individual promoters. Transcription factors are sequence-specific DNA-binding proteins, which bind to specific consensus sequences at or near to target promoters to up- or down-regulate transcription in response to specific signals (Ishihama, 2000; Lee *et al*., 2012). There are more than 300 genes in the *E. coli* genome coding for DNA binding proteins, which may represent the *E. coli* set of transcription factors, 35% of which are specifically activators of transcription, 43% repressors of transcription and 22% are dual regulators (Pèrez-Rueda and Collado-Vides, 2000). The promoters under control of one transcription factor are referred to as a regulon, however the majority of *E. coli* transcription factors act at only one or a few promoters. In contrast to this, seven global regulator proteins (CRP, IHF, FIS, FNR, NarL, ArcA and Lrp) have large regulons and, together, directly modulate the activity of 51% of *E. coli* promoters (Gottesman, 1984; Martinez-Antonio and Collado-Vides, 2003).

Binding of transcription factors at a target promoter can activate transcription from that promoter, however binding of transcription factors can also act to repress transcriptional activity. Activators of transcription typically function by direct protein-protein interaction

with the holoenzyme either to recruit RNAP to the promoter or by assisting open complex formation. Other activators function through alteration of promoter DNA conformation to allow more efficient binding by the RNAP holoenzyme (Ebright and Busby, 1995; Hochschild and Dove, 1998; Lee *et al*., 2012). Activation of transcription at many promoters works by a single transcription factor acting via one of 3 simple mechanisms often referred to as; class I or class II activation, or activation by induction of a conformational change in the target promoter (Lee *et al*., 2012; Figure 1.8). However, at many promoters, several factors interact and integrate different intra- or extra-cellular signals into transcriptional responses.

In the case of simple class I activation, the activator protein binds to a target DNA site upstream of the promoter -35 element (Figure 1.8a). This allows the activator to recruit holoenzyme to the promoter via direct protein-protein contact with the RNAP αCTD. The αCTD is connected to the αNTD by a flexible linker, which permits variation in the position of the activator binding site. However, activation can be lost if the binding site does not reside on the same face of the DNA helix as the -10 and -35 promoter elements (Gaston *et al*., 1990; Busby and Ebright, 1999; Browning and Busby, 2004). One of the best characterised examples of class I activation is the *lac* promoter, which is part of the paradigm for genetic regulation first used to describe the operon model by Jacob and Monod (1961). The *lac* promoter controls expression of genes required for the transport and metabolism of lactose in the absence of glucose. The three genes under control of this promoter are transcribed as a single mRNA and are *lacZ*, *lacY* and *lacA*, which encode the following proteins, β-galactosidase, *lac* permease and thiogalactoside transacetylase, respectively. Transcription of the *lacZYA* operon is driven by σ^{70} RNAP holoenzyme and is activated by the cyclic AMP receptor protein (CRP) binding at a DNA site centred at position -61.5 bp (Dickson *et al*., 1975; Simpson, 1985; Schmitz, 1981; Reznikoff, 1992; Figure 1.9). The CRP protein acts via a class I mechanism, directly interacting with the αCTD of RNAP to recruit the holoenzyme to the promoter (Meiklejohn and Gralla, 1985; Ebright, 1993). The CRP protein is only active

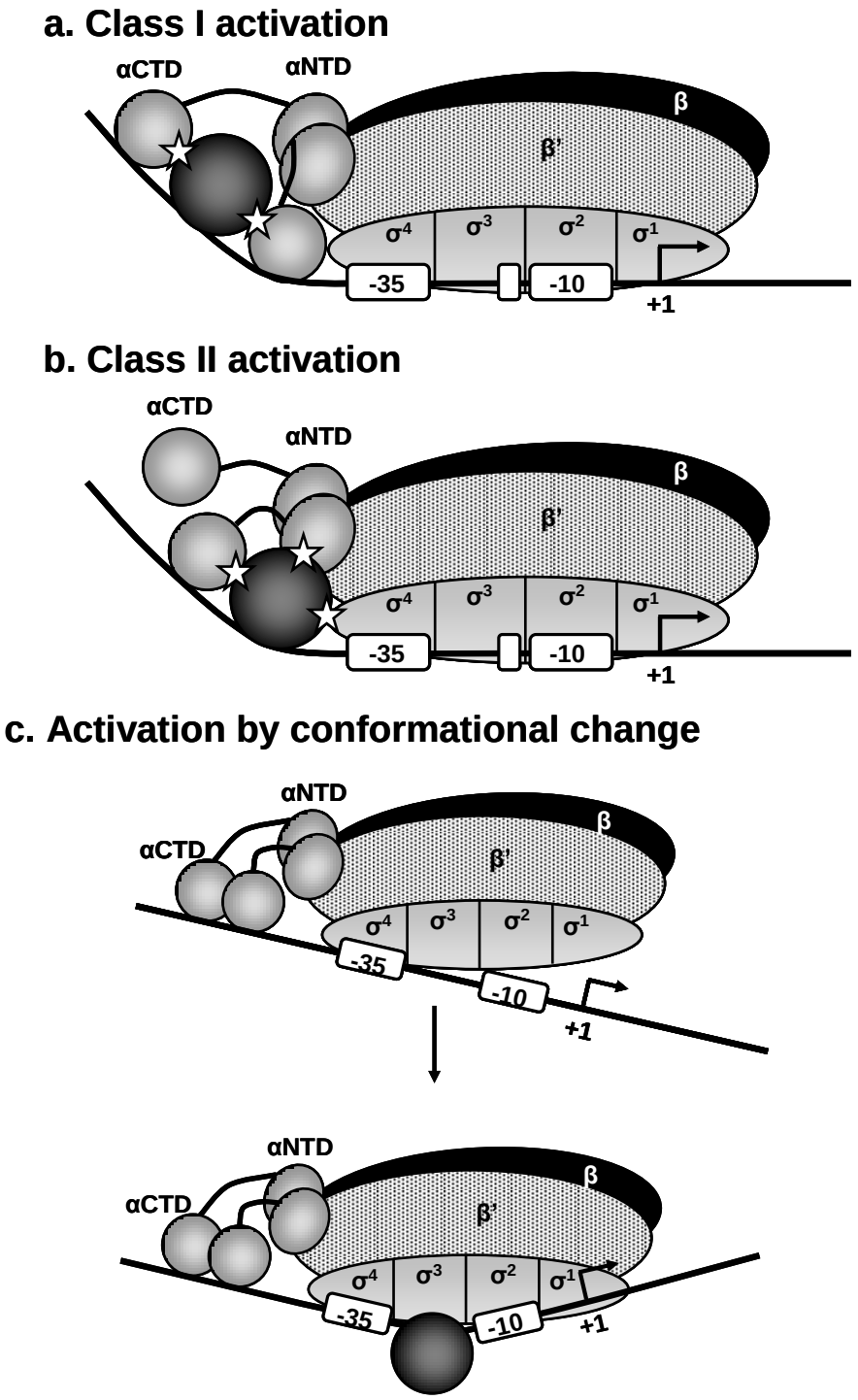

Figure 1.8: Simple mechanisms of transcription activation

Simple mechanisms of activation by a transcription factor. **a.** Class I activation. The activator protein (black sphere) binds a DNA site upstream of the -35 element (white box), usually at position -61.5 bp, and contacts (white stars) one or both α subunit caboxy-terminal-domains (αCTD) to assist in recruitment of RNAP to the promoter. **b.** Class II activation. The activator binds close to RNAP, usually at position -41.5 bp, and can make contacts with one of the αCTD subunits, the αNTD subunit or the σ subunit of the holoenzyme to assist in recruitment. **c.** Activation by conformational change. The activator binds to a site between the -35 and -10 promoter elements, inducing a conformational change in the promoter DNA, which allows full interaction of RNAP with the promoter. Figure adapted from Lee *et al*. (2012).

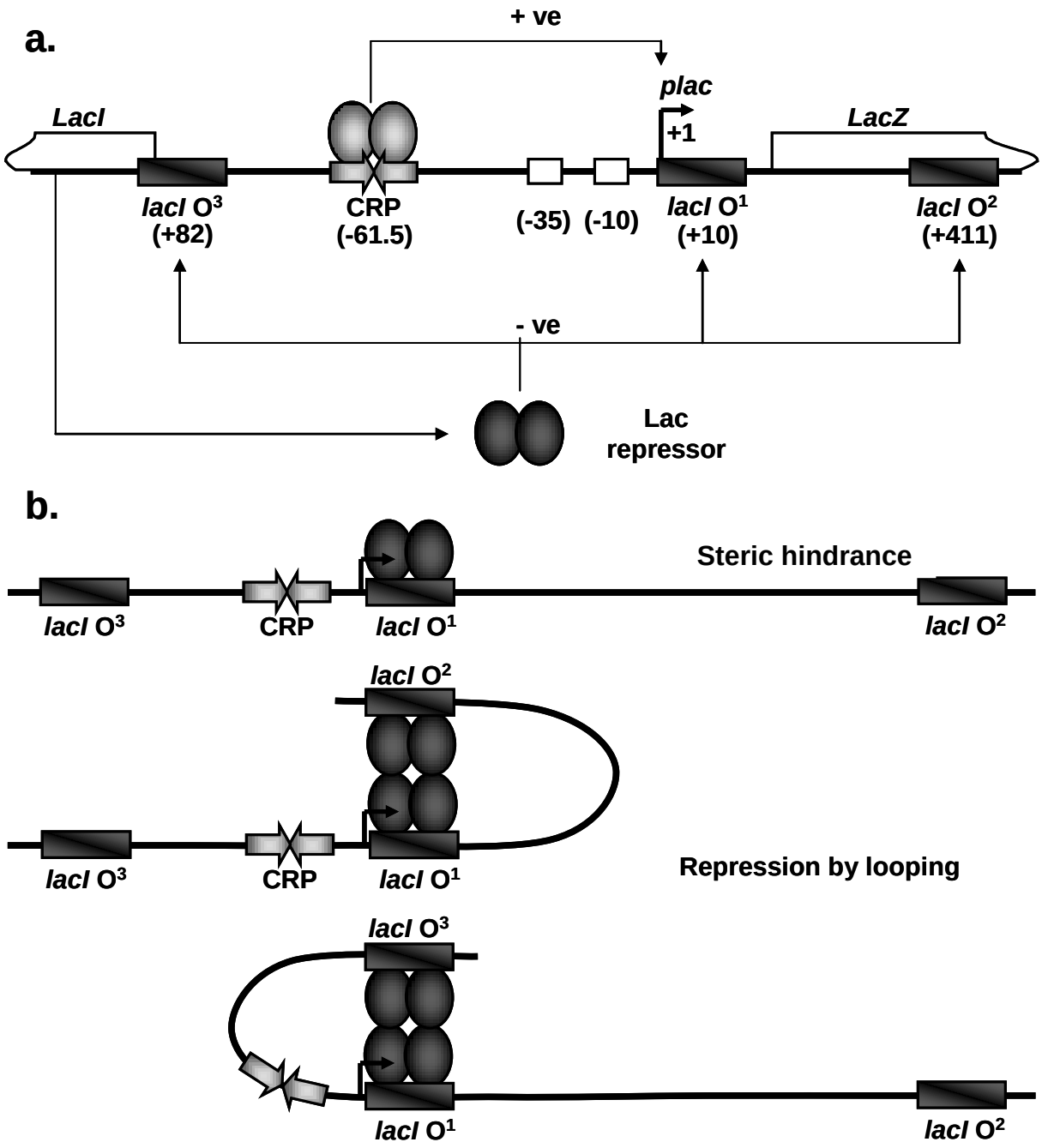

Figure 1.9: Regulation of the *lacZYA* promoter region by CRP and Lac repressor

a. Schematic representation of the *lacZYA* upstream regulatory region. The 3' end of the *lacI* and 5' end of *lacZ* genes (white boxes on top of DNA strand (black line)) flank the regulatory region. The *lac* promoter is activated by CRP (grey spheres) binding to a target DNA site (inverted grey arrows) centred at position -61.5 bp. The -35 and -10 promoter elements are represented by white boxes and the transcription start site by the bent arrow labelled +1. Promoter activity is repressed by binding of the Lac repressor (black sphere) to different combinations of the 3 DNA *lac* operator sites (black boxes). **b.** Repression of the *lacZYA* operon. The Lac repressor can bind to the O^1 binding site, overlapping the transcription start site, therefore blocking RNAP by steric hindrance. The Lac repressor can also bind as a tetramer to interact with O^1 and one of the auxiliary operator sites, O^2 or O^3, to cause looping. This blocks RNAP access to the promoter elements, therefore repressing transcription.

when bound by the inducer molecule cAMP, the production of which is repressed in the presence of glucose. Therefore activation of the *lac* promoter is directly linked to the absence of glucose through the interaction of cAMP, CRP and the *lacZYA* regulatory region (Pastan and Perlman, 1970; Busby and Ebright, 1999).

Class II activation is unlike class I activation in that there is less flexibility in the positioning of the activator binding site. This is because class II activation not only requires contact with the α-subunit, but also direct contact of the activator with domain 4 of the σ^{70} subunit of RNAP (Figure 1.8b). Therefore the binding site is usually centred at position -41.5 bp, relative to the transcription start point, overlapping the promoter -35 element (Gaston *et al*., 1990; Busby and Ebright, 1997; reviewed in Dove *et al*., 2003). During class II activation, the activator may also contact other subunits of the RNAP holoenzyme, for example the FNR (regulator of fumarate and nitrate reductase) protein can also make contact with the αNTD along with contacts to the αCTD and σ^{70} (Blake *et al*., 2002; Figure 1.8b).

In contrast to class I and class II activation, activation by conformational change usually requires binding of the activator protein to the spacer region between the -10 and -35 promoter elements instead of to sites upstream of the -35 element (Figure 1.8c). The spacing of the core -10 and -35 promoter elements is not optimal at these promoters, thus upon binding of the UP and -35 elements by the RNAP holoenzyme, the -10 element is misplaced and open complex formation cannot continue. Binding of the activator to a target site within the spacer region induces a twist that brings the -10 promoter element back into alignment with the -35 (Figure 1.8c). This allows promoter melting around the transcription start site through interaction of σ region 2 with the -10 element and subsequent RNA synthesis (reviewed in Brown *et al*., 2003 and Lee *et al*., 2012).

Transcription factors can also repress promoter activity by simple or complex mechanisms through interaction of one or many proteins in response to intra- or extra-cellular signals. Simple mechanisms of repression by a protein include the following; steric hindrance of either RNAP or an activator, DNA looping, or modulation of an activator, which all reduce the level of transcription initiation at the target promoter (reviewed in Browning and Busby, 2004; Figure 1.9b). Repression by modulation of an activator functions by direct protein-protein interactions, between the repressor and activator proteins, to stop interaction with RNAP holoenzyme (Browning and Busby, 2004).

An excellent example of a promoter repressed by steric hindrance and DNA looping is the previously mentioned paradigm of transcription regulation, the *lac* promoter. The *lac* promoter is repressed by the protein product of the *lacI* gene, which is located immediately upstream of the *lacZYA* regulatory region and is under transcriptional control of a weak constitutively active promoter (Jacob and Monod, 1961; Calos, 1978; Wilson, 2007; Figure 1.9b). The Lac repressor is a target for the inducer allo-lactose, which binds the repressor protein and reduces its affinity for the DNA binding site, therefore relieving repression of the *lacZYA* operon (Jobe and Bourgeois, 1972). In the absence of inducer, the Lac repressor binds the primary operator site (O^1) which is situated directly over the transcription start site, therefore sterically hindering RNAP binding to promoter DNA (Figure 1.9b). However, binding to this site alone is insufficient for complete repression of transcriptional activity. Two auxiliary "pseudo-operators" (O^2 and O^3) are situated within the *lacZ* and *lacI* genes and lac repressor tetramers bound at these sites can form "repression loops" by interaction with the primary operator O^1 (Figure 1.9b). Formation of DNA loops allows full repression of the operon by exclusion of the transcription initiation machinery from the promoter (Reznikoff *et al*., 1974; Oehler *et al*., 1990; Oehler *et al*., 1994; Wilson *et al*, 2007).

1.4 Effects of chromosome structure on transcription

1.4.1 Regulation of transcription initiation by nucleoid associated proteins

Three of the seven global regulators of transcription, FIS, IHF and Lrp, are also defined as NAPs. Recent ChIP-chip studies by Grainger *et al*. (2006) have shown that the majority of binding sites for the NAPs H-NS, FIS and IHF, are positioned within intergenic regions. Therefore it is unsurprising that NAPs not only structure the chromosome, but also affect transcription on a genome-wide scale, performing both roles simultaneously (reviewed in Dillon and Dorman, 2010). The NAP, IHF, of *E. coli* binds to a specific target consensus sequence and introduces a sharp 160˚ U-turn into the DNA, centred around the target binding site (Rice *et al*., 1996; Figure 1.3c). Binding, and bending, by IHF is essential for the regulation of approximately 120 genes in *E. coli* (Arfin *et al*., 2000). IHF can directly interact with the RNAP holoenzyme complex to regulate transcription. However, the majority of regulation, by IHF, is through modulation of promoter architecture to allow interaction between RNAP and distally bound activators (reviewed in Goosen and van de Putte, 1995). IHF can also function through another mechanism, which involves redistribution of energy stored in the IHF binding site to assist in promoter open-complex formation. For example, the *E. coli ilvGMEDA* operon is activated by binding of IHF, which displaces torsional energy stored in the negatively-supercoiled IHF binding site. Displaced torsional energy is transferred to the neighbouring promoter region, which is prone to base-pairing disruption. The single-stranded bubble is then formed around the transcription start site, therefore allowing promoter open complex formation (Sheridan *et al*., 1998; reviewed in Dillon and Dorman, 2010).

The FIS protein, like IHF, can also affect transcription initiation through a range of different mechanisms, including direct interaction with RNAP and manipulation of local superhelicity (reviewed in Browning and Busby, 2010). For example, binding of the FIS protein to

helically-phased sites, positioned upstream of the *tyrT* promoter, stabilises a DNA microloop, constraining a microdomain of negative superhelicity at the promoter. Stabilisation of local negative superhelicity facilitates open complex formation and therefore transcription of *tyrT* (Auner *et al*., 2003; Dillon and Dorman, 2010). As at the *tyrT* promoter, local DNA superhelicity can affect the activity of many promoters throughout the *E. coli* chromosome. Variation in local negative superhelicity (over or under-twisting) can affect the position of promoter elements with respect to each other, much like the mechanism of activation by conformational change, as described in section 1.3.3 (Lee *et al*., 2012; Figure 1.8c). Whether this conformational change will affect transcription initiation positively or negatively depends upon the specific structure of individual promoter elements.

1.4.2 DNA supercoiling effects on transcription

Peter *et al*. (2004) utilised inhibitors of DNA gyrase and Topo IV to reduce the negative superhelical density of the *E. coli* chromosome, causing the molecule to become more relaxed. Microarray analysis was used to measure changes in transcriptional output. This work revealed that transcription of 106 genes was up-regulated and 200 genes were down-regulated in response to inhibition of DNA gyrase. Blot *et al*. (2006) utilised a microarray-based approach to assess the combined effects of global relaxation of chromosomal superhelicity and mutations in the *fis* and *hns* genes. This work confirmed that the transcriptome of *E. coli* is sensitive to relaxation of negative supercoiling, but also demonstrated that the spatial distribution of superhelicity is dependent upon NAPs. As described in section 1.1., the nucleoid is structured into dynamic, topologically isolated, supercoiled domains, which were recently found to be approximately 10 kb in size (Postow, 2004; Figure 1.1). NAPs have long been proposed as the organising barriers to supercoil diffusion, due to their DNA architectural role, which is in agreement with the work of Blot *et*

al. (2006). These barriers were potentially identified by the work of Vora *et al*. (2009) that sought to identify the protein occupancy landscape of the *E. coli* genome. Formaldehyde cross-linking was used to stabilise protein/DNA complexes within the cell, which were released by lysis and sonication then reduced to ~ 50 bp footprints with DNase I treatment. Protein/DNA complexes were then extracted from the interface between the aqueous and organic phases of a phenol-chloroform extraction. Sections of DNA bound by large quantities of protein were then de-crosslinked and identified by use of microarray, which identified extended protein occupancy domains (EPODs) of the chromosome (Vora *et al*., 2009). These are large regions of the chromosome, between 1 and ~ 14 kb, contiguously bound by protein. This data was compared to transcription profiling data and RNAP ChIP-chip data presented by Grainger *et al*. (2005) to demonstrate that the EPODs were either highly expressed, in which case most of the protein could be RNAP, or transcriptionally silent (tsEPODs) (Vora *et al*., 2009). Comparison of the tsEPOD data with the NAP ChIP-chip study of Grainger *et al*. (2006) revealed that tsEPODs were enriched for NAP binding sites. Therefore the tsEPODs are suggested to be the organising centres of the chromosome, acting as barriers to diffusion of negative superhelicity.

1.4.3 Transcription factories and uneven distribution of RNA polymerase

A review by Cook, (2010) proposed that RNAP could also be the driving force behind formation of topologically isolated, superhelical chromosomal loops. Cook based this hypothesis primarily on the observation that the *E. coli* nucleoid is unfolded after treatment with rifampicin, an inhibitor of RNAP, or with RNase (Dworsky and Schaechter, 1973; Kavenoff and Bowen, 1976; Cook, 2010). Cook proposed a model of organisation for all species in which RNAP is static, pulling the DNA template through the complex instead of tracking along it, and clustered into "transcription factories" (Figure 1.10). Transcription

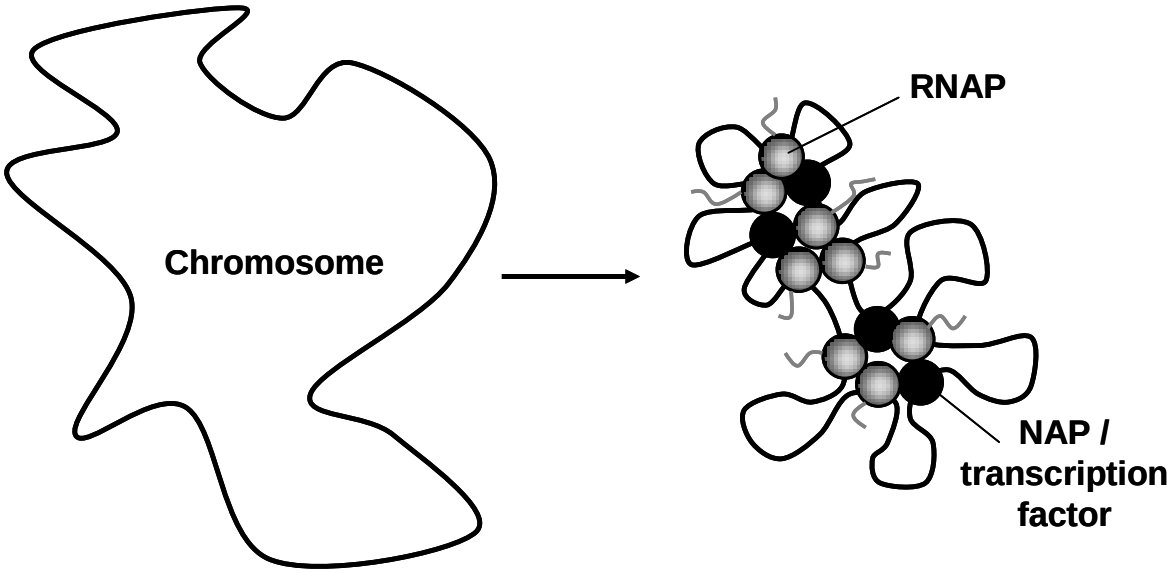

Figure 1.10: Model of chromosome structure by formation of transcription factories

Aggregation of actively transcribing RNA polymerase molecules (grey spheres) and nucleoid associated proteins (NAPs), or transcription factors, would lead to compaction of the chromosome. Stationary, inactive, polymerase molecules bound to DNA would also serve to compact the chromosome. This model of compaction would be dynamic and potentially contribute to the regulation of transcription as promoters are pulled into closer proximity to a transcription factory. Adapted from Cook, (2002).

factories could include transcription factors, topoisomerases and RNAP molecules gathered in one specific locus. Evidence for transcription factories is observed during fast growth of *E. coli*, in which fluorescently labelled RNAP forms distinct clusters on the periphery of the nucleoid (Cabrera and Jin, 2003; reviewed in Jin and Cabrera, 2006; Figure 1.11a). However, no obvious clusters of RNAP are observed during moderate growth (Jin and Cabrera, 2006; Bakshi *et al*., 2012). Chromosomal superhelicity can be affected by the process of transcription, with the actively transcribing elongation complex causing positive supercoiling ahead and negative supercoiling behind the complex (Wu *et al*., 1988; Tsao *et al*., 1999). Therefore transcription factories would greatly influence chromosomal architecture, due to their tethering effect and the effect on DNA superhelicity. Transcription factories could also play a role in the variation in local DNA superhelical density throughout the chromosome, which was reported by Postow, (2004). Therefore activity of a supercoiling-sensitive promoter could be affected by changes in location within the chromosome and promoters in closer proximity to a transcription factory would be more likely to initiate transcription.

1.4.4 The localisation of transcription and translation within the cell

The hypothesis that some chromosomal positions may be more permissive for transcription than others has been around since the early years of study into *E. coli* nucleoid structure. Ryter and Chang (1975) visualised the sites of active transcription in cells supplemented with radioactive uracil by high resolution autoradiography. This study revealed that the sites of active RNA synthesis are located on the nucleoid periphery and far out into the cytoplasm. However, this work found no sign of transcription within the nucleoid area and the authors proposed that transcription would occur on genes spread out into the cytoplasm (Figure 1.11a).

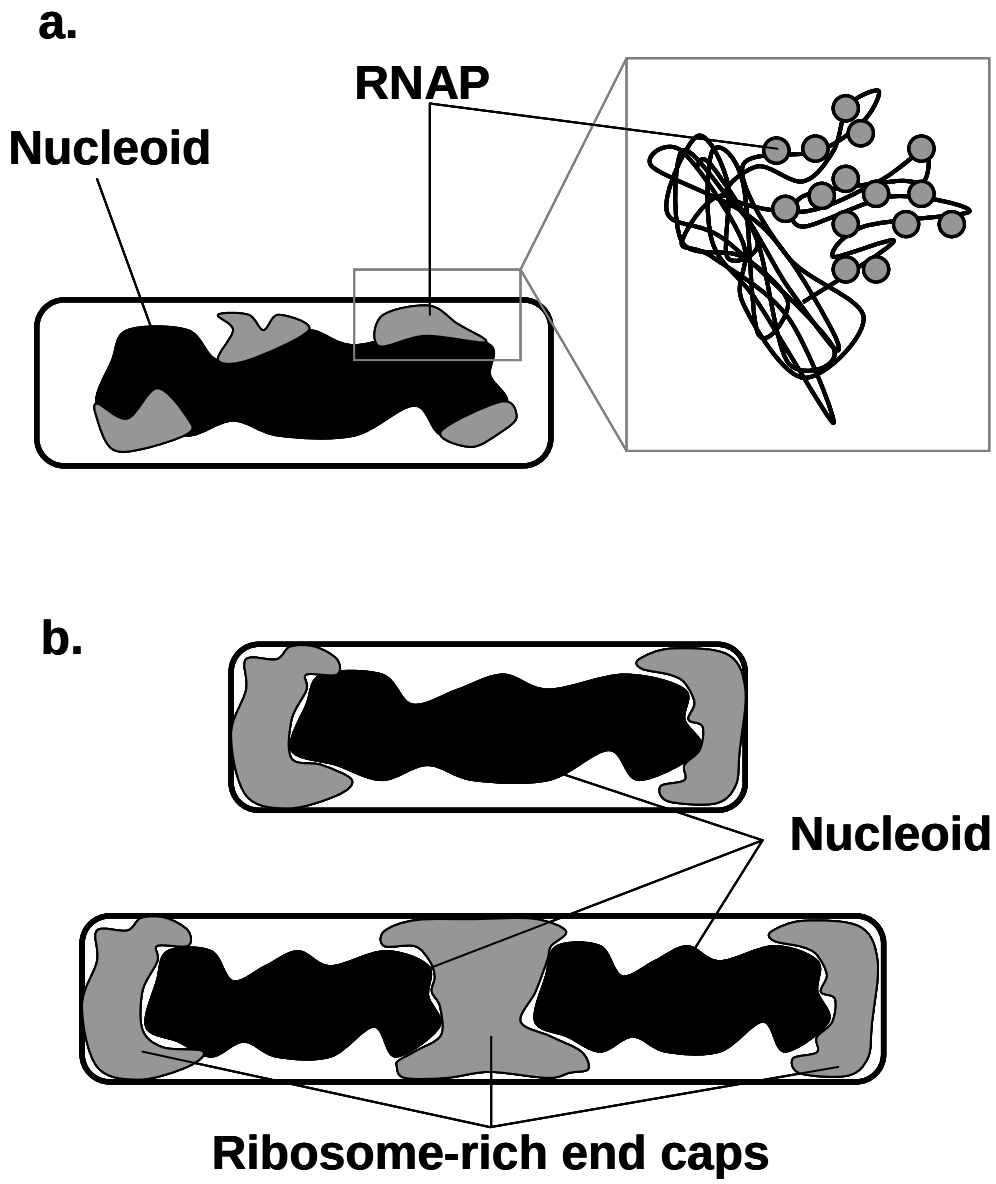

Figure 1.11: Uneven distribution of RNA polymerase and ribosomes in *E. coli*

a. Schematic representation of fluorescence microscopy images of GFP-tagged RNA polymerase (RNAP) in fast growing *E. coli*. RNAP aggregates into clusers (grey areas) situated on the nucleoid (black area) periphery and out into the cytoplasm. The model to explain this suggests that DNA loops are extended into the cytoplasm during transcription by clustered RNAP molecules. Therefore loops located far fro each other in the genome are brought into close proximity by active transcription. Figure adapted from Jin *et al*. (2006). **b.** Schematic representation of super-resolution microscopy images of fluorescently labelled ribosomes, collected by Bakshi *et al*. (2012). Ribosomes are observed to be strongly segregated from the nucleoid and form clusters at the cell poles. Distributions of RNAP and ribosomes were observed to extend up to the cytoplasmic membrane, however RNAP was not observed near the membrane of the end-caps (Bakshi *et al*., 2012).

The work of Ryter and Chang (1975) correlates with recent studies by Cabrera and Jin, (2003) and Bakshi *et al*. (2012), which visualised the position of fluorescently labelled RNAP using fluorescence microscopy and super-resolution microscopy, respectively. The super-resolution study by Bakshi *et al*. (2012) found that RNAP localised to the nucleoid lobes and also extended throughout the cytoplasm to the cytoplasmic membrane. However, due to the cell being a three-dimensional shape, the study could not distinguish whether large numbers of RNAP molecules were within or merely on top of the nucleoid in the images produced. When taken with the observations of Ryter and Chang (1975), these data suggest that RNAP is excluded from the nucleoid. Further to this, the nucleoid is a space devoid of ribosomes, as reviewed by Robinow and Kellenberger (1994), and recently demonstrated by super-resolution microscopy studies of fluorescently labelled ribosomes in *E. coli* (Bakshi *et al*., 2012). The work of Bakshi *et al*. (2012) discovered a strong segregation between the nucleoid and ribosome rich regions, with the ribosomes forming dense clusters at the cell poles. This would suggest that genes positioned near to the nucleoid periphery would be more accessible to the translation machinery than those in the centre of the compacted nucleoid. Therefore certain positions within the *E. coli* chromosome are expected to be more conducive to transcription initiation than others.

1.4.5 Previous studies on position-dependent modulation of promoter activity

The effect of position within the chromosome on promoter activity has been investigated by only a few research groups (Beckwith *et al*., 1966; Schmid and Roth, 1987; Sousa *et al*., 1997), with one recent study being published during the course of this Ph.D. thesis (Block *et al*., 2012). Beckwith *et al*. (1966) utilised recent advances, which allowed transposition of *E. coli* genes to new chromosomal locations, to study the transposed *lac* operon. Expression of the *lac* operon from an F'*lac* episome integrated at 10 different chromosomal loci, some of

which were in both orientations, was studied during growth on minimal medium. Beckwith *et al*. reported up to two-fold differences in expression of the *lac* genes and an increase in gene expression with proximity to the origin of replication. Later work showed that during rapid bacterial growth, the chromosome may have multiple replication points, which all initiate from the single origin of replication (Caro and Berg, 1968; Wolf *et al*., 1968; Figure 1.12). Therefore the average dosage of chromosomal positions is expected to be different, with more copies of the origin proximal area than that of the replication terminus (Figure 1.12). This effect is referred to as gene dosage and the results of Beckwith *et al*. (1966) correlate with the increased gene copy number expected by the Helmstetter and Cooper gene dosage model (Helmstetter and Cooper, 1968).

Schmid and Roth (1987) demonstrated similar results for the *his* operon in *S. typhimurium*. The first four genes and the promoter of the *his* operon, *hisGDCB*, flanked by Tn*10* insertion elements, was translocated to 16 known Tn*10* insertion sites throughout the chromosome. Expression of the *hisD* gene product was measured during different growth conditions and the results correlated with that expected due to gene dosage effects, expression was reduced, as expected, at slower growth rates. Sousa *et al*. (1997) also sought to assess the effects of position within the chromosome on promoter activity. The regulator of the *nah* genes of *P. putida* was fused to *lacZ* and carried by a mini-Tn*5* transposon to various locations on the *E. coli* chromosome. Like the studies of Beckwith *et al*. (1966) and Schmid and Roth (1987) an increase in gene expression, with increasing proximity to the origin of replication, was noted. This work also attempted to measure local DNA supercoiling by inserting the supercoiling sensitive *gyrB* promoter, fused to *lacZ*, into each of the chromosomal loci. No correlation between position-based modulation of gene expression and local superhelicity could be made. However this work did not consider that the *gyrB* promoter may also be sensitive to other local chromosomal factors, such as NAP binding.

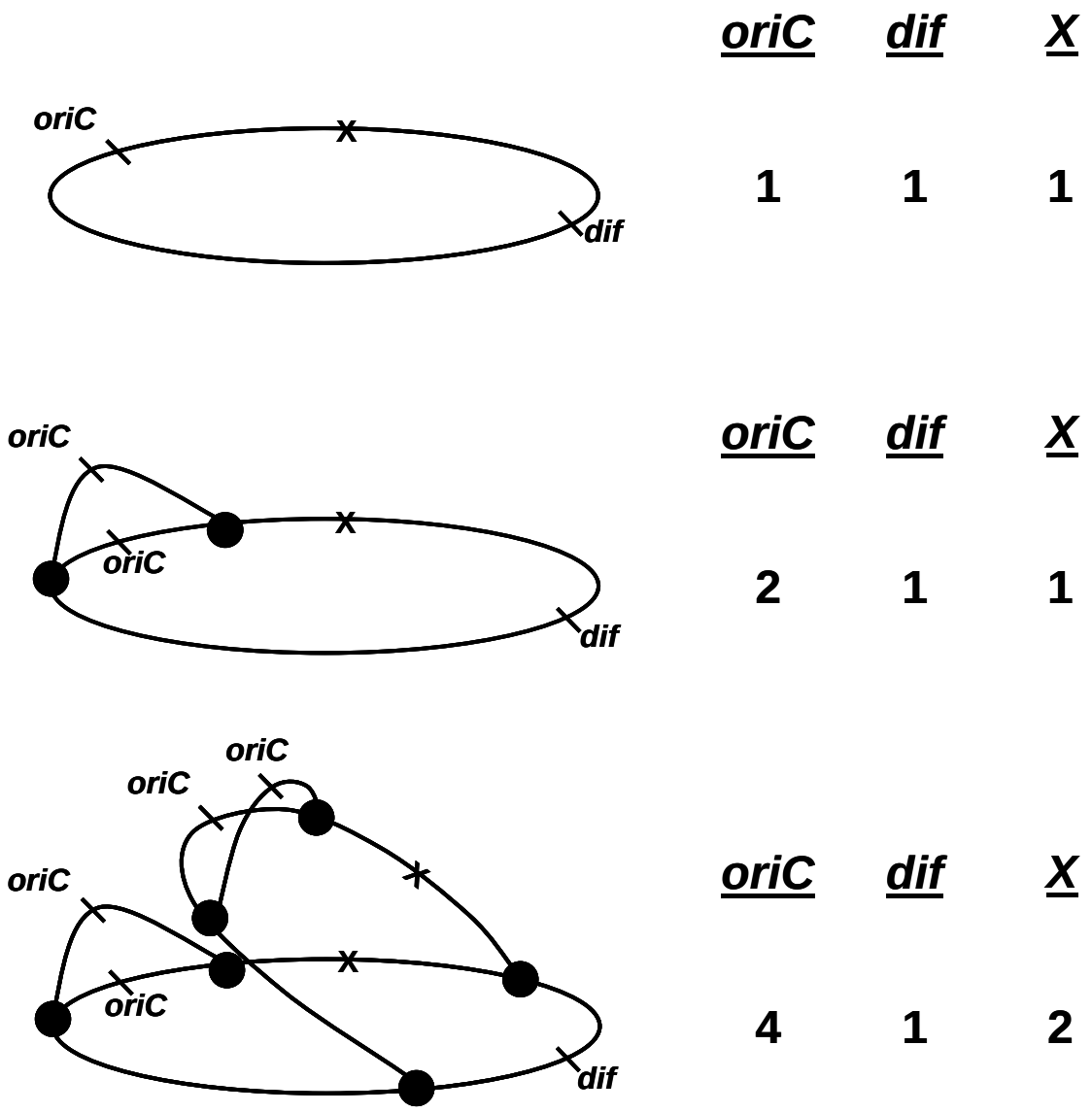

Figure 1.12: Schematic representation of the gene dosage model

Schematic representation of *E. coli* chromosome replication and the gene dosage model. The oval represents the *E. coli* chromosome, with the positions of the origin of replication (*oriC*), *dif* site and random position X marked. The *dif* is involved in separation of sister chromosomes after replication. Early in the replication process the replication forks (black circles) have not reached position X, but have duplicated the origin proximal region, therefore creating more copies of *oriC* than position X or the *dif* site. During rapid growth multiple replication forks can form, therefore in the scenario at the bottom of the figure there are four copies of *oriC* and two copies of position X for each copy of the *dif* site. Figure adapted from http://www.science-projects.com/V10A.htm.

The most recent attempt to assess position-dependent modulation of gene expression was completed by Block *et al* (2012). The exceptionally strong P_L promoter of phage lambda was modified, by Lutz and Bujard (1997), to be regulated by the Lac repressor, therefore allowing induction of promoter activity by the addition of IPTG to the medium. The modified P_L promoter, fused to the *yfp* gene, was flanked by transcription terminators and inserted into the coding region of six different pseudogenes and at two reference loci within the chromosome of *E. coli* (Block *et al*., 2012). Gene expression was analysed during various growth rates, however the pattern of expression was as expected from the Helmstetter and Cooper gene dosage model (Helmstetter and Cooper, 1968). This experiment was repeated without the flanking transcription terminators and some anomalies to the gene dosage model were discovered. However, anomalies to the gene dosage model were attributed to "local sequence effects", such as clashing elongation complexes, mRNA degradation or reduced translation (Block *et al*., 2012). The work of Block *et al*. (2012) also sought to analyse the effect of varying distance between the gene encoding a transcription factor and the target promoter, however no effect was detected.

These studies used approaches to move the same promoter and reporter to various chromosomal locations to answer the question of whether location affects transcription level. However several features of these experiments need to be highlighted. For example, the promoter elements in all of the above experiments were flanked by the vehicles used to insert them into the genome, thus, the *lac* promoter was contained within the F'*lac* episome, the *hisGDCB* operon was flanked by two Tn*10* insertions and the *psaI-lacZ* reporter was contained within Tn*5*. All of these elements, along with the general large size of the inserts, could serve to shield the regulatory elements under study from location-specific modulation effects. Another shortcoming of these studies is that the insertion location of the Tn*5* and Tn*10* elements is biased, therefore not allowing direct targeting of the probe to specific locations. Further to this, the inserted pieces of DNA are relatively large, thus potentially

altering the local chromosomal architecture purely by their presence (Beckwith *et al*., 1966; Schmid and Roth, 1987; Sousa *et al*., 1997).

1.5 Aims and outline of the project

A study was required in which a small promoter::reporter probe would be inserted into the genome, at several locations, utilising a recent homologous recombination technique to allow greater accuracy in targeting and choice of insertion locus. The non-coding DNA between convergent genes should be targeted, with a natural *E. coli* promoter fused to a reporter, in order to minimise disruption to neighbouring genomic processes, such as transcription. Therefore chapter 3 describes an investigation into the effects of position within the chromosome of *E. coli* on promoter activity during exponential growth and the relation of gene dosage to position-based modulation of promoter activity was also determined. The effect of position within the chromosome on promoter activity has not previously been assessed during stationary phase, however *E. coli* do experience conditions of substrate limitation and starvation (Beckwith *et al*., 1966; Schmid and Roth, 1987; Sousa *et al*., 1997; Block *et al*., 2012; Navarro-Llorens *et al*., 2010). Consequently, a second aim of the project was to measure position-based effects on promoter activity during stationary phase.

The third aim of this project was to describe the mechanisms of position-based modulation of promoter activity. Therefore chapter 4 describes studies into potential mechanisms of position-dependent variation of promoter activity, including orientation with respect to replication, neighbouring promoter activity, local superhelical density and positioning within tsEPODs. Finally, plasmid-based reporter gene technology is utilised on a regular basis for the characterisation of promoters. Therefore chapter 5 describes experiments to compare the activities of various promoters located either on the chromosome of *E. coli* or on a plasmid.

Chapter 2:

Materials and Methods

2.1 Suppliers

Unless otherwise stated, all chemicals were supplied by Sigma-Aldrich, BDH Prolabo or Fisher-Scientific. Restriction endonucleases, calf alkaline phosphatase, T4 DNA Ligase and Phusion DNA polymerase were purchased from New England BioLabs. Biomix Red DNA polymerase, cDNA Synthesis Kit, Sensimix Hi-ROX Kit and dNTP mix were supplied by Bioline. Enzymes were used as described by the manufacturer and in the buffers supplied by the manufacturer. Oligodeoxyribonucleotides were supplied by Alta Bioscience, University of Birmingham.

2.2 Bacterial growth media

2.2.1 Liquid media

All liquid media were made using distilled water, sterilised by autoclaving for 20 min at 120°C and 15 psi, and stored at room temperature. One litre of LB medium contained: 20 g tryptone; 10 g yeast extract and 10 g NaCl dissolved in distilled water. M9 minimal medium was made as a 10 x solution, which contained: 60 g Na_2HPO_4; 30 g KH_2PO_4; 5 g NaCl; 10 g NH_4Cl; 10 mg Biotin and 10 mg thiamine dissolved in distilled water. 10 x M9 minimal medium was then adjusted to pH7.4 before autoclaving. 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 $CaCl_2$ and 0.5 ml 20% casamino acids were mixed aseptically and made up to 100 ml with sterile distilled water. M9 minimal salts medium was always supplemented with 0.3% fructose as the carbon source and was supplemented with 0.2% melibiose where required.

One litre of minimal salts medium (MS) contained: 4 g KH_2PO_4; 10.5 g K_2HPO_4; 1 g $(NH_4)_2SO_4$; 0.05 g $MgSO_4.7H_2O$; 0.6 g trisodium citrate. The medium was then supplemented with 10 ml ammonium molybdate (1 mM), 10 ml sodium selenate (1 mM) and 10 ml of *E. coli* sulphur free salts to a total of 1 litre. The composition of one litre of *E. coli* sulphur free salts was: 82 g $MgCl.7H_2O$; 10 g $MnCl_2.4H_2O$; 4 g $FeCl_2.6H_2O$ and 1 g CaCl, supplemented with 20 ml of concentrated HCl per litre of distilled water. Minimal medium was supplemented with 10% LB, 20 mM fumarate, 20 mM TMAO and 0.4% (v/v) glycerol as a carbon source. SOC medium is a rich media used in the recovery step of *E. coli* competent cell transformations to maximise the transformation efficiency. Ready-made SOC solution was purchased from Sigma-Aldrich.

2.2.2 Solid media

All solid media were made using 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 was prepared by adding 23 g nutrient agar powder (Difco) to distilled water to a total volume of one litre before autoclaving. LB agar was prepared by the addition of 20 g tryptone, 10 g yeast extract, 10 g NaCl and 10 g agar to distilled water to a total volume of one litre. 50 g MacConkey lactose agar (Difco) was added to distilled water to a total volume of one litre before sterilisation. One litre of lactose M9 minimal salts agar contained: 100 ml 10 x M9 salts; 1 g α-lactose and 15 g bacteriological agar (Oxoid).

2.2.3 Antibiotics and other supplements

All stock solutions of antibiotics and other supplements were sterile filtered through 0.2 µm syringe filters and stored at -20°C. For stock solution, 40 mg/ml ampicillin was dissolved in sterile distilled water. Stock solutions of kanamycin were 50 mg/ml in sterile distilled water. Chloramphenicol and tetracycline were each dissolved in methanol at concentrations of 25 mg/ml and 10mg/ml, respectively. X-gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside) (Bioline), the chromogenic substrate for β-galactosidase, was stored at a concentration of 20 mg/ml in dimethylformamide. The gratuitous inducer of the *lac* operon, IPTG (isopropyl β-D-1-thiogalactopyranoside) (Bioline), was stored at 100 mM in sterile distilled water. The DNA gyrase inhibitor, novobiocin, was stored at 100 mg/ml in sterile distilled water.

To select for bacterial strains carrying antibiotic resistance genes, antibiotics were added to liquid or solid media, after autoclaving, at the following final concentrations: 80 µg/ml ampicillin, 35 µg/ml tetracycline, 50 µg/ml kanamycin, 25 µg/ml chloramphenicol. Blue-white colour selection of Lac^+ and Lac^- colonies was achieved by the addition of 40 µg/ml X-gal to molten LB agar, after autoclaving. To induce expression of the *lac* operon during blue-white selection, IPTG was added to media to a final concentration of 0.1 mM. IPTG and novobiocin were added to media after sterilisation and cooling as required and specified in results chapters.

2.3 Bacterial strains and plasmids

2.3.1 Bacterial strains and growth conditions

The bacterial strains used in this study are listed in Table 2.1. Strains were stored as glycerol stocks in LB media supplemented with 15% glycerol at -80°C. Strains were streaked onto the appropriate agar plate and incubated overnight at 37°C before use. Overnight cultures were

prepared by inoculation of 5-10 ml of the appropriate sterile medium with a single fresh colony and incubation at 37°C with aeration for 14-16 h (LB) or up to 20 h (minimal medium). Growth of liquid cultures was monitored by measuring the optical density at 650 nm or 620 nm, using a Helios Gamma Spectrophotometer or a Labsystems Multiskan MS plate reader, respectively (Thermo Fisher Scientific Inc.).

2.3.2 Plasmids

Plasmids used in this study are listed in table 2.2. Plasmid maps are shown in figures 2.1-2.6. Plasmid pRW50 carrying various promoter fragments as transcription fusions to *lacZ* was used to quantify promoter activities by measuring β-galactosidase activity. Derivatives of plasmids pKH3 and pKH5 carrying various promoter fragments as transcription fusions to *lacZ* were used to insert these fusions into the chromosome by gene gorging. Plasmid pKH5 was also used to create pJB3, in which the *lacI* and *lacZ* homology regions are replaced by cloning sites and the translation initiation region fused to the emerald *gfp* gene in place of *lacZ*. Plasmid pJB15 was created from pJB3 with the *lacI* and *lacZ* homology regions cloned into the multiple cloning sites. The *lac* promoter region, from position -92 bp to +122 bp relative to the transcription start site, was also fused to emerald *gfp*. Plasmids pJB16 and pJB17 were created from pJB15 to target the *lac28::egfp* fusion to the *ara* and *mel* loci, however the gene gorging method was not very effective without red/white selection. Therefore plasmids pJB20 and pJB22 were created by cloning the donor fragments from pJB16 and pJB17 between the *SceI* recognition sites of the gene doctoring plasmid pDOC-C. Plasmids based on pJB22 carrying various homology regions to the *E. coli* chromosome were used to insert the *lac28::gfp* fusion at various chromosomal loci in single copy. Plasmid pCP20 was used in the removal of the *kan* gene from the chromosome.

Table 2.1: *E. coli* strains used in this study (continued on pages 47-49)

Strain	Description[1]	Origin
MG1655	*E. coli* K12 strain	Blattner *et al.* (1997)
RLG221	*E. coli* K-12 *Δlac ΔrecA* used for manipulating plasmids	R. Gourse
XL1-Blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F' *proAB lacI^{q}ZΔ M15* Tn*10* (Tet^R)]	Stratagene
KH000	MG1655 *lacI*$^-$	Hollands (2009)
KH001	Derivative of KH000 in which the *lac* promoter has been replaced with the cloning site from pKH3	Hollands (2009)
KH002	Derivative of KH001 carrying a chromosomal TB10a::*lacZ* translation fusion at the *lac* locus	Hollands (2009)
BRY01	Derivative of KH001 in which the *lac* promoter has been replaced with a *phcp::lacZ* transcription fusion	This study
BRY03	Derivative of KH001 in which the *lac* promoter has been replaced with a *CC-41.5::lacZ* transcription fusion	This study
BRY05	Derivative of KH001 in which the *lac* promoter has been replaced with a *CC-61.5::lacZ* transcription fusion	This study
BRY07	Derivative of MG1655 in which the *lac* promoter region is replaced by a *kan* gene and *lac00::egfp* fusion	This study
BRY08	Derivative of MG1655 in which the *lac* promoter region is replaced by a *kan* gene and *lac28::egfp* fusion.	This study
BRY09	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 72143 bp (*ara* locus)	This study
BRY11	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 4342915 bp (*mel* locus)	This study
BRY12	Derivative of BRY08, from which the *kan* gene has been removed. Position 365529 - 365652 bp (*lac+O2* locus)	This study
BRY13	Derivative of BRY09, from which the *kan* gene has been removed. Position 72143 bp (*ara* locus)	This study
BRY15	Derivative of BRY11, from which the *kan* gene has been removed. Position 4342915 bp (*mel* locus)	This study
BRY17	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 72143 bp in the opposite orientation to that of BRY09. *yabI<kan lac28::gfp>thiQ* (*Inverse ara* locus)	This study

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

Strain	Description[1]	Origin
BRY18	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 238142 bp (*yafT* locus)	This study
BRY19	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 313681 bp (*eaeH* locus)	This study
BRY21	Derivative of BRY17, from which the *kan* gene has been removed. Position 72143 bp (*Inverse ara* locus)	This study
BRY22	Derivative of BRY18, from which the *kan* gene has been removed. Position 238142 bp (*yafT* locus)	This study
BRY23	Derivative of BRY19, from which the *kan* gene has been removed. Position 313681 bp (*eaeH* locus)	This study
BRY26	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 2314948 bp (*rcs* locus)	This study
BRY27	Derivative of BRY26, from which the *kan* gene has been removed. Position 2314948 bp (*rcs* locus)	This study
BRY28	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 1005714 bp (*ycb* locus)	This study
BRY29	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 1606129 bp (*tam* locus)	This study
BRY30	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 3104995 bp (*nupG* locus)	This study
BRY31	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 3982359 bp (*asl* locus)	This study
BRY32	Derivative of BRY28, from which the *kan* gene has been removed. Position 1005714 bp (*ycb* locus)	This study
BRY33	Derivative of BRY29, from which the *kan* gene has been removed. Position 1606129 bp (*tam* locus)	This study
BRY34	Derivative of BRY30, from which the *kan* gene has been removed. Position 3104995 bp (*nupG* locus)	This study
BRY35	Derivative of BRY31, from which the *kan* gene has been removed. Position 3982359 bp (*asl* locus)	This study
BRY36	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 4342915 bp in the opposite orientation to that of BRY11. *melB<gfp::lac28 kan>yjdF* (*Inverse mel* locus)	S. E. Russell (Unpublished)

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

Strain	Description[1]	Origin
BRY37	Derivative of BRY36, from which the *kan* gene has been removed. Position 4342915 bp (*Inverse mel* locus)	S. E. Russell (Unpublished)
BRY38	Derivative of MG1655, in which the lac promoter and the start of the lacZ gene (including *lacI O2*) was replaced with a *kan* gene and *lac28::egfp* fusion (*lac* locus)	This study
BRY39	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 365131-365652 bp in the opposite orientation to that of BRY38. *lacZ<kan lac28::gfp>lacI* (*Inverse lac* locus)	This study
BRY40	Derivative of BRY38, from which the *kan* gene has been removed. Position 365131 - 365652 bp (*lac* - O^2 locus)	This study
BRY41	Derivative of BRY39, from which the *kan* gene has been removed. Position 365131 - 365652 bp (*Inverse lac* locus)	This study
BRY44	Derivative of BRY27 in which the *lac* promoter has been replaced by the pKH3 cloning site by gene gorging (Lac^-)	This study
BRY47	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 2987820 bp (*yqe* locus)	This study
BRY48	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 3132854 bp (*pitB* locus)	This study
BRY51	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 229046 bp (*dkgB* locus)	This study
BRY52	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 237530-239300 bp (*ΔyafT* locus)	This study
BRY53	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 312940-315678 bp (*ΔeaeH* locus)	This study
BRY54	Derivative of BRY47, from which the *kan* gene has been removed. Position 2987820 bp (*yqe* locus)	This study
BRY55	Derivative of BRY48, from which the *kan* gene has been removed. Position 3132854 bp (*pitB* locus)	This study
BRY58	Derivative of BRY51, from which the *kan* gene has been removed. Position 229046 bp (*dkgB* locus)	This study
BRY59	Derivative of BRY52, from which the *kan* gene has been removed. Position 237530 - 239300 bp (*ΔyafT* locus)	This study

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

Strain	Description[1]	Origin
BRY60	Derivative of BRY53, from which the *kan* gene has been removed. Position 312940 - 315678 bp (*ΔeaeH* locus)	This study
BRY61	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 2984331 - 2995806 bp (*Δyqe* locus)	This study
BRY62	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 3131588 - 3133385 bp (*ΔpitB* locus)	This study
BRY66	Derivative of MG1655 carrying a *kan* gene and *lac28::egfp* fusion at position 2509358 bp (*mntH* locus)	This study
BRY68	Derivative of BRY61, from which the *kan* gene has been removed. Position 2984331 - 2995806 bp (*Δyqe* locus)	This study
BRY69	Derivative of BRY62, from which the *kan* gene has been removed. Position 3131588 - 3133385 bp (*ΔpitB* locus)	This study
BRY73	Derivative of BRY66, from which the *kan* gene has been removed. Position 2509358 bp (*mntH* locus)	This study

[1] Position refers to location of the insert in the *E. coli* genome in base pairs (bp) with respect to the coordinate-system origin, as described by www.EcoCyc.org (Keseler *et al*., 2011).

Table 2.2: Plasmids used in this study (continued on pages 51-54)

Plasmid	Description[1]	Origin
pRW50	Low copy number, broad host range *lac* expression vector used for cloning EcoRI-HindIII promoter fragments as fusions to *lacZYA* (Tet^R) (Figure 2.1)	Lodge *et al.* (1992)
pRW50/*phcp*3831NS	pRW50 derivative carrying an *Eco*RI-*Hin*dIII *hcp* regulatory region fragment, from –125 bp to +42 bp, in which the NsrR binding site has been knocked out.	Chismon (2011)
pRW500	pRW50 derivative in which the *Eco*RI-*Sac*I fragment has been replaced with the cloning site and *lacZ Eco*RI-*Sac*I fragment from pKH3. (Tet^R) (Figure 2.2)	This study
pRW500/*phcp3831NS*	pRW500 derivative carrying an *Eco*RI-*Bam*HI *hcp* regulatory region fragment from -125 bp to +11 bp as a transcription fusion to *lacZ.*	This study
pRW500/*CC-41.5*	pRW500 derivative carrying the *CC-41.5* promoter from -76 bp to +1 bp cloned as a transcription fusion to *lacZ*	This study
pRW500/*CC-61.5*	pRW500 derivative carrying the *CC-61.5* promoter from -96 bp to +1 bp cloned as a transcription fusion to *lacZ*	This study
pRW500/TB10a	pRW500 derivative carrying the *melR* promoter from-100 to +52 cloned as a translation fusion to *lacZ*	This study
pRW500/*plac*	pRW500 derivative carrying the *lac* promoter on a *Eco*RI-*Sac*I fragment naturally fused to *lacZ* from -92 to the natural *Sac*I site in *lacZ*	This study
pACBSR	Mutagenesis plasmid for use in gene gorging (Cm^R) (Figure 2.3)	Scarab Genomics
pCP20	Temperature sensitive plasmid encoding FLP recombinase (Amp^R Cm^R) (Figure 2.4)	Cherepanov and Wackernagel (1995)
pKH3	Derivative of pBR322. Donor plasmid for construction of chromosomal promoter::*lac* fusions by gene gorging (Amp^R) (Figure 2.5)	Hollands (2009)

Table 2.2: Plasmids used in this study (continued)

Plasmid	Description[1]	Origin
pKH5	Derivative of pKH3 containing a *kan* gene upstream of the promoter cloning site. Donor plasmid for construction of chromosomal promoter::*lac* fusions by gene gorging (Amp^R) (Figure 2.5)	Hollands (2009)
pDEX-G	Gene doctoring plasmid, which allows homology regions to be cloned either side of *gfp* with flanking SceI sites.	(Lee *et al*., 2009)
pDOC-C	Gene doctoring plasmid used as a source of the pEX100T backbone for pJB gene doctoring plasmids (Amp^R*sacB*)	(Lee *et al*., 2009)
pJB3	Derivative of pKH5 in which the *lac* homology regions are replaced with cloning sites and the translation initiation region fused to *gfp*.	This study
pJB4	Derivative of pJB3 in which a *lac00* promoter fragment is cloned into the *Eco*RI-*Hin*dIII sites.	This study
pJB6	Derivative of pJB4 in which homology regions for the *lacI* and *lacZ* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively	This study
pJB15	Derivative of pJB6 in which the *lac00* promoter fragment is replaced by the *lac28* promoter fragment. Used to create the BRY08 strain	This study
pJB16	Derivative of pJB15 in which the homology regions for the *thiQ* and *yabI* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively	This study
pJB18	Derivative of pJB15 in which the homology regions for the *melB* and *yjdF* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively	This study
pJB20	Derivative of pDOC-C, in which the donor fragment from pJB16 has been cloned between the *I-SceI* recognition sites. Used to create the BRY09 strain (Figure 2.6)	This study
pJB22	Derivative of pDOC-C, in which the donor fragment from pJB18 has been cloned between the *I-SceI* recognition sites. Used to create the BRY11 strain	This study

Table 2.2: Plasmids used in this study (continued)

Plasmid	Description[1]	Origin
pJB27	Derivative of pJB22, in which the homology regions for the *lacZ* and *lacI* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY16 strain	This study
pJB28	Derivative of pJB22, in which the homology regions for the *yabI* and *thiQ* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY17 strain	This study
pJB29	Derivative of pJB22 in which the homology regions for the *yafT* and *ykfM* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY18 strain	This study
pJB30	Derivative of pJB22 in which the homology regions for the *eaeH* pseudogene are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY23 strain	This study
pJB32	Derivative of pJB22 in which the homology regions for the *rcsB* and *rcsC* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY26 strain	This study
pJB37	Derivative of pJB22 in which the homology regions for the *ycbW* and *ycbX* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY28 strain	This study
pJB38	Derivative of pJB22 in which the homology regions for the *tam* and *yneE* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY29 strain	This study
pJB39	Derivative of pJB22 in which the homology regions for the *nupG* and *speC* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY30 strain	This study
pJB40	Derivative of pJB22 in which the homology regions for the *aslB* and *aslB* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY31 strain	This study

Table 2.2: Plasmids used in this study (continued)

Plasmid	Description[1]	Origin
pJB42	Derivative of pJB22 in which the homology regions for the *lacI* and *lacZ -02* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY38 strain	This study
pJB43	Derivative of pJB22 in which the homology regions for the *lacZ -02* and *lacI* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY39 strain	This study
pJB45	Derivative of pJB22 in which the homology regions for the *yjdF* and *melB* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY36 strain	S. E. Russell (Unpublished)
pJB53	Derivative of pJB22 in which the homology regions for the *yqeJ* and *yqeL* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY47 strain	This study
pJB54	Derivative of pJB22 in which the homology regions for the *yghT* and *pitB* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY48 strain	This study
pJB57	Derivative of pJB22 in which the homology regions for the *aspU* and *dkgB* genes are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites respectively. Used to create the BRY51 strain	This study
pJB58	Derivative of pJB22 in which the homology regions for the Δ*yafT* strain are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY52 strain	This study
pJB59	Derivative of pJB22 in which the homology regions for the Δ*eaeH* strain are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY53 strain	This study
pJB66	Derivative of pJB22 in which the homology regions for the Δ*yqe* strain are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY61 strain	This study
pJB67	Derivative of pJB22 in which the homology regions for the Δ*pitB* strain are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY62 strain	This study

Table 2.2: Plasmids used in this study (continued)

Plasmid	Description[1]	Origin
pSB3	Derivative of pJB22 in which the homology regions for the *mntH* strain are cloned into the *Mfe*I-*Xma*I and *Nhe*I-*Sac*I sites. Used to create the BRY73 strain	S. Bevan (unpublished)

[1] - and + numbers refer to the size of the promoter fragment in base pairs (bp) with respect to the transcription start site.

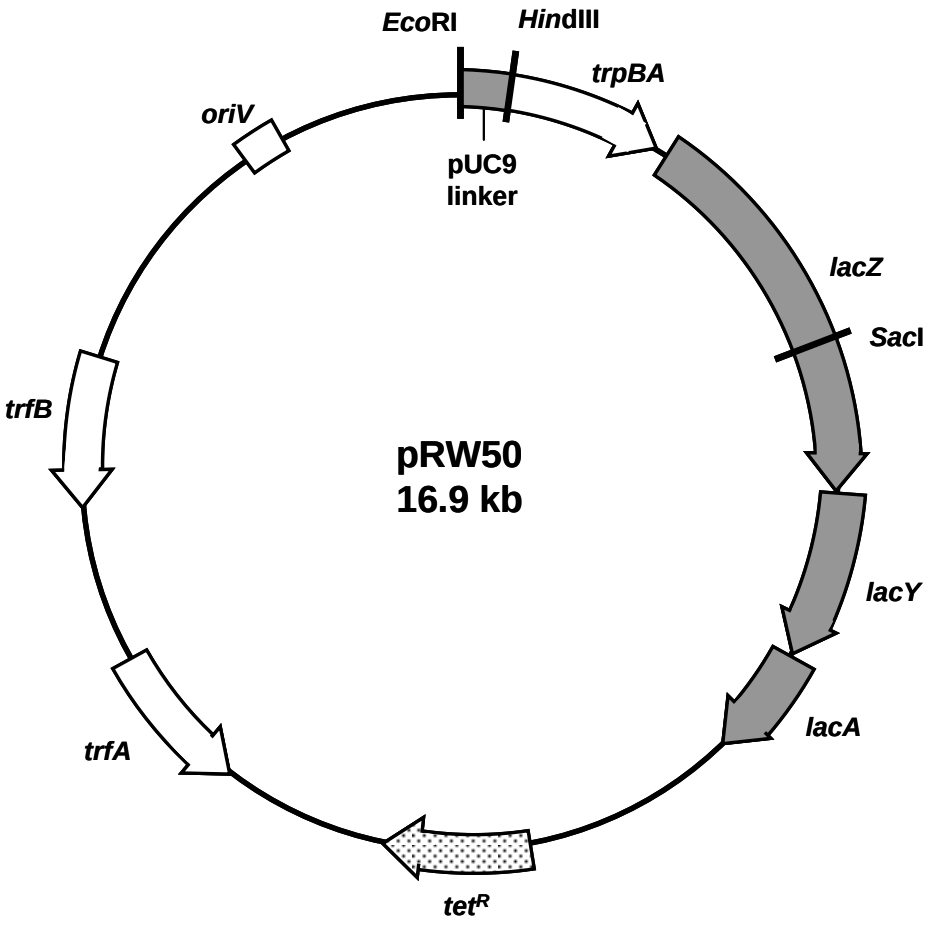

Figure 2.1: Map of the *lac* expression vector pRW50

The pUC9 linker (shaded grey) allows *Eco*RI-*Hin*dIII promoter fragments to be cloned as translation fusions to the *trpBA* fragment (white) and *lacZYA* operon. Therefore the promoter of interest drives expression of the *lac* genes (grey). The tetracycline resistance gene (tet^R), plasmid replication genes *trfA* and *trfB* and the origin of replication (*oriV*) are also indicated.

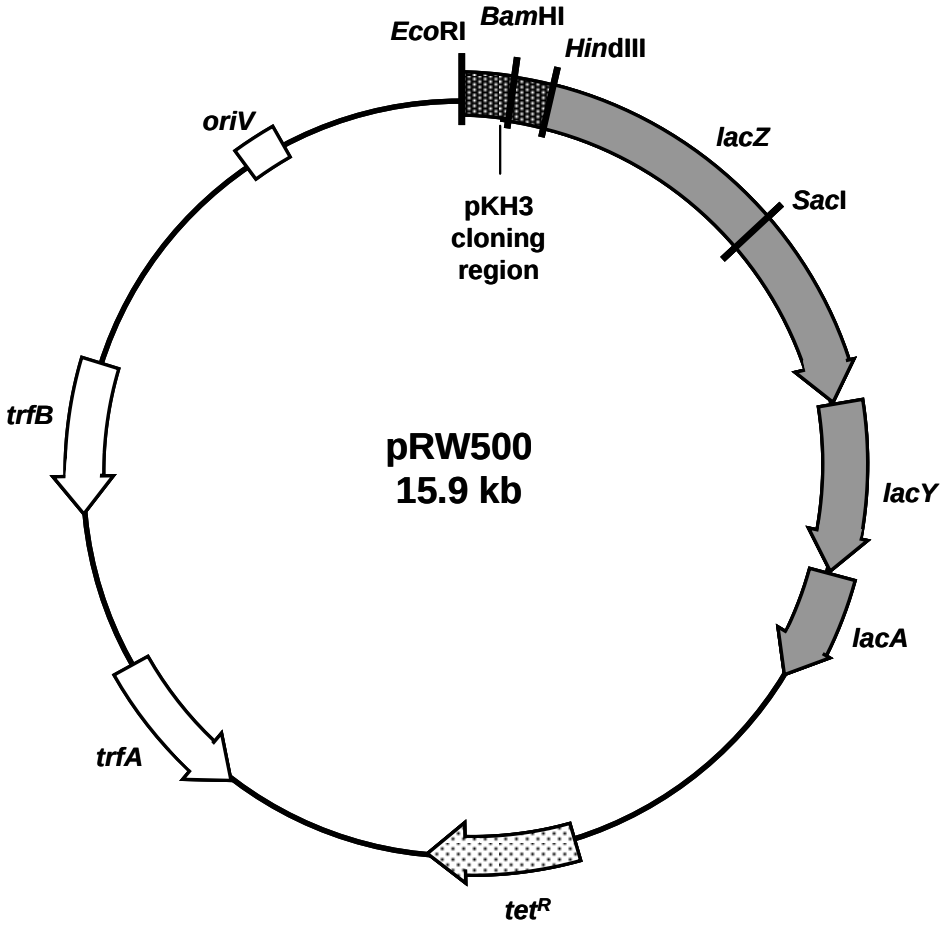

Figure 2.2: Map of the *lac* expression vector pRW500

pRW500 carries the pKH3 cloning region fused directly to the *lacZYA* operon. The plasmid is a derivative of pRW50 (Lodge *et al*., 1992) in which the *trpBA* fragment has been removed by replacement of the *Eco*RI-*Sac*I fragment with that encoded by the chromosome of the KH001 strain (Hollands, 2009). Promoter fragments are cloned into the pKH3 cloning region as transcription (*Eco*RI-*Bam*HI) or translation (*Eco*RI-*Hin*dIII) fusions to the *lacZYA* operon (grey), therefore allowing the promoter of interest to drive expression of the *lac* genes. The tetracycline resistance gene (tet^R), plasmid replication genes *trfA* and *trfB* and the origin of replication (*oriV*) are also indicated.

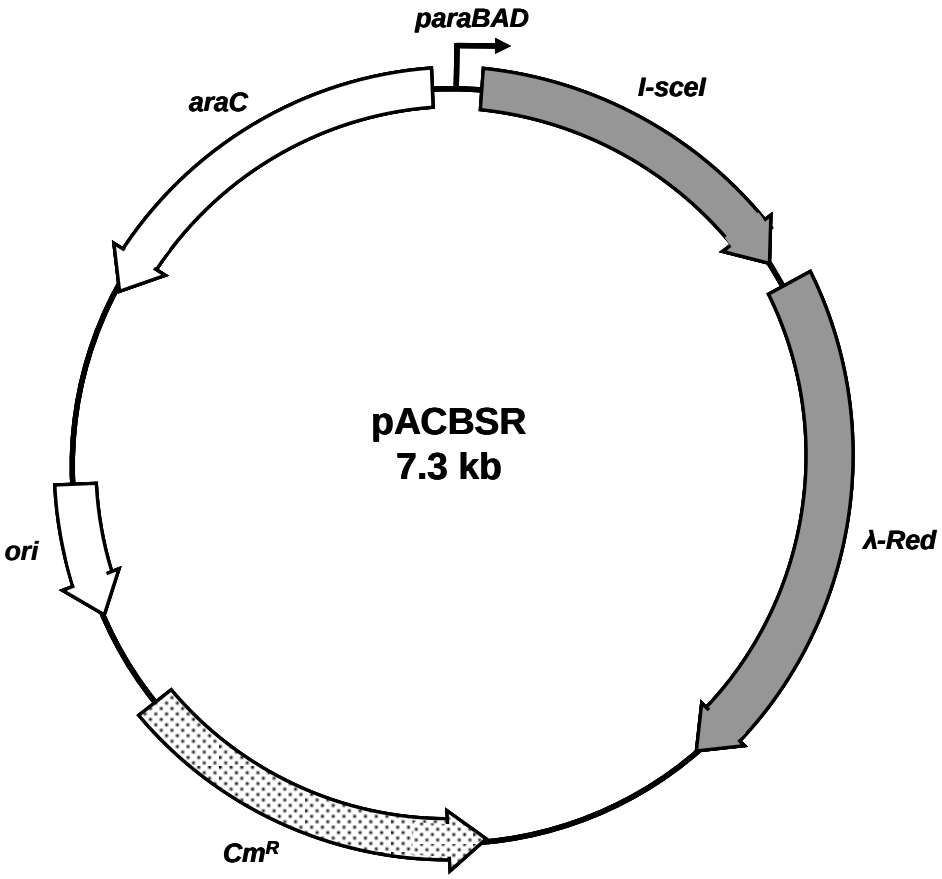

Figure 2.3: Map of the mutagenesis plasmid pACBSR

pACBSR is used as a mutagenesis plasmid in gene gorging and gene doctoring. Genes encoding the *Saccharomyces cerevisiae* I-SceI endonuclease and λ-Red recombination proteins (grey) are under transcriptional control of the arabinose inducible *araBAD* (*paraBAD*) promoter. The λ-Red system consists of three genes: *exo, bet* and *gam*, however for simplicity the boundaries are not shown. The *araC* gene, which encodes a regulator of the *araBAD* promoter, chloramphenicol resistance gene (Cm^R) and origin of replication (*ori*) are also indicated.

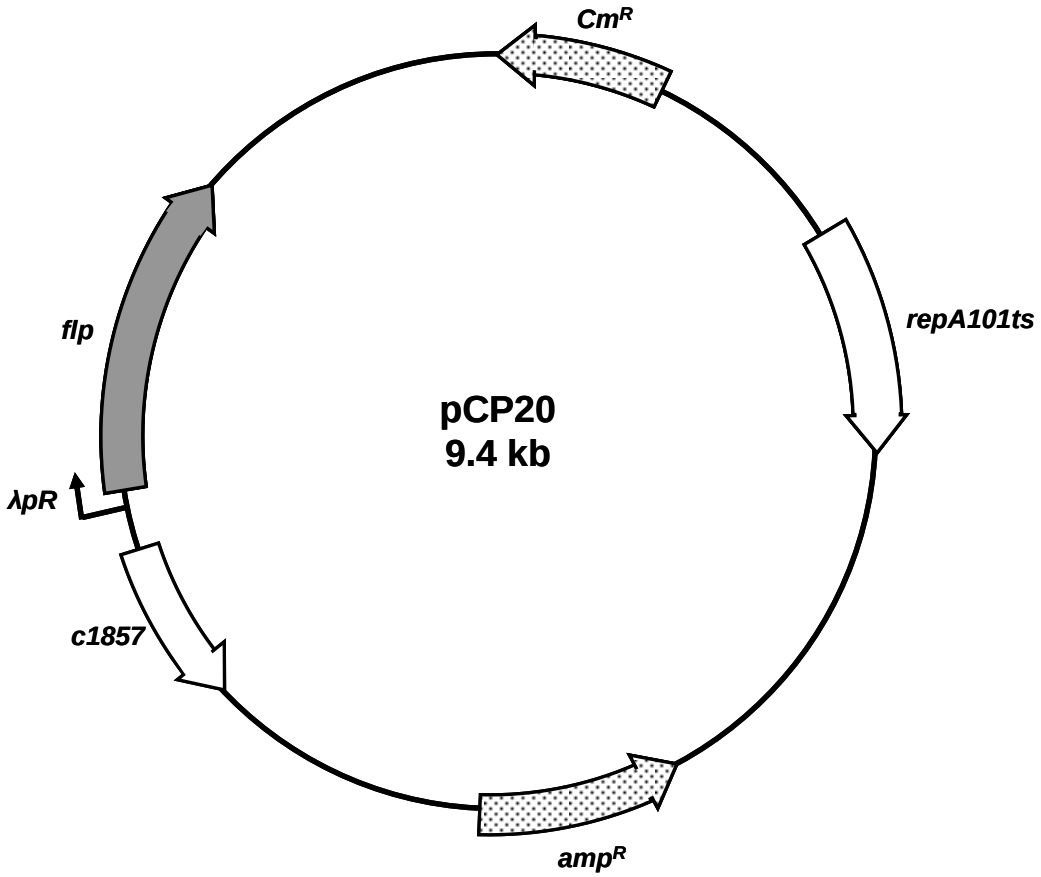

Figure 2.4: Map of pCP20

pCP20 is used to express FLP recombinase to allow recombination of FRT sites to remove chromosome-encoded kanamycin resistance cassette. The *flp* recombinase (grey) is under control of the λpR promoter, which is regulated by the phage λ temperature-sensitive repressor *c1857*. Replication of the plasmid is dependent on the temperature-sensitive origin *repA101ts*. Chloramphenicol (Cm^R) and ampicillin (amp^R) resistance genes are also indicated.

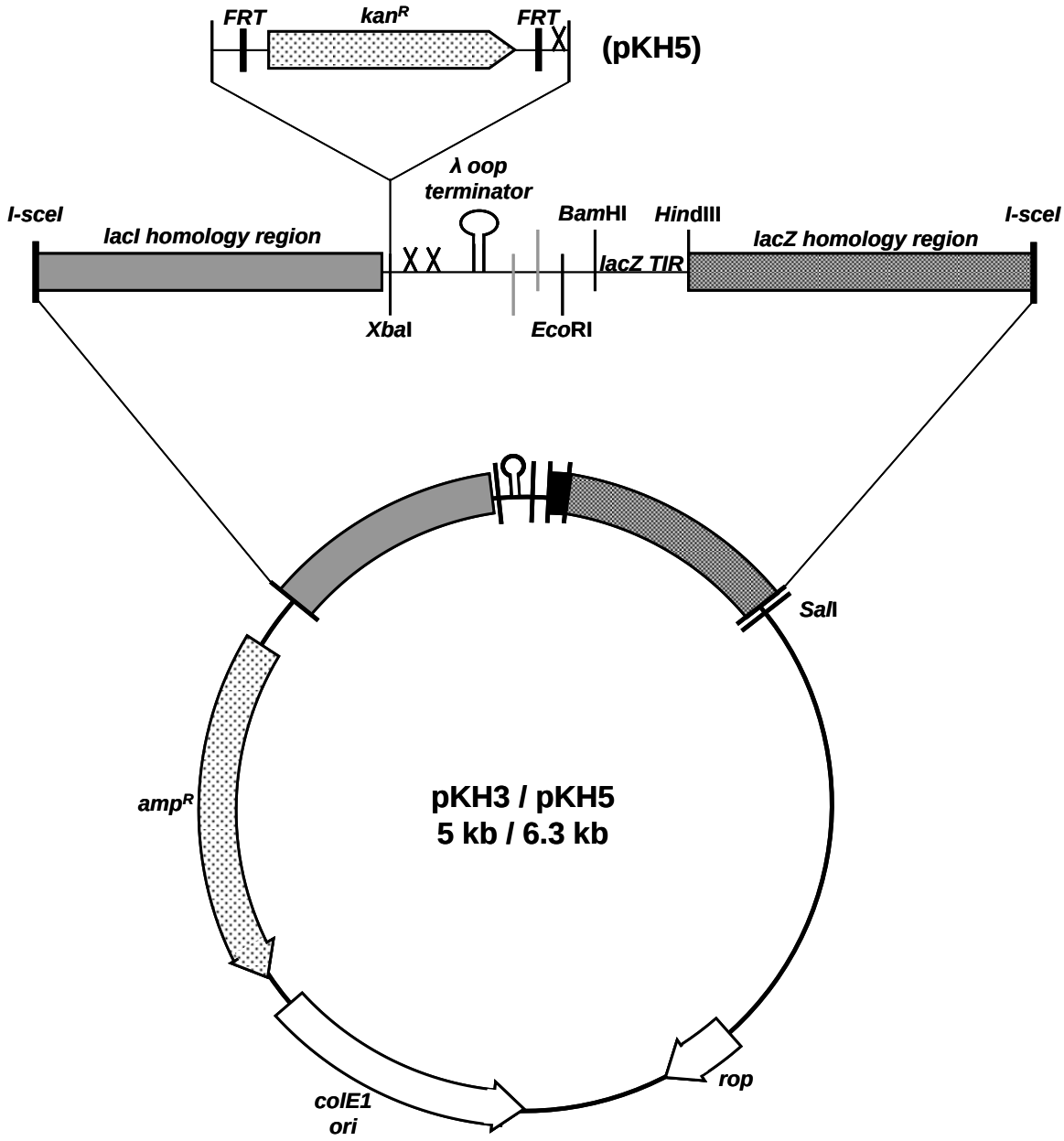

Figure 2.5: Map of the pKH3 and pKH5 gene gorging donor plasmids

pKH3 is a gene gorging donor plasmid carrying a promoter cloning site flanked by homology regions the 3' end of *lacI* (grey) and the 5' end of *lacZ* (black dotted), which are flanked by 18 bp DNA recognition sites for the *Saccharomyces cerevisiae* I-SceI endonuclease. Promoter fragments are cloned as transcription (*Eco*RI-*Bam*HI) or translation (*Eco*RI-*Hin*dIII) fusions to the *lacZ* 5' homology region, the former of which utilises the *lacZ* translation initiation region (*lacZ TIR*). The λ *oop* terminator is indicated by a stem-loop structure and is positioned upstream of the promoter cloning site to prevent transcriptional read-through. pKH5 is a derivative of pKH3 in which a kanamycin resistance gene (kan^R) has been inserted at the *Xba*I restriction site. The kan^R gene is flanked by FRT (flippase recognition target) sites, which allow removal of the kanamycin resistance cassette after transfer of the promoter::*lacZ* fusion to the chromosome. Stop codons are also included in all three reading frames to prevent translational read-through (X). The ampicillin resistance gene (amp^R), origin of replication (*colE1 ori*) and plasmid replication gene, *rop*, are also shown.

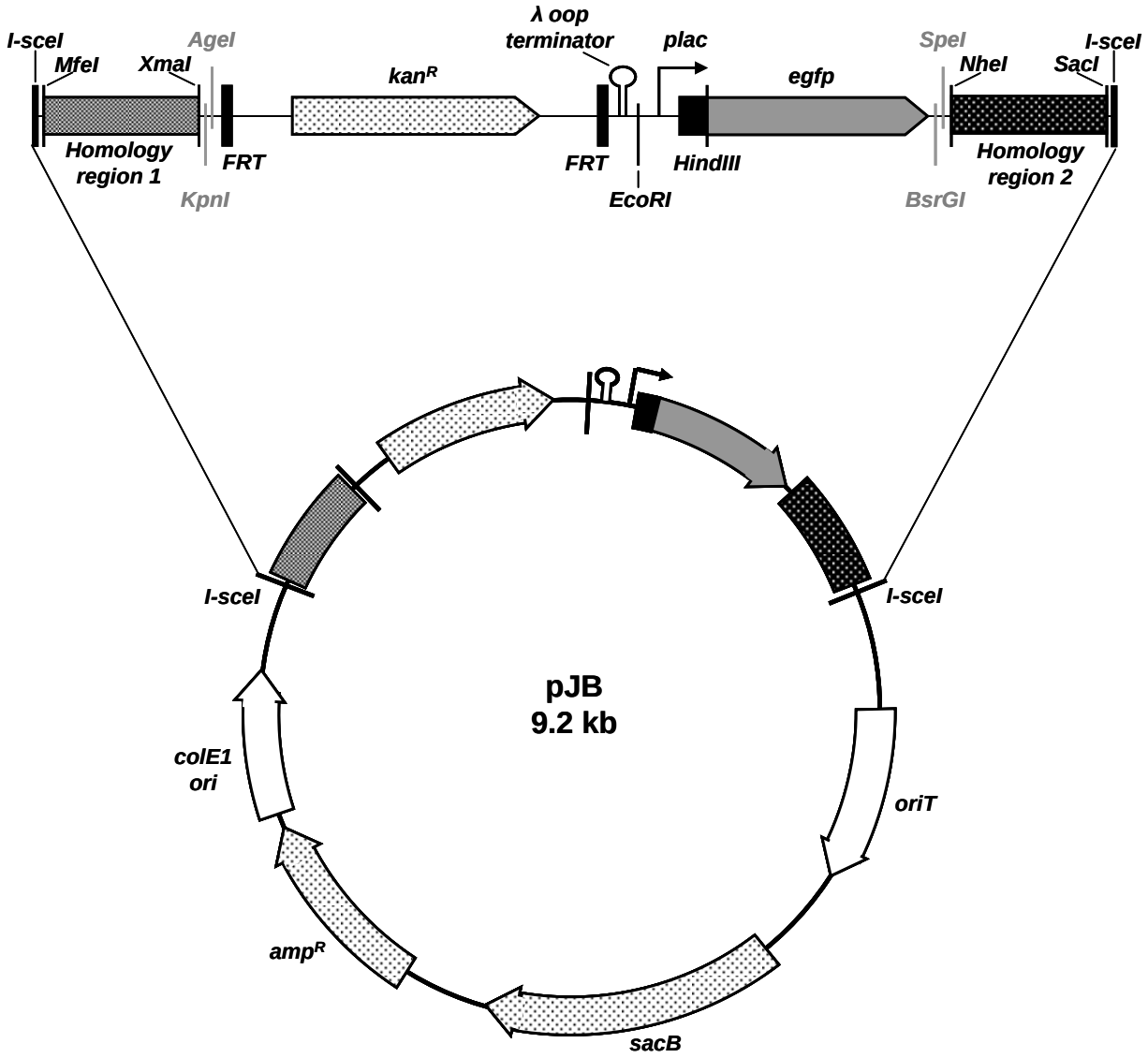

Figure 2.6: Map of pJB gene doctoring donor plasmid

The pJB series of gene doctoring plasmids are derivatives of the pDOC-C gene doctoring plasmid (Lee *et al*., 2009) and are used to insert the *lac28::egfp* fusion at any desired position of the *E. coli* chromosome. Approximately 500 bp regions of homology, flanking the desired position of insertion on the chromosome, are cloned into the multiple cloning sites which flank the *lac28::egfp* fusion. The FRT flanked kanamycin resistance cassette (kan^R) and λ *oop* transcription terminator (Stem-loop) of pKH5 (Hollands, 2009) are included upstream of the fusion to allow selection of recombinants and prevention of transcriptional read-through from neighboring operons. The plasmid carries the *sacB* sucrose sensitivity gene (*sacB*) allowing selection against retention of the donor plasmid. The ampicillin resistance gene (amp^R) and origins of replication (*colE1 ori, oriT*) are also shown.

Figure 2.7 Sequence of the pJB plasmid *I-SceI* donor molecule

Figure on page 62. The base sequence of the pJB plasmid *I-SceI* donor molecule. The *I-SceI* DNA target recognition sites are double underlined and labelled. Restriction endonuclease target sites are boxed and labelled. The FLP recognition targets (FRT) DNA site are shaded grey and labelled. ATG start codon and open reading frames are in capitals with the start and stop codons in bold and underlined. The kanamycin resistance (kan^R) and *egfp* genes are in capitals and labelled. The *lacZ* start codon and the *lacZ* gene are shown in bold capitals with the Shine-Dalgarno sequence underlined. The promoter -10, -35 are underlined, with the transcription start site (+1) bold and underlined. The *lacI* operator sequences are boxed and labelled and the DNA site for CRP is double underlined. Additional stop codons in different reading frames were included and are denoted by an X and underlining.

I-SceI *Mfe*I *Xma*I *Kpn*I *Age*I

tagggataacagggtaatcaattgtatgcccgggttgtaggtaccaactaccggtggctgg

FRT

Agctgcttcgaagttcctatactttctagagaataggaacttcggaataggaacttcaagatc
ccctcacgctgccgcaagcactcagggcgcaagggctgctaaaggaagcggaacacgtagaac
ttaagggaattgccagctggggcgccctctggtaaggttgggaagccctgcaaagtaaactgg
atggctttcttgccgccaaggatctgatggcgcaggggatcaagatctgatcaagagacagga

kan^{R}

tgaggatcgtttcgc**ATG**ATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTG
GAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTC
CGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAAT
GAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCT
GTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAG
GATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGG
CGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAG
CGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAG
GGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGCGAGGATCTC
GTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGA
TTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGT
GATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCC
GCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTC**TGA**gcgggactc
tggggttcgaaatgaccgaccaagcgacgcccaacctgccatcacgagatttcgattccaccg
ccgccttctatgaaaggttgggcttcggaatcgttttccgggacgccggctggatgatcctcc
agcgcggggatctcatgctggagttctcgcccaccccagcttcaaaagcgctctgaagttcct

FRT X X X

atactttctagagaataggaacttcggaataggaactaaggtaagctagactaagtaattcag

λ-oop terminator *Eco*RI *LacI O*^{3}

aa**cgctcggt**tgcc**gccgggcg**tttttt attagatctctcgaggaattcggcagtgagcgca

CRP *-35*

acgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttccgg

-10 ***+1*** *LacI O*^{1} *SD* *lacZ*

ctcgtatgttgtgtgg**A**attgtgagcggataacaatttcacacaggaaacagct**ATG**ACCATG
ATTACGGATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA

*Hin*dIII *egfp*

CTTAATCGCCTTaagctt**AGC**AAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTC
GAGCTGGACGGCGACGTAAACGGCcACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCC
ACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTCGTGACCACCTTGACCTACGGCGTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAG
CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTC
AAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAG
TACAACTACAACAGCCACAAGGTCTATATCACCGCCGACAAGCAGAAGAACGGCATCAAGGTG
AACTTCAAGACCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAG
AACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCC
GCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCC

X X

GCCGGGATCACTCTCGGCATGGACGAGCTTTACAAG**TAA**gtaactaacgaatgtacagccata

*Spe*I *Nhe*I *Sac*I *I-SceI*

actagtacttgctagcgcacaagagctctagggataacagggtaat

Figure 2.7 Sequence of the pJB plasmid *I-SceI* donor molecule

2.4 Gel electrophoresis

2.4.1 Agarose gel electrophoresis of DNA

DNA fragments greater than 500 bp in length were analysed using agarose gel electrophoresis. 0.8-1.2% solutions of agarose in 1 x TAE buffer (40 mM Tris acetate, 2 mM Na_2EDTA) (National Diagnostics) were heated to 100°C for 2 min in a microwave to dissolve the agarose. Agarose solutions were then were cooled to approximately 50°C and poured onto a gel casting plate. DNA samples were mixed in a 1:1 ratio with DNA loading dye (0.025% bromophenol blue; 0.025% xylene cyanol F; 20% glycerol; 10 mM Tris, pH 7.5; 1 mM EDTA) before being loaded into the wells. The DNA markers used were the 100 bp and 1 kb DNA ladders (New England BioLabs), which were diluted 6-fold in DNA loading dye, supplied by the manufacturer. Samples were subjected to electrophoresis in 1 x TAE buffer at 3-5 V/cm for 30 – 45 min. Gels were stained in a solution of 0.5 μg/ml ethidium bromide (BioRad) for 15 min, viewed and photographed under 300 nm ultraviolet light using a gel documentation system (Bio-Rad). However, to reduce damage to DNA fragments to be purified by extraction, gels were stained with SYBRSAFE (Invitrogen) and visualised using a blue light box.

2.4.2 Chloroquine agarose gel electrophoresis of DNA

Chloroquine agarose gel electrophoresis was used to monitor the superhelicity of the reporter plasmid pBR322. 300 ml of 1% agarose solution containing 2 x TBE (178 mM Tris borate pH 8.3, 4 mM Na_2EDTA) (National Diagnostics) was heated to 100°C in a microwave to dissolve agarose. The solution was allowed to cool to approximately 50°C then supplemented with 2.5 μg/ml chloroquine. The solution was poured into a large gel casting tank to give a gel that was approximately 1.5 cm thick with wells of ~20 μl capacity. Plasmid DNA samples (1 μg of

DNA in 20 µl final volume) were mixed in a 1:1 ratio with 6 x DNA loading dye (New England BioLabs) and loaded into the wells. The samples were subjected to electrophoresis in 2 x TBE at 3 V/cm for 24 h. After electrophoresis, gels were washed repeatedly with deionised water for approximately 2 h before being stained with 1 µg/ml ethidium bromide (BioRad) for approximately 30 min. Gels were briefly washed again with water then viewed and photographed under ultraviolet light of 300 nm using a gel documentation system (Bio-Rad). The intensity of each lane was plotted against distance travelled (relative front) to measure migration and resolution of different plasmid topoisomers.

2.4.3 Polyacrylamide gel electrophoresis of DNA

DNA fragments smaller than 1 kb were analysed by polyacrylamide gel electrophoresis. Polyacrylamide gels contained 7.5% (w/v) stock acrylamide (ProtGel) (National Diagnostics), 4% glycerol and 1 x TBE. Gels were polymerised by the addition of 0.01 volumes of 10% (w/v) ammonium persulphate and 0.001 volumes TEMED (N,N,N′,N′-tetramethylethylenediamine). Samples were mixed in a 1:1 ratio with DNA loading dye before being loaded into the sample wells. The DNA markers used were the 100 bp and 1 kb DNA ladders (New England BioLabs), which were diluted 6-fold in DNA loading dye, supplied by the manufacturer. Samples were subjected to electrophoresis in 1 x TBE at 30-40 mA for 30 min to 4 hours, then stained in ethidium bromide solution and visualised under UV light as described in section 2.5.1.

2.5 Extraction and purification of nucleic acids

2.5.1 Phenol/chloroform extraction of DNA

Contaminating proteins were removed from DNA solutions by phenol/chloroform extraction. DNA solutions were mixed with an equal volume of phenol/chloroform/isoamylalcohol

solution (composition 25/24/1 v/v, pH 8.0) (Fisher Scientific), mixed by vortex for 15 s and the aqueous and organic phases separated by centrifugation for 3 min at ~18000 x *g*. The aqueous layer, which contained the DNA, was transferred to a fresh tube. The remaining organic phase was then mixed with an equal volume of TE buffer (10 mM Tris pH 8.0, 0.5 mM EDTA) by vortex and centrifuged for 3 min at ~18000 x *g*. The DNA-containing aqueous layer was combined with that removed previously and concentrated by ethanol precipitation.

2.5.2 Ethanol precipitation of DNA

DNA solutions were concentrated by ethanol precipitation. 0.1 volumes of 3 M sodium acetate (pH 5.2) and 2.5-3 volumes of ice-cold 100% ethanol were added to the DNA solution. 1 µl 20 mg/ml glycogen was also added to solutions containing DNA fragments smaller than 500 bp. Samples were incubated at –20°C for 30 min, then centrifuged for 15 min at 4°C at ~18000 x *g*. After removal of the supernatant, the pellet was washed with 1 ml ice-cold 70% ethanol and centrifuged for 10 min at 4°C at ~18000 x *g*. The supernatant was removed again and the pellet dried for 10-15 minutes under vacuum. The dry pellet was resuspended in the desired buffer, as indicated.

2.5.3 Purification of DNA using QIAquick PCR Purification kit

DNA was purified after PCR or restriction digestion using the QIAquick PCR Purification Kit (Qiagen), as described by the manufacturer. Purified DNA was eluted from QIAquick columns in 50 µl sterile distilled water.

2.5.4 Extraction of DNA fragments from agarose gels

DNA samples were subjected to electrophoresis in 0.8-1.2% agarose gels as described in section 2.4.1. DNA bands were excised from the gel and eluted using the QIAquick Gel Extraction Kit (Qiagen) as described by the manufactuer. DNA fragments were eluted from QIAquick columns in 50 µl sterile distilled water.

2.5.5 Electroelution of DNA fragments from polyacrylamide gels

Small DNA fragments (<1 kb) were purified by the process of electroelution (Maniatis *et al.*, 1982) after PCR or restriction digestion. DNA samples were subjected to electrophoresis using a 7.5% polyacrylamide gel as described in section 2.4.3. DNA bands were excised from the gel and placed into 6.3 mm dialysis tubing (Medicell International Ltd.), filled with 0.1 x TBE buffer and sealed. Elution of DNA was achieved by passing a current of 30 mA through the gel for 20-30 mins. The DNA-containing buffer was then removed from the dialysis bag and transferred to a microcentrifuge tube. DNA was extracted from the buffer by phenol/chloroform extraction and ethanol precipitation as described in sections 2.5.1 and 2.5.2.

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

An overnight culture of the strain carrying the plasmid to be extracted was grown in 5 ml LB medium supplemented with the appropriate antibiotic. Plasmid DNA was extracted from overnight culture using the QIAprep Spin Miniprep kit (Qiagen), as described by the manufacturer. Plasmid DNA was eluted from the QIAprep column in 50 µl (high copy number plasmids) or 25 µl (low copy number plasmids) sterile distilled water.

2.5.7 Purification of genomic DNA

To extract genomic DNA, an overnight culture of the appropriate *E. coli* strain was grown in 5 ml M9 minimal salts medium, supplemented with 0.3% fructose. 10 ml M9 minimal medium (plus fructose) was inoculated with 200 µl overnight culture, and the culture grown under appropriate culture conditions to an OD_{650} of 0.4-0.5. Cells were pelleted from 5 ml aliquots of culture by centrifugation for 20 min at ~2700 x *g*, 4°C. Cell pellets were stored at –80°C. Genomic DNA was then extracted from bacteria using the illustra bacteria genomicPrep Mini Spin Kit (GE Healthcare), as described in the manufacturer's protocol, eluted in 200 µl of elution buffer and stored at -20°C. An aliquot was digested with *Hin*dIII to check the purity of the sample and the DNA concentration determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific).

2.5.8 Preparation of total cellular RNA

Total cellular RNA was extracted for use in qRT-PCR using the QIAGEN RNeasy mini kit. Solutions for use in work with RNA were treated with diethyl pyrocarbonate (DEPC) to inhibit RNase activity. DEPC was added to solutions to a final concentration of 0.1%, and left overnight in a fume hood to allow the DEPC to evaporate before autoclaving. An overnight culture of the appropriate strain was grown in 5 ml M9 minimal salts medium, supplemented with 0.3% fructose, which were prepared as previously described. 10 ml M9 minimal medium (plus fructose) was inoculated with 200 µl overnight culture, and the culture grown under appropriate culture conditions to an OD_{650} of 0.4-0.5. Three 2.5 ml aliquots of culture were each mixed with 5 ml RNAlater (Ambion) to stabilise RNA, incubated at room temperature for 5 min, and the cells pelleted by centrifugation for 20 min at ~2700 x *g*, 4°C. Cell pellets were stored at –80°C. The RNA purification was completed using the Qiagen RNeasy Mini

Kit as described by the manufacturer. RNA was eluted from each column in 30 µl RNase-free water, and the two eluates for each sample were combined. To remove any DNA contamination the Qiagen RNase-Free DNase Set was used for on-column DNase digestion during the RNA purification, as described by the manufacturer. RNA concentrations and A260/A280 were determined using a NanoDrop ND-1000 spectrophotometer (Thermo Scientific) and an aliquot analysed by agarose gel electrophoresis to check the integrity of the RNA. RNA samples were stored at -80°C.

2.6 Bacterial transformations

2.6.1 Preparation of competent cells using the calcium chloride method

1 ml of an overnight culture of the strain to be transformed was used to inoculate 50 ml LB medium, which was incubated at 37°C with aeration until mid-exponential phase (OD_{650} 0.3-0.5). The culture was then incubated for 10 min on ice after which the cells were harvested by centrifugation for 5 min at ~3400 x *g* at 4°C. The cell pellet was resuspended in 25 ml ice-cold calcium chloride (100 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5), incubated for 10 min on ice after which the cells harvested by centrifugation for 5 min at ~3400 x *g* at 4°C. The cells were resuspended in 3 ml ice-cold buffer freeze-thaw buffer (100 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.5, 15% glycerol) and the cells kept on ice for 24 h before use. Competent cells were stored as 200 µl aliquots at –80°C.

2.6.2 Preparation of competent cells using the rubidium chloride method

The rubidium chloride method was used to maximize the transformation efficiency of competent cells, needed for difficult cloning procedures or double transformations. The

desired strain was grown as described in section 2.6.1. and the culture incubated on ice for 10 mins before harvest of the cells by centrifugation for 5 min at ~3400 x *g* at 4°C. The supernatant was discarded and the cell pellet resuspended in 1/2.5 volume of ice-cold TFB1 buffer (30 mM potassium acetate, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM RbCl, 15% glycerol, pH 5.8 with 1 M acetic acid). Cells were incubated on ice for 10 min and centrifuged at ~3400 x *g* for 5 min at 4°C. Cells were resuspended in 1/25 volume ice-cold TFB2 buffer (10 mM PIPES, pH 6.5, 75 mM $CaCl_2$, 10 mM RbCl, 15% glycerol, pH 6.5 with 1 M KOH) and incubated on ice for 1 h. Cells were stored as 200 µl aliquots in sterile microfuge tubes at -80°C.

2.6.3 Transformation of competent cells with plasmid DNA

The desired strain was made competent by the previously described procedures after which 50-100 µl competent cells were mixed with 100 ng plasmid DNA on ice, and incubated on ice for 45-60 min. Cells were heat-shocked at 42°C for 90 s and returned to ice for 2 min. 1 ml 37°C SOC medium was added to the cells, which were incubated for 30-60 minutes at 37°C with aeration. Cells were harvested by centrifugation at ~18000 x *g*, and then resuspended in approximately 100 µl of the supernatant. Cells were spread onto nutrient agar or MacConkey agar supplemented with required antibiotic(s), and incubated overnight at 37°C.

2.7 Recombinant DNA techniques

2.7.1 Routine PCR

Polymerase chain reaction (PCR) was carried out using Finnzymes Phusion™ High-Fidelity DNA polymerase (New England BioLabs) or Biotaq polymerase (Bioline) in the buffer

supplied with the relevant enzyme. Reaction conditions for different enzymes were adjusted according to manufacturer's instructions. Phusion polymerase was used for PCR throughout the study whilst Biotaq polymerase was used mainly for colony PCR and difficult PCR. dNTPs (Bioline) were used at a final concentration of 1 mM (0.25 mM each) and the appropriate oligonucleotide primers were used at a final concentration of 1 µM each. PCR reaction mixes were made up to 50 µl in sterile distilled water and the relevant buffer, supplied by the manufacturer. DNA amplification was performed in an oil-free thermal cycler (GeneAmp® PCR System, Applied Biosystems) and cycling conditions are shown in table 2.3.

2.7.2 Colony PCR

Colony PCR was used to amplify homology regions and promoter fragments from the chromosome and to screen for the presence of chromosomal insertions/deletions. 100 µl sterile distilled water was inoculated with one single fresh colony and heated to 100°C for 10 min. Cell debris was harvested by centrifugation for 1 min at ~18000 × *g*. 10 µl of the genomic DNA-containing supernatant was used in a 50 µl Biotaq polymerase PCR as described in section 2.7.1.

2.7.3 Site directed mutagenic PCR

Site directed mutagenic PCR was used to introduce point mutations in a DNA sequence of interest. To do this, the desired base change was incorporated into an oligonucleotide primer, which was used to amplify the DNA fragment carrying desired mutations. To create point mutations in the *Bsr*GI site encoded by the pJB2 plasmid, primers were designed to contain the desired mutation in the middle of the primer (~10-15 bp of correct sequence either side)

Table 2.3: Standard PCR cycle

An annealing temperature (T_A°C) of 3-5°C below the melting temperature of the primers was used. The extension time (X min) was calculated based on PCR product length and rate of extension of the polymerase used (according to the manufacturer's instructions).

Temperature	Time		Purpose
94°C	5 min		"hot start"
94°C	30 s		Melting
T_A°C	30 s	30-35 cycles	Annealing
72°C	X min		Extension
72°C	5 min		Final extension

and anneal to the same sequence on opposite strands of the plasmid. These primers were used in routine PCR using Phusion polymerase with annealing temperature of 55°C for 1 minute and extension time of 5 minutes, to allow the entire plasmid to be amplified, with 12 cycles of the reaction. Template plasmid DNA was digested with the methylation-sensitive restriction endonuclease *Dpn*I, which cleaves only when the target site is methylated, for 1 h at 37°C. The *ΔrecA E. coli* XL-1 Blue strain was then transformed with the PCR product by the rubidium chloride method, described in section 2.6.2, to allow recovery and amplification of plasmid DNA carrying the desired mutation. Several candidates were screened for the desired mutation using primer D55668 (Table 2.4).

2.7.4 Restriction digestion of DNA

Purified PCR product or plasmid miniprep DNA solutions were digested by the addition of 40 units restriction enzyme (New England BioLabs) to 50 µl purified DNA solution in a final volume of 60 µl of the appropriate reaction buffer, supplied by the manufacturer. Reactions were incubated for 3 h at 37°C. Digested plasmid DNA for use in cloning was treated with 20 units calf alkaline phosphatase (CIP) (New England BioLabs) for a further 1 h to remove terminal 5' phosphate groups, in order to prevent religation of vector DNA. Digested and CIP treated DNA was purified as previously described in section 2.5.

2.7.5 DNA ligations

The concentration of DNA fragments to be ligated was estimated by agarose gel electrophoresis and comparison with molecular weight markers, such as the 100 bp and 1 kb DNA ladders (New England BioLabs). The optimum molar ratio of vector:insert for ligation

Table 2.4: Oligonucleotides used for sequencing of inserts

Name	Sequence (5' – 3')	Use
D54890	TGAAGGGCAATCAGCTGTTG	Anneals to 3' end of *lacI* gene, upstream of cloning site in pKH3. Used for sequencing inserts in pKH3.
D55668	GAGCGGCGACGATAGTCATG	Anneals downstream of *Sal*I site in pBR322. Used for sequencing inserts in pBR322 during construction of pKH3 and pJB plasmids.
D10520	CCCTGCGGTGCCCCTCAAG	Anneals upstream of *Eco*RI site in pRW50. Used for sequencing and amplification of inserts in this vector.
D56613	CTTGATGTCTCTGACCAGAC	Anneals within the *lacI* sequence outside of the pKH3 homology. Used for screening gene gorging candidates by PCR amplification.
D56614	TTATGCAGCAACGAGACGTC	Anneals within the *lacZ* sequence outside of the pKH3 homology. Used for screening gene gorging candidates by PCR amplification.
D68556	TTTACGTCGCCGTCCAG	Anneals downstream of the start codon of *gfp*. Used for sequencing promoter inserts in pJB plasmid derivatives and BRY strains.
D58793	GGATGTGCTGCAAGG	Sequencing primer to check homology regions inserted between *Nhe*I and *Sac*I. Binds downstream of *I-sceI* site
D58794	TATGCTTCCGGCTCG	Sequencing primer to check homology regions inserted between *Mfe*I and *Xma*I. Binds downstream of *I-sceI* site

was found to be between 1:1 and 1:3, therefore these ratios were used routinely throughout the study, however individual reactions occasionally required further optimisation. Restriction digested and alkaline phosphatase treated vector DNA was mixed with restriction digested insert DNA and 1 µl T4 DNA ligase (New England BioLabs) in a final reaction volume of 20 µl T4 DNA ligase buffer, which was supplied by the manufacturer. Reactions were incubated for 20 min at room temperature or 16 h at 4˚C. The rubidium chloride method of transformation, described in sections 2.6.2 and 2.6.3, was used to transform 50 µl RLG221 rubidium chloride competent cells with 10 µl of each ligation reaction. Candidate recombinant plasmids were screened by colony PCR, described in section 2.7.2, using oligonucleotide primers flanking the insertion site, therefore detecting a change in size by gel electrophoresis. Successful candidates were further screened for the presence of the correct insert by sequencing using oligonucleotide primers flanking the insertion site, as described in section 2.7.6.

2.7.6 DNA sequencing

Sequencing of plasmid and PCR products was completed by the Functional Genomics and Proteomics Laboratory, University of Birmingham, UK. Plasmid templates were prepared by "mini-prep", as described in section 2.5.6, after which ~100 ng of plasmid miniprep was mixed with 1 µM sequencing primer in a final volume of 10 µl. PCR product were purified as described in section 2.5.3, after which 1-100 ng (dependent on fragment size) purified PCR product was mixed with 1 µM sequencing primer in a final volume of 10 µl. Oligonucleotides used for sequencing inserts in plasmids are listed in Table 2.4.

2.8 Gene doctoring

2.8.1 Gene doctoring method

A general protocol for insertion of a DNA donor fragment into the *E. coli* chromosome by gene doctoring is described here with slight modifications from the original protocol (Lee *et al*., 2009). This method requires the use of a plasmid which carries kanamycin resistance in the donor fragment, with ampicillin resistance and the sucrose sensitivity gene carried by the rest of the plasmid. Details of target site selection and donor plasmid construction can be found in section 2.9. The strain to be modified by gene doctoring was co-transformed with the mutagenesis plasmid, pACBSR (Cm^R; Figure 2.3), and the donor plasmid (Kan^R, Amp^R, *sacB*; Figure 2.6), and co-transformants were selected on nutrient agar supplemented with chloramphenicol and ampicillin. Maintenance of a functional *sacB* gene on the donor plasmid was tested by restreaking single co-transformant colonies onto agar supplemented with kanamycin and 5% sucrose and then agar supplemented with kanamycin and chloramphenicol. A single sucrose sensitive co-transformant colony was suspended in 500 µl LB medium, supplemented with ampicillin and chloramphenicol, and incubated at 37˚C with aeration for ~3 h until the culture was turbid. Cells were harvested by centrifugation at ~18000 x *g*, and then washed in 0.1 x LB by resuspension. This process was repeated twice to remove any residual antibiotics, as no selection is required during induction of the mutagenesis machinery. Cells were then suspended in 500 µl 0.1 x LB supplemented with 0.3% L-arabinose to induce expression of the I-SceI endonuclease and λ-red homologous recombination proteins, which are encoded by the mutagenesis plasmid pACBSR. 0.1 x LB medium was used as no further growth was required during the mutagenesis procedure. The culture was incubated at 37˚C with aeration for a further ~3 h, after which 125 µl of the culture was spread onto each of 4 nutrient agar plates supplemented with kanamycin and 5% sucrose. Efficiency of donor plasmid linearisation was tested by spreading 100 µl of a 1/1000

dilution of the culture, taken before and after addition of arabinose to the culture medium, on nutrient agar and nutrient agar supplemented with ampicillin. Cells in which the donor plasmid has been linearised will not be able to grow in the presence of ampicillin, therefore comparison of the number of ampicillin resistant colonies pre- and post-induction demonstrates the efficiency of donor plasmid linearisation. All plates were incubated at room temperature until visible colonies were detected.

2.8.2 Screening of gene doctoring candidates

Kanamycin resistant, sucrose insensitive colonies were restreaked onto nutrient agar supplemented with ampicillin and nutrient agar supplemented with kanamycin and 5% sucrose to check for loss of the donor plasmid. Ampicillin sensitive candidates were screened for the presence of the chromosomal insert by colony PCR, as described in section 2.7.2, using oligonucleotide primers designed to bind to regions of the chromosome flanking the regions of homology used for recombination. PCR products demonstrating the presence of the insert were further screened by DNA sequencing, as described in section 2.7.6, after which the candidates were screened for chloramphenicol sensitivity to confirm loss of the mutagenesis plasmid.

2.8.3 Excision of FRT flanked *kan* gene from the chromosomal insert

Selection of candidates carrying the chromosomal insert requires a kanamycin resistance marker to be incorporated with the desired insert. The kanamycin resistance cassette is flanked by directly repeated FRT sites, which are targets for FLP recombinase, therefore allowing excision of the cassette by expression of FLP recombinase within the cell from the temperature sensitive plasmid pCP20 (Figure 2.4). Successful, chloramphenicol sensitive,

gene doctoring candidates were transformed with pCP20 by the calcium chloride method, described in sections 2.6.1 and 2.6.3, and selected by growth at 30˚C (permissive temperature) on nutrient agar supplemented with ampicillin. For each strain, 4 transformants were restreaked on non-selective nutrient agar and grown overnight at 37˚C (non-permissive temperature) to induce expression of FLP recombinase and plasmid loss. Individual colonies were then restreaked onto nutrient agar with and without kanamycin or ampicillin and incubated at 37˚C overnight. Candidates sensitive to kanamycin and ampicillin were further screened for loss of the kanamycin resistance gene by colony PCR using oligonucleotide primers positioned as described in section 2.8.2. The PCR product was ~1.3 kb smaller upon loss of the kanamycin resistance cassette, but still larger than that from the starting strain. Excision of the *kan* gene was further confirmed by DNA sequencing of the PCR product.

2.9 Insertion of the *lac28::egfp* fusion at different chromosomal loci

2.9.1 Overview of strategy

Gene doctoring is used to transfer a promoter::*egfp* fusion to different chromosomal loci to allow the effect of chromosomal position on promoter activity to be analysed by measuring expression of the eGFP reporter gene. To achieve this, a master donor plasmid was created to allow the fusion of any promoter fragment to *egfp*, and the cloning of regions of homology to the target chromosome into sites flanking this fusion on the plasmid. Target sites for insertion were selected and regions of homology to the chromosome cloned into the donor plasmid. Gene doctoring, as described in section 2.8, was then used to transfer the *lac28::egfp* fusion to the desired target locus on the *E. coli* K-12 MG1655 chromosome.

2.9.2 Construction of the donor plasmid pJB

The pJB gene doctoring master donor plasmid was constructed by modifying the gene gorging vector pKH5 to replace the *lacI* homology region with a multiple cloning site and the *lacZ* homology region with *egfp* and a multiple cloning site. The *I-sceI* donor fragment was then cloned into the gene doctoring vector pDOC-C (Lee *et al*., 2009). The oligonucleotide primers used in this construction are listed in table 2.5. Primers D67847 and D67848 were used in inverse PCR, with pKH5 as the template, to remove the *lacI* homology region from pKH5 and incorporate target DNA sites for the restriction endonucleases *Mfe*I and *Age*I (Figure 2.8). D67847 anneals to the region containing the *I-sceI* target site upstream of the *lacI* homology region with the 3' end oriented away from the *lacI* fragment and the 5' end incorporating a target site for *Mfe*I. D67848 anneals downstream of the *lacI* fragment in pKH5 and incorporates a target site for *Age*I. The PCR product was purified using the QIAquick PCR purification kit, digested with the restriction endonucleases *Mfe*I and *Age*I and further purified by extraction from a 0.8% agarose gel using the QIAquick Gel Extraction Kit. Two oligonucleotides, D67849 and D67876, were annealed to form an *Mfe*I-*Age*I fragment encoding target sites for the restriction endonucleases *Xma*I and *Kpn*I. 20 µl each of 10 µM D67849 and D67876, which are complementary primers that form an *Mfe*I-*Age*I fragment, were mixed and heated to 65˚C for 10 minutes then cooled to room temperature to allow the oligonucleotides to anneal (Figure 2.8). The digested and purified vector was then ligated with the *Mfe*I-*Age*I fragment and used to transform *E. coli* RLG221 by the calcium chloride method to produce plasmid pJB1 (Figure 2.8). The insert in pJB1 was confirmed by test restriction digest with *Xma*I and *Kpn*I and agarose gel electrophoresis.

The second stage of master donor plasmid construction required the *egfp* gene to be cloned in place of the *lacZ* fragment in pJB1 to produce plasmid pJB2 (Figure 2.8). The *egfp* gene was amplified by PCR from pDEX-G using primers D67961 and D67962, which incorporate a

Table 2.5: DNA oligonucleotides used for construction of pJB plasmids

Name	Sequence (5' – 3') [1]	Use
D67847	TACGCAATTGATTACCCTGTTATCCCTAG	Anneals to the *Sce*I site upstream of the *lacI* homology in pKH5 and incorporates an *Mfe*I site. Used with D67848 to create pJB1 by inverse PCR.
D67848	GCTACGACCGGTGGCTGGAGCTGCTTCGAAG	Anneals downstream of the *lacI* homology in pKH5 and incorporates an *Age*I site. Used with D67847 to create pJB1 by inverse PCR.
D67849	AATTGTATGCCCCGGGTTGTAGGTACCAACTA	Complementary to D67876 and encodes the pJB1 multiple cloning site (Figure 2.8).
D67876	CCGGTAGTTGGTACCTACAACCCGGGGCATAC	Complementary to D67849 and encodes the pJB1 multiple cloning site (Figure 2.8).
D67961	TACGAAGCTTAGCAAGGGCGAGGAGCTG	Anneals to the 5' end of *gfp* in pDEX-G and incorporates a *Hin*dIII site. Used to create pJB2
D67962	TACGGTCGACTTGTATGTGTACATTCGTTAGTTACTTACTTGTACAGCTCGTCCATG	Anneals to the 3' end of *gfp* in pDEX-G and incorporates two extra stop codons in each reading frame, a *Bsr*GI site and a *Sal*I site. Used to create pJB2 (Figure 2.8).
D68276	CTCGGCATGGACGAGCTTACAAGTAAGTAACTAAC	Anneals to the 3' end of *gfp* and incorporates a G→T silent mutation of the *Bsr*GI site. Used to create pJB2ΔB.
D68277	GTTAGTTACTTACTTGTAAAGCTCGTCCATGCCGAG	Anneals to the 3' end of *gfp* and incorporates a G→T silent mutation of the *Bsr*GI site. Used to create pJB2ΔB.
D68085	GTACAGCCATAACTAGTACTTGCTAGCGCACAAGAGCTCTAGGGATAACAGGGTAATG	Complementary to D68086 and encodes the pJB3 multiple cloning site (Figure 2.9).
D68086	TCGACATTACCCTGTTATCCCTAGAGCTCTTGTGCGCTAGCAAGTACTAGTTATGGCT	Complementary to D68085 and encodes the pJB3 multiple cloning site (Figure 2.9).
D66948	TATAGAATTCGGGCAGTGAGCGCAACGC	Anneals upstream of O^3 in the *lac* promoter region and incorporates an *Eco*RI site. Used to create the *lac00* and lac28 fragments.

Table 2.5: DNA oligonucleotides used for construction of pJB plasmids (continued)

Name	Sequence (5' – 3') [1]	Use
D68443	GCCGAAGCTTCATAGCTGTT TCCTGTGTG	Anneals to the *lac* promoter translation initiation region and incorporates a *Hin*dIII site downstream of the ATG. Used with D66948 to amplify *lac00* to create pJB4.
D69482	GCCGAAGCTTAAGGCGATTA AGTTGGG	Anneals to *lacZ* to incorporate a *Hin*dIII site immediately downstream of codon 28 to create the *lac28* fragment.

[1] Target sites for restriction endonucleases are underlined.

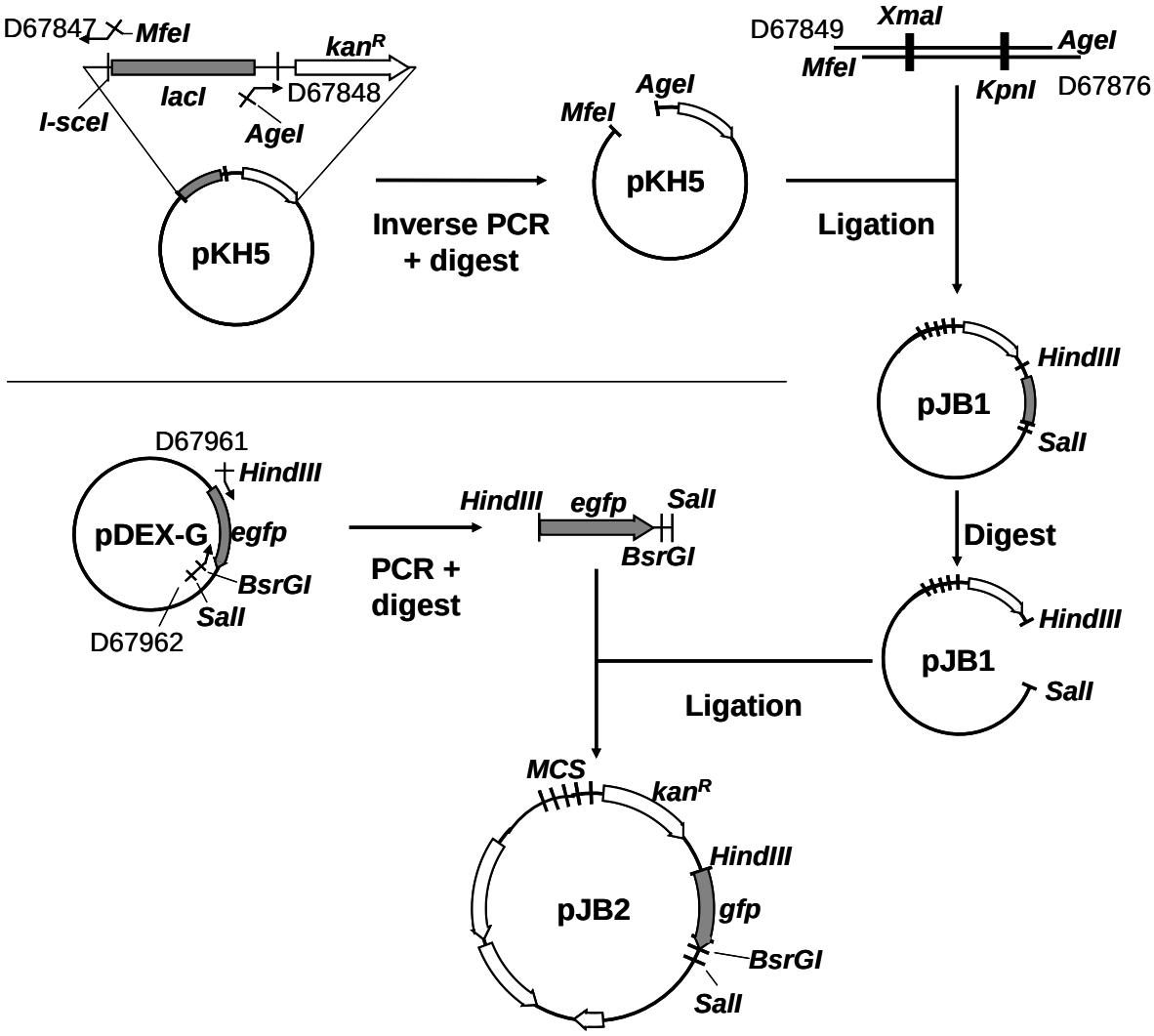

Figure 2.8: Schematic overview of pJB2 construction

For details see text (section 2.9.2). Inverse PCR was used to replace the *lacI* homology region of pKH5 with an *Mfe*I and *Age*I endonuclease target sites. Oligonucleotides were annealed to form a *MfeI-Age*I fragment containing a multiple cloning site (MCS), which was cloned into *Mfe*I-*Age*I digested pKH5. The resulting plasmid (pJB1) was then digested with the endonucleases *Hin*dIII and *Sal*I, after which the *egfp* gene of plasmid pDEX-G was cloned into this site to form pJB2. Black circles represent plasmid DNA with the plasmid name in the centre. Genes are represented by block arrows with the direction of transcription depicted by the arrow and the gene name labelled. The position of restriction endonuclease target sites of interest are represented by black lines perpendicular to the plasmid or PCR DNA lines. Oligonucleotides are represented by thin arrow lines with the primer code number (DNNNNN). Cloning steps are represented by thick black arrows with the procedure labelled.

*Hind*III target site at the 5’ end of the gene and stop codons with a downstream *Bsr*GI target site at the 3’ end. The PCR product was then purified using the QIAquick PCR Purification Kit and digested using the restriction endonucleases *Hin*dIII and *Sal*I. pJB1 was also digested with *Hin*dIII and *Sal*I, after which both vector and digested PCR product were purified by extraction from a 0.8% agarose gel. *Hind*III/*Sal*I digested pJB1 was then ligated with the *Hin*dIII/*Sal*I fragment carrying the *egfp* gene and used to transform calcium chloride competent *E. coli* RLG221 cells to produce plasmid pJB2.

The next stage in plasmid construction required the insertion of a multiple cloning site and a target for the I-SceI endonuclease downstream of the *egfp* gene in pJB2. However, a target recognition site for the restriction endonuclease *Bsr*GI was overlooked at the 3’ end of the*egfp* gene, therefore this first had to be deactivated by site directed mutagenesis. The mutagenesis primers D68276 and D68277 incorporate a G to T silent mutation within the *Bsr*GI site and anneal to complementary sequences with the mutation positioned in the centre. The primers were used in a Phusion polymerase PCR reaction, after which the template plasmid DNA was digested by adding 1µl *Dpn*I enzyme (NEB) to the PCR reaction, which was incubated at 37˚C for a further 1 h, leaving only the un-methylated PCR product. The PCR product was then used to transform rubidium chloride competent *E. coli* XL-1 Blue cells and transformants were selected on nutrient agar supplemented with ampicillin. The mutation was checked by sequencing using D55668 (Table 2.4). The complementary oligonucleotides, D68085 and D68086, encode a *Bsr*GI-*Sal*I fragment carrying target sites for the restriction endonucleases *Spe*I, *Nhe*I, *Sac*I and I-SceI. These primers were annealed as previously described for the pJB1 multiple cloning site and cloned into *Bsr*GI-*Sal*I digested pJB2 to produce plasmid pJB3 (Figure 2.9). The insert was confirmed by sequencing using D55668 (Table 2.4).

The *E. coli lacZYA* promoter is well characterised and known to be a strong natural promoter, therefore the regulatory region of *lacZYA* was cloned into pJB3 to drive expression of the *egfp*

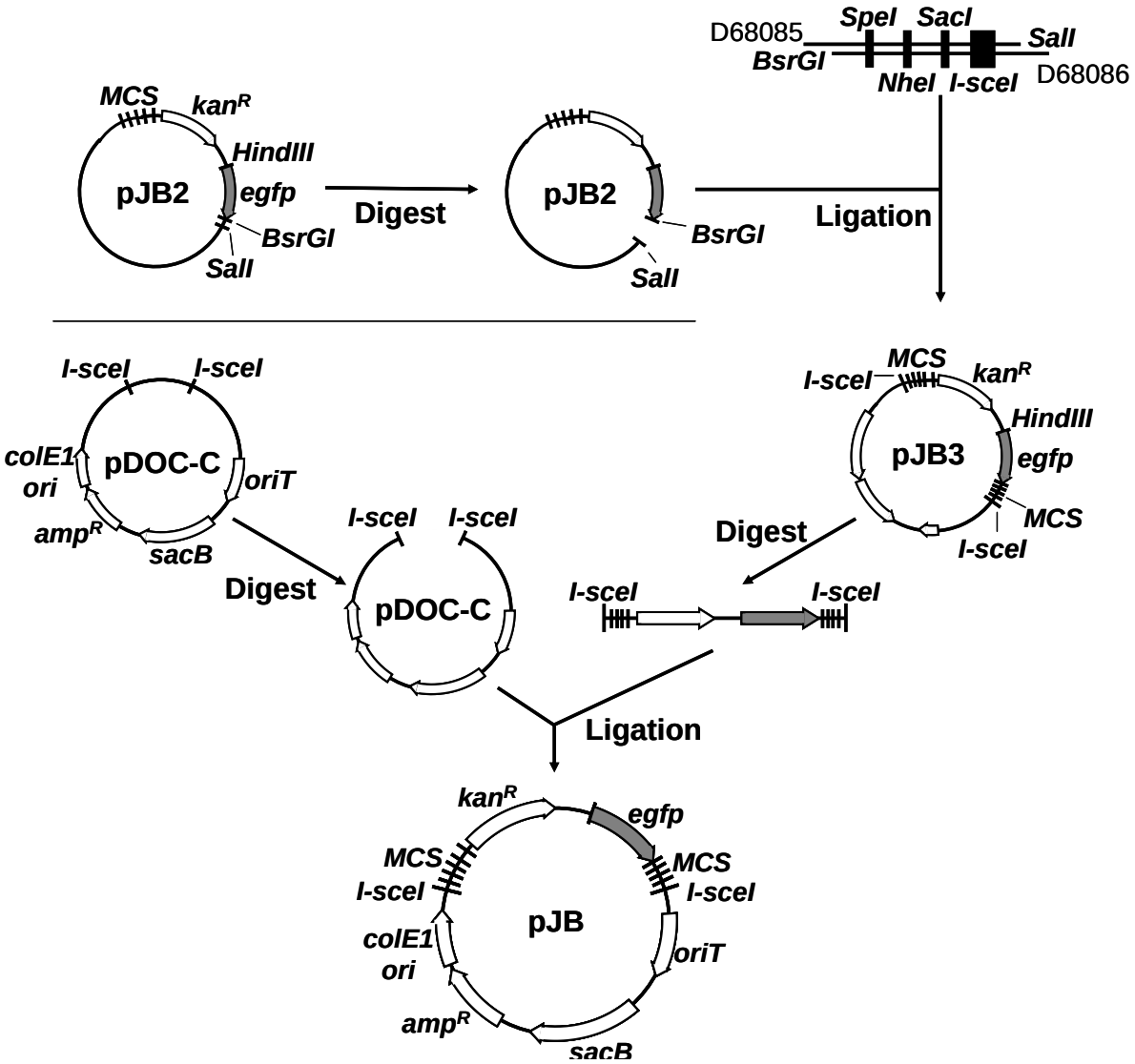

Figure 2.9: Schematic overview of pJB construction

For details see text (section 2.9.2). Oligonucleotide primers were annealed to form a *Bsr*GI-*Sal*I fragment containing a multiple cloning site (MCS), which was then cloned into *Bsr*GI-*Sal*I digested pJB2 plasmid to form pJB3. The *egfp* containing *I-sceI* fragment of pJB3 was then cloned into *I-sce*I digested pDOC-C plasmid to form the master gene doctoring plasmid pJB. Black circles represent plasmid DNA with the plasmid name in the centre. Genes are represented by block arrows with the direction of transcription depicted by the arrow and the gene name labelled. The position of restriction endonuclease target sites of interest are represented by black lines perpendicular to the plasmid or PCR DNA lines. Oligonucleotides are represented by thin arrow lines with the primer code number (DNNNNN). Cloning steps are represented by thick black arrows with the procedure labelled.

gene. The *lacZYA* regulatory region, from position -93 bp to +41 bp relative to the transcription start site and including the *lacZ* ATG, was amplified by PCR from *E. coli* K-12 genomic DNA using primers D68498 and D66948, which incorporate target sites for the endonucleases *Eco*RI and *Hin*dIII. The resulting promoter fragment was referred to as *lac00* and was then digested with *Eco*RI and *Hin*dIII, after which the promoter fragment was purified by extraction from a 0.8% agarose gel using the QIAquick Gel Extraction Kit. The digested and purified promoter fragment was then ligated into *Eco*RI/*Hin*dIII digested and CIP treated pJB3 to produce plasmid pJB4. The presence and integrity of the promoter fragment insert were confirmed by DNA sequencing using primer D68556 (Table 2.4). The insert was checked by sequencing using D68556. *E. coli* K-12 KH000 cells carrying pJB4 were then checked for expression of *egfp* by fluorescence microscopy, however the cells were found to be non-fluorescent (see chapter 3). Therefore a *lac* promoter fragment carrying part of the *lacZ* gene was cloned into the donor plasmid in place of *lac00*. The *lacZYA* regulatory region, from position -93 bp to +122 bp relative to the transcription start site and including the first 28 codons of the *lacZ* gene, was amplified by PCR from *E. coli* K-12 genomic DNA using primers D68498 and D69482, which incorporate target sites for the endonucleases *Eco*RI and *Hin*dIII. The resulting promoter fragment was referred to as *lac28* and cloned into *Eco*RI-*Hin*dIII digested pJB3 plasmid to produce plasmid pJB15, as previously described for the *lac00* fragment.

Finally, the *I-sceI* fragment of pJB15 was cloned into pDOC-C to produce a plasmid capable of being used to transfer the donor molecule to the *E. coli* chromosome through the method of gene doctoring (Figure 2.9). The pJB15 and pDOC-C plasmids were both digested with the endonuclease I-SceI, however pDOC-C was further treated with CIP to remove terminal 5' phosphate groups, in order to prevent re-ligation of vector DNA. Both vector and insert were then purified by extraction from 0.8% agarose gels using the QIAquick Gel Extraction Kit, after which I-SceI digested pDOC-C was ligated with the pJB15 *I-sceI* fragment and used to

transform calcium chloride *E. coli* RLG221 to produce the pJB master donor plasmid (Figure 2.6).

2.9.3 Selection of target sites for *lac28::egfp* insertion

For the *lac28::gfp* fusion to be inserted at different locations around the *E. coli* K-12 MG1655 chromosome, the sites were selected and targeted specifically to minimise disruption to local chromosomal processes. The positions targeted are represented on the circular map presented in figure 2.10 and detail of each insertion site, including position of genes and relevant transcription elements, is represented in figures 2.11-2.24. Intergenic regions between convergent genes were targeted where possible and rho-independent transcription terminators left undisturbed in order to minimise disruption of local transcriptional activity and chromosomal architecture. The exceptions to this were the *dkgB* locus, which is immediately downstream of the *rrsHrrlHrrfH* and *aspU* genes, but upstream of the *dkgB* regulatory region, and the *eaeH* locus, which is situated within the non-expressed *eaeH* pseudogene. Homology regions (HR) for gene doctoring were selected to insert the fusion without replacement of any natural *E. coli* K-12 MG1655 DNA, except in the case of the *ΔtsEPOD* strains, in which the region of high protein binding was intentionally replaced by the *lac28::egfp* fusion. HRs were also selected so as not to include any recognition sites for the *Mfe*I, *Xma*I, *Nhe*I and *SacI* restriction endonucleases. Identification of local sequence elements was enabled by analysis of data provided by EcoCyc at www.ecocyc.org (Keseler *et al*., 2011).

2.9.4 Cloning of regions of homology to the *E. coli* chromosome into the pJB plasmid

In order to insert the *lac28::egfp* fusion at the chromosomal loci described in section 2.9.3, the homology regions selected, which are represented in figure 2.11-2.24, were cloned into

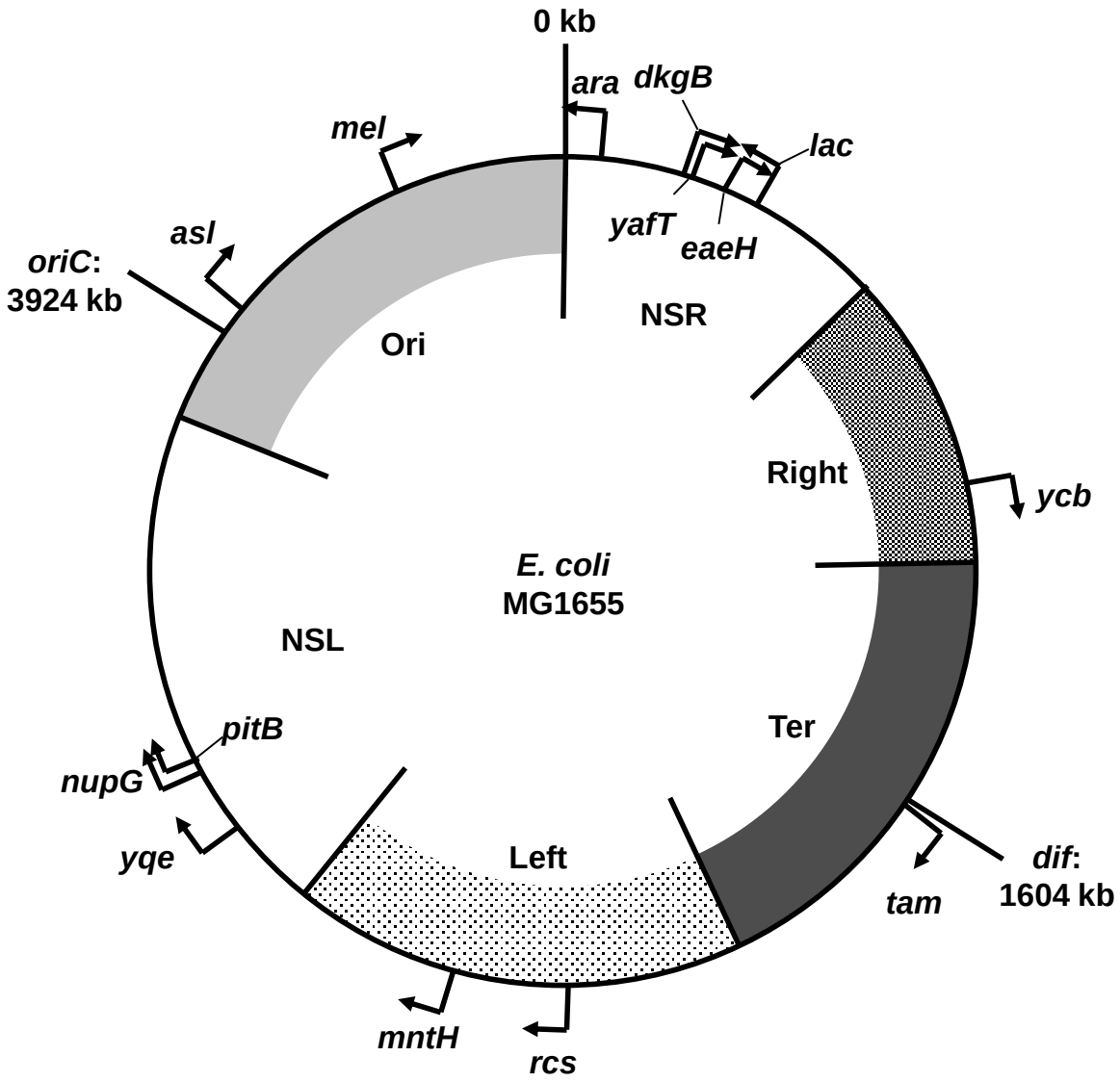

Figure 2.10: Map of the *E. coli* MG1655 chromosome and *lac28::egfp* insertion sites

The black circle represents the *E. coli* MG1655 circular chromosome of total size 4639675 bp. The grey shaded blocks and two white gaps represent the six macrodomains as defined by Valens *et al*. (2004) and are labelled. The position of *oriC* and the *diff*, the direct opposite locus, are marked with black lines and labelled with their positions in kb relative to the co-ordinate system origin. Arrows indicate the position of sites of insertion for the *lac28::egfp* construct and the direction of the arrows indicates the orientation of the promoter. Sites of insertion are labelled according to the names of neighbouring genes.

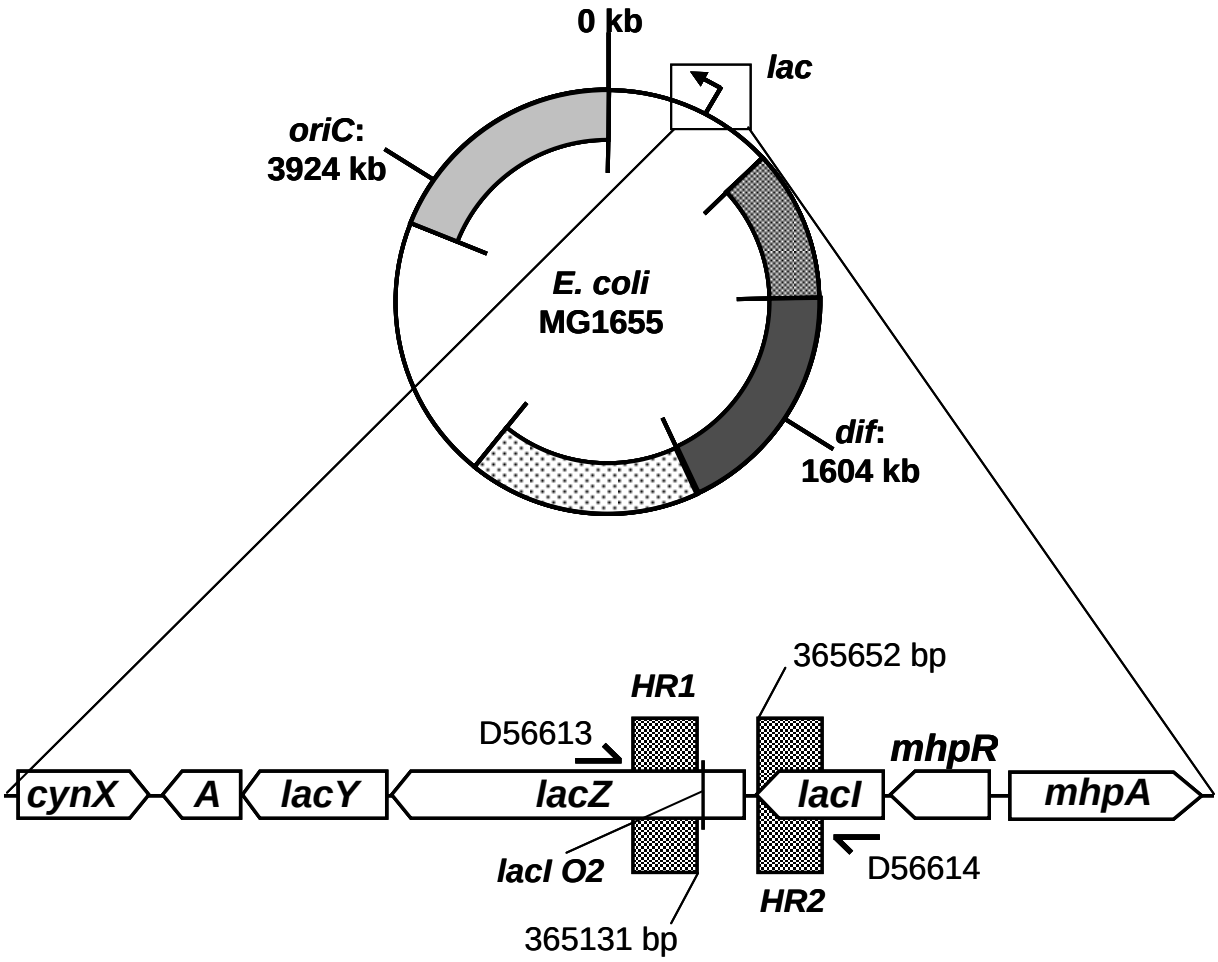

Figure 2.11: Schematic diagram of the *lac* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *lac* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *lac* locus insertion site, with genes represented by white block arrows. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY40, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*, with the exact positions in bp with respect to the co-ordinate system origin. Due to the position of the HRs, gene doctoring will cause replacement of the region 365131-365652 bp, which contains the *lacZYA* regulatory region, with the *lac28::egfp* fusion. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D56613 and D56614). Approximately to scale.

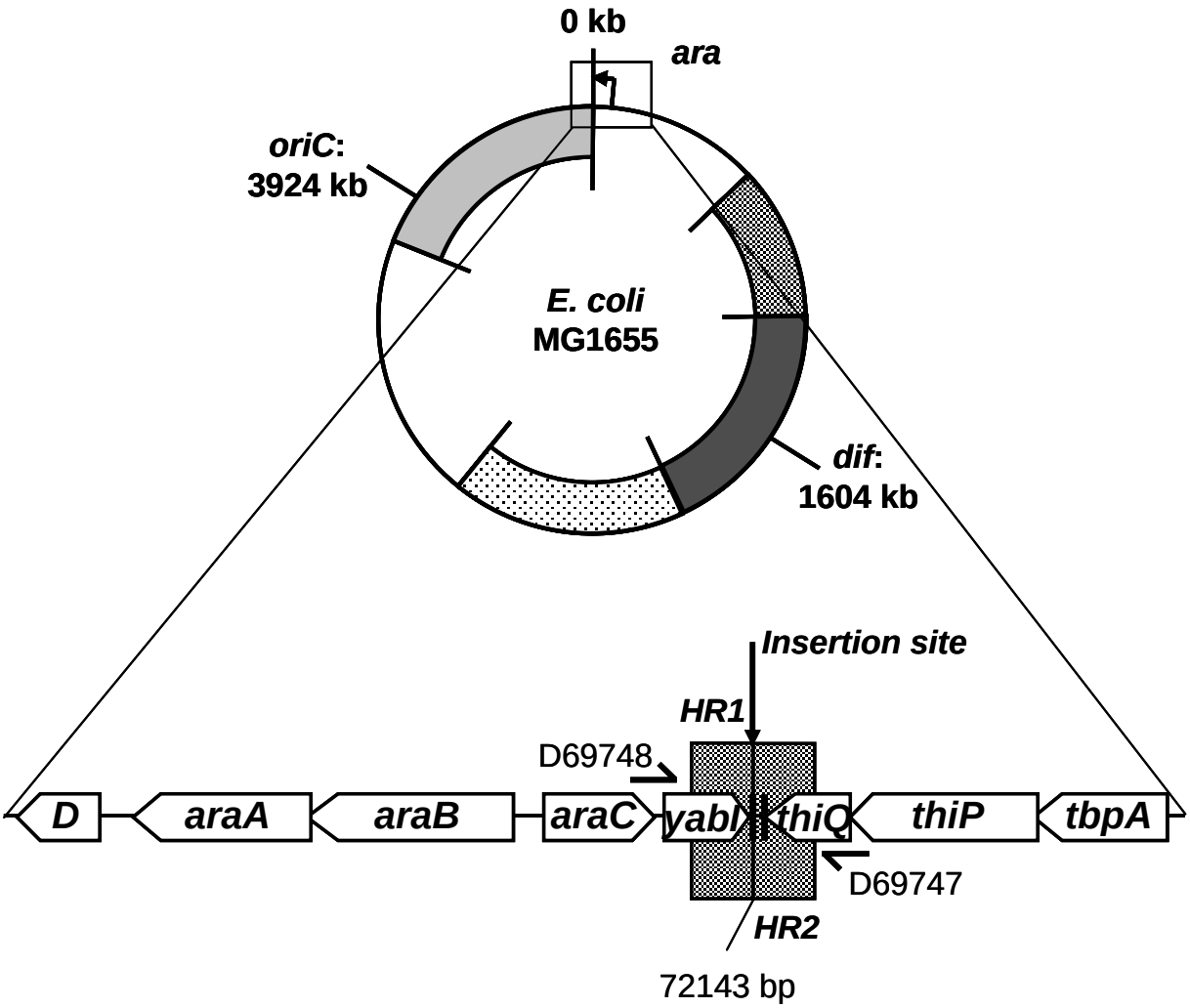

Figure 2.12: Schematic diagram of the *ara* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *ara* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *ara* locus insertion site, with genes represented by white block arrows. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strains BRY13 and BRY21, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (72143 bp) is labelled "insertion site" and is given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D69748 and D69747). Approximately to scale.

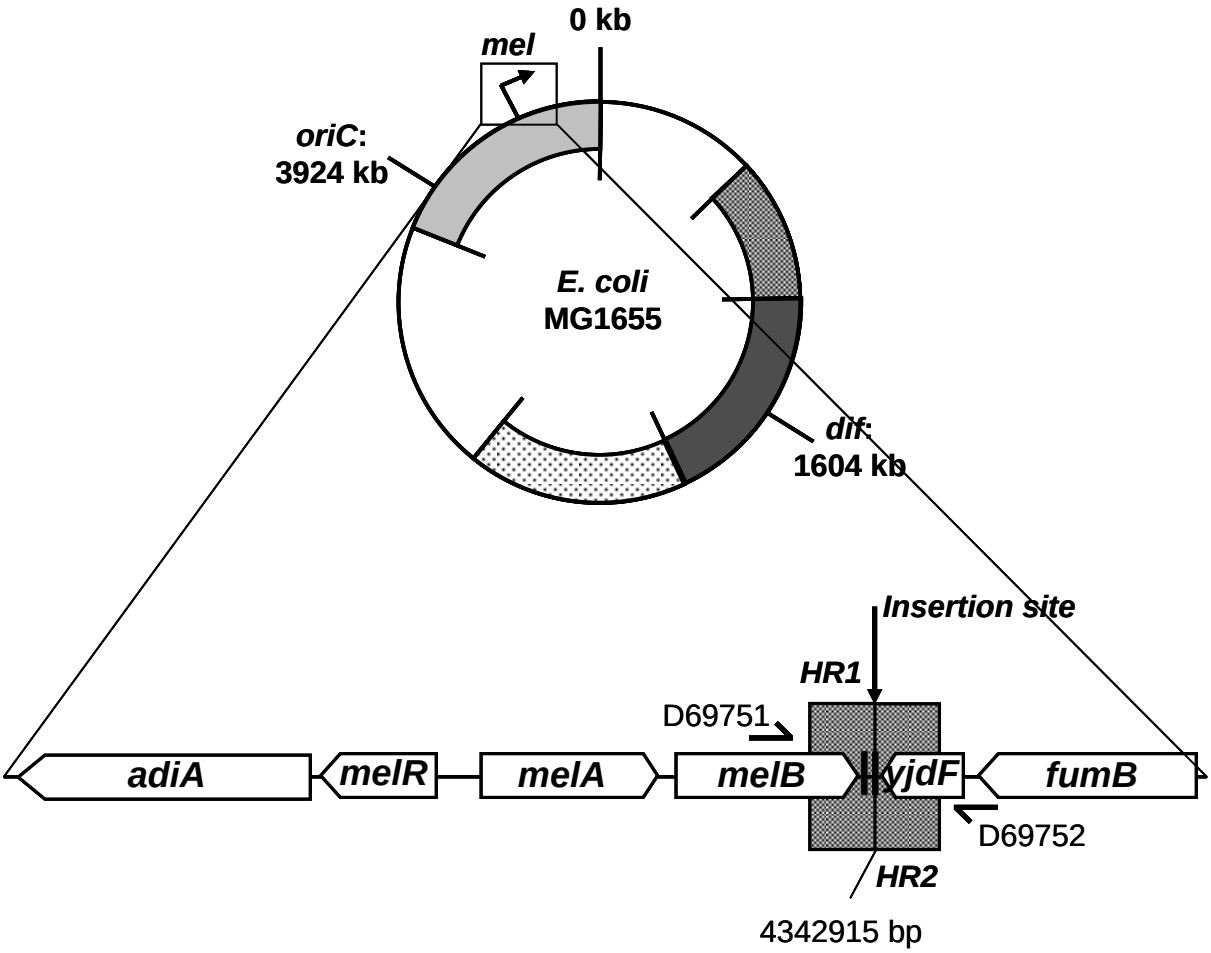

Figure 2.13: Schematic diagram of the *mel* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *mel* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *mel* locus insertion site, with genes represented by white block arrows. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strains BRY15 and BRY37, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (4342915 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D69751 and D69752). Approximately to scale.

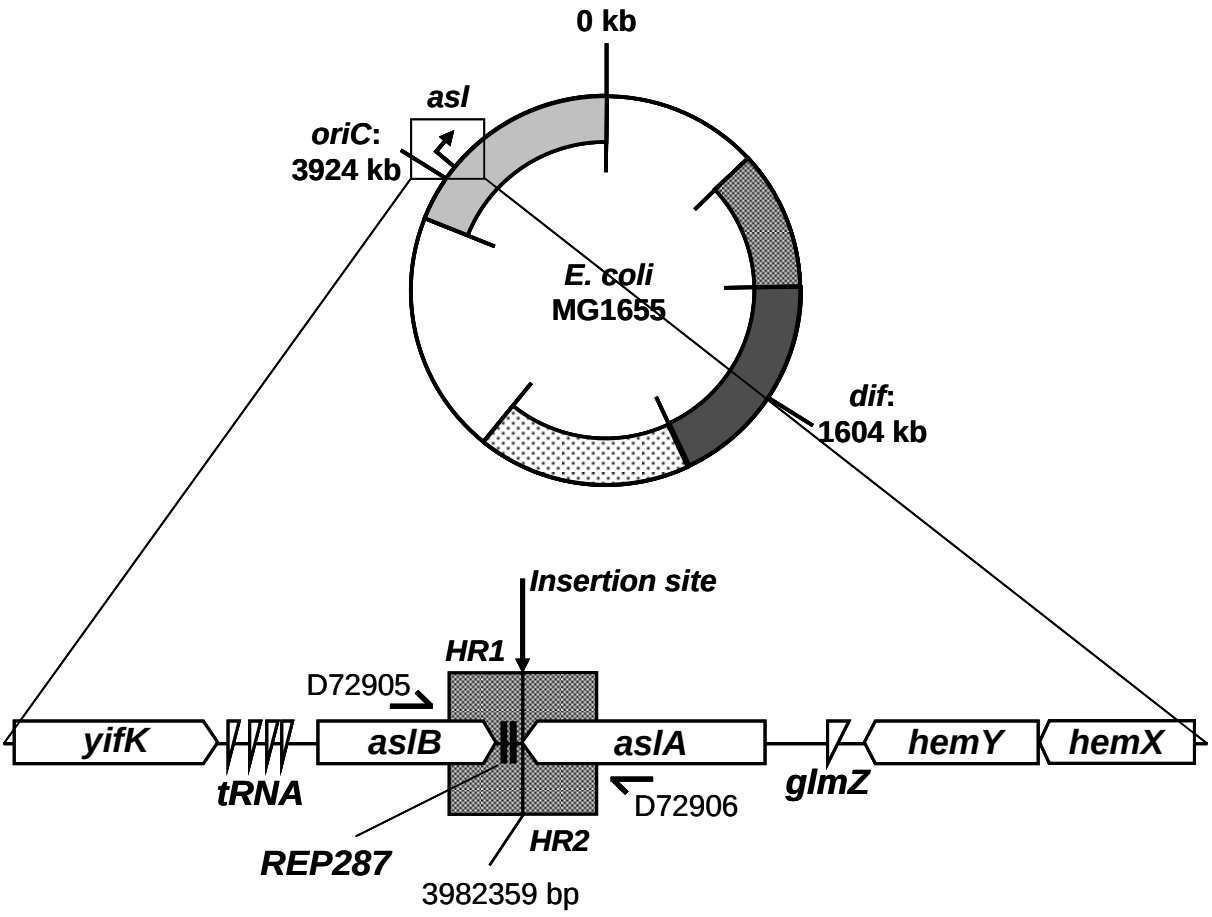

Figure 2.14: Schematic diagram of the *asl* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *asl* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *asl* locus insertion site, with genes represented by white block arrows. White triangles represent RNA coding genes. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY35, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (3982359 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D72905 and D72906). Approximately to scale.

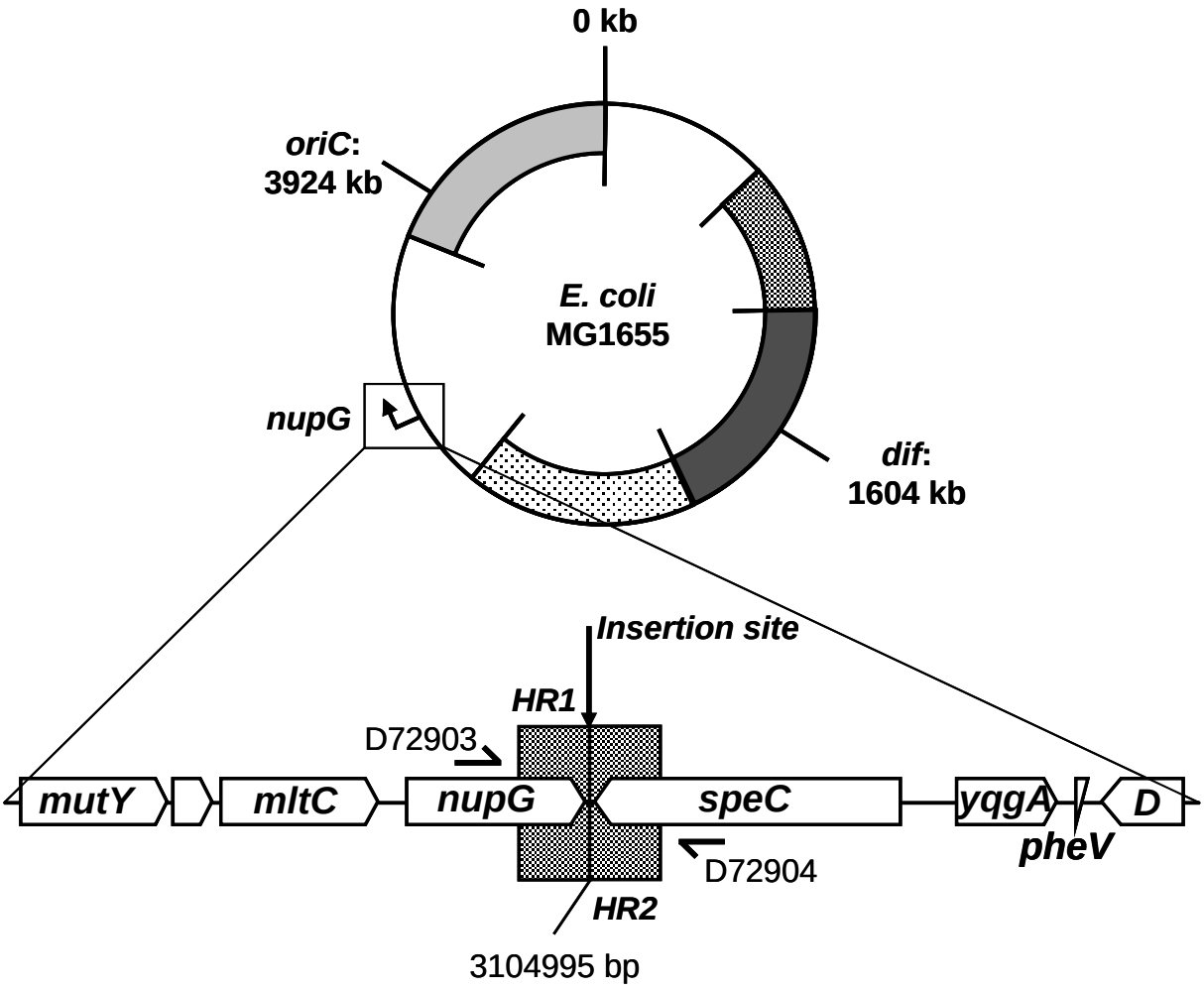

Figure 2.15: Schematic diagram of the *nupG* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *nupG* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *nupG* locus insertion site, with genes represented by white block arrows. White triangles represent RNA coding genes. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY34, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (3104995 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D72903 and D72904). Approximately to scale.

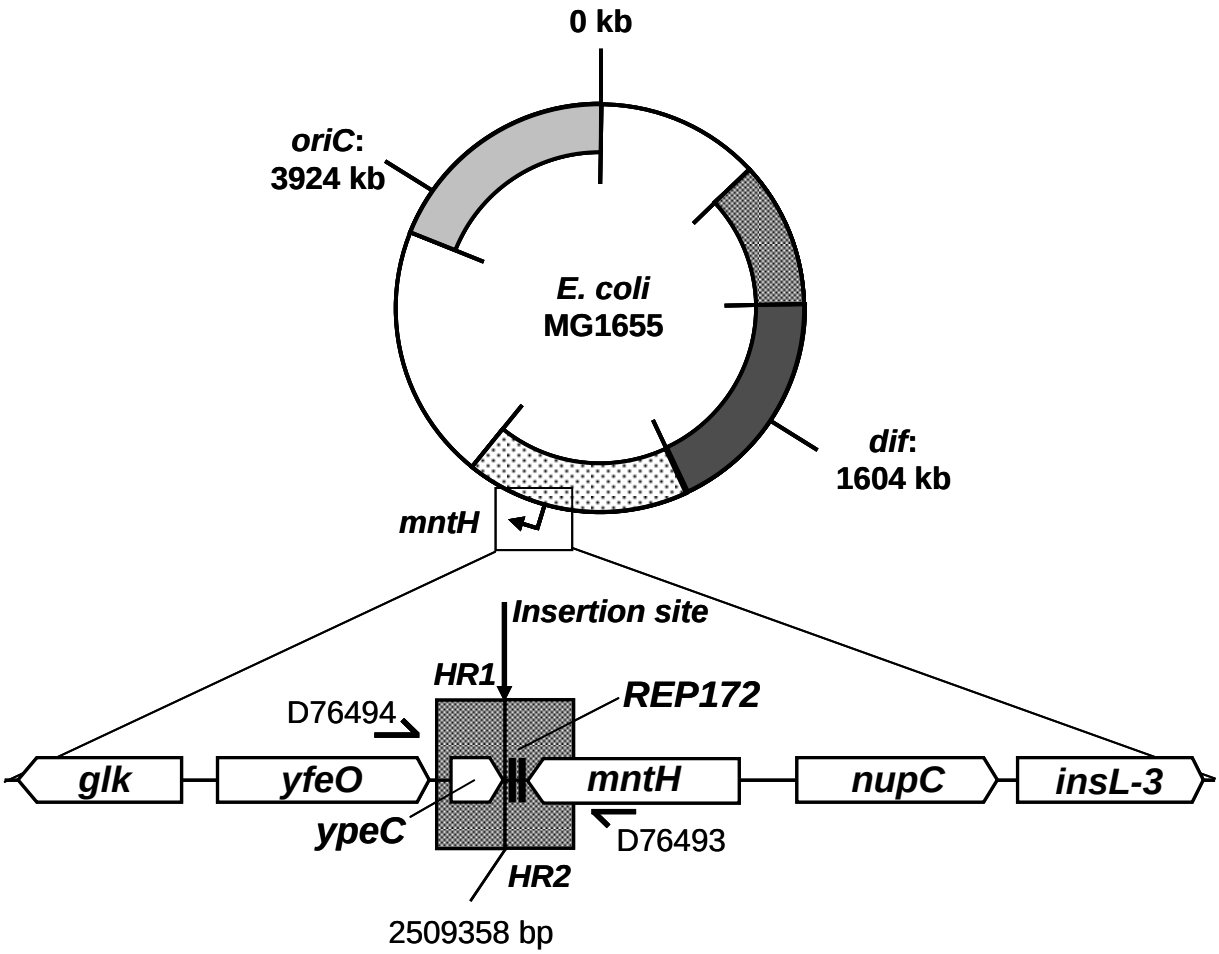

Figure 2.16: Schematic diagram of the *mntH* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *mntH* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *mntH* locus insertion site, with genes represented by white block arrows. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY73, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (2509358 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D76494 and D76493). Approximately to scale.

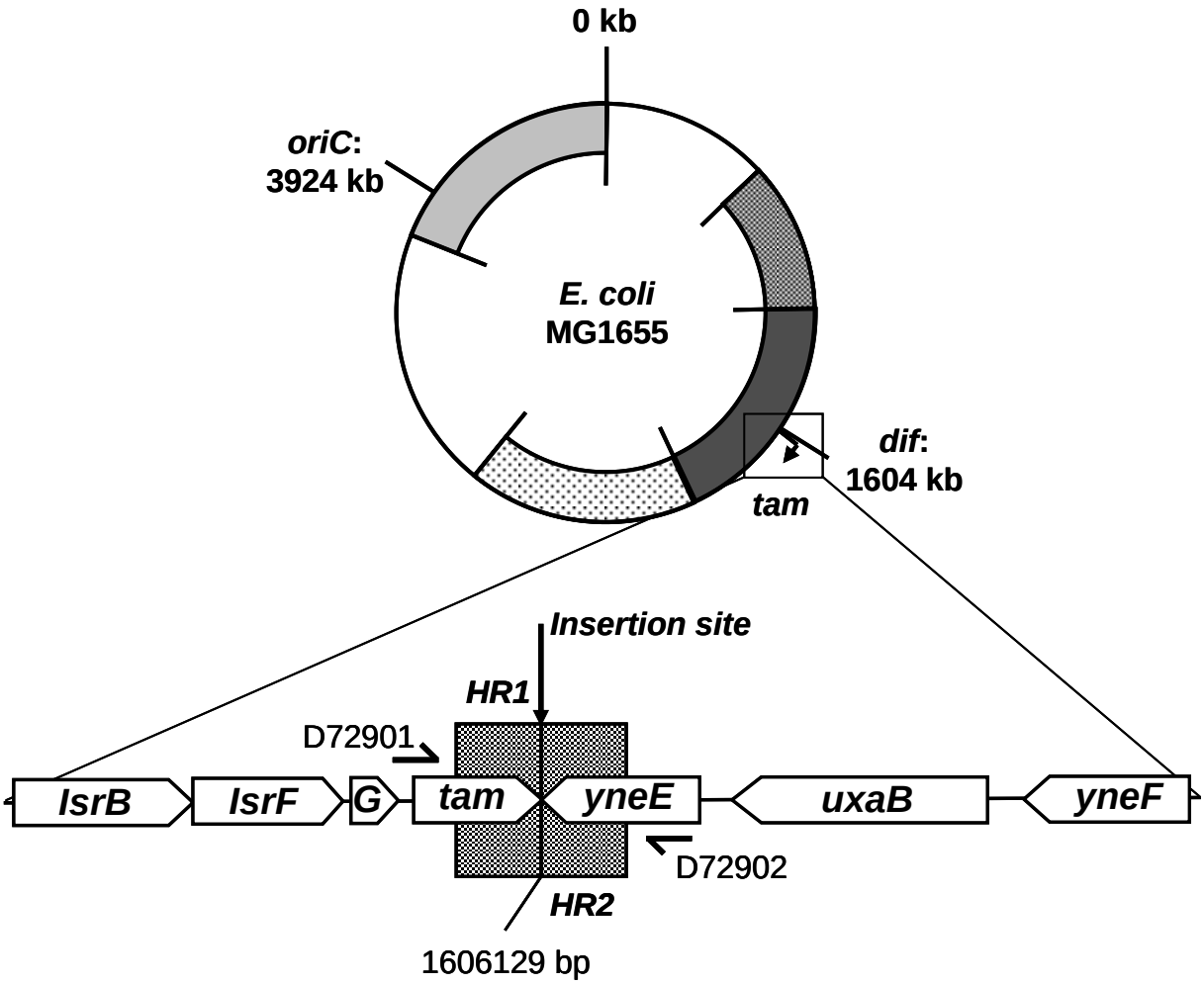

Figure 2.17: Schematic diagram of the *tam* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *tam* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *tam* locus insertion site, with genes represented by white block arrows. The position of homology regions, used for insertion of the *lac28::egfp* fusionin strain BRY33, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (1606129 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D69751 and D69752). Approximately to scale.

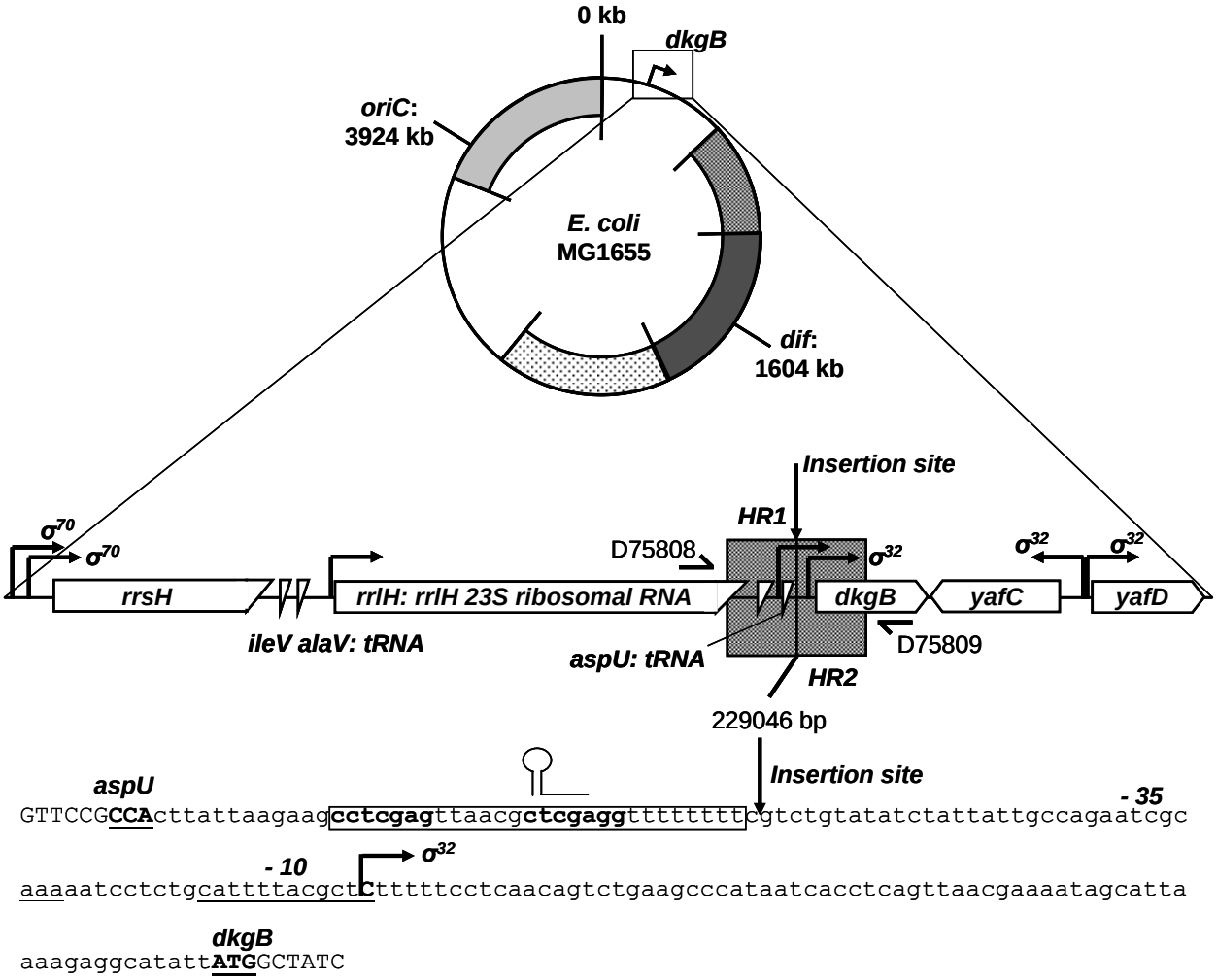

Figure 2.18: Schematic diagram and DNA sequence of the *dkgB* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *dkgB* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *dkgB* locus insertion site, with genes represented by white block arrows. Genes encoding RNA are indicated by the ▭ symbol. The position of promoters is indicated by the black arrows with the promoter type also indicated (σ^{70} or σ^{32}). The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY58, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (229046 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D75808 and D75809). Approximately to scale. Also shown is the DNA base sequence of the *dkgB* insertion site, with RNA or protein coding regions in uppercase. The promoter -10 and -35 regions are underlined and the transcription start site in uppercase with an arrow. The *aspU* Rho-independent transcription terminator is boxed and labelled with a stem-loop structure.

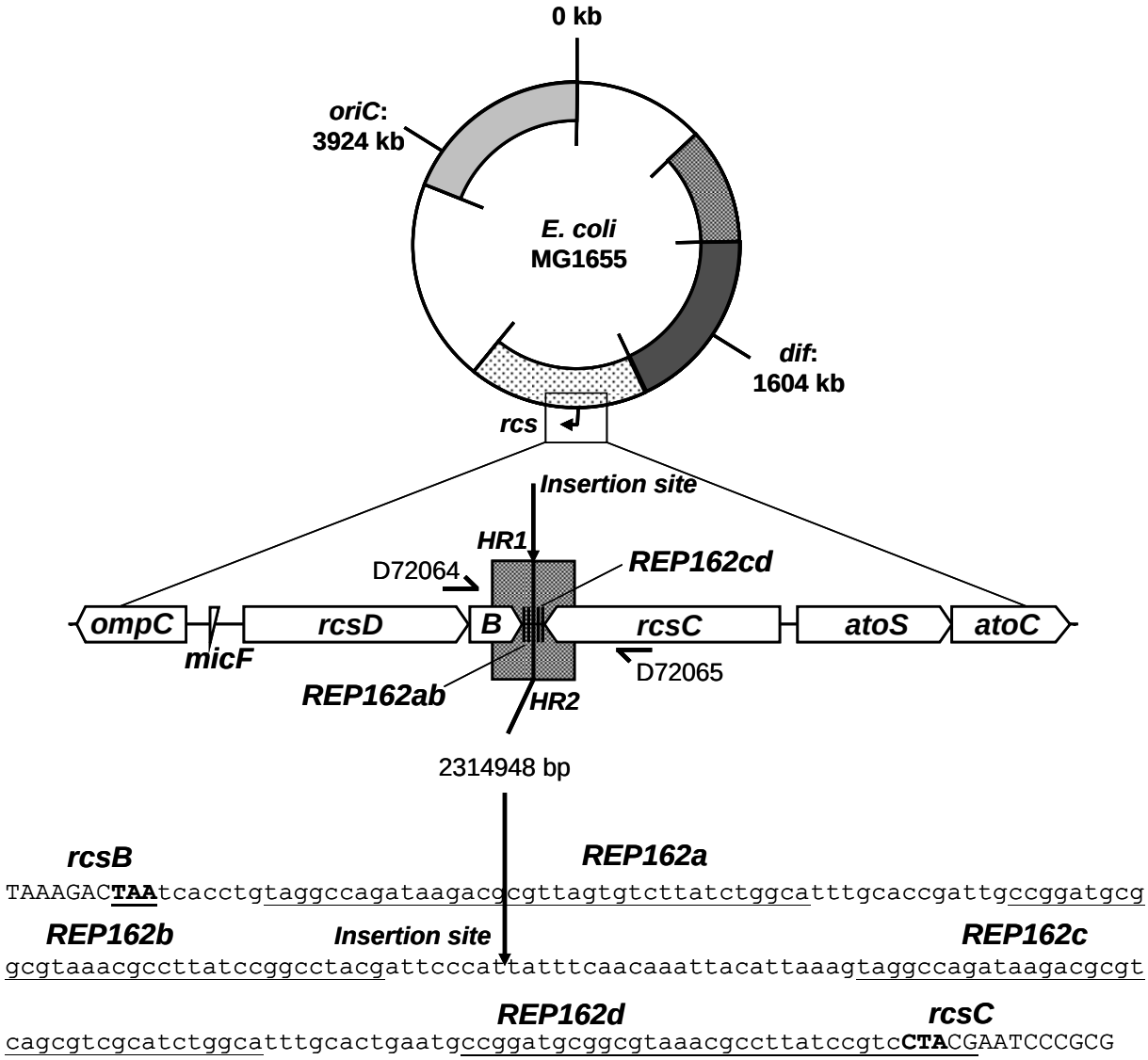

Figure 2.19: Schematic diagram and DNA sequence of the *rcs* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *rcs* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *rcs* locus insertion site, with genes represented by white block arrows. White triangles represent RNA coding genes. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY27, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (2314948 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The position of REP elements at the target insertion site is represented by vertical black lines. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D75808 and D75809). Approximately to scale. Also shown is the DNA base sequence of the *rcs* insertion site, with protein coding regions in uppercase. The promoter -10 and -35 regions are underlined and the transcription start site in uppercase with an arrow. The REP element sequences are underlined and labelled.

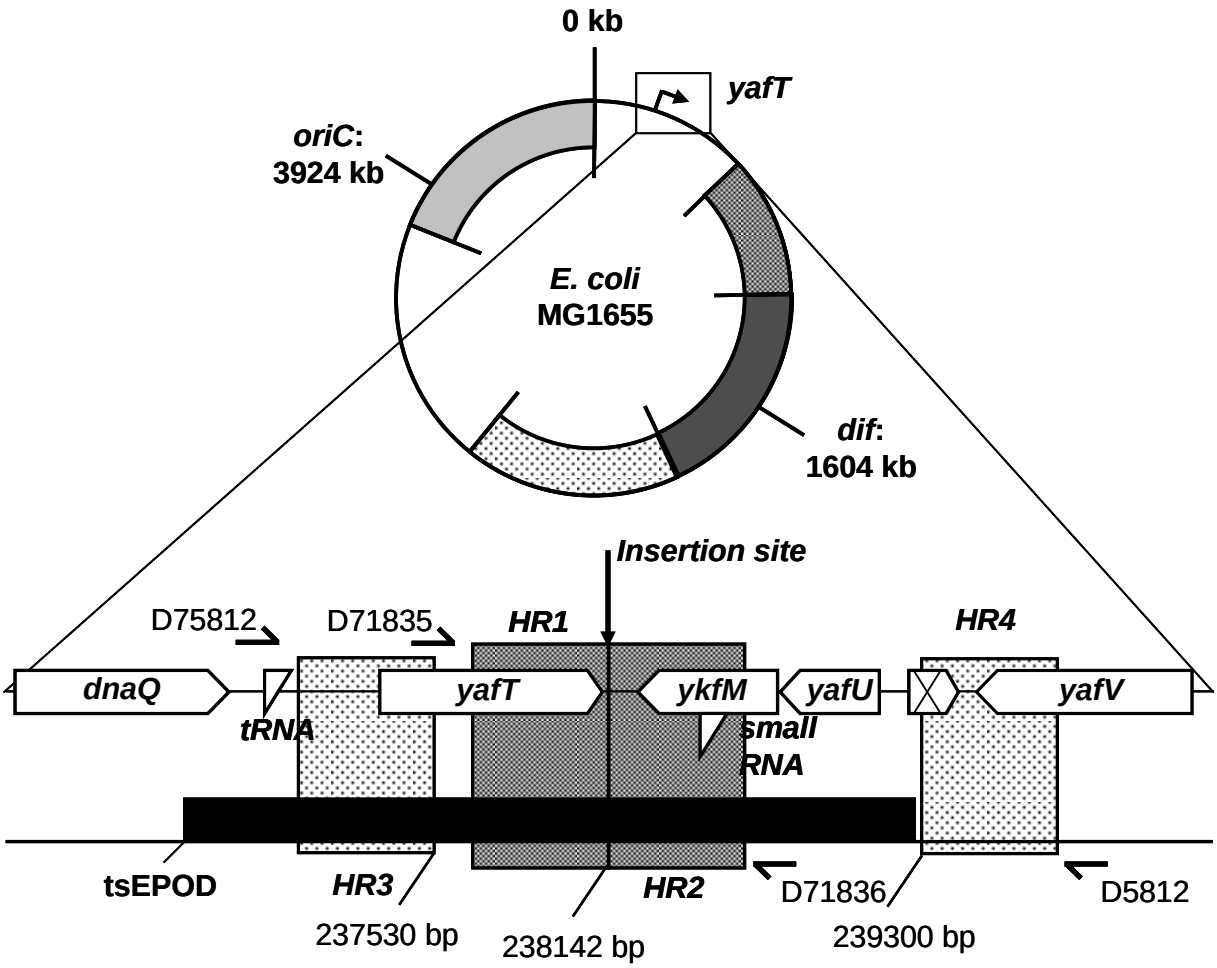

Figure 2.20: Schematic diagram of the *yafT* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *yafT* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *yafT* locus insertion site, with genes represented by white block arrows. White triangles represent RNA coding genes. Crossed boxes represent pseudogenes. The position of transcriptionally silent extended protein occupancy domains (tsEPOD) is indicated by the black box. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY22, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*, with the exact positions in bp with respect to the co-ordinate system origin. Use of *HR3* and *HR4* for gene doctoring will cause replacement of the region between the HRs (237530-239300 bp) with the *lac28::egfp* fusion, creating the ΔtsEPOD strain BRY59. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (DNNNNN). Approximately to scale.

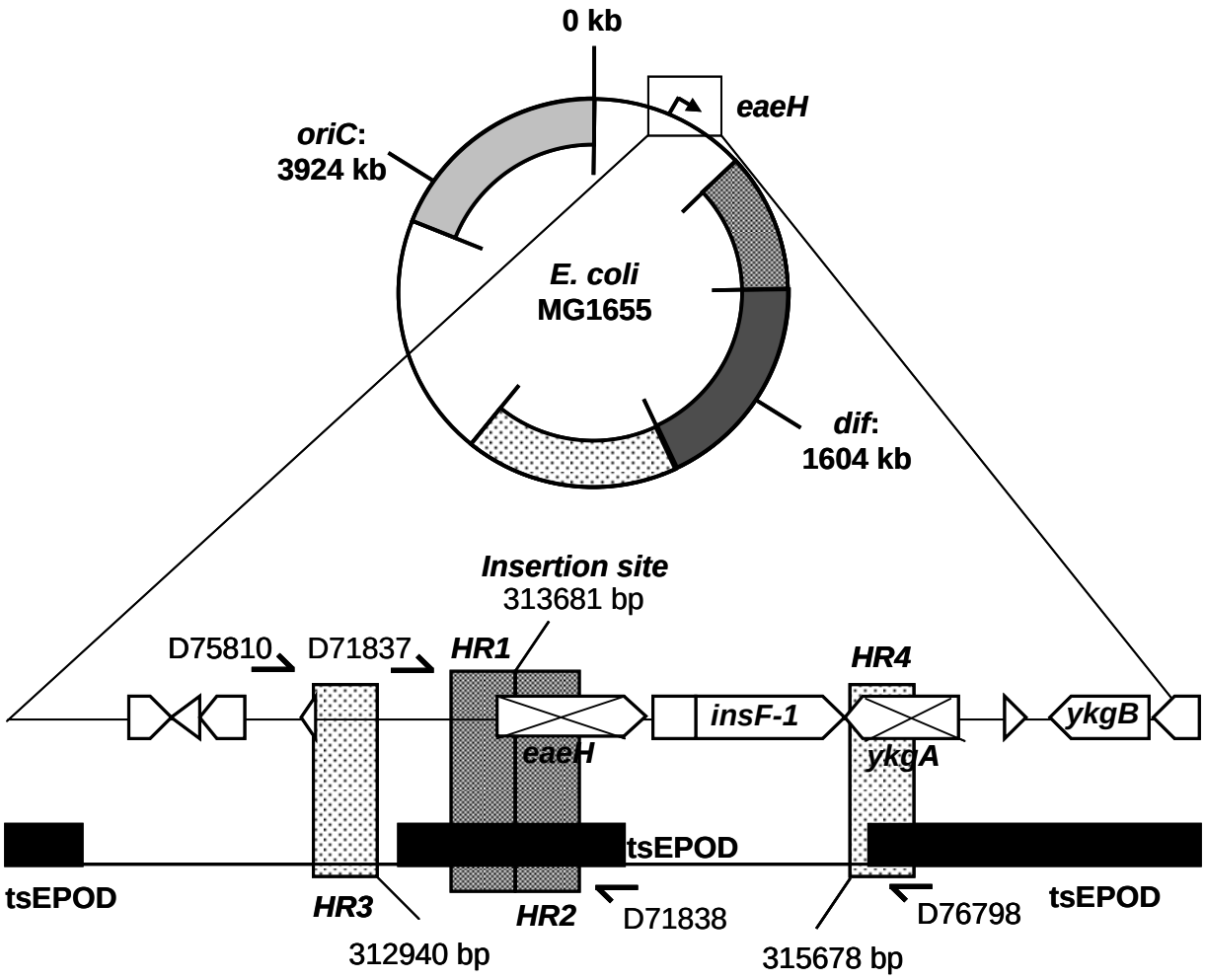

Figure 2.21: Schematic diagram of the *eaeH* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *eaeH* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *eaeH* locus insertion site, with genes represented by white block arrows. Crossed arrows represent pseudogenes. The position of transcriptionally silent extended protein occupancy domains (tsEPOD) is indicated by the black box. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY23, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*, with the exact positions in bp with respect to the co-ordinate system origin. Use of *HR3* and *HR4* for gene doctoring will cause replacement of the region between the HRs (312940-315678 bp) with the *lac28::egfp* fusion, creating the ΔtsEPOD strain BRY60. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (DNNNNN). Approximately to scale.

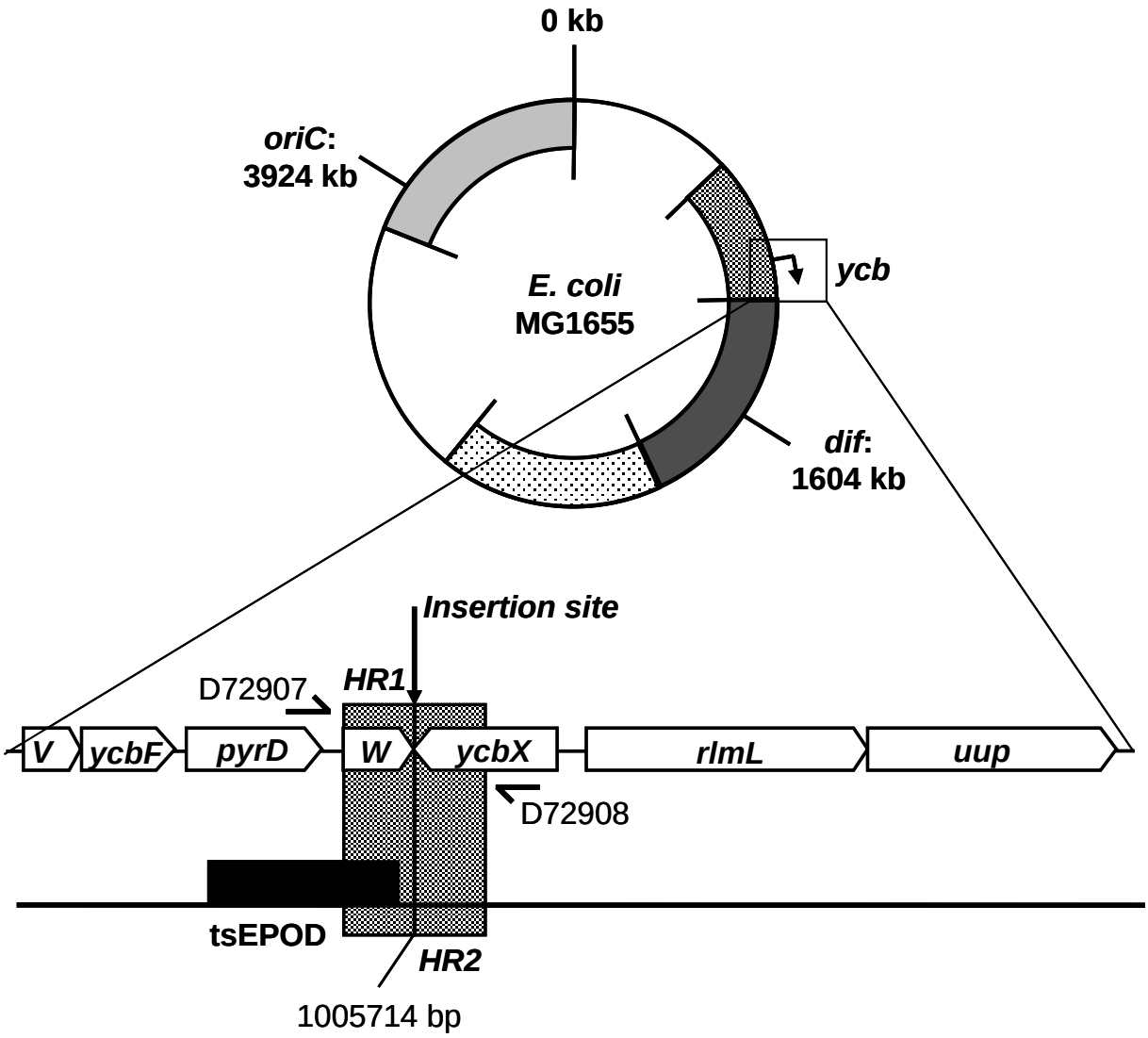

Figure 2.22: Schematic diagram of the *ycb* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *ycb* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *ycb* locus insertion site, with genes represented by white block arrows. The position of transcriptionally silent extended protein occupancy domains (tsEPOD) is indicated by the black box. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY32, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*. The exact position of *lac28::egfp* insertion (1005714 bp) is labelled "insertion site" and given in bp, with respect to the co-ordinate system origin. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (D72907 and D72908). Approximately to scale.

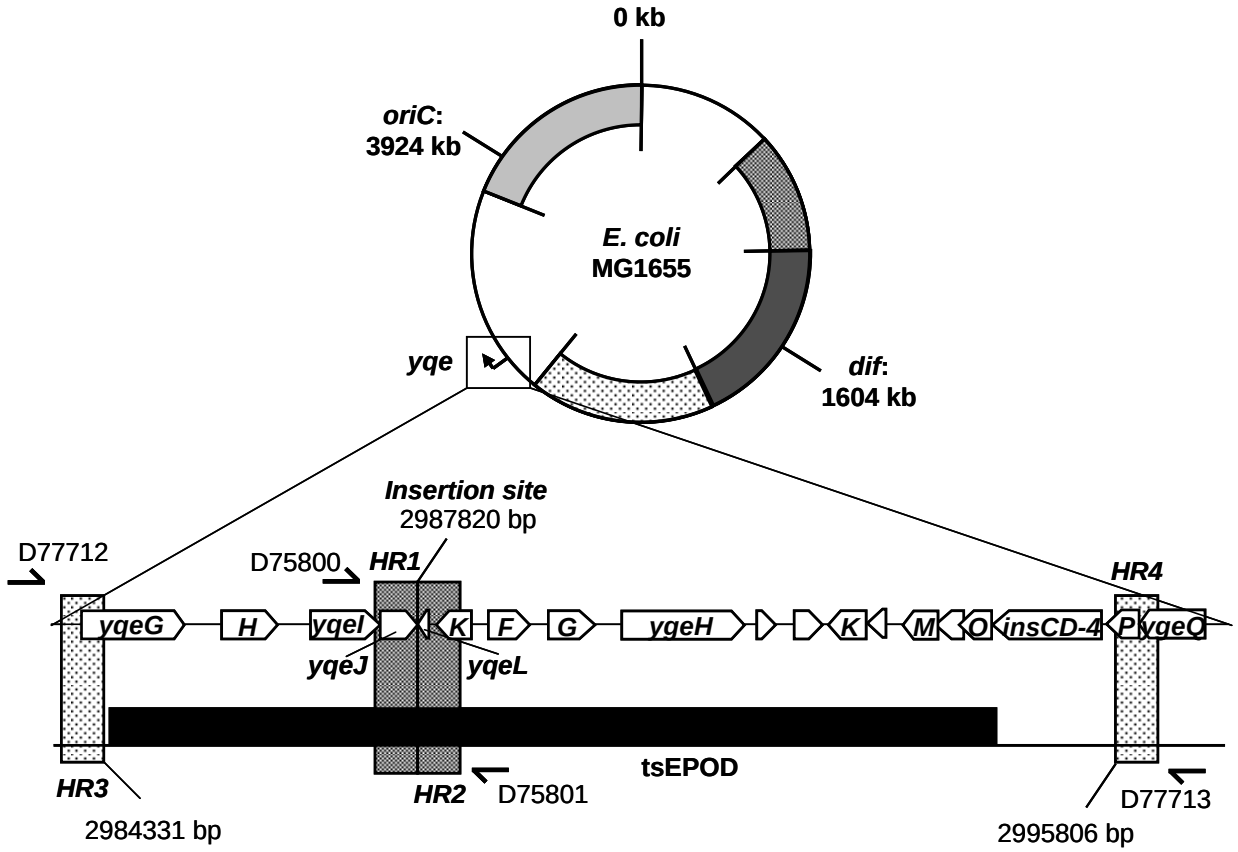

Figure 2.23: Schematic diagram of the *yqe* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *yqe* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *yqe* locus insertion site, with genes represented by white block arrows. The position of transcriptionally silent extended protein occupancy domains (tsEPOD) is indicated by the black box. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY54, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*, with the exact positions in bp with respect to the co-ordinate system origin. Use of *HR3* and *HR4* for gene doctoring will cause replacement of the region between the HRs (2984331-2995806 bp) with the *lac28::egfp* fusion, creating the ΔtsEPOD strain BRY68. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (DNNNNN). Approximately to scale.

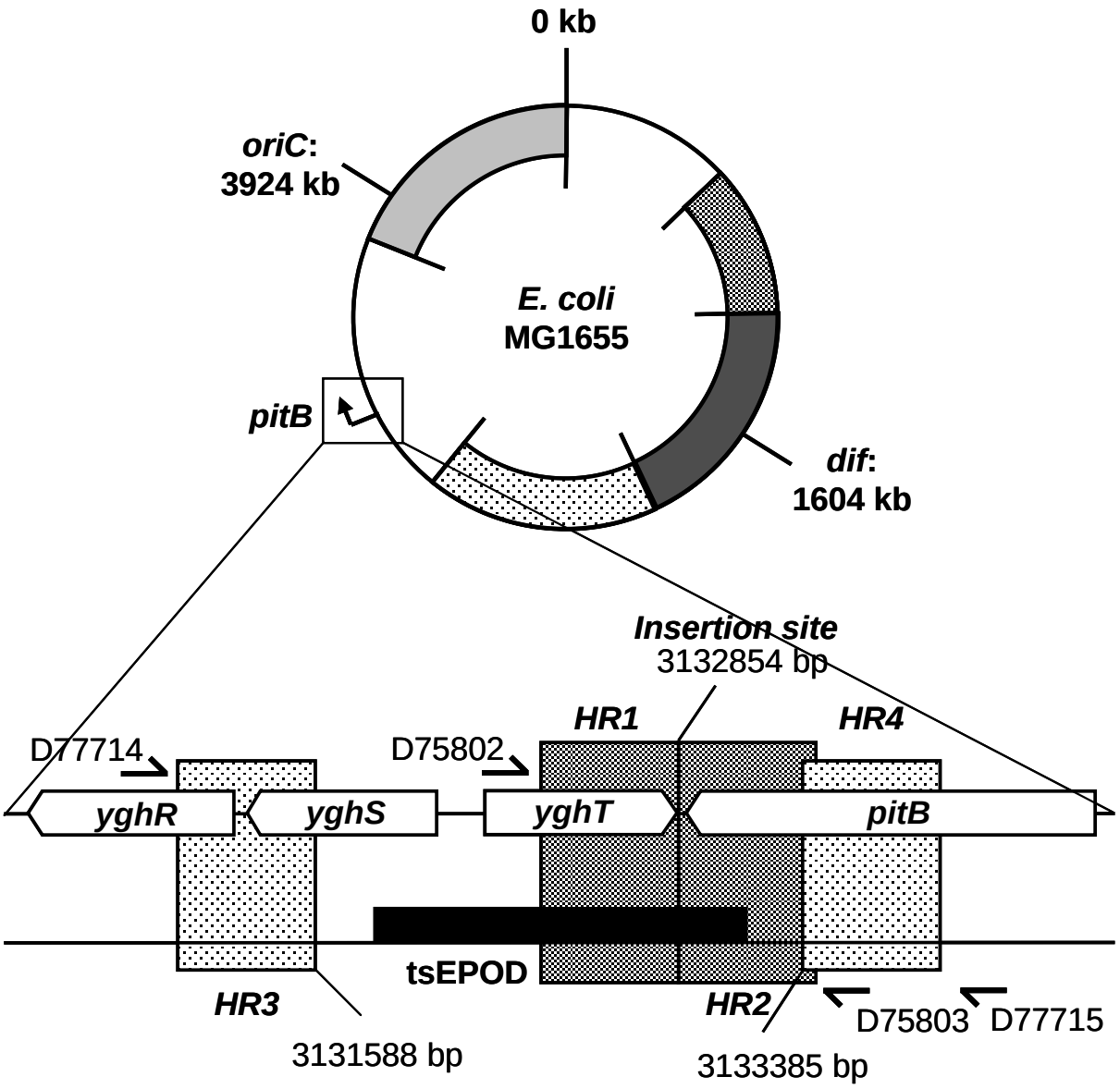

Figure 2.24: Schematic diagram of the *pitB* locus *lac28::egfp* fusion insertion site

The black circle represents the *E. coli* MG1655 chromosome and the grey shaded blocks and two white gaps represent the six macrodomains (see figure 2.10). The position of the *pitB* locus *lac28::egfp* fusion insertion site is represented by the arrow and label. Also shown is the organisation of genes at the *pitB* locus insertion site, with genes represented by white block arrows. The position of transcriptionally silent extended protein occupancy domains (tsEPOD) is indicated by the black box. The position of homology regions, used for insertion of the *lac28::egfp* fusion in strain BRY55, by gene doctoring, is represented by the grey shaded boxes, labelled *HR1* and *HR2*, with the exact positions in bp with respect to the co-ordinate system origin. Use of *HR3* and *HR4* for gene doctoring will cause replacement of the region between the HRs (3131588-3133385 bp) with the *lac28::egfp* fusion, creating the ΔtsEPOD strain BRY69. The approximate position of oligonucleotide primers used to PCR check/sequence the chromosomal insert is represented by black arrows labelled with the primer code (DNNNNN). Approximately to scale.

the master donor pJB plasmid to allow the transfer of the *lac28::egfp* fusion to the chromosome of *E. coli* K-12 MG1655 by the gene doctoring method. The upstream HR of each site, referred to as HR1, was amplified from *E. coli* K-12 MG1655 genomic DNA by PCR, using primers that incorporate an upstream *Mfe*I and downstream *Xma*I target site (Table 2.6). The HR1 PCR product was purified using the QIAquick PCR Purification Kit, after which it was digested with the restriction endonucleases *Mfe*I and *Xma*I and further purified by use of the QIAquick PCR Purification Kit. The pJB22 gene doctoring plasmid was used as stock for all target sites, as excision of the ~ 500 bp HR before cloning of the new HR was easily detectable by agarose gel electrophoresis. pJB22 was digested with *Mfe*I and *Xma*I, treated with CIP to prevent re-ligation and purified by excision from a 0.8% agarose gel using the QIAquick Gel Extraction Kit. *Mfe*I-*Xma*I digested pJB22 was then ligated with the *Mfe*I-*Xma*I HR1 fragment before being used to transform calcium chloride competent *E. coli* RLG221. The presence of the correct insert was checked by DNA sequencing with the primer D58794 (Table 2.4).

The resulting plasmid was assigned a pJB number and stored at -20˚C, after which a separate stock of the plasmid was digested with the restriction endonucleases *Nhe*I and *Sac*I, treated with CIP and purified by extraction from a 0.8% agarose gel. The downstream HR, HR2, was amplified from *E. coli* K-12 MG1655 genomic DNA by PCR using primers which incorporate a target site for *Nhe*I upstream and *Sac*I downstream of the fragment (Table 2.6). The resulting HR2 PCR fragment was purified and digested with *Nhe*I and *Sac*I before further purification using the QIAquick PCR Purification Kit. *Nhe*I-*Sac*I digested plasmid containing HR1 was then ligated with the *Nhe*I-*Sac*I HR2 PCR product before being used to transform calcium chloride competent *E. coli* RLG221 to produce the finished donor plasmid. The presence of the correct HR2 in the finished gene doctoring donro plasmid was confirmed by DNA sequencing using primer D58793 (Table 2.4).

Table 2.6: Oligonucleotides used to amplify homology regions

Name	Sequence (5' – 3') [1]	Use [2]
D69231	GCCGCAATTGCCGGGATT GAAACTGAACG	Used to amplify the *thiQ* homology region and introduce an *Mfe*I site. Used with D69232. (*ara*)
D69232	GCCGCCCGGGCGACGCTT GCCGCGTCTTATC	Used to amplify the *thiQ* homology region and introduce an *Xma*I site. Used with D69231. (*ara*)
D69233	GCCGGAGCTCCTGAACAT GCGTTGCATCAAC	Used to amplify the *yabI* homology region and introduce a *Sac*I site. Used with D69234. (*ara*)
D69234	GCCGGCTAGCCATCAGGC AACCCGCAC	Used to amplify the *yabI* homology region and introduce an *Nhe*I site. Used with D69233. (*ara*)
D71381	GCCGGAGCTCCCGGGATT GAAACTGAACG	Used to amplify the *thiQ* homology region and introduce a *Sac*I site. Used with D71382. (*Inverse ara*)
D71382	GCCGGCTAGCCGACGCTT GCCGCGTCTTATC	Used to amplify the *thiQ* homology region and introduce an *Nhe*I site. Used with D71381. (*Inverse ara*)
D71383	GCCGCCCGGGCATCAGGC AACCCGCAC	Used to amplify the *yabI* homology region and introduce an *Xma*I site. Used with D71384. (*Inverse ara*)
D71384	GCCGCAATTGCTGAACAT GCGTTGCATCAAC	Used to amplify the *yabI* homology region and introduce an *Mfe*I site. Used with D71383. (*Inverse ara*)
D69239	GCCGCAATTGGATCCAGA TCCATCTGCTGG	Used to amplify the *melB* homology region and introduce an *Mfe*I site. Used with D69240. (*mel*)
D69240	GCCGCCCGGGTTCCTTATC CGGTCTACAAAATTTG	Used to amplify the *melB* homology region and introduce an *Xma*I site. Used with D69239. (*mel*)
D69241	GCCGGAGCTCCGTTAACG CCGCTGCTCTATACGC	Used to amplify the *yjdF* homology region and introduce a *Sac*I site. Used with D69242. (*mel*)
D69242	GCCGGCTAGCTTCACGCC GCATCCGGC	Used to amplify the *yjdF* homology region and introduce an *Nhe*I site. Used with D69241. (*mel*)
D72989	GCCGGAGCTCGATCCAGA TCCATCTGCTGG	Used to amplify the *melB* homology region and introduce a *Sac*I site. Used with D72876. (*Inverse mel*)
D72876	GCCGGCTAGCTTCCTTATC CGGTCTACAAAATTTG	Used to amplify the *melB* homology region and introduce an *Nhe*I site. Used with D72989. (*Inverse mel*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D72988	GCCGCAATTGCGTTAACG CCGCTGCTCTATACGC	Used to amplify the *yjdF* homology region and introduce an *Mfe*I site. Used with D72878. (*Inverse mel*)
D72878	GCCGCCCGGGTTCACGCC GCATCCGGC	Used to amplify the *yjdF* homology region and introduce an *Xma*I site. Used with D72988. (*Inverse mel*)
D71369	GCCGCAATTGATATTATAC TGTTTCCAC	Used to amplify the *yafT* homology region and introduce an *Mfe*I site. Used with D71370. (*yafT*)
D71370	GCCGCCCGGGTCCTGGCG GCAGATATAATTTATTATT G	Used to amplify the *yafT* homology region and introduce an *Xma*I site. Used with D71369. (*yafT*)
D71371	GCCGGCTAGCATTTCTGGC GGCAATAATAC	Used to amplify the *ykfM* homology region and introduce an *Nhe*I site. Used with D71372. (*yafT*)
D71372	GCCGGAGCTCGAGTTAAA CCTCTATATTC	Used to amplify the *ykfM* homology region and introduce a *Sac*I site. Used with D71371. (*yafT*)
D71373	GCCGCAATTGATCCATAA AAAATATATTGG	Used to amplify the upstream *eaeH* homology region and introduce an *Mfe*I site. Used with D71374. (*eaeH*)
D71374	GCCGCCCGGGGAAAAAGA ACCTGAACAGAG	Used to amplify the upstream *eaeH* homology region and introduce an *Xma*I site. Used with D71373. (*eaeH*)
D71375	GCCGGCTAGCCACTCGCT GTCACCTTTACC	Used to amplify the downstream *eaeH* homology region and introduce an *Nhe*I site. Used with D71376. (*eaeH*)
D71376	GCCGGAGCTCGCACCAAC ACCAATGCG	Used to amplify the downstream *eaeH* homology region and introduce a *Sac*I site. Used with D71375. (*eaeH*)
D71881	GGACCAATTGATTGTTCTG ACTATGAACAACAACC	Used to amplify the *rcsB* homology region and introduce an *Mfe*I site. Used with D71882. (*rcs*)
D71882	GGACCCCGGGATGGGAAT CGTAGGCCG	Used to amplify the *rcsB* homology region and introduce an *Xma*I site. Used with D71881. (*rcs*)
D71883	GGACGCTAGCTATTTCAA CAAATTACATTAAAGTAG G	Used to amplify the rcsC homology region and introduce an *Nhe*I site. Used with D71884. (*rcs*)
D71884	GGACGAGCTCGCGGTCAG CGATAATGAC	Used to amplify the *rcsC* homology region and introduce a *Sac*I site. Used with D71883. (*rcs*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D72684	CGGCCAATTGAAGAGCAC GATCGAATGATG	Used to amplify the *ycbW* homology region and introduce an *Mfe*I site. Used with D72685. (*ycb*)
D72685	CGGCCCCGGGTTAGACTG CCTGTTCGAGG	Used to amplify the *ycbW* homology region and introduce an *Xma*I site. Used with D72684. (*ycb*)
D72686	CGGCGCTAGCCTAACGCG CCAACTTAAG	Used to amplify the *ycbX* homology region and introduce an *Nhe*I site. Used with D72687. (*ycb*)
D72687	CGGCGAGCTCCTAGCCGC TGTATTTTCACC	Used to amplify the *ycbX* homology region and introduce a *Sac*I site. Used with D72686. (*ycb*)
D72688	CGGCCAATTGCTGGCAAC CGGTACAGG	Used to amplify the *tam* homology region and introduce an *Mfe*I site. Used with D72689. (*tam*)
D72689	CGGCCCCGGGTTTACTCCA TACGCCGGG	Used to amplify the *tam* homology region and introduce an *Xma*I site. Used with D72688. (*tam*)
D72690	CGGCGCTAGCTATCACGT CAGCTGGTAATG	Used to amplify the *yneE* homology region and introduce an *Nhe*I site. Used with D72691. (*tam*)
D72691	CGGCGAGCTCTATCTCAA GACTGAAGATCTTCAG	Used to amplify the *yneE* homology region and introduce a *Sac*I site. Used with D72690. (*tam*)
D72692	CGGCCAATTGCAATCATC ATGTCGATTTC	Used to amplify the *nupG* homology region and introduce an *Mfe*I site. Used with D72693. (*nupG*)
D72693	CGGCCCCGGGTAATTAGT GGCTAACCGTC	Used to amplify the *nupG* homology region and introduce an *Xma*I site. Used with D72692. (*nupG*)
D72694	CGGCGCTAGCCGCAAAGA AAAACGGGTC	Used to amplify the *speC* homology region and introduce an *Nhe*I site. Used with D72695. (*nupG*)
D72695	CGGCGAGCTCGTGTTGCC GAGCGTTTATAAC	Used to amplify the *speC* homology region and introduce a *Sac*I site. Used with D72694. (*nupG*)
D72695	CGGCGAGCTCGTGTTGCC GAGCGTTTATAAC	Used to amplify the *speC* homology region and introduce a *Sac*I site. Used with D72694. (*nupG*)
D72696	CAAACAATTGCAGAAATG ATCG	Used to amplify the *aslB* homology region and introduce an *Mfe*I site. Used with D72697. (*asl*)
D72697	CGGCCCCGGGGAGATCTG CCTTTGCCG	Used to amplify the *aslB* homology region and introduce an *Xma*I site. Used with D72696. (*asl*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D72698	CGGCGCTAGCAGCGATAG CGCCGGCTTAG	Used to amplify the *aslA* homology region and introduce an *Nhe*I site. Used with D72699. (*asl*)
D72699	CGGCGAGCTCGGAAAGGG ATGATCCAACCG	Used to amplify the *aslA* homology region and introduce a *Sac*I site. Used with D72698. (*asl*)
D72369	GCCGGCTAGCCTGGCTAC AGGAAGGCC	Used to amplify the *lacZ* homology region and introduce an *Nhe*I site. Used with D72370. (*lac*)
D72370	GCCGGAGCTCGTTTCGGG TTTTCGACG	Used to amplify the *lacZ* homology region and introduce a *Sac*I site. Used with D72369. (*lac*)
D68504	TACGCAATTGCCATGAAG ACGGTACGCG	Used to amplify the *lacI* homology region and introduce an *Mfe*I site. Used with D68505. (*lac*)
D68505	TACGCCCGGGTCACTGCC CGCTTTCCAG	Used to amplify the *lacI* homology region and introduce an *Xma*I site. Used with D68504. (*lac*)
D72371	GGACCAATGGTTTCGGGT TTTCGACG	Used to amplify the *inverse lacZ* homology region and introduce an *Mfe*I site. Used with D72372. (*Inverse lac*)
D72372	GGACCCCGGGCTGGCTAC AGGAAGGCC	Used to amplify the *inverse lacZ* homology region and introduce an *Xma*I site. Used with D72371. (*Inverse lac*)
D71377	TACGGAGCTCCCATGAAG ACGGTACGCG	Used to amplify the *inverse lacI* homology region and introduce a *Sac*I site. Used with D71378. (*Inverse lac*)
D71378	TACGGCTAGCTCACTGCCC GCTTTCCAG	Used to amplify the *inverse lacI* homology region and introduce an *Nhe*I site. Used with D71377. (*Inverse lac*)
D68765	GGACGCTAGCATGACCAT GATTACGGATTC	Used to amplify the *lacZ (+O_2)* homology region and introduce an *Nhe*I site. Used with D68766. (*lac +O_2*)
D68766	GGACGAGCTCCACATATC CTGATCTTCCAG	Used to amplify the *lacZ (+O_2)* homology region and introduce a *Sac*I site. Used with D68765. (*lac +O_2*)
D75746	GCTCAATTGATGAGGCTT ATCTGACGC	Used to amplify the *mntH* homology region and introduce an *Mfe*I site. Used with D75747. (*mntH*)
D75747	GCTCCCGGGGCCAATGGA GCACAATGC	Used to amplify the *mntH* homology region and introduce an *Xma*I site. Used with D75746. (*mntH*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D75748	GCTGCTAGCGGACGCGTTTAATGGCG	Used to amplify the *ypeC* homology region and introduce an *Nhe*I site. Used with D75749. (*mntH*)
D75749	GCTGAGCTCGTGCTGGTGGTAACACG	Used to amplify the *ypeC* homology region and introduce a *Sac*I site. Used with D75748. (*mntH*)
D74762	CGGCCAATTGCCTGACTCCTTGAGAGTCC	Used to amplify the *aspU* homology region and introduce an *Mfe*I site. Used with D74763. (*dkgB*)
D74763	CGGCCCCGGGGAAAAAAAACCTCGAGCG	Used to amplify the *aspU* homology region and introduce an *Xma*I site. Used with D74762. (*dkgB*)
D74764	CGGCGCTAGCGTCTGTATATCTATTATTGCCAG	Used to amplify the *dkgB* homology region and introduce an *Nhe*I site. Used with D74765. (*dkgB*)
D74765	CGGCGAGCTCAAATACCGATCTCACGC	Used to amplify the *dkgB* homology region and introduce a *Sac*I site. Used with D74764. (*dkgB*)
D74746	CGGCCAATTGGTGATTGATTATAAAAAAAAC	Used to amplify the *yqeJ* homology region and introduce an *Mfe*I site. Used with D74747. (*yqe*)
D74747	CGGCCCCGGGATGTATTTTAGTTTACCTTGC	Used to amplify the *yqeJ* homology region and introduce an *Xma*I site. Used with D74746. (*yqe*)
D74748	CGGCGCTAGCAACAAAAAGATGGAACCTCG	Used to amplify the *yqeL* homology region and introduce an *Nhe*I site. Used with D74749. (*yqe*)
D74749	CGGCGAGCTCCACGAAAAAGAAACCAAGG	Used to amplify the *yqeL* homology region and introduce a *Sac*I site. Used with D74748. (*yqe*)
D74750	CGGCCAATTGCAAAAAAATTGCCTGGACCAG	Used to amplify the *yghT* homology region and introduce an *Mfe*I site. Used with D74751. (*pitB*)
D74751	CGGCCCCGGGCTGGCGATCGGGCATTTC	Used to amplify the *yghT* homology region and introduce an *Xma*I site. Used with D74750. (*pitB*)
D74752	CGGCGCTAGCGACCGGGCATTTTCAGG	Used to amplify the *pitB* homology region and introduce an *Nhe*I site. Used with D74753. (*pitB*)
D74753	CGGCGAGCTCGAAACTGCCAGGCGTCAG	Used to amplify the *pitB* homology region and introduce a *Sac*I site. Used with D74752. (*pitB*)
D75191	GGACCAATTGGATTTGCTGTACTTTATTC	Used to amplify the upstream *ΔeaeH* homology region and introduce an *Mfe*I site. Used with D75192. (*ΔeaeH*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D75192	GGACCCCGGGATATCTATTATTCTCCCG	Used to amplify the upstream *ΔeaeH* homology region and introduce an *Xma*I site. Used with D71373. (*ΔeaeH*)
D76797	GCCGGCTAGCGGCCTGTGTCCATATTACG	Used to amplify the downstream homology *ΔeaeH* region and introduce an *Nhe*I site. Used with D71376. (*ΔeaeH*)
D76798	GCCGGAGCTCCACCGCGAAATCTATGC	Used to amplify the downstream *ΔeaeH* homology region and introduce a *Sac*I site. Used with D71375. (*ΔeaeH*)
D75195	GGACCAATTGTGAGTTCAGAGAGCCGC	Used to amplify the upstream *ΔyafT* homology region and introduce an *Mfe*I site. Used with D75196. (*ΔyafT*)
D75196	GGACCCCGGGATCCTCTTCGGATACGG	Used to amplify the upstream *ΔyafT* homology region and introduce an *Xma*I site. Used with D75195. (*ΔyafT*)
D75197	GCCGGCTAGCGTTAAGACGTAAGATGCG	Used to amplify the downstream homology *ΔyafT* region and introduce an *Nhe*I site. Used with D75198. (*ΔyafT*)
D75198	GCCGGAGCTCTGCTACGACTTACGTTTTC	Used to amplify the downstream *ΔyafT* homology region and introduce a *Sac*I site. Used with D75197. (*ΔyafT*)
D76801	GGACCAATTGAAATATGCCTGAGCAGC	Used to amplify the upstream *Δyqe* homology region and introduce an *Mfe*I site. Used with D76802. (*Δyqe*)
D76802	GGACCCCGGGTTCCGTAAGTGAGTTGG	Used to amplify the upstream *Δyqe* homology region and introduce an *Xma*I site. Used with D76801. (*Δyqe*)
D76803	GCCGGCTAGCTTGTCTGGAGATTCAGG	Used to amplify the downstream homology *Δyqe* region and introduce an *Nhe*I site. Used with D76804. (*Δyqe*)
D76804	GCCGGAGCTCCAAGCAATCAACACTGG	Used to amplify the downstream *Δyqe* homology region and introduce a *Sac*I site. Used with D76803. (*Δyqe*)
D76805	GGACCAATTGATCGTTTGATTTTGTCG	Used to amplify the upstream *ΔpitB* homology region and introduce an *Mfe*I site. Used with D76806. (*ΔpitB*)

Table 2.6: Oligonucleotides used to amplify homology regions (continued)

Name	Sequence (5' – 3') [1]	Use [2]
D76806	GGACCCCGGGAAAAATGTTAAGGCAGC	Used to amplify the upstream *ΔpitB* homology region and introduce an *Xma*I site. Used with D76805. (*ΔpitB*)
D76807	GTTTCGCTAGCTTCGCGG	Used to amplify the downstream homology *ΔpitB* region and introduce an *Nhe*I site. Used with D76808. (*ΔpitB*)
D76808	GCCGGAGCTCAATGCTGGTACTGGTGG	Used to amplify the downstream *ΔpitB* homology region and introduce a *Sac*I site. Used with D76807. (*ΔpitB*)

[1] Target sites for restriction endonucleases are underlined.

[2] The chromosomal locus for which the homology region and primer are used to target is italicised and in brackets.

2.9.5 Construction of strains carrying *lac28::egfp* at different chromosomal positions

The pJB donor plasmids, described in section 2.9.4, were used to transfer the *lac28::egfp* fusion to the specifically targeted chromosomal loci, described in section 2.9.3, by the gene doctoring method. *E. coli* K-12 MG1655 cells were co-transformed, by the rubidium chloride method described in section 2.6.2, with the mutagenesis plasmid pACBSR and the relevant pJB donor plasmid. Transformants were selected on nutrient agar supplemented with chloramphenicol and ampicillin, and used to transfer the *lac28::egfp* fusion to the desired location on the chromosome of *E. coli* K-12 MG1655 by the gene doctoring method, described in section 2.8. Candidates were screened for the presence of the insert by colony PCR, as described in section 2.7.2, using oligonucleotide primers designed to bind to regions of the chromosome flanking the homology regions used (Table 2.7). Successful candidates were assigned a BRY strain number and stored as glycerol stocks at -80˚C (Table 2.1). The kanamycin resistance cassette was then excised from the chromosome using FLP recombinase, as described in section 2.8.3. Candidates were then screened for the loss of the *kan* gene by colony PCR, as previously described. The integrity of the *lac28* promoter fragment on the chromosome was then checked by purifying the PCR product using the QIAquick PCR Purification Kit, after which the *lac28* promoter region was sequenced using primer D68556 (Table 2.4). Successful candidates were assigned a BRY number and stored as glycerol stocks at -80˚C (Table 2.1).

2.10 Fluorescence assays

2.10.1 Exponential growth phase fluorescence assays

Fluorescence assays were used to measure the level of fluorescence as a function of OD_{620} from strains carrying chromosomally encoded *lac28::egfp* fusions to determine the level of

Table 2.7: Oligonucleotides used for sequencing and/or PCR screening of inserts

Name	Sequence (5' – 3')	Use [1]
D54890	TGAAGGGCAATCAGCTGTTG	Anneals to 3' end of *lacI* gene, upstream of cloning site in pKH3. Used for sequencing inserts in pKH3.
D55668	GAGCGGCGACGATAGTCATG	Anneals downstream of SalI site in pBR322. Used for sequencing inserts in pBR322 during construction of pKH3.
D10520	CCCTGCGGTGCCCCTCAAG	Anneals upstream of *Eco*RI site in pRW50. Used for sequencing and amplification of inserts in this vector.
D56613	CTTGATGTCTCTGACCAGAC	Anneals within the *lacI* sequence outside of the pKH3 homology. Used for screening gene gorging candidates by PCR amplification.
D56614	TTATGCAGCAACGAGACGTC	Anneals within the *lacZ* sequence outside of the pKH3 homology. Used for screening gene gorging candidates by PCR amplification.
D68556	TTTACGTCGCCGTCCAG	Anneals downstream of the start codon of *gfp*. Used for sequencing promoter inserts in pJB3 derivatives and BRY strains.
D69747	GTCGCACAGAACATCGG	Anneals to *thiQ* gene outside of the homology regions used. Used for PCR screening with D69748. (*ara*)
D69748	TCGCTGGTCATTTCTGAAG	Anneals to *yabI* gene outside of the homology regions used. Used for PCR screening with D69747. (*ara*)
D69751	TATCGCCTCAATGGTGACA	Anneals to *melB* gene outside of the homology regions used. Used for PCR screening with D69752. (*mel*)
D69752	TTGCCACCGCCAGAC	Anneals to *yjdF* gene outside of the homology regions used. Used for PCR screening with D69751. (*mel*)
D71835	GGAAACCATTATGCAGGAGG	Anneals to *yafT* gene outside of the homology regions used. Used for PCR screening with D71836. (*yafT*)
D71836	TTTCCCTGCATTCAATGC	Anneals to *ykfM* gene outside of the homology regions used. Used for PCR screening with D71835. (*yafT*)

Table 2.7: Oligonucleotides used for sequencing and/or PCR screening of inserts

Name	Sequence (5' – 3')	Use [1]
D71837	AATTTACAGTCCGATGAAGG	Anneals to *eaeH* gene upstream of the homology regions used. Used for PCR screening with D71838. (*eaeH*)
D71838	TTTCAGATAATCGCGCC	Anneals to *eaeH* gene upstream of the homology regions used. Used for PCR screening with D71837. (*eaeH*)
D72064	GCGCCATTTCCCAAGCC	Anneals to *rcsB* gene upstream of the homology regions used. Used for PCR screening with D72065. (*rcs*)
D72065	CGCTCTGCCGTCAACGG	Anneals to *rcsC* gene upstream of the homology regions used. Used for PCR screening with D72064. (*rcs*)
D72901	GTGCTTTGCCAGACTGC	Anneals to *tam* gene outside of the homology regions used. Used for PCR screening with D72902. (*tam*)
D72902	TGACATTACGCAAACAGCC	Anneals to *yneE* gene outside of the homology regions used. Used for PCR screening with D72901. (*tam*)
D72903	ATCCGATGTTTGCCAGC	Anneals to *nupG* gene outside of the homology regions used. Used for PCR screening with D72904. (*nupG*)
D72904	GCATAATGAACAGCATATTGAGG	Anneals to *speC* gene outside of the homology regions used. Used for PCR screening with D72903. (*nupG*)
D72905	CACTATGTTTATCCGCAATATCG	Anneals to *aslB* gene outside of the homology regions used. Used for PCR screening with D72906. (*asl*)
D72906	TCGCGTACCGACTTTCG	Anneals to *aslC* gene outside of the homology regions used. Used for PCR screening with D72905. (*asl*)
D72907	GGCAGATGCGAATTAAACC	Anneals to *ycbW* gene outside of the homology regions used. Used for PCR screening with D72908. (*ycb*)
D72908	TCGCATTGGTGATGTGG	Anneals to *ycbX* gene outside of the homology regions used. Used for PCR screening with D72907. (*ycb*)

Table 2.7: Oligonucleotides used for sequencing and/or PCR screening of inserts

Name	Sequence (5' – 3')	Use [1]
D76493	GCTACAGCTGCGGCGGC	Anneals to *mntH* gene outside of the homology regions used. Used for PCR screening with D76494. (*mntH*)
D76494	GCGGCAATAACCGTTTCTTGCG	Anneals to *ypeC* gene outside of the homology regions used. Used for PCR screening with D76493. (*mntH*)
D75808	AGTGGACGCATCACTGG	Anneals to *aspU* gene outside of the homology regions used. Used for PCR screening with D75809. (*dkgB*)
D75809	CCTCATCTTTCAGGGCC	Anneals to *dkgB* gene outside of the homology regions used. Used for PCR screening with D75808. (*dkgB*)
D75800	AGCGCATTTGTCATAGG	Anneals to *yqeJ* gene outside of the homology regions used. Used for PCR screening with D75801. (*yqe*)
D75801	TTTCTCTCATTTGATTAGAGC	Anneals to *yqeL* gene outside of the homology regions used. Used for PCR screening with D75800. (*yqe*)
D75802	ACACCTCCATTAATTGC	Anneals to *yghT* gene outside of the homology regions used. Used for PCR screening with D75803. (*pitB*)
D75803	ACACTACCTGCAACAGC	Anneals to *pitB* gene outside of the homology regions used. Used for PCR screening with D75802. (*pitB*)
D75810	ACCTGACAGCGTGTATTCC	Check primer for *ΔeaeH*, which anneals outside of the homology regions used. Used for PCR screening. (*ΔeaeH*)
D75812	CATTACAAACGGAGTCTGG	Check primer for *ΔyafT*, which anneals upstream of the homology regions used. Used for PCR screening with D75813. (*ΔyafT*)
D75813	AAAACCGTTGACGAAGG	Check primer for *ΔyafT*, which anneals downstream of the homology regions used. Used for PCR screening with D75812. (*ΔyafT*)
D77712	GTCTTTCATCACAACCTCG	Check primer for *Δyqe*, which anneals upstream of the homology regions used. Used for PCR screening with D77713. (*Δyqe*)

Table 2.7: Oligonucleotides used for sequencing and/or PCR screening of inserts

Name	Sequence (5' – 3')	Use [1]
D77713	GGTCTTTACCTTGATCTCC	Check primer for *Δyqe*, which anneals downstream of the homology regions used. Used for PCR screening with D77712. (*Δyqe*)
D77714	GGAGATTTGGTTTTCATGC	Check primer for *ΔpitB*, which anneals upstream of the homology regions used. Used for PCR screening with D77715. (*ΔpitB*)
D77715	TTGCGCTGATTGTTTCC	Check primer for *ΔpitB*, which anneals downstream of the homology regions used. Used for PCR screening with D77714. (*ΔpitB*)

[1] The chromosomal locus for which the primer was used to amplify is italicised and in brackets

eGFP expression and *lac28* promoter activity. Strains carrying the chromosome-encoded *lac28::egfp* fusion were grown overnight at 37°C with aeration in M9 minimal salts media, supplemented with 0.3% fructose. Density of the overnight culture was determined by measuring OD_{620} and then used to sub-culture into 5 ml of the same medium for each strain/condition to a final OD_{620} ~ 0.03. Where required, the inducers IPTG and melibiose were added to the growth medium, to the final concentrations stated in the results chapters, when subcutured. Cultures were incubated at 37°C with aeration until mid-exponential phase (OD_{620} of 0.3-0.5). 250 µl samples of each culture were aliquoted into a sterile, black, optically clear bottomed, 96-well, Corning Costar 3603 plate (Thermo Scientific). Fluorescence at excitation wavelength 485 nm and emission wavelength 510 nm was measured for an integration time of 1 second using a Thermo Fluoroskan Ascent FL fluorometer (Thermo Scientific) after a 10 second shake step at 600 rpm. The OD_{620} of the samples was then measured using a Labsystems Multiskan MS plate reader (Thermo Scientific). Each experiment consisted of a minimum of 3 biological replicates for each strain tested and experiments were repeated on at least 2 separate occasions. The starting strain, *E. coli* K-12 MG1655, was grown under the same conditions as the test strains to measure auto-fluorescence of the cells, however no fluorescence was usually detected at the optical density used. The activity of the *lac28* fragment promoter was derived as fluorescence/OD_{620} to represent specific fluorescence of the culture, with mean and standard deviation calculated for each strain/condition. The *in vivo* half-life of the GFP protein is estimated to be greater than 24 hours when expressed in *E. coli* (Andersen *et al*., 1998). Hence, expressed eGFP protein is expected to accumulate during the time course used in these experiments, therefore representing the expression over the whole growth period used.

2.10.2 Stationary phase fluorescence assays

Fluorescence assays were also used to measure the activity of the *lac28::egfp* fusion during stationary phase. Strains carrying the chromosome-encoded *lac28::egfp* fusion were grown overnight at 37°C with aeration in M9 minimal salts media, supplemented with 0.3% fructose. The culture was split into two 2.5 ml aliquots and 250 μM IPTG added to one of the cultures to induce activity of the *lac28* fragment promoter. Cultures were then incubated for a further 5 hours after the addition of inducer, after which 250 μl samples of each culture were aliquoted into a sterile, black, optically clear bottomed, 96-well, Corning Costar 3603 plate (Thermo Scientific). Fluorescence at excitation wavelength 485 nm and emission wavelength 510 nm was measured for an integration time of 1 second using a Thermo Fluoroskan Ascent FL fluorometer (Thermo Scientific) after a 10 second shake step at 600 rpm. Samples were diluted 1 in 10 in phosphate buffered saline and the OD_{620} was measured using a Labsystems Multiskan MS plate reader (Thermo Scientific). The activity of the *lac28* fragment promoter was then derived by fluorescence/OD_{620} to represent specific fluorescence of the culture. The non-induced cultures demonstrated a fluorescence signal at the high optical density used in the experiment due to auto-fluorescence of the cells, therefore the fluorescence/OD_{620} value for the un-induced sample was subtracted from that of the induced sample to subtract background auto-fluorescence. This background fluorescence did vary slightly, but was always less than the starting MG1655 strain. Each experiment consisted of a minimum of 3 biological replicates for each strain tested and experiments were repeated on at least 2 separate days, with mean and standard deviation calculated for each strain/condition.

2.11 Measurement of mRNA or gDNA by Quantitative real-time PCR

Quantitative real-time PCR (qRT-PCR) was used to measure the copy number of the *egfp*

gene and to measure the levels of *egfp* mRNA in *E. coli* BRY strains. Total cellular RNA and genomic DNA were extracted from the same cultures, which were grown with and without inducer, by the methods described in sections 2.5.7. and 2.5.8. Total RNA was used as template to synthesise *egfp* and *polA* cDNA using the cDNA Synthesis Kit (Bioline), as described by the manufacturer's protocol. The *egfp* QuantiTect Primer Assay (Qiagen) was used as the experimentally validated primer set and the D63693 and D63694 primers, which are designed to amplify the *polA* reference gene mRNA. To verify the absence of contaminating DNA in the samples, regular PCR was done with the RNA samples as the template. cDNA or gDNA was used as template in qRT-PCR, which was performed using an ABI Prism® 7000 Sequence Detection System (Applied Biosystems) and the SensiMix SYBR Hi-ROX Kit (Bioline), as described by the manufacturer's protocol. Relative expression levels of the *egfp* target gene were determined by normalising reaction threshold cycle (C_T) values to that of the *polA* reference gene. ΔC_T values for un-induced (no IPTG) cultures were used as calibrators ($\Delta\Delta C_T = \Delta C_{Tinduced} - \Delta C_{Tun\text{-}induced)}$ for analysis of results by the relative quantification method ($2^{-\Delta\Delta C_T}$), which was used with standard curves (Livak and Scmittgen, 2001). Each reaction was repeated three times for each of three separate biological replicates to yield mean and standard deviation for each experiment. gDNA samples were analysed by the same method, however the ΔC_T values for the BRY33 strain cultures, which carry the *egfp* gene at the *tam* locus, were used as calibrators for analysis of results by the relative quantification method ($2^{-\Delta\Delta C_T}$).

2.12 Gene gorging

The gene gorging method of recombineering was adapted by Hollands (2009) to allow replacement of the *E. coli* chromosomal *lacZYA* promoter with any promoter of choice. This method was used to create a range of chromosome encoded promoter*::lacZYA* fusions, to

allow comparison of promoter activities when encoded by the chromosome or a multi-copy plasmid. Hence, a general protocol for replacement of the *lacZYA* regulatory region with a promoter fragment of interest on the *E. coli* chromosome by gene gorging is described here with slight modifications from the original protocol (Hollands, 2009). This method requires the use of the pKH3 plasmid which carries ampicillin resistance and the starting strain *E. coli* K-12 KH001, a derivative of MG1655 which carries a deletion in the DNA binding domain of *lacI*, and in which the *lac* promoter has been knocked out. The KH001 strain was co-transformed with the mutagenesis plasmid, pACBSR (Cm^R; Figure 2.3), and the pKH3 donor plasmid carrying the promoter of interest (Amp^R; Figure 2.5), and co-transformants were selected on MacConkey lactose agar supplemented with chloramphenicol and ampicillin. A single co-transformant colony was suspended in 500 µl LB medium, supplemented with ampicillin and chloramphenicol, and incubated at 37˚C with aeration for ~3 h until the culture was turbid. Cells were harvested by centrifugation at ~18000 x *g*, and then washed in 0.1 x LB by resuspension. This process was repeated twice to remove any residual antibiotics, as no selection is required during induction of the mutagenesis machinery. Cells were then suspended in 500 µl 0.1 x LB supplemented with 0.3% L-arabinose to induce expression of the I-SceI endonuclease and λ-red homologous recombination proteins, which are encoded by the mutagenesis plasmid pACBSR. 0.1 x LB medium was used as no further growth was required during the mutagenesis procedure. The culture was incubated at 37˚C with aeration for a further ~3 h, after which 100 µl of a 10^{-6} dilution of the culture was spread onto each of 10-15 non-selective MacConkey lactose agar plates and incubated overnight at 37˚C. Recombinant colonies displayed a Lac^+ phenotype in a Lac^- background. Candidates were screened for the presence of the chromosomal insert by colony PCR, as described in section 2.7.2, and sequencing using primers D56613 and D56614, which anneal to regions flanking the *lacI* and *lacZ* HRs (Table 2.7).

Efficiency of donor plasmid linearisation was tested by spreading 100 µl of a 1/1000 dilution of the culture, taken before and after addition of arabinose to the culture medium, on nutrient agar and nutrient agar supplemented with ampicillin. Cells in which the donor plasmid has been linearised will not be able to grow in the presence of ampicillin, therefore comparison of the number of ampicillin resistant colonies pre- and post-induction demonstrates the efficiency of donor plasmid linearisation. Candidates were screened for chloramphenicol sensitivity to confirm loss of the mutagenesis plasmid, after which a BRY strain number was assigned and the strain stored as a glycerol stock at -80˚C.

2.13 Construction of chromosomal promoter::*lacZ* fusions by gene gorging

2.13.1 Construction of promoter::*lacZ* fusions in the donor plasmid pKH3

The *hcp*3831NS, *CC-41.5* and *CC-61.5* promoter fragments were cloned to form transcription fusions to *lacZ* in pKH3 to create donor vectors for gene gorging (Table 2.2). To do this, a *Bam*HI recognition site was introduced downstream of the promoter transcription start sites (upstream of the translation initiation region), so as not to include the translation initiation regions of these promoters. Promoters were amplified by PCR from pRW50 vectors containing the relevant promoter, using the relevant primers, as listed in table 2.8. PCR products were purified using a QIAquick PCR purification kit then digested using *Eco*RI and *Bam*HI restriction endonucleases. The *CC* promoter transcription fusions were cloned using *Eco*RI upstream and *Bgl*II downstream, due to the promoter containing an internal *Bam*HI recognition site, therefore these digests were done using *Eco*RI and *Bgl*II. The digested fragments were then purified by electroelution from a 7.5% polyacrylamide gel and cloned into *Eco*RI/*Bam*HI digested pKH3 vector, which had been CIP treated and purified by extraction from a 0.8% agarose gel. The promoter fragment inserts were confirmed by

Table 2.8: Oligonucleotides used for amplification of promoter fragments

Name	Sequence (5' – 3') [1]	Use
D65217	GCCGGGATCCTCCTTAATA TACATGTTTAAGG	Anneals to the *phcp*3831NS fragment to incorporate a *Bam*HI site 10 bp downstream of the transcription start site.
D65683	GCCGAGATCTATATCAGA ATTATGGCAG	Anneals to the 3' end of the *CC* promoter fragments to incorporate a *Bgl*II site 1 bp downstream of the transcription start site.
D66433	CGAGGAATTCGAACTCGG TACCC	Anneals to the 5' end of the *CC* promoter fragments and introduces a G→A mutation in the *Sac*I site immediately downstream of the *Eco*RI site.
D66948	TATAGAATTCGGGCAGTG AGCGCAACGC	Anneals upstream of O^3 in the *lac*ZYA regulatory region and incorporates an *Eco*RI site. Used to create pRW500/*plac*.
D69482	GCCGAAGCTTAAGGCGAT TAAGTTGGG	Anneals to *lacZ* to incorporate a *Hin*dIII site immediately downstream of codon 28 to create the *lac28* promoter fragment.
D54890	TGAAGGGCAATCAGCTGT TG	Anneals to 3' end of *lacI* gene, upstream of cloning site in pKH3. Used for sequencing inserts in pKH3.
D66432	CGCCACCATCCAGTGCAG	Anneals downstream of the *Sac*I restriction site in *lacZ*. Used in creating pRW500 derivatives.

[1] Target sites for restriction endonucleases are underlined.

sequencing using primer D54890 (Table 2.7). Successful candidates were stored at -20˚C.

2.13.2 Construction of chromosomal promoter::*lacZ* fusion by gene gorging

The *hcp*3831NS, *CC-41.5* and *CC-61.5::lacZ* transcription fusions were transferred to the chromosome of *E. coli* strain KH001 co-transformed with the mutagenesis plasmid, pACBSR, and the relevant pKH3 derivative by the gene gorging method described in section 2.12. Recombinant colonies were identified by a red phenotype (Lac^+) in a white background (Lac^-) on MacConkey lactose agar. Candidates were screened by colony PCR, as described in section 2.7.2, and sequencing using primers D56613 and D56614 (Table 2.7). Successful candidates were assigned a BRY strain number and stored as glycerol stocks at -80˚C and on MacConkey lactose plates (Table 2.1).

2.14 Construction of promoter::*lacZYA* fusions in pRW500

2.14.1 Construction of pRW500

The *lac* expression vector pRW50 contains a *trpBA* fragment fused to *lacZYA*, creating a promoter*::trpBA::lacZYA* fusion, whereas the chromosomal promoter *lacZ* fusions created by gene gorging, as described in section 2.13.2, do not contain the *trpBA* fragment. Therefore the *trpBA* fragment of pRW50 was replaced with the pKH3 cloning region, which is present on the chromosome of strain KH001, to make promoter::*lacZ* fusions, encoded by the plasmid, comparable with that of the chromosomal copies, described in section 2.13.2 (Figure 2.25). The pKH3 promoter cloning region and 5' end of *lacZ* were amplified by PCR using *E. coli* K-12 KH001 genomic DNA as template and using primers D54890 and D66432 (Table 2.8). Primer D54890 anneals to the 3' end of the *lacI* gene, upstream of the promoter cloning site in

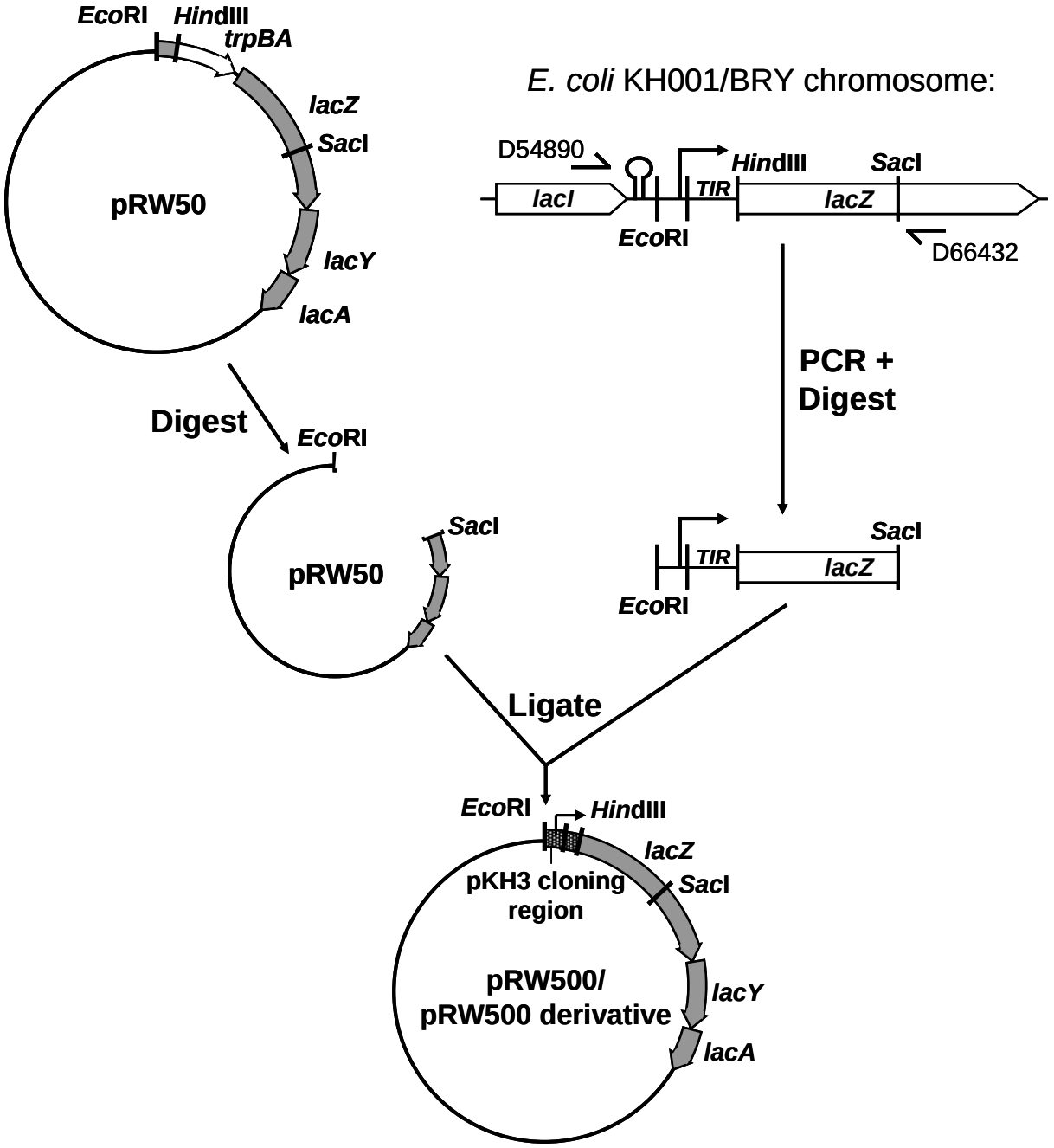

Figure 2.25: Schematic overview of pRW500 and pRW500 derivative construction

For details see text (section 2.14). The pKH3 cloning site*::lacZ* chromosomal fusion was amplified from *E. coli* KH001 genomic DNA. This was then cloned into *Eco*RI-*Sac*I digested pRW50 to replace the *trpBA* fragment and create plasmid pRW500. The same procedure was used to create pRW500 derivatives carrying promoter*::lacZ* fusions, except *E. coli* MG1655, KH002, BRY01, BRY03 or BRY05 genomic DNA was used as template for the PCR. Black circles represent plasmid DNA with the plasmid name in the centre. Genes are represented by block arrows with the direction of transcription depicted by the arrow and the gene name labelled. The position of restriction endonuclease target sites of interest are represented by black lines perpendicular to the plasmid or PCR DNA lines. Oligonucleotides are represented by thin arrow lines with the primer code number (D54890 and D66432). Cloning steps are represented by thick black arrows with the procedure labelled.

KH001, whereas primer D66432 anneals downstream of a unique target site for the restriction endonuclease *Sac*I, which is within the *lacZ* gene. The PCR product was purified using the QIAquick PCR Purification Kit then digested with the endonucleases *Eco*RI and *Sac*I before further purification by extraction from a 0.8% agarose gel. The plasmid pRW50 was digested with *Eco*RI and *Sac*I, after which the larger fragment was purified by extraction from a 0.8% agarose gel, using the QIAquick Gel Extraction Kit. *Eco*RI-*Sac*I digested pRW50 was then ligated with the *Eco*RI-*Sac*I PCR fragment before being used to transform calcium chloride competent *E. coli* RLG221 to produce the plasmid pRW500 (Figure 2.2). The presence of the correct insert of was screened by DNA sequencing using primer D10520 (Table 2.4).

2.14.2 Construction of promoter*::lacZYA* fusions in pRW500

The chromosomal promoter*::lacZYA* fusions described in section 2.13.2 were amplified by PCR to allow direct cloning of each fusion into pRW500 and comparison between chromosome-encoded and plasmid-borne promoter activities (Figure 2.25). The *lac, melR (TB10a), phcp3831NS, CC-41.5* and *CC-61.5* promoter*::lacZYA* fusions were amplified by colony PCR from *E. coli* K-12 MG1655, KH002, BRY01, BRY03 or BRY05 strains using primers D54890 and D66432 (Table 2.8). PCR products were purified using the QIAquick PCR Purification Kit, after which they were digested with the restriction endonucleases *Eco*RI and *Sac*I before further purification by extraction from a 0.8% agarose gel. Plasmid pRW500 was prepared by *Eco*RI-*Sac*I digest, CIP treatment and purification by extraction from a 0.8% agarose gel. *Eco*RI-*Sac*I digested pRW500 was then ligated with the *Eco*RI-*Sac*I PCR fragments, carrying the promoter*::lacZ* fusions, before being used to transform calcium chloride competent *E. coli* RLG221. Candidates were screened for the presence of the correct promoter*::lacZ* fusion by DNA sequencing using primer D54890 (Table 2.8). Successful candidates were stored at -20˚C (Table 2.2).

2.15 β-galactosidase assays

β-galactosidase assays were used to measure the activity of promoter*::lacZ* fusions encoded either by the chromosome or plasmid during a range of growth conditions. Strains carrying the relevant promoter*::lacZ* fusion were grown overnight at 37°C with aeration in either LB media, M9 minimal salts media or minimal salts medium, as defined in section 2.2.1 and supplemented with the relevant anitbiotics, sugars or inducers. Density of the overnight culture was determined by measuring OD_{650} and then used to sub-culture into 5 ml of the relevant medium for each strain/condition to a final OD_{620} ~ 0.03. Cultures were incubated at 37°C with aeration until mid-exponential phase (OD_{620} of 0.3-0.5). Each culture was lysed by adding 2 drops each of toluene and 1% sodium deoxycholate, mixing by vortex for 15 s and aerating for 30 min at 37 °C.

The β-galactosidase activity of each culture was assayed by addition of 100 µl of each culture lysate to 2.5 ml Z buffer (10 mM KCl, 1 mM $MgSO_4.7H_2O$, 60 mM Na_2HPO_4, 30 mM $NaH_2PO_4.2H_2O$ supplemented with 2.7 ml β-mercaptoethanol per litre of distilled water, adjusted to pH 7) supplemented with 13 mM 2-Nitrophenyl β-D-galactopyranoside (ONPG). The reaction was incubated at 37˚C until a yellow colour developed, after which the reaction was stopped by adding 1 ml 1 M sodium carbonate. Absorbance of the reaction at OD_{420} was measured with a "no lysate" control reaction subtracted as background. β-galactosidase activity was calculated in Miller units using the following calculation:

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

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)

1000/4.5 = 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 (in ml)

Each experiment consisted of a minimum of 3 biological replicates for each strain tested and experiments were repeated on at least 2 separate occasions, with mean and standard deviation being calculated for each strain/condition.

2.16 α-galactosidase activity assays

α-galactosidase assays were used to measure expression of the *melAB* operon, which encodes genes required for transport and utilisation of the sugar melibiose. Strains used were deficient in *lacY* expression, due to the *lacY* gene product interfering with transport of melibiose, the inducer of the *melAB* operon. The relevant *E. coli* strains were grown overnight with aeration in 5 ml M9 minimal salts media at 30°C, due to the temperature sensitivity of the melibiose symporter, MelB. Density of the overnight culture was determined by measuring OD_{650} and then used to sub-culture into 5 ml M9 minimal salts medium, supplemented with 0.2% melibiose where stated, to a final OD_{620} ~ 0.03. Cultures were incubated at 30°C with aeration until mid-exponential phase (OD_{620} of 0.3-0.5). Optical density of the culture at OD_{650} and OD_{410} was measured and recorded. 50 µg/ml chloramphenicol was added to each culture and incubated for 10 mins with aeration to arrest protein production. 60 µl 0.1 M 4-nitrophenyl α-D-galactopyranoside was added directly to 2 ml of each culture and incubated at 30˚C with

aeration until a yellow colour developed, after which the reaction was stopped by addition of 160 µl 0.5 M EDTA and 167 µl 3 M sodium carbonate. The final OD_{410} was recorded and α-galactosidase activities calculated as follows:

$$\alpha\text{-galactosidase activity} = \frac{2.5 \times 2.387 \times \Delta OD_{410}}{t \times 0.0045 \times 2 \times OD_{650}}$$

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).

2.387 = final assay volume (ml)

0.0045 = factor for conversion of OD_{410} into nmol p-nitrophenyl (PNP), based on 1 nmol ml^{-1} PNP having an OD_{410} of 0.0045

t = incubation time (min)

v = volume of culture added (in ml)

Each experiment consisted of a minimum of 3 biological replicates for each strain tested and experiments were repeated on at least 2 separate occasions, with mean and standard deviation being calculated for each strain/condition. Mean activity for the un-induced cultures was subtracted from that of the induced cultures as background.

2.17 Fluorescence microscopy

Cells were harvested by centrifugation at ~18000 x *g* from 1.5 ml of the appropriate liquid culture. Cells were then washed 3 times with PBS. Cells were harvested by centrifugation and all supernatant was removed using a pipette. The cell pellet was resuspended in 5 µl PBS and pipetted into the centre of the slide and a cover slip applied. Slides were imaged using a Nikon Eclipse 90i microscope using a FITC filter for visualising GFP.

Chapter 3:

Position-dependent modulation of promoter activity in *Escherichia coli*

3.1 Introduction

3.1.1 Position-based modulation of promoter activity

The expression level of any bacterial gene is set by modulation of RNAP binding to defined target sequences within the upstream regulatory element, in response to specific chemical or physical stimuli. This is achieved by transcription factor binding in regulatory regions and interacting with the multi-subunit RNAP holoenzyme to regulate transcription initiation (reviewed by Lee *et al*., 2012). However, changes to the physiological status of the cell, such as growth phase or nutrient limitation, can also affect the global transcription profile through NAP binding (Browning *et al*., 2010; Dillon and Dorman, 2010) or variations of superhelical density (Peter *et al*., 2004; Blot *et al*., 2006; Travers and Muskhelishvili, 2007). These global effects on gene expression, and the compaction of the genome to form the nucleoid, suggest that position within the folded chromosome could also affect the level of transcription, as suggested by Schmid and Roth (1987). However, to date, the effect of chromosomal position on gene expression has barely been investigated, with previous studies being limited in experimental design by technologies and knowledge available at the time (Beckwith *et al*., 1966; Schmid and Roth, 1987; Sousa *et al*., 1997). Therefore, in the work presented in this chapter, I aimed to utilise the gene doctoring method of recombineering (Lee *et al*., 2009) to insert one relatively small promoter and reporter gene at different loci throughout the *E. coli* K-12 chromosome.

3.1.2 Gene doctoring: chromosome engineering by homologous recombination

The gene doctoring method is an enhanced protocol for the incorporation of mutant alleles into the bacterial chromosome by λ-Red mediated homologous recombination (Lee *et al*., 2009). The Red recombinase system of bacteriophage λ exploits three proteins, encoded by

the *gam*, *bet* and *exo* genes, which facilitate the recombination of linear double-stranded DNA molecules. In gene doctoring, the *Saccharomyces cerevisiae* I-SceI endonuclease is used to generate a linear DNA fragment (the donor molecule) *in vivo*. Expression of the Gam protein inhibits degradation of this fragment by the host RecBCD complex. The Exo protein is a 3' to 5' exonuclease that acts on the donor molecule to create 3' single stranded overhangs to allow binding by the single stranded DNA binding protein, Bet. The Bet protein then facilitates strand invasion and recombination at the site of homology in the host chromosome, with the assistance of the host RecA protein (Murphy, 1998).

The gene doctoring method uses a two-plasmid system based on the gene gorging method of Herring *et al*. (2003) in which the target DNA, containing regions of homology to the chromosome, is delivered to the cell on a stable, high copy number plasmid (Figure 3.1). The donor DNA, carrying the regions of homology, is flanked by DNA recognition sites for the I-SceI endonuclease, of which there are none in the *E. coli* chromosome. The I-SceI endonuclease and the λ-Red genes are under control of an arabinose inducible promoter encoded by the mutagenesis plasmid pACBSR, which is co-transformed with the donor plasmid. Therefore, upon addition of arabinose to the growth medium, the donor plasmid is linearised by the I-SceI endonuclease and the λ-Red machinery facilitates recombination of the donor fragment with the *E. coli* chromosome. Recombination is selected by incorporation of a kanamycin resistance cassette on the donor fragment, which is flanked by DNA recognition sites for the FLP recombinase. Subsequent expression of the FLP recombinase from the pCP20 plasmid allows recombination of the FLP recombinase target sites resulting in removal of the *kan* cassette. Further to this, the *sacB* sucrose sensitivity gene is included in the donor plasmid, outside of the donor fragment, therefore allowing counter-selection against retention of the donor plasmid (Figure 3.1). The gene doctoring method was utilised in this study to incorporate a promoter and reporter gene at various chromosomal loci.

Figure 3.1: Diagram of gene doctoring recombineering

Figure on page 130. a. Recipient *E. coli* MG1655 is co-transformed with the pACBSR recombineering plasmid and the pJB donor plasmid. Addition of arabinose to the growth medium induces expression of I-SceI endonuclease and the λ-Red recombination machinery from the pACBSR plasmid. The donor plasmid is linearised by the I-SceI endonuclease to allow λ-Red mediated recombination of the donor fragment with the homologous regions of the chromosome. Recombinants should have lost the *sacB* gene and retained the *kan* cassette, therefore are selected by ability to grow on solid media supplemented with kanamycin and sucrose. Recombinants are transformed with the temperature sensitive pCP20 plasmid encoding FLP-recombinase, which recognises sites flanking the kanamycin resistance cassette allowing removal of the cassette. Recombinants are verified by PCR amplification of genomic DNA and sequencing. **b.** Schematic diagram of pACBSR and pJB donor plasmids. Arrows represent genes and indicate the direction of transcription. Approximately to scale.

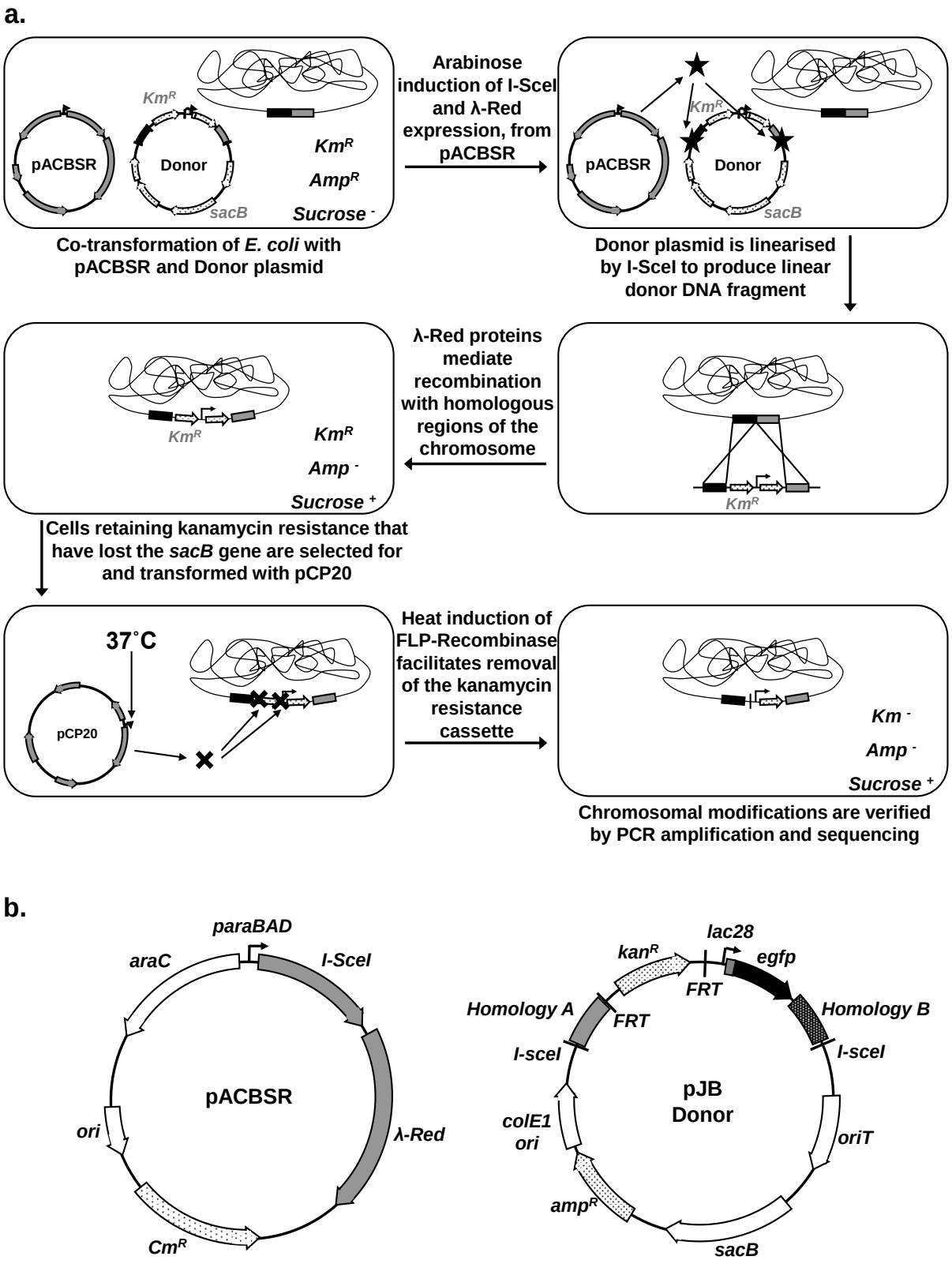

Figure 3.1: Diagram of gene doctoring recombineering

3.2 Construction of donor plasmids and preliminary experiments

The green fluorescent protein (GFP) of the jellyfish *Aequorea victoria* offers an exceptionally useful tool for the study of biological systems. Since the discovery of GFP in 1962, this protein has been modified to produce an enormous variety in colour, brightness, maturation and codon optimisation, making it applicable to a vast number of applications (Shimomura *et al*., 1962; Shaner *et al*., 2005). The requirement for a small reporter gene, for this experiment, led to the selection of the 714 bp *egfp* gene, which is codon optimised and encodes a brighter version of the *A. victoria* GFP, also demonstrating faster maturation in bacteria than the wild type protein (Tsien, 1998).

The gene doctoring method requires the use of a plasmid to donate the insert and flanking homology regions, therefore a master donor plasmid for the fusion of any promoter to the *egfp* gene was created. The donor plasmid was based on the *I-SceI* flanked donor fragment of the pKH5 gene gorging plasmid cloned into the pDOC-C gene doctoring plasmid (Figure 2.5; Figure 2.8; Figure 2.9). This resulted in the pJB3 master donor plasmid. The plasmid contains the kanamycin resistance cassette and λ oop transcription terminator, of the pKH5 plasmid, to allow selection of recombinants and prevent read-through from upstream operons. The *lacI* homology region was replaced with a multiple cloning site and the *lacZ* homology replaced with the *egfp* gene, followed by a downstream multiple cloning site (Figure 2.8; Figure 2.9). For details of plasmid construction and plasmid maps please see section 2.9.2 and figures 2.5 and 2.6.

The *lac* promoter was chosen to drive transcription of *egfp* because it is a very well characterised, strong and natural *E. coli* promoter. Therefore the *lac00* fragment, containing the endogenous *E. coli lac* promoter region from position -93 bp to +41 bp relative to the transcription start site, was amplified by PCR. This fragment includes the *lacZ* ATG and was fused to the *egfp* gene by cloning the fragment into the *Eco*RI/*Hin*dIII site of the gorging

vector pJB3, to create plasmid pJB4 (Figure 3.2a). The pJB4 plasmid was used to transform the *E. coli* MG1655 *lacI*$^{-}$ derivative strain, KH000, which was then visualised by fluorescence microscopy (Table 2.1; Figure 3.3b). The fluorescent signal detected from cells, transformed with the pJB4 plasmid, was very low and required a long 900 ms exposure for detection. This suggested that expression of eGFP by the *lac00* fragment was poor. Experiments reported by Kudla *et al*. (2009) demonstrated that expression of a promoter::*gfp* fusion could be significantly increased by addition of a 28 codon leader sequence with weak secondary mRNA structure. Therefore, the first 28 codons of the *lacZ* gene were included in the promoter fragment. The *lac28* fragment, containing the endogenous *E. coli lac* promoter region from position -93 bp to +122 bp, relative to the transcription start site, was amplified by PCR from genomic DNA. This fragment was then fused to the *egfp* gene, of the pJB3 plasmid, by cloning into the *Eco*RI/*Hin*dIII site, to create the pJB15 plasmid (Figure 3.2b). Visualisation of *E. coli* KH000, transformed with pJB15, by fluorescence microscopy with a 10 ms exposure, showed cells with a similar fluorescence intensity to that of pJB4 transformed cells, visualised with a 900 ms exposure. These results indicated approximately a 90-fold increase in expression upon addition of the first 28 codons of the *lacZ* gene to the promoter fragment (Figure 3.3). Therefore this fusion was used for insertion into the chromosome.

The pJB15 plasmid carries stop codons in each of the three reading frames and the λ *oop* transcription terminator upstream of the *lac28* fragment, to prevent translational and transcriptional read-through from neighbouring genes. Upstream of the terminator is the kanamycin resistance gene flanked by flippase recognition target (FRT) sites, which when removed by FLP recombinase, leaves behind one copy of the 34 bp FRT site. The *egfp* gene also has two stop codons in the next two reading frames after the *egfp* stop codon to prevent any translational read-through (Figure 2.6). Oligonucleotide primers used in the construction of these plasmids and fragments are listed in table 2.5 (Materials and Methods). The *I-SceI*

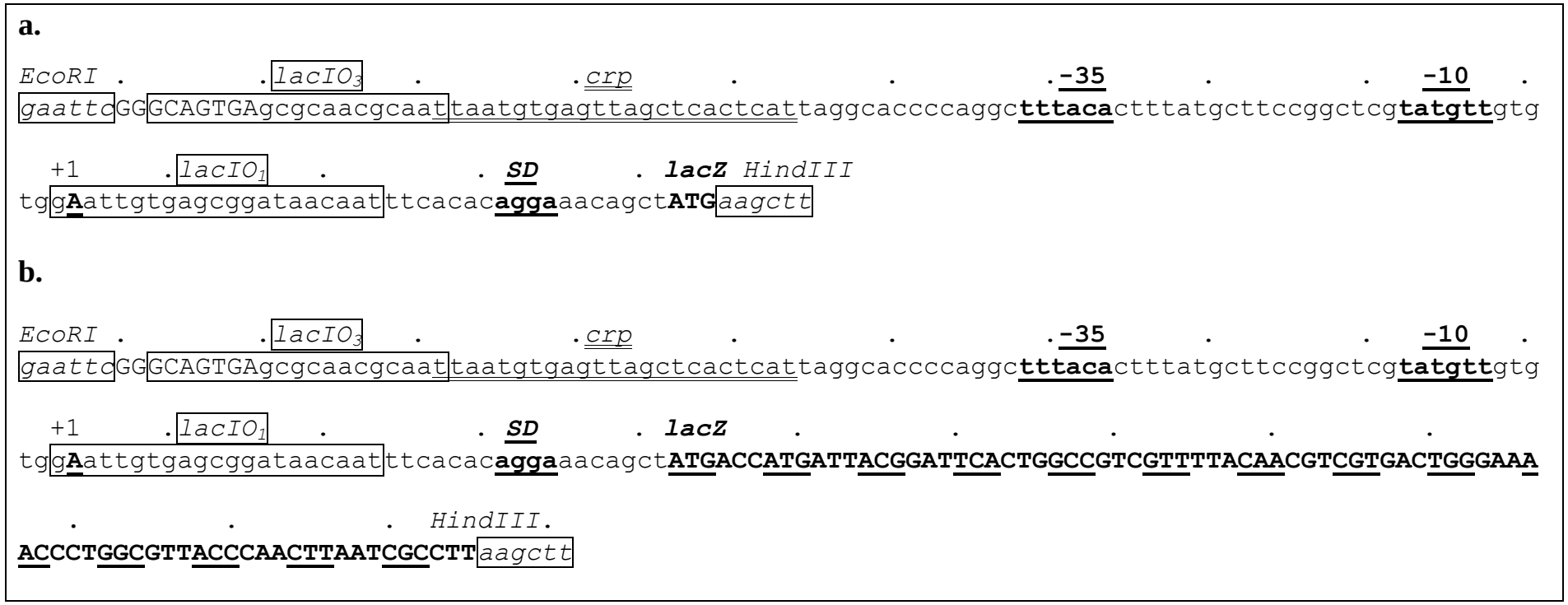

Figure 3.2: Sequence of the *lac00* and *lac28* promoter fragments

a. The base sequence of the *lac00 Eco*RI-*Hin*dIII promoter fragment containing the *lacZYA* regulatory element from – 93 bp to + 41 bp relative to the transcription start site. **b.** The sequence of the *lac28 Eco*RI-*Hin*dIII promoter fragment containing the *lacZYA* regulatory element from -93 bp to + 122 bp relative to the transcription start site. The *lacZ* start codon and the *lacZ* gene are shown in bold capitals with the Shine-Dalgarno sequence underlined and bold. The promoter -10, -35 and transcription start sites are in bold and underlined. The *lacI* operator sequences are boxed and labelled and the DNA site for CRP is double underlined. Restriction endonuclease target sites are boxed, italicised and labelled.

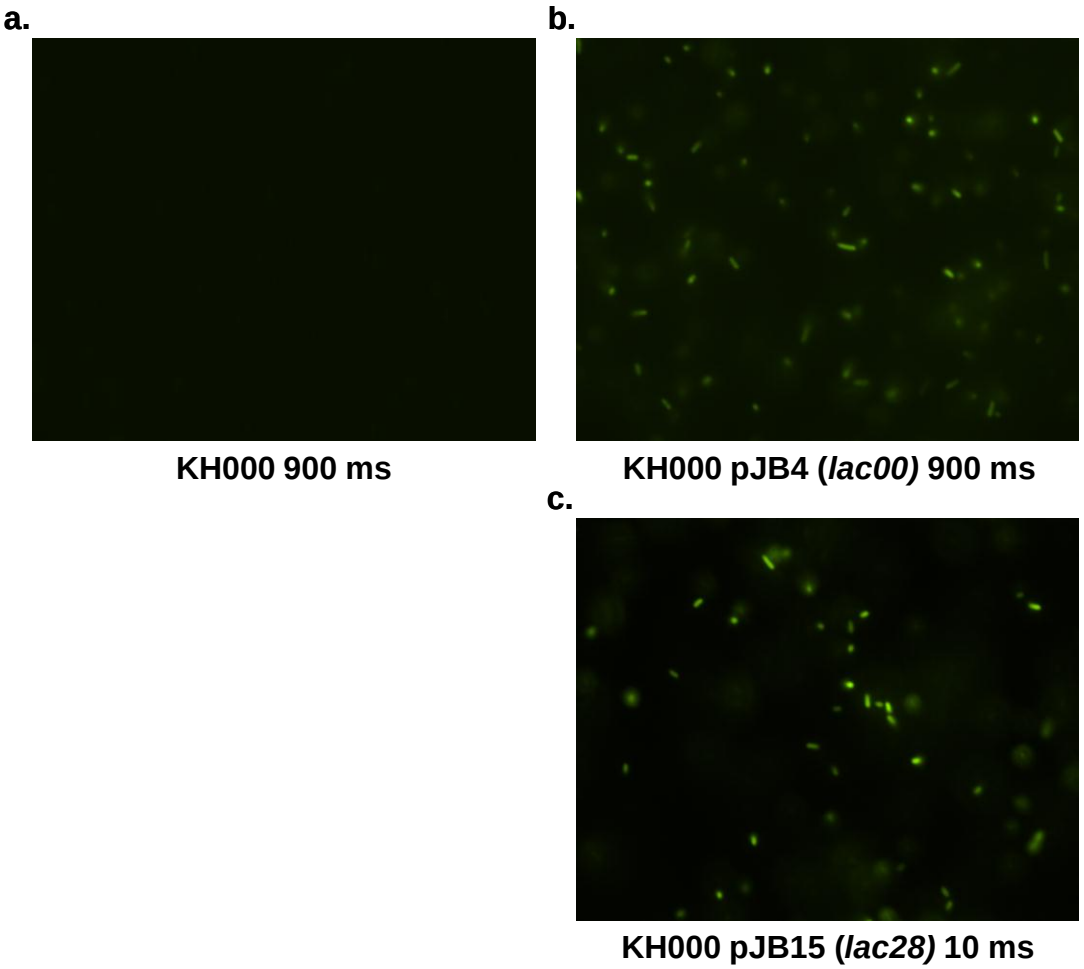

Figure 3.3: Expression of eGFP by cells carrying either *lac00::egfp* or *lac28::egfp*

Images were captured using a Nikon Eclipse 90i and eGFP fluorescence visualised using a FITC filter. **a.** The $LacI^-$ KH000 strain was visualised by fluorescence microscopy with a 900 ms exposure and showed no signal. **b.** The KH000 strain transformed with pJB4 (*lac00*::*egfp*) was visualised by fluorescence microscopy with a 900 ms exposure and showed a signal, indicating eGFP expression **c.** The KH000 strain transformed with pJB15 (*lac28*::*egfp*) was visualised by fluorescence microscopy with a 10 ms exposure and showed a similar signal intensity to that of KH000 pJB4 with a 900 ms exposure, therefore indicating a ~ 90-fold increase in expression.

donor fragment of pJB15 was cloned into the gene doctoring plasmid pDOC-C, to allow the fragment to be inserted into the *E. coli* K-12 MG1655 chromosome at target loci (Figure 2.9).

Insertion of *egfp* into the *E. coli* chromosome was targeted to non-coding DNA sites between convergent genes to minimise disruption of local chromosomal architecture and gene expression. Approximately 500 bp of homology either side of each target locus was amplified by PCR from MG1655 genomic DNA, using pairs of primers that incorporate either *Mfe*I and *Xma*I, or *Nhe*I and *Sac*I restriction endonuclease target sites at the ends of the PCR products (Figure 3.4). Oligonucleotide primers used for homology region amplification are listed with a description of usage in table 2.6. The PCR product, encoding the upstream homology region, was then digested with *Mfe*I and *Xma*I before being ligated into *Mfe*I-*Xma*I digested, CIP treated and gel purified pJB22 (Figure 2.9). The resulting plasmid was checked by sequencing and used to repeat the procedure for the downstream homology region with *Nhe*I and *Sac*I restriction endonucleases (Figure 3.4). *E. coli* K-12 MG1655 was then transformed with the resulting plasmid and the mutagenesis plasmid pACBSR, after which the *lac28::egfp* fusion was inserted into the target site by the gene doctoring method as described in section 2.8 (Figure 3.1).

3.3 Insertion of the *lac28::egfp* fusion at different chromosomal loci

To assess the effect of chromosomal location on gene expression, a series of gene doctoring donor plasmids, each carrying the *lac28::egfp* fusion, was created to allow the insertion of the fusion at various chromosomal loci. The *lac* target locus was the first selected in order to find the activity of the fusion at the wild-type location for the promoter. The chromosomal *lac* gene regulatory element, including the LacI O^2 binding site, was replaced with the *lac28::egfp* fusion. Homology regions were selected to allow replacement of the *lacZYA* gene regulatory

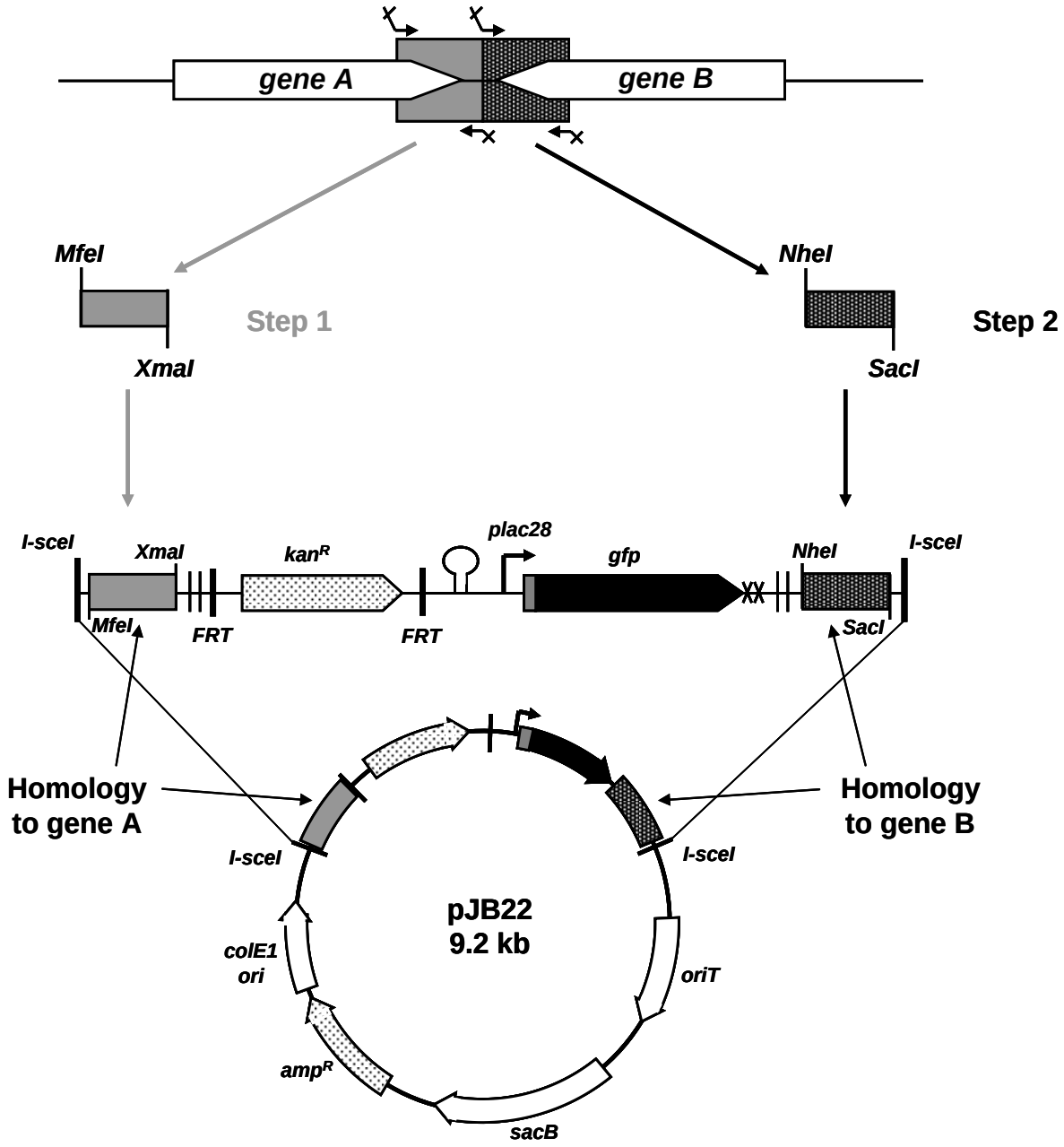

Figure 3.4: General protocol for construction of gene doctoring donor plasmids

Basic protocol used to make all donor plasmids for insertion of the *lac28::egfp* fusion into the *E. coli* chromosome by gene doctoring. Arrows represent genes, while black lines denote intergenic regions (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. DNA modifying enzyme target sites are represented by vertical black lines and are labelled. The arrow represents the position and orientation of the *lac28* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. **Step 1.** The gene A homology region is amplified by PCR from MG1655 genomic DNA, digested with *Mfe*I and *Xma*I then ligated into *Mfe*I-*Xma*I digested pJB22. **Step 2.** The downstream gene B homology region is then amplified by PCR to incorporate *NheI* and *SacI* DNA sites and cloned into the *Nhe*I-*Sac*I site of the resulting plasmid.

region from position -84 bp to +436 bp, with the *lac28::egfp* fusion, and cloned into the pJB22 gene doctoring donor plasmid as described in section 3.2 (Figure 3.4). Gene doctoring was then used to insert the fusion at the *lac* locus in the genome of *E. coli* K-12 MG1655. The resulting strain was referred to as BRY40 (Table 2.1).

Homology regions were then selected to target the reporter to non-coding regions between convergent genes at different chromosomal positions. Homology regions were amplified by PCR then cloned into the pJB22 gene doctoring plasmid, as described in section 3.2 (Figure 3.4). pJB gene doctoring donor plasmids, listed in table 2.2, were used to place the identical *lac28::egfp* fusion at 13 other chromosomal positions in *E. coli* K-12 MG1655: *ara* (72143 bp), *mel* (4342915 bp), *yafT* (238142 bp), *eaeH* (313681 bp), *rcs* (2314948 bp), *ycb* (1005714 bp), *tam* (1606129 bp), *nupG* (3104995 bp), *asl* (3982359 bp), *yqe* (2987820), *pitB* (3132854 bp), *dkgB* (229046 bp) and *mntH* (2509358 bp) (Figures 2.11-2.24; Figure 3.5). Target loci were named based on a neighbouring gene and exact positions of insertion within the *E. coli* chromosome are given in bp, with respect to the co-ordinate system origin, as defined by EcoCyc (Keseler *et al*., 2011). Resulting strains were assigned BRY numbers, e.g BRY40, see table 2.1.

The 13 chromosomal positions were targeted to allow incorporation of the fusion at a diverse range of chromosomal loci, which represent at least one target site in every macrodomain, as defined by Valens *et al*. (2004). Several loci were targeted to allow investigation of suspected position-specific effects. For example, selection of the *yafT, eaeH, ycb, yqe* and *pitB* positions allowed investigation of promoter activity from within tsEPODs. Also, the *asl* and *tam* loci were targeted to allow comparison of expression near to the origin of replication with that near to the furthest point from the origin of replication on the chromosomal map. Further to this, the *mel* and *dkgB* loci were chosen to allow the analysis of the effects of neighbouring gene expression. The remaining target loci were selected to obtain better coverage of the

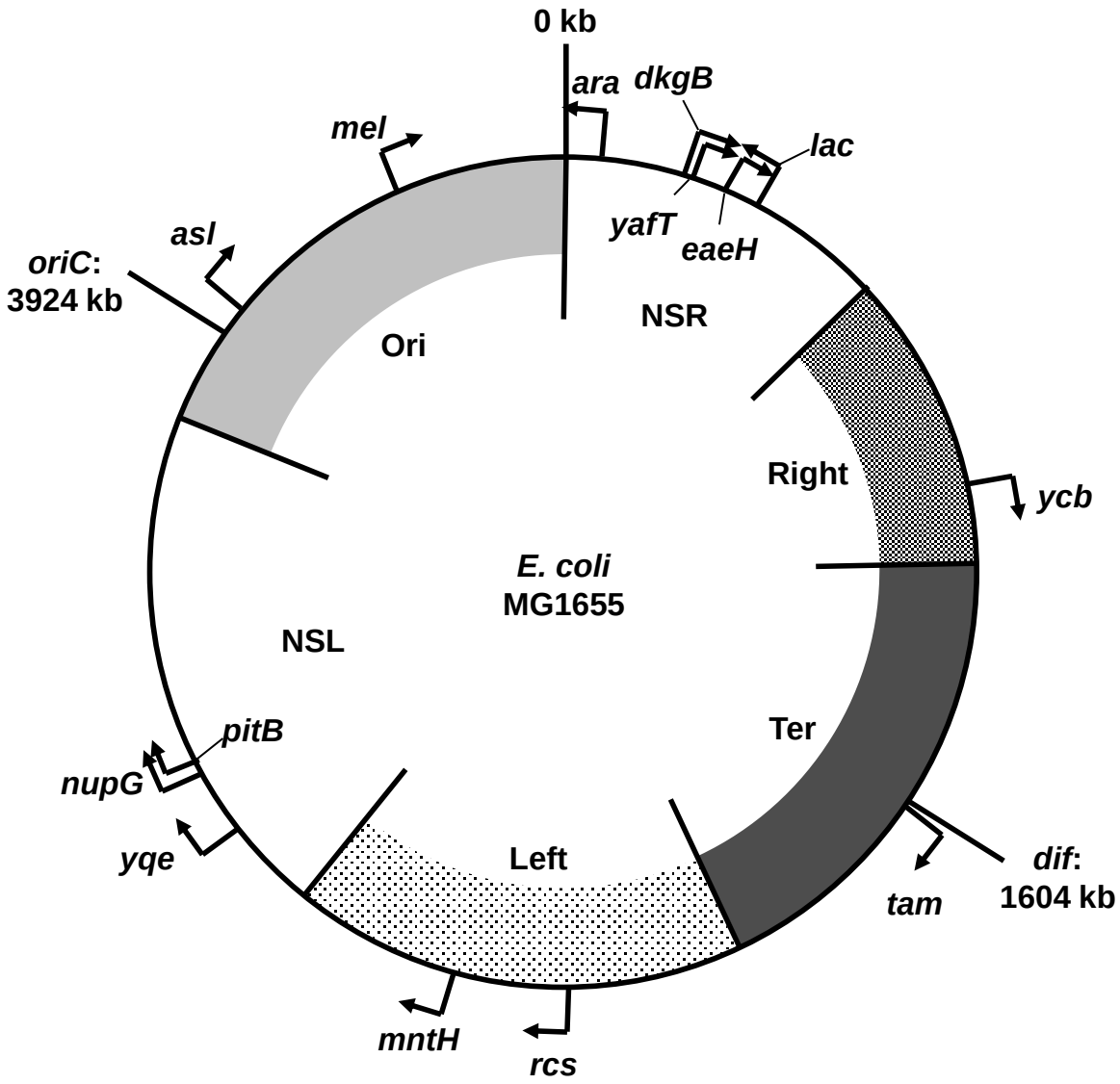

Figure 3.5: Map of the *E. coli* MG1655 chromosome and positions of *lac28::egfp* insertion

The black circle represents the *E. coli* MG1655 circular chromosome of total size 4639675 bp. The grey shaded blocks and two white gaps represent the six macrodomain as defined by Valens *et al*. (2004) and are labelled. The position of *oriC* and the *dif*, the direct opposite locus are marked with black lines and labelled with their positions in kb relative to the co-ordinate system origin. Arrows indicate the position of sites of insertion for the *lac28::egfp* construct and the direction of the arrows indicates the orientation of the promoter. Sites of insertion are labelled according to the names of neighbouring genes.

chromosome. Where possible, no genetic material was removed from the chromosome in order to minimise disruption to local chromosomal processes, such as transcription. The target site of reporter gene insertion was positioned to avoid disturbance to Rho-independent terminators of neighbouring gene transcription, where necessary. Insertion of *egfp* at the *eaeH* locus was within a pseudogene, as opposed to between convergent genes. However, this should have minimum consequences, due to the *eaeH* pseudogene not being expressed. Also, insertion at the *dkgB* locus was not between convergent genes. The fusion was targeted to a position upstream of the *dkgB* gene regulatory element and downstream of the ribosomal RNA coding operon *rrsHrrlHrrfH* and *aspU* gene. Orientation of the *lac28::egfp* fusion on the left arm of replication at the *tam, rcs, mntH, yqe, nupG* and *pitB* opposes that of replication. Insertion of *egfp* at the *ara* and *lac* loci on the right arm of replication also is opposed to replication. However, transcription of *egfp* at all the other loci was co-directional with DNA replication, therefore giving a representation of both orientations with respect to replication.

3.4 Validation of eGFP fluorescence as a measure of mRNA level

Regulation of gene expression in *E. coli* can also occur post-transcriptionally through mechanisms such as regulation of translation or mRNA stability. The experiments completed in this chapter utilised the *lac28*::*egfp* fusion to measure effects of chromosomal position on promoter activity by quantification of eGFP fluorescence. Therefore, to check that this is a valid measure of *egfp* mRNA level, with no interference from regulation of translation, the mRNA product was measured after expression at 3 different positions and compared to the eGFP fluorescence from those same cultures. The BRY33, BRY40 and BRY34 strains, carrying *egfp* at the *tam, lac* and *nupG* loci, were selected because they demonstrated a wide range of eGFP expression levels (Figure 3.6).

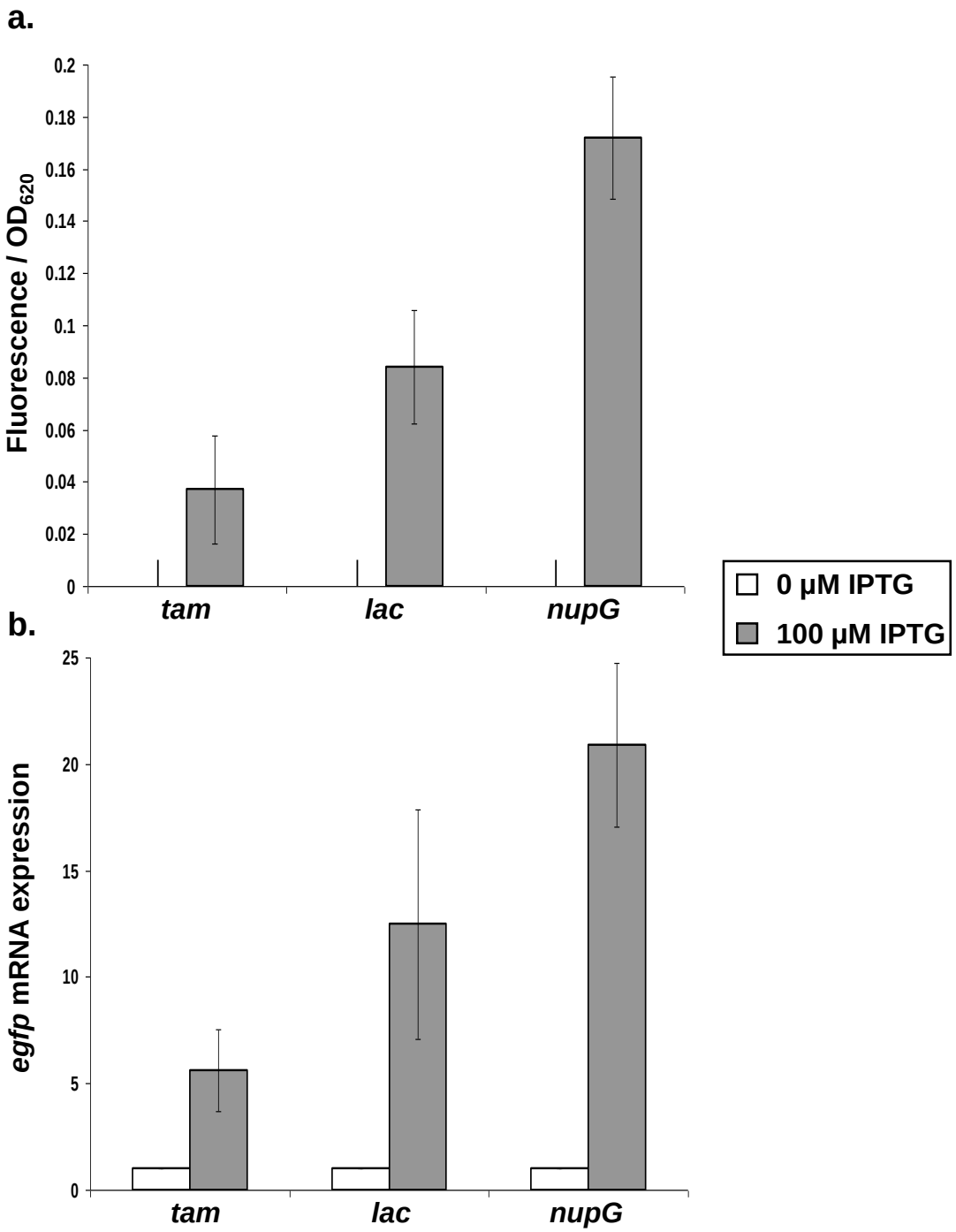

Figure 3.6: Comparison of eGFP fluorescence and *egfp* mRNA levels in different constructs

a. Fluorescence/OD_{620} measured in strains BRY33 (*tam*), BRY40 (*lac*) or BRY34 (*nupG*). Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. **b.** *egfp* mRNA expression upon induction of the *lac28* promoter::*egfp* construct from the cultures described in (a.). Data are normalised to the un-induced culture. Data show that GFP fluorescence patterns and *egfp* mRNA levels correlate.

Expression of the *egfp* target gene at the *tam*, *lac* and *nupG* loci was assayed by measuring fluorescence during mid-late exponential growth on M9 minimal salts medium, supplemented with 0.3% fructose as the carbon source. Experiments were performed in the presence or absence of 100 μM IPTG, due to the *lac* promoter being inducible (see later). Total RNA was isolated from each of these cultures in parallel. Quantitative real-time PCR (qRT-PCR) was used to measure the level of *egfp* mRNA expressed in strains carrying the fusion at the *tam*, *lac* and *nupG* loci. Oligonucleotide primers specific for the target *egfp* gene (Qiagen) and an internal control, *polA*, were used to amplify cDNA before being used in the qRT-PCR reaction. Expression of *egfp* at the *tam* locus, which is furthest from the origin of replication, was approximately 60% lower than that at the *lac* locus (Figure 3.6). However, expression of *egfp* at the *nupG* locus gave more than 2-fold greater expression than that at the *lac* locus. The fluorescence of strains BRY33, BRY40 and BRY34 correlated well with the fold changes in *egfp* mRNA level measured by qRT-PCR upon induction with the inducer IPTG (Figure 3.6). These results demonstrate that measurement of eGFP fluorescence is a valid measure of the mRNA levels within the cell. Therefore translational efficiency is not altered by change of promoter::*egfp* position within the chromosome.

3.5 Promoter activity during mid-logarithmic growth

3.5.1 The effect of chromosome position on promoter activity

To measure the effect of position within the chromosome, and local genetic context, on promoter activity, *egfp* expression was assayed in *E. coli* strains carrying the *lac28::egfp* fusion at the different chromosomal positions, described in section 3.3 (Table 2.1; Figure 3.5). Fluorescence of strains was measured during mid-late exponential growth on M9 minimal salts medium supplemented with 100 μM IPTG to induce transcription of *egfp*, where stated.

Expression of the *egfp* gene was derived from the fluorescence at 510 nm divided by OD_{620} of at least three separate experiments to give the mean and standard deviation. The starting *E. coli* K-12 MG1655 strain was included as a control for auto-fluorescence, however the control samples and the un-induced cultures registered no fluorescent signal. Therefore no auto-fluorescence was detected at the OD_{620} measured and no fluorescent signal is detected in the absence of the inducer, IPTG. These controls suggest that no transcriptional read-through is occurring from neighbouring operons.

The data, summarised in figure 3.7, show that during exponential growth, promoter activity is not only dependent on the sequence of the promoter region, but also upon position within the chromosome. Position is shown to have an effect on *egfp* expression, giving up to a 310-fold variation in promoter activity across the chromosomal loci tested, regardless of the *lac28* fragment and *egfp* reporter gene having the same sequences at all loci (Figure 3.7). Positions that were uniquely better or worse for promoter activity were found across the entire genome. The greatest activity was found at the *asl* locus, which is closest to the origin of replication and gave a signal that was approximately 2.3-fold greater than that of the strain carrying *egfp* at the *lac* locus. Similar levels of activity to that at the *asl* locus were found at the *nupG* locus, which is positioned on the left arm of replication, within the NSL macrodomain (Figure 3.7). The weakest expression of *egfp* was at the *ara, yafT, tam, mntH* and *pitB* target loci, which reside within the NSR, Ter, Left and NSL macrodomains (Figure 3.7). These results suggest that position within the *E. coli* chromosome does affect promoter activity and that these effects can be large.

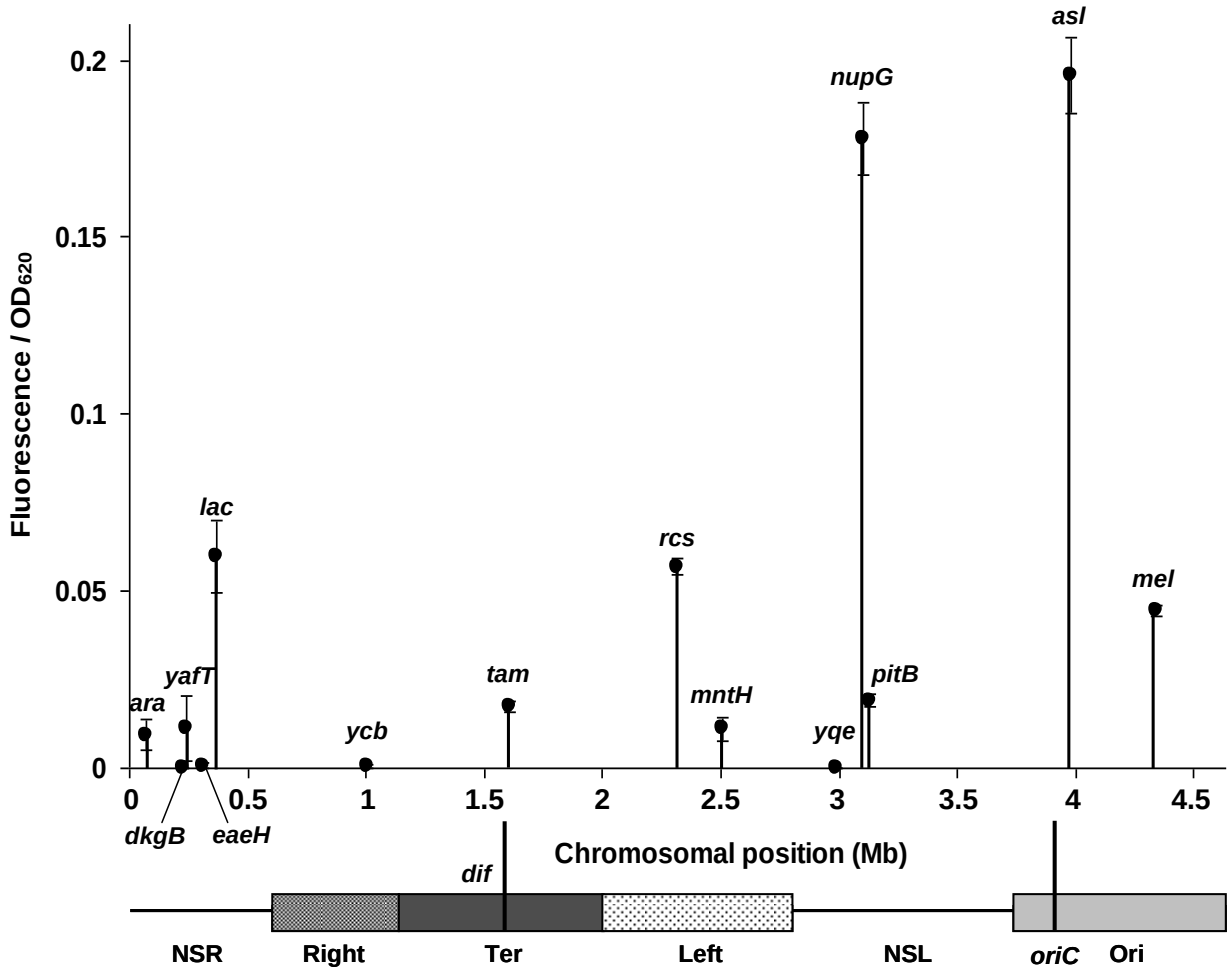

Figure 3.7: Expression of *egfp* at different chromosomal positions during mid-exponential growth

Expression of the *lac28*::*egfp* fusion during mid-exponential growth (OD_{620} ~ 0.4) at 37 ˚C in M9 minimal medium supplemented with 100 μM IPTG, the inducer of the *lac28* promoter. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the sample and is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. *E. coli* MG1655 and each of the strains in the absence of inducer were also assayed, however no fluorescence signal was detected, therefore indicating no auto-fluorescence or un-induced expression. These data are not included for the purpose of clarity. The position of each point on the x-axis indicates chromosomal position in Mb with respect to the co-ordinate system origin and is to scale. The positions of the six macrodomains are represented by the grey shaded boxes with *oriC* and the *dif* indicated by vertical black lines.

3.5.2 Position-dependent modulation of promoter activity does not correlate with macrodomains

The *E. coli* chromosome is organised into six distinct macrodomains, two of which have greater mobility than the rest. These macrodomains occupy distinct positions within the cell and this organisation may be an influential factor in position-dependent modulation of promoter activity (Valens *et al*., 2004; Espeli *et al*., 2008). The *rcs* and *mntH* positions are both within the Left macrodomain, however *egfp* expression at *rcs* was approximately 5-fold greater than that of the *mntH* locus (Figure 3.8). The reporter gene was also targeted to two loci within the ORI macrodomain, the *asl* and *mel* loci. Activity at the *asl* locus was approximately 4.5-fold greater than that of the promoter activity at the *mel* locus, regardless of the fact that they reside within the same macrodomain (Figure 3.8). Further to this, the *ara, dkgB* and *lac* target loci are all positioned within the NSR macrodomain, however activity of the *lac28::egfp* fusion at the *dkgB* locus was undetectable and expression of *egfp* at the *ara* locus was approximately 6.5-fold weaker than that at the *lac* locus (Figure 3.8). These results indicate that strength of promoter activity is unlikely to correlate with macrodomain location.

3.5.3 Transcription silencing within tsEPODs

The *yafT, eaeH, ycb, yqe* and *pitB* target loci are positioned within regions of the chromosome that are highly occupied by protein binding, and have extremely low levels of transcription of genes within the domain (Figure 3.5). The work of Vora *et al*. (2009) defined these regions as transcriptionally silent extended protein occupancy domains (tsEPODs) and demonstrated that they are distributed across the entire chromosome. Expression of *egfp* at the *eaeH* and *ycb* tsEPOD loci was extremely weak, giving a signal that was < 1% of that from the *lac* locus or approximately 310-fold weaker than that seen from the best expressed locus,

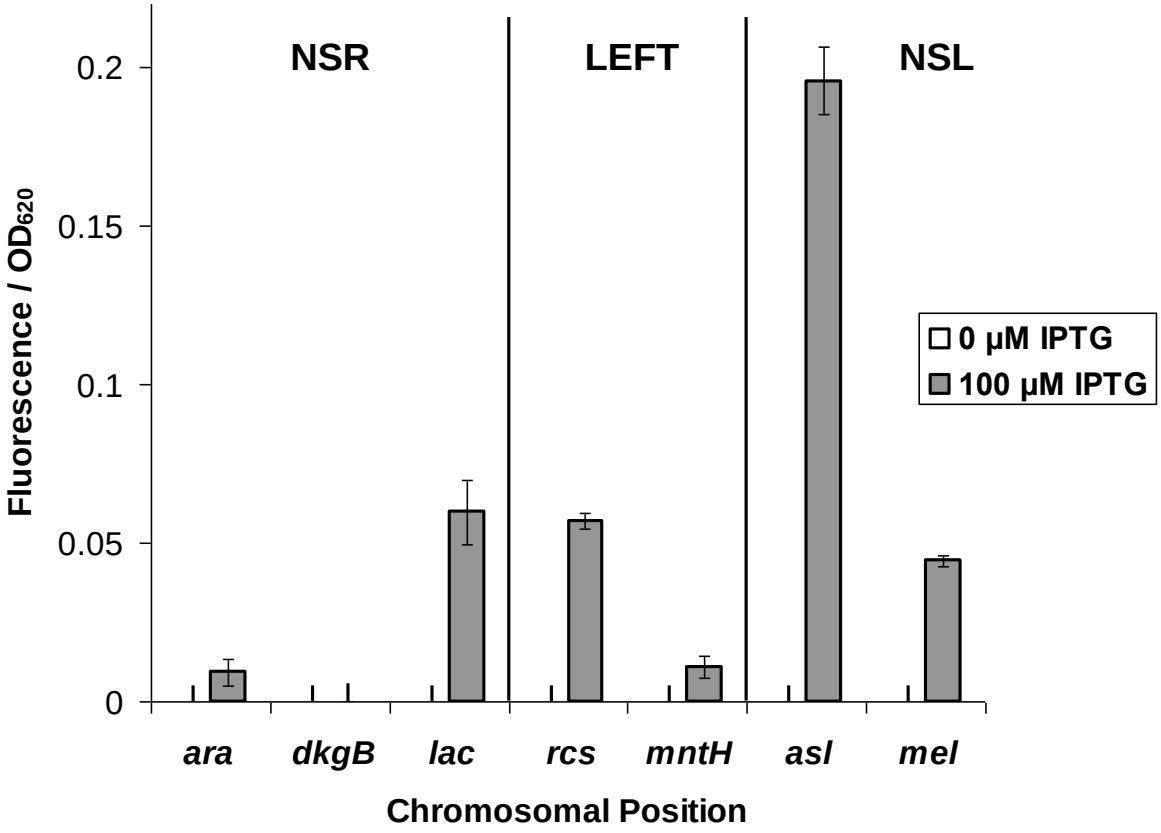

Figure 3.8: Comparison of promoter activity within different macrodomains

Fluorescence/OD_{620} measured in strains BRY13 (*ara*), BRY58 (*dkgB*), BRY40 (*lac*), BRY27 (*rcs*), BRY73 (*mntH*), BRY34 (*asl*), or BRY15 (*mel*), carrying *lac28::egfp* at the designated chromosomal locus. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon, IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Loci within the NSR, Left and NSL macrodomains are indicated by lines and labels. Data show that expression of *egfp* varies drastically within macrodomains, therefore suggesting no correlation between promoter activity and position with respect to macrodomains.

asl (Figure 3.7; Figure 3.9). Activity of the *lac28* fragment promoter at the *yqe* locus was undetectable, as the strain carrying *egfp* at this locus gave no fluorescent signal in the presence of the inducer IPTG (Figure 3.9). However, *lac28* fragment promoter activity at the *yafT* and *pitB* tsEPOD loci was stronger than that of the other tsEPOD loci, giving activity similar to that of the promoter at the *tam* locus, near to the *dif* site. The activity at the *yafT* and *pitB* loci was approximately 20% and 30% as strong as that at the wild-type *lac* locus, respectively (Figure 3.9). These results suggest that the transcriptionally silent profile of tsEPODs could be due to the nature of the domain, as opposed to merely containing poor promoters. Expression at tsEPODs is analysed in further detail in section 4.5.

3.5.4 The effect of proximity to the origin of replication

Position within a chromosome has been reported to affect gene expression in several studies, which identified proximity to the origin as a key factor (Beckwith *et al*., 1966; Chandler and Pritchard, 1975; Schmid and Roth, 1987; Sousa *et al*., 1997; Block *et al*., 2012). Analysis of position-dependent variation of promoter activity demonstrates a similar pattern of increased activity with proximity to the origin of replication (Figure 3.7; Figure 3.10). Expression of *egfp* at the *lac, tam, rcs, nupG* and *asl* loci showed increasing activity with proximity to the origin of replication, with the greatest activity at the *asl* locus, nearest the origin, and the weakest at the *tam* locus, near the *dif* site (Figure 3.10). However, activity of the *lac28* fragment promoter did not conform to this expected pattern of gene expression at several of the loci tested.

Activity of the reporter at the *mel, ara* and *dkgB* loci was significantly less than at the *lac* locus, in spite of these three loci being positioned closer to the origin of replication than the *lac* locus (Figure 3.11). Also, *egfp* positioned at the *mntH* locus was closer to the origin than

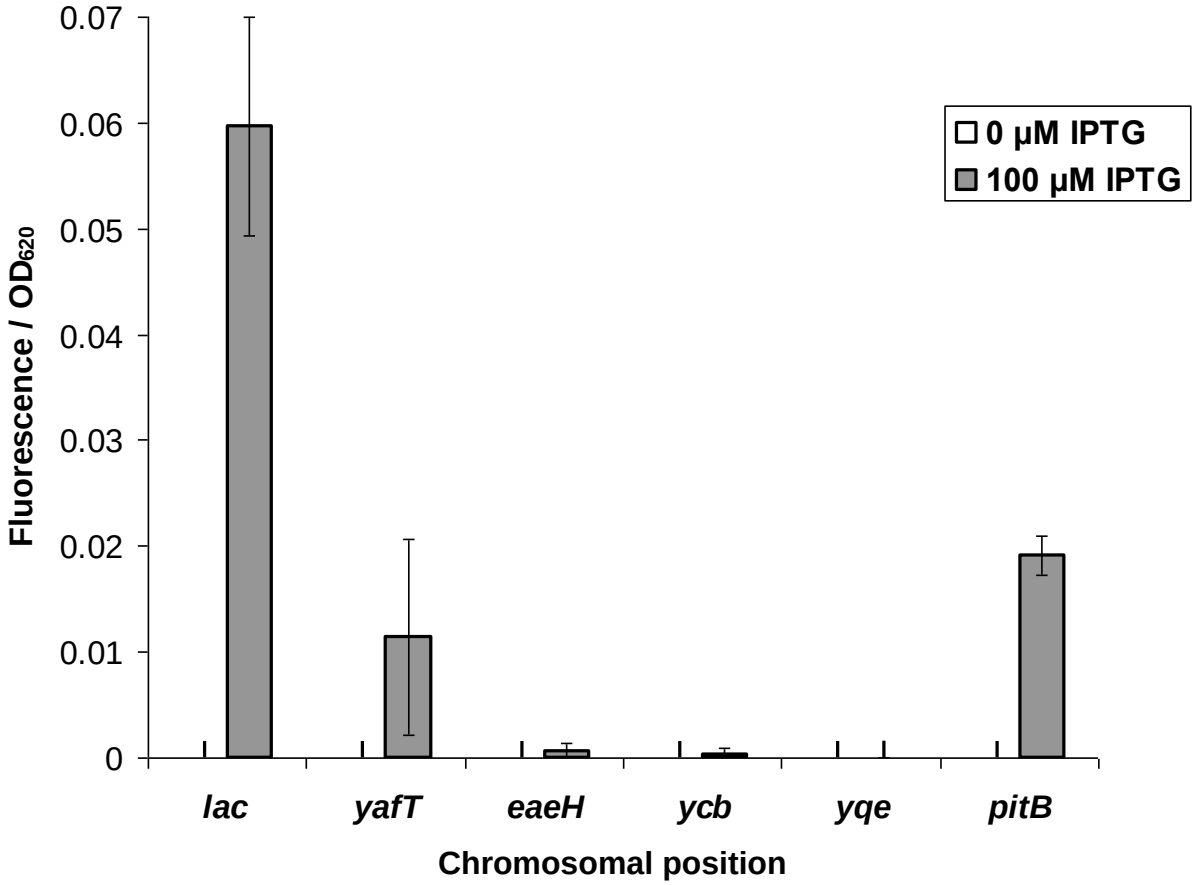

Figure 3.9: Silencing of promoter activity at tsEPOD loci

Fluorescence/OD_{620} measured in strains BRY40 (*lac*), BRY22 (*yafT*), BRY23 (*eaeH*), BRY32 (*ycb*), BRY54 (*yqe*) or BRY55 (*pitB*), carrying *lac28::egfp* at the designated chromosomal locus. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon, IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Data show that promoter activity at tsEPOD loci is exceptionally weak, however expression at the *yafT* and *pitB* loci is strongest.

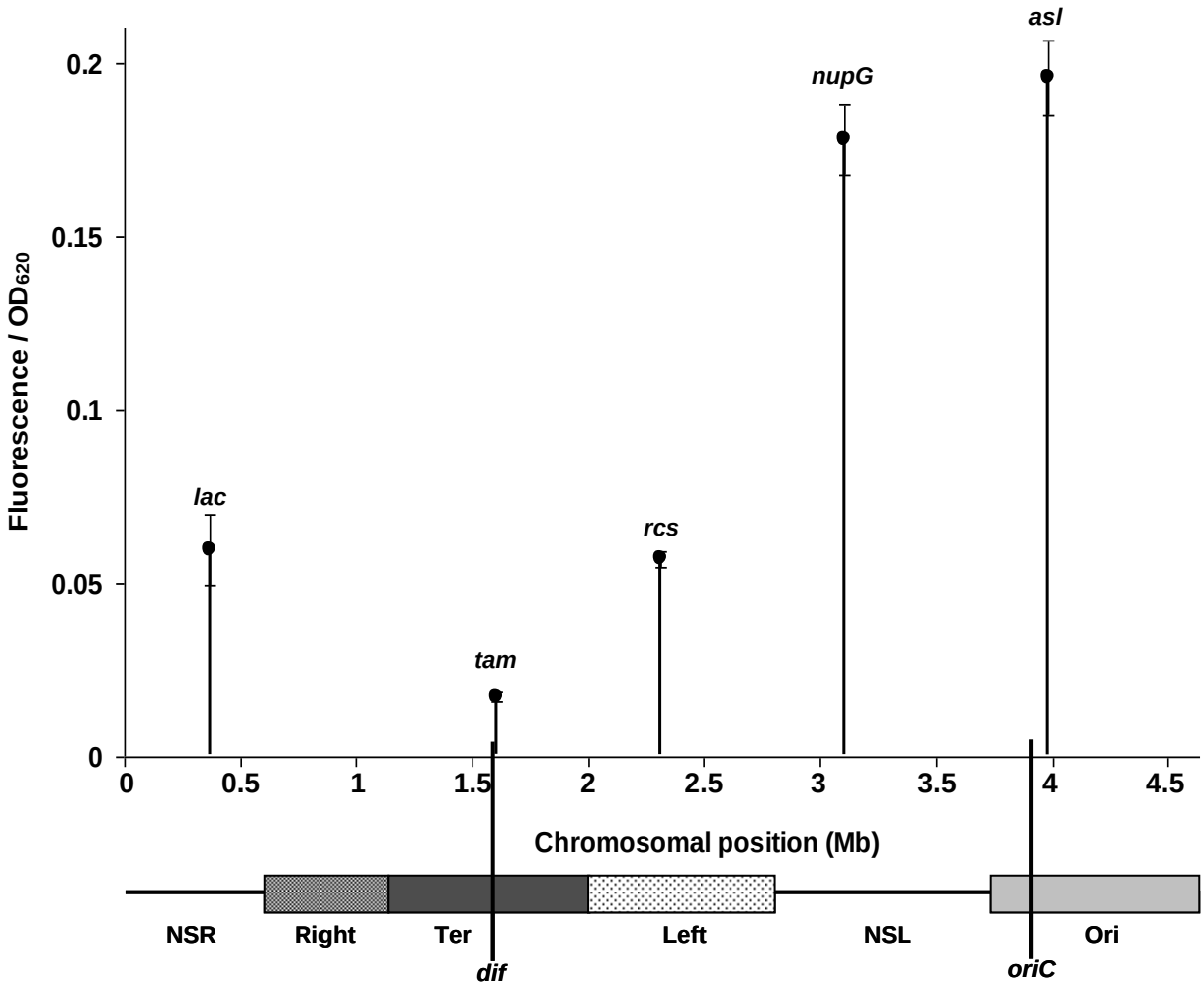

Figure 3.10: Chromosomal loci with increasing promoter activity with proximity to *oriC*

Expression of the *lac28* promoter::*egfp* fusion during mid-exponential growth (OD_{620} ~ 0.4) at 37 ˚C in M9 minimal medium supplemented with 100 μM IPTG, the inducer of the *lac28* promoter. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the sample and is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Wild type *E. coli* MG1655 and each of the strains in the absence of inducer were also assayed, however no fluorescence signal was detected, therefore indicating no auto-fluorescence or un-induced expression. These data are not included for the purpose of clarity. Position of each point on the x-axis indicates chromosomal position in Mb with respect to the coordinate systme origin and is to scale. The positions of the six macrodomains are represented by the grey shaded boxes with *oriC* and the *dif* indicated by vertical black lines. Data show the chromosomal positions that correlate with the pattern of increasing *egfp* expression with proximity to the origin of replication.

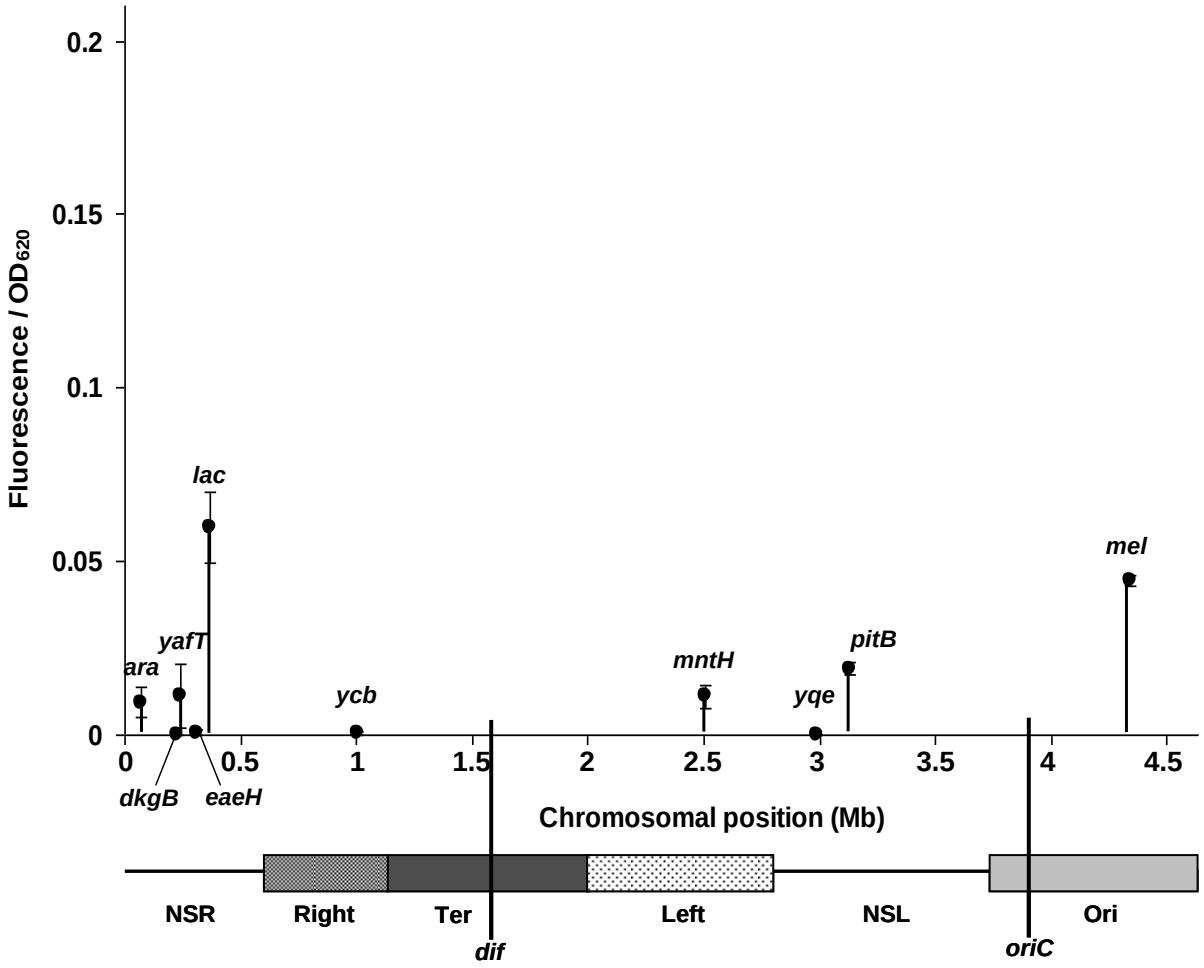

Figure 3.11: Chromosomal loci which do not show increasing promoter activity with proximity to *oriC*

Expression of the *lac28* promoter::*egfp* fusion during mid-exponential growth (OD_{620} ~ 0.4) at 37 ˚C in M9 minimal medium supplemented with 100 μM IPTG, the inducer of the *lac28* promoter. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the sample and is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Wild type *E. coli* MG1655 and each of the strains in the absence of inducer were also assayed, however no fluorescence signal was detected, therefore indicating no auto-fluorescence or un-induced expression. These data are not included for the purpose of clarity. Position of each point on the x-axis indicates chromosomal position in Mb with respect to the coordinate system origin and is to scale. The positions of the six macrodomains are represented by the grey shaded boxes with *oriC* and the *dif* indicated by vertical black lines. Data show the chromosomal positions that do not correlate with the pattern of increasing *egfp* expression with proximity to the origin of replication.

the *rcs* locus, however promoter activity at the *mntH* locus was approximately 5-fold weaker than that at the *rcs* locus (Figure 3.11). Further to this, the distance of the *nupG* locus from the origin of replication is similar to that of positions tested on the right arm of replication, however promoter activity at the *nupG* locus was far greater than that of the *lac* locus. The extremely low levels of *egfp* expression reported at tsEPOD target loci also break the pattern of increasing activity with proximity to the origin of replication, due to the target positions being spread throughout the NSR, Right and NSL macrodomains (Figure 3.11).

Insertion of *egfp* at the *asl* and *tam* loci allowed a comparison between promoter activity at positions near to the origin of replication or on the opposite side of the chromosome (Figure 3.5). Activity at the *asl* locus was approximately 2.3-fold greater than at the wild type *lac* locus, whereas expression at the tam locus was approximately 30% of activity at the *lac* locus. These results revealed an 11-fold increase in promoter activity near to the origin of replication compared to that from the *tam* locus, near to the *dif* site (Figure 3.12). Activity of the promoter at the origin of replication was expected to be stronger due to gene dosage effects, as demonstrated by Beckwith *et al*. (1966), Schmid and Roth (1987) and Sousa *et al*. (1997). However, an 11-fold variation was far greater than expected from copy number differences between Ori and Ter (Bremer and Dennis, 1996; Figure 3.12). This variation was investigated further in section 3.5.5, section 3.9 and chapter 4.

3.5.5 Gene dosage effects

Previous work has demonstrated that, in fast growing bacteria, gene expression increases by up to ~ 6-fold with increasing proximity to the origin of replication (*oriC*) due to increased copies of this part of the chromosome. This effect is referred to as gene dosage and is reduced with slower growth rates (Beckwith *et al*., 1966; Schmid and Roth, 1987; Sousa *et al*., 1997;

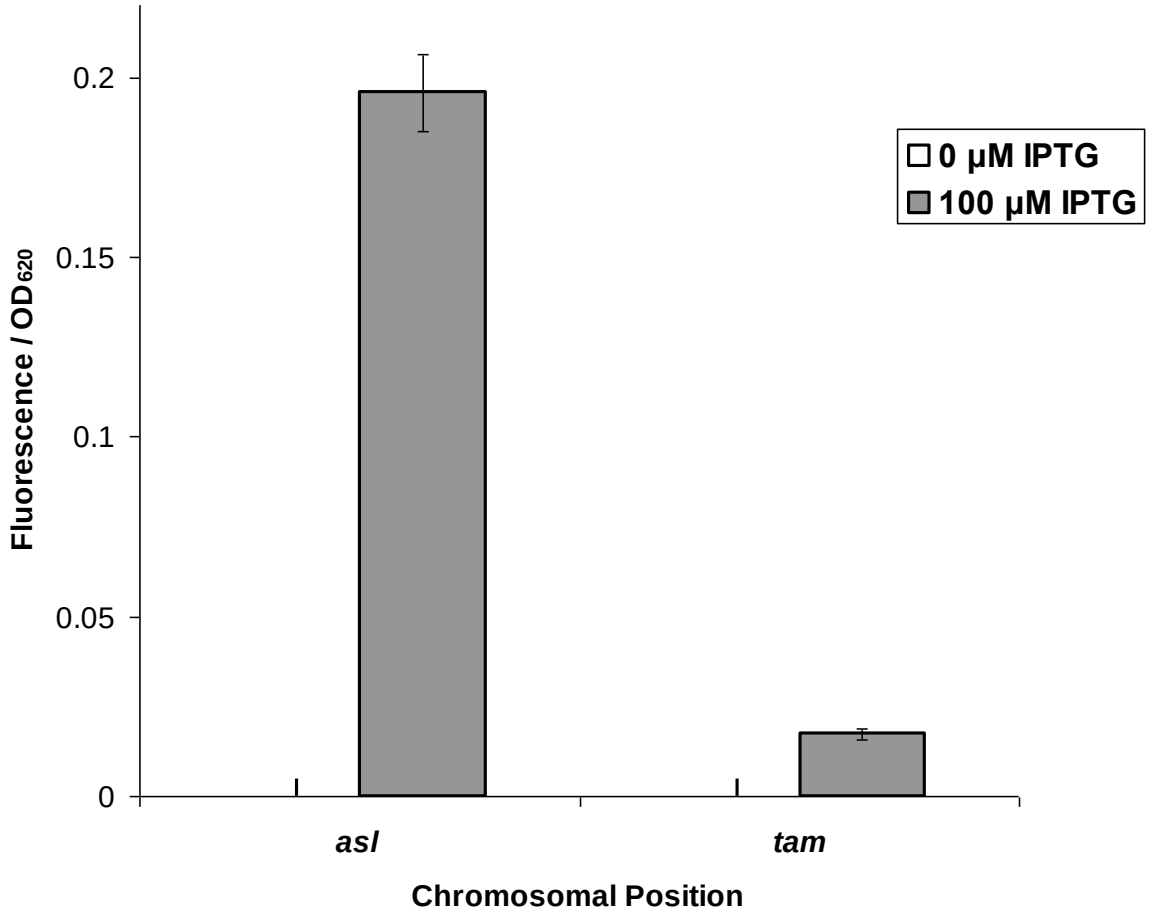

Figure 3.12: Promoter activity at loci near to *oriC* and the *dif* site

Fluorescence/OD_{620} measured in strains BRY35 (*asl*) or BRY33 (*tam*), carrying *lac28::egfp* at the designated chromosomal locus. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Data show that promoter activity at the *asl* locus, near to the origin of replication, was approximately 11-fold greater than expression at the *tam* locus, near to the *dif* site.

Block *et al*., 2012). Similar effects were reported in section 3.5.4, with the difference between expression of *egfp* near to *oriC* and the *dif* site being approximately 11-fold, during exponential growth (Figure 3.7). These data are in support of gene dosage being the contributing factor to this variation. However, the 11-fold difference between expression at the *asl* and *tam* loci was far greater than the copy number difference expected even, during the fastest growth rates (Helmstetter and Cooper, 1968; Bremer and Dennis, 1996). Therefore the gene dosage ratio of *egfp* at the *tam*, *lac*, *nupG* and *asl* loci, in the BRY33, BRY40, BRY34 and BRY35 strains, was tested in my experimental conditions by comparing the quantity of *egfp* gene genomic DNA in each of the four strains (Table 2.1; Figure 3.13).

Genomic DNA was isolated during mid-late exponential growth on M9 minimal salts medium. Quantitative real-time PCR was used to measure the copy number of the *egfp* gene in each strain by using the isolated DNA as template, primers for the *egfp* gene (Qiagen) and primers for the internal control *polA*, the location of which does not change between strains. The samples from the BRY33 strain cultures, carrying *egfp* at the *tam* locus, are taken as the calibrator in the analysis, thus the gene dose ratio of the *tam* locus is taken as 1. Each data point is an average of three separate growth experiments, for which three separate qRT-PCR reactions were completed. Errors bars represent one standard deviation from the mean (Figure 3.13).

The gene dose ratio of *egfp* at the *asl* locus, with respect to the *tam* locus, was approximately 1.4, therefore representing approximately 1.4 copies of the origin of replication for every 1 copy of the *dif* site (Figure 3.13). The results presented in figure 3.13 demonstrate that the copy number of the *egfp* gene at the *lac* locus, in the NSR macrodomain, is similar to that of the *tam* locus. However, the *nupG* locus, which is in the NSL macrodomain, has a copy number closer to that of *oriC*. This result could be due to the position on the left arm of replication being slightly over-represented compared to the similar position on the right

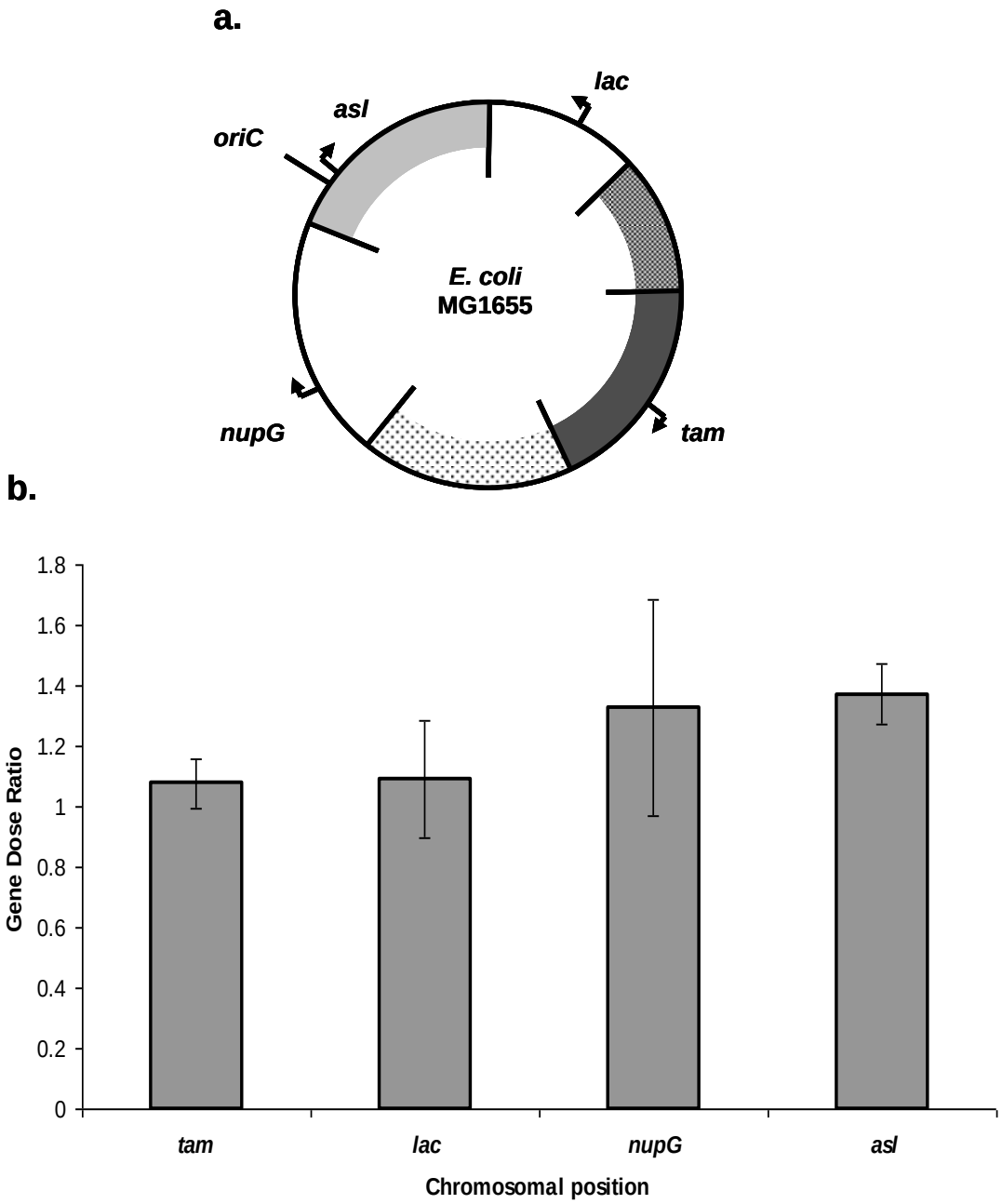

Figure 3.13: Gene dose ratio of the *tam, lac, nupG* and *asl* loci

a. Circular map of the *E. coli* MG1655 chromosome with the positions of the six macrodomains represented by grey boxes or white spaces and the position of *oriC* labelled and marked with a black line. Position and orientation of the *lac28* promoter::*egfp* fusion in strains BRY33 (*tam*), BRY34 (*nupG*), BRY 40 (*lac*) and BRY35 (*asl*) are represented by the black arrows (positions approximately to scale). **b.** Gene dose ratio of the *egfp* gene in the BRY33, BRY34, BRY40 and BRY35 strains grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium. Gene dose ratio is a representation of the relative quantity of *egfp* gene in genomic DNA samples, with the *tam* locus taken as 1, determined through the use of quantitative real-time PCR.

(Figure 3.13). The doubling time in the growth conditions used for these experiments is approximately 54 minutes during exponential growth. This data is in approximate agreement with that presented by Bremer and Dennis (1996), which states that there is approximately 1.8 copies of the origin for every 1 copy of the *ter*, during growth with a doubling time of approximately 60 minutes. Therefore these results clearly demonstrate that the 11-fold difference between the activity of the *lac28* fragment promoter, at the *tam* and *asl* loci, can not be explained by the effect of gene dosage, which in this experiment can only account for approximately 1.4-fold variations in activity.

3.6 Position-dependent repression of promoter activity by LacI

3.6.1 The *lac* operon

Transcription of the *lac* operon is driven by the σ^{70} RNAP holoenzyme and activated by CRP binding at position -61.5 bp, relative to the transcription start site, to recruit RNAP by a class I mechanism (Gaston *et al*., 1990; Busby and Ebright, 1999; Figure 1.9). Activity of the *lac* promoter is repressed by the upstream *lacI* gene product, in the absence of lactose, through binding of the repressor to the primary operator site (O^1), which is situated directly over the transcription start site. Two secondary operators, O^2 and O^3, are situated one within the *lacZ* gene and the other overlapping the 3' terminus of *lacI* gene, respectively. Interaction between Lac repressor tetramers, bound at these sites, forms "repression loops" that allow full repression of the operon (Reznikoff *et al*., 1974; Oehler *et al*., 1990; Wilson *et al*, 2007; Figure 1.9). The action of Lac repressor can be inhibited by the gratuitous inducer IPTG, which can enter the cell independently at high concentrations, but requires transport by the *lacY* gene product at low concentrations (Marbach and Bettenbrock, 2012). I have investigated the effect of changing *lac28::egfp* target chromosomal position on repression by the LacI protein.

3.6.2 The effect of target position on LacI-dependent repression

All the strains described in section 3.4 carry the *lacI* gene at the wild type locus under control of its natural promoter, whereas the position of the repressible target *lac28* fragment was at different locations. Therefore, the effect of varying the target position on repression by LacI was tested by measuring fluorescence of strains carrying the *lac28::egfp* fusion at the *mel, ara, ycb, tam, rcs, nupG* and *asl* positions during mid-late exponential growth on M9 minimal salts medium supplemented with varying concentrations of IPTG. The strains carrying the fusion at the tsEPOD loci were excluded from this experiment due to the low or undetectable levels of promoter activity at these loci during exponential growth. The BRY40 strain, carrying *egfp* at the *lac* locus, was also excluded from this experiment due to disruption of the endogenous *lacZYA* operon by the *lac28::egfp* insertion.

The *lac28* fragment promoter was fully induced by the addition of 40 μM IPTG to the growth medium when positioned at all the loci tested, except for at the *ycb* locus. Activity of the promoter at the *ycb* locus was only induced to 40% of maximum expression with 40 μM and required 70 μM IPTG for full induction (Figure 3.14; Figure 3.15). Unlike at all other positions, activity of the promoter at the *ycb* locus was not detectable at concentrations lower than 40 μM IPTG (Figure 3.14). However, expression at the *ycb* locus is very low when fully induced, therefore this may be explain the lack of detectable activity at low concentrations. While the *lac28* fragment promoter was fully induced at all other loci with 40 μM, induction of the promoter was not the same at all loci in the presence of 10 μM IPTG (Figure 3.14; Figure 3.15; Figure 3.16). The positions tested and percentage induction of the *lac28* fragment promoter, in the presence of 10 μM IPTG, are represented on the circular map of the *E. coli* chromosome in figure 3.16. Activity of the *lac28* fragment promoter demonstrated the tightest repression at the *ycb* and *ara* loci, which are positioned closest to the *lacI* gene (Figure 3.16). Hence, growth in the presence of 10 μM IPTG induced promoter activity at the *ara* locus to

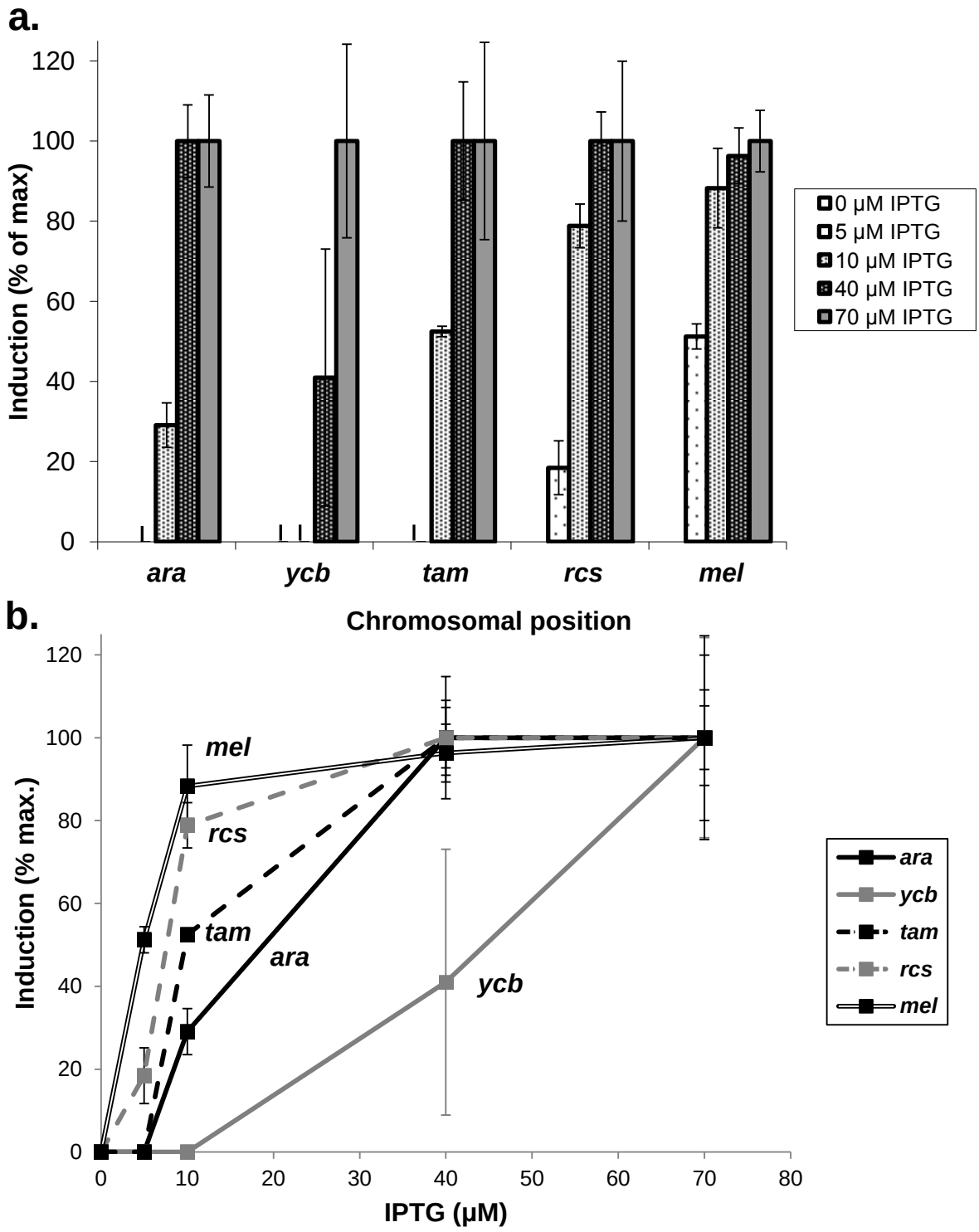

Figure 3.14: Induction of *lac28* promoter activity at different chromosomal loci

Fluorescence/OD_{620} measured in strains BRY13 (*ara*), BRY32 (*ycb*), BRY33 (*tam*), BRY27 (*rcs*), or BRY15 (*mel*), in which the *lac28::egfp* fusion is carried at the designated position. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium and the inducer of the *lac* operon, IPTG. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments and means are normalised to the maximum level of expression for each locus. Error bars represent one standard deviation from the mean. The results demonstrate that repression of *egfp* expression is not equal at all loci.

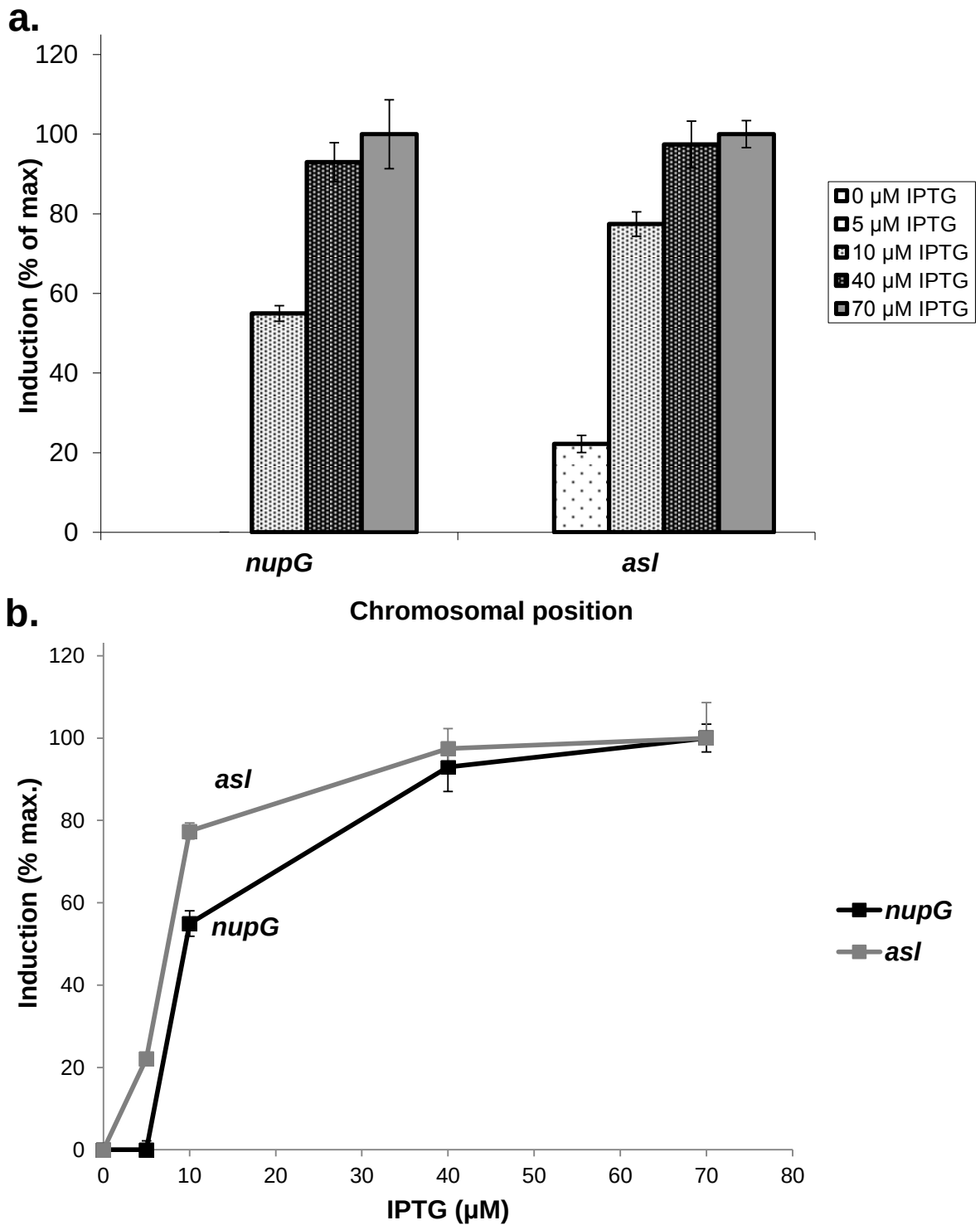

Figure 3.15: Induction of *lac28* promoter activity at the *nupG* and *asl* chromosomal loci

Fluorescence/OD_{620} measured in strains BRY34 (*nupG*) and BRY35 (*asl*), in which the *lac28::egfp* fusion is carried at the designated position. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon, IPTG. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments and means are normalised to the maximum level of expression for each locus. Error bars represent one standard deviation from the mean. The results demonstrate that repression of *egfp* expression is greater at the *nupG* locus than at the *asl* position.

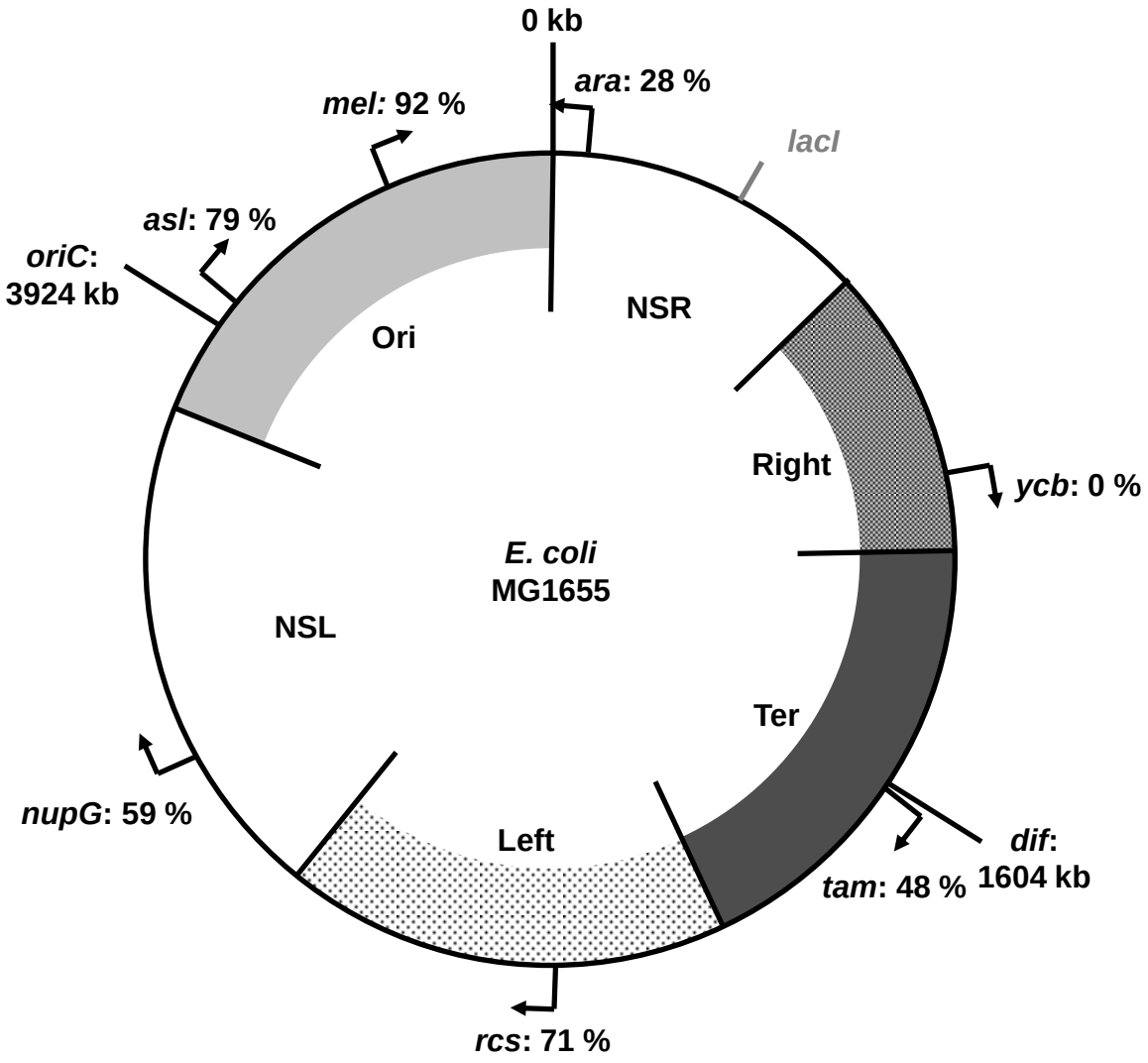

Figure 3.16: Variation in *lac* promoter level of induction with chromosomal position

Induction of the *lac* promoter in the presence of 10 μM IPTG when positioned at different chromosomal loci. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon, IPTG. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments and means are normalised to the maximum level of expression for the promoter at each locus. The black circle represents the *E. coli* MG1655 circular chromosome and the grey shaded blocks and two white gaps represent the six macrodomains, as defined by Valens *et al*. (2004). The position of *oriC* and the *dif*, the direct opposite locus are marked with black lines and labelled with their positions in kb relative to the co-ordinate system origin. Arrows indicate the position of sites of insertion for the *lac28::egfp* construct and the direction of the arrows indicates the orientation of the promoter. The *lacI* gene is at the wild type position in all strains and is marked by the grey line. Results demonstrate that repression of *lac* promoter activity is affected by position of the promoter within the chromosome.

approximately 30% of the maximum achievable activity at that position (Figure 3.14; Figure 3.16). These results show that repression of the *lac28* fragment promoter is not equal at all loci.

Repression of the promoter was weakest at the *mel* locus, with activity reaching approximately 50% of maximum in the presence of 5 μM IPTG, whereas, the addition of 10 μM IPTG to the growth medium achieved greater than 90% induction of the promoter at the *mel* locus (Figure 3.15). The promoter at the *asl* locus, like the *mel* locus, is positioned within the Ori macrodomain and is also poorly repressed by LacI (Figure 3.15; Figure 3.16). Growth in the presence of 5 μM IPTG induced activity of the promoter at the *asl* locus by approximately 20%, whereas 10 μM IPTG induced activity of the promoter to approximately 79% of the maximum (Figure 3.15). These results demonstrate that some positions within the genome are less conducive to repression by LacI and suggest that access to the Ori macrodomain by the Lac repressor could be limited.

Fully induced promoter activity at the *nupG* and *asl* positions was similar, therefore allowing comparison between induction of the promoter when maximum activity is similar. However, regardless of the fact that maximum activity is similar, activity at these two loci in the presence of low IPTG concentrations differed. Growth in the presence of 5 μM IPTG induced activity of the promoter at the *asl* locus by approximately 20%, whereas expression at the *nupG* locus was undetectable (Figure 3.15). Expression of the reporter gene was also induced to a greater level at the *asl* locus in the presence of 10 μM IPTG, giving approximately 75% of maximum activity, whereas promoter activity at the *nupG* locus reached approximately 55% of its maximum (Figure 3.15). This result suggests that the repressive action of the LacI protein is greater at the *nupG* locus, which is located in the NSL macrodomain, than at the *asl* locus, which is located in the Ori macrodomain, regardless of the fact that maximum promoter activity is similar at the two loci. These results all suggest that location of the target promoter

within the cell can affect the efficiency of repression by the LacI repressor protein, which is in disagreement with the results obtained by Block *et al*. (2012). This differential repression could be due to unique features of the specific loci targeted, such as variations in superhelicity, however this would require further investigation.

3.6.3 The effect of Lac permease on induction of the *lac28* fragment promoter

Insertion of the *lac28::egfp* fusion at the *lac* locus, in strain BRY40, causes disruption of *lacZYA* expression due to replacement of the wild-type *lac* promoter (Figure 2.11). Consequently, expression of the *lacY* gene product, Lac permease, is depleted in the BRY40 strain, therefore potentially having an effect on IPTG transport (Marbach and Bettenbrock, 2012). Hence, this strain was not used in the experiments described in section 3.6.2., as induction of the *lac28* promoter at the *lac* locus is not comparable to the other loci tested. To test the effect of LacY depletion on induction of the *lac28* fragment promoter, activity of the promoter at the *rcs* locus was assayed with and without *lacY* expression and compared to that at the *lac* locus. To do this, the *lacZYA* regulatory region in strain BRY27 was replaced with the pKH3 empty cloning site, using the gene gorging method described in section 2.12, to stop expression of the *lacY* gene. The resulting $LacY^-$ strain, carrying the *lac28::egfp* fusion at the *rcs* locus, is referred to as BRY44.

Induction of *lac28::egfp* activity was assayed by measuring fluorescence in the presence of varying concentrations of IPTG during mid-late exponential growth of the *lac* (BRY40), *rcs* (BRY27) and *rcsΔlacY* (BRY44) strains. Induction of the *lac28* fragment promoter, at the *lac* locus, required addition of at least 40 μM IPTG to the growth medium and required 100 μM IPTG for full induction (Figure 3.17). In contrast, activity of the promoter at the *rcs* locus, in the presence of LacY, was induced at 5 μM IPTG and reached maximum induction in the

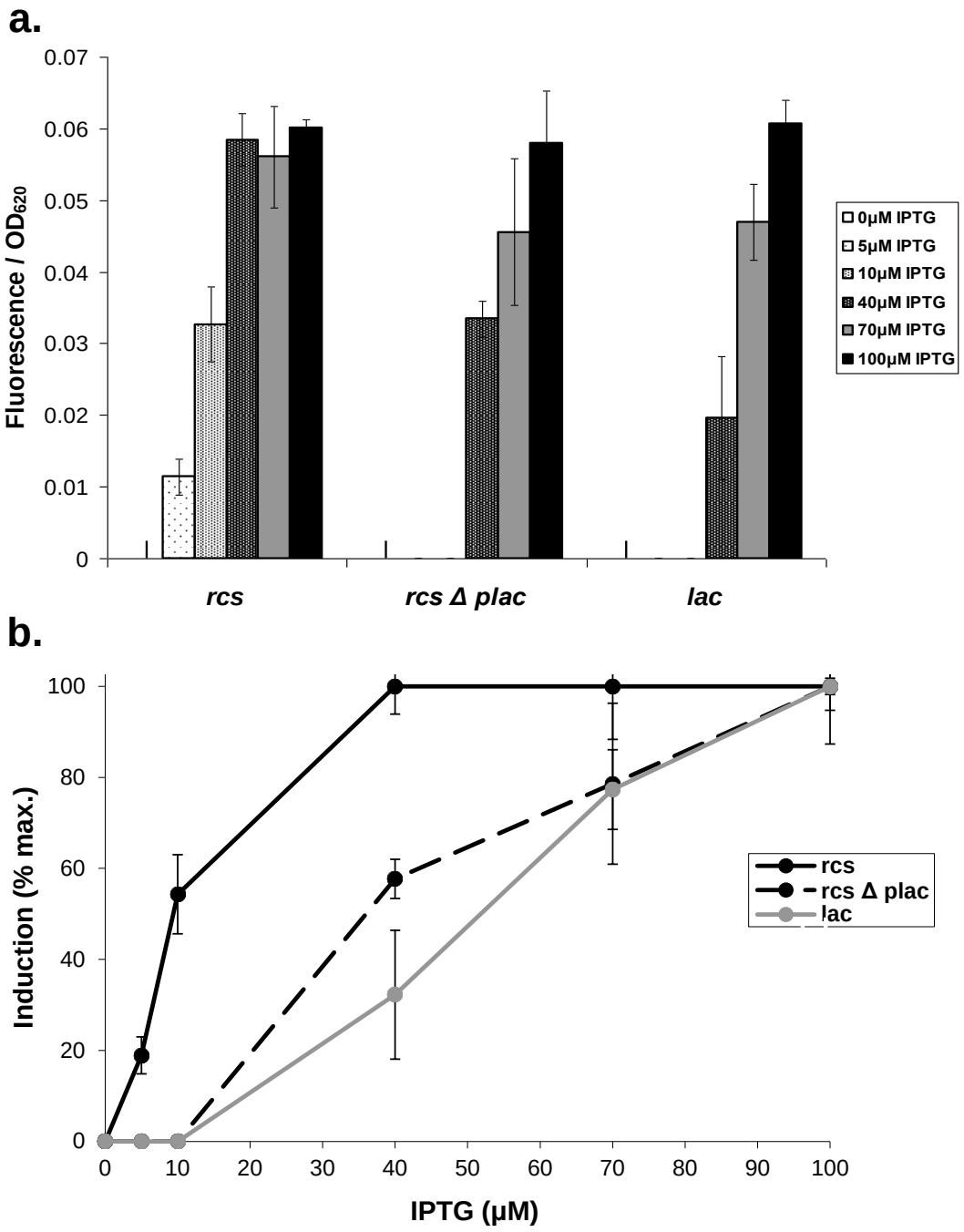

Figure 3.17: Effect of LacY on *lac28* fragment promoter induction at the *rcs* locus

Fluorescence/OD_{620} measured in strains BRY27 (*rcs*), BRY44 (*rcs* Δ *lacY*) or BRY40 (*lac*), in which BRY27 and BRY44 carry the *lac28* promoter::*egfp* at the *rcs* locus and the BRY40 strain at the *lac* locus. The wild type *lac* promoter has been replaced with the pKH3 cloning site by the gene gorging method, in strain BRY44, in order to knock out *lacZYA* expression. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with IPTG. Data shown are averages of fluorescence / OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. The results demonstrate that the Δ *plac* mutation causes induction of the promoter at the *rcs* locus to behave as that at the *lac* locus, however repression was still greater at the *lac* locus.

presence of 40 μM IPTG (Figure 3.17). These results demonstrate that repression of the *lac28* promoter is relieved by lower concentrations of inducer when positioned at the *rcs* locus compared to that at the *lac* locus. Deletion of the *lacZYA* regulatory region abolished induction of the promoter at the *rcs* locus at low IPTG concentrations and caused the promoter to behave in a similar way to that at the *lac* locus. Induction of *lac28* fragment promoter activity, at the *rcs* locus in the absence of LacY, required at least 40 μM IPTG and full induction was achieved by the presence of 100 μM IPTG, as at the *lac* locus (Figure 3.17). These data suggest that IPTG is transported into the cell at lower concentrations by the *lacY* gene product, however depleting *lacY* expression has no effect in the presence of higher IPTG concentrations.

While the pattern of induction was similar at the *rcs* and *lac* loci in the absece of Lac permease, repression of *egfp* expression was stronger at the *lac* locus, which is immediately downstream of the *lacI* gene (Figure 3.17). Addition of 40 μM IPTG induced the activity of the *lac28* fragment promoter, at the *rcs* (*ΔlacY*) locus, to approximately 58% of its maximum, whereas the promoter at the *lac* locus only reached 32% of maximum (Figure 3.17). These results demonstrate that repression of the *lac* promoter, by LacI, is tighter at the *lac* locus, which is immediately downstream of the *lacI* gene, than at the *rcs* locus. This provides further evidence to suggest that the chromosomal position of a target promoter can affect LacI-dependent repression (Figure 3.17).

3.6.4 The effect of *lacI* O^2 on *egfp* expression at the *lac* locus

The *lac* promoter is repressed, by the *lacI* gene product, through steric hindrance and looping between Lac repressor dimers bound at 2 of three possible operator sites (Reznikoff *et al*., 1974; Oehler *et al*., 1990; Wilson *et al*., 2007: Figure 1.9). The *lac* operator O^2 is located

401 bp downstream of the primary operator O^1 and was not included in the *lac28* fragment, in order to keep the size of the fragment to a minimum (Figure 3.2b). Creation of the BRY40 strain, carrying *egfp* at the *lac* locus, used a homology region cloned from position 399 bp to 898 bp, relative to the *lacZ* start codon, therefore *lacI* O^2 was not present on the chromosome of this strain (Figure 3.18a).

In a complementary experiments, the long-range effects of *lacI* O^2 on *lac28::egfp* activity were tested by insertion of the *lac28::egfp* fusion at the *lac* locus while maintaining the O^2 operator on the chromosome. This was achieved by using a homology region from position 1 bp to 611 bp of the *lacZ* gene, therefore maintaining *lacI* O^2 on the chromosome (Figure 3.18b). The *lac* operator O^2 on the chromosome of the resulting strain, referred to as BRY12, is centred 1245 bp downstream from the O^1 binding site of the *lac28* fragment, due to insertion of the *lac28::egfp* fusion. Note that this differs from the situation at the endogenous *lac* promoter in *E. coli* K-12 MG1655, where *lac* O^1 and O^2 are 401 bp apart.

The effect of O^2 on repression of the *lac28* fragment promoter, at the *lac* locus, was assayed by comparing fluorescence of the BRY12 (*lac* $+O^2$) and BRY40 (*lac*) strains during mid-late exponential growth on M9 minimal salts, supplemented with varying concentrations of the inducer IPTG. The presence of O^2 on the chromosome causes the fluorescence to be decreased, therefore indicating an increased level of repression (Figure 3.19). The absence of O^2 increased expression by approximately 40% at the two higher concentrations of 70 μM and 100 μM IPTG. However the threshold of inducer required to give a signal appeared to be unaffected by the presence of the *lac* operator O^2, as no signal was achieved at IPTG concentrations lower than 40 μM (Figure 3.19). These data suggest that the *lac* operator O^2 binding site has an effect on the repression of *egfp* expression at the *lac* locus, however the majority of the repression remains. This data is in agreement with results presented by Oehler *et al*. (1994). Cooperation of operator sites has previously been shown to have smaller effects

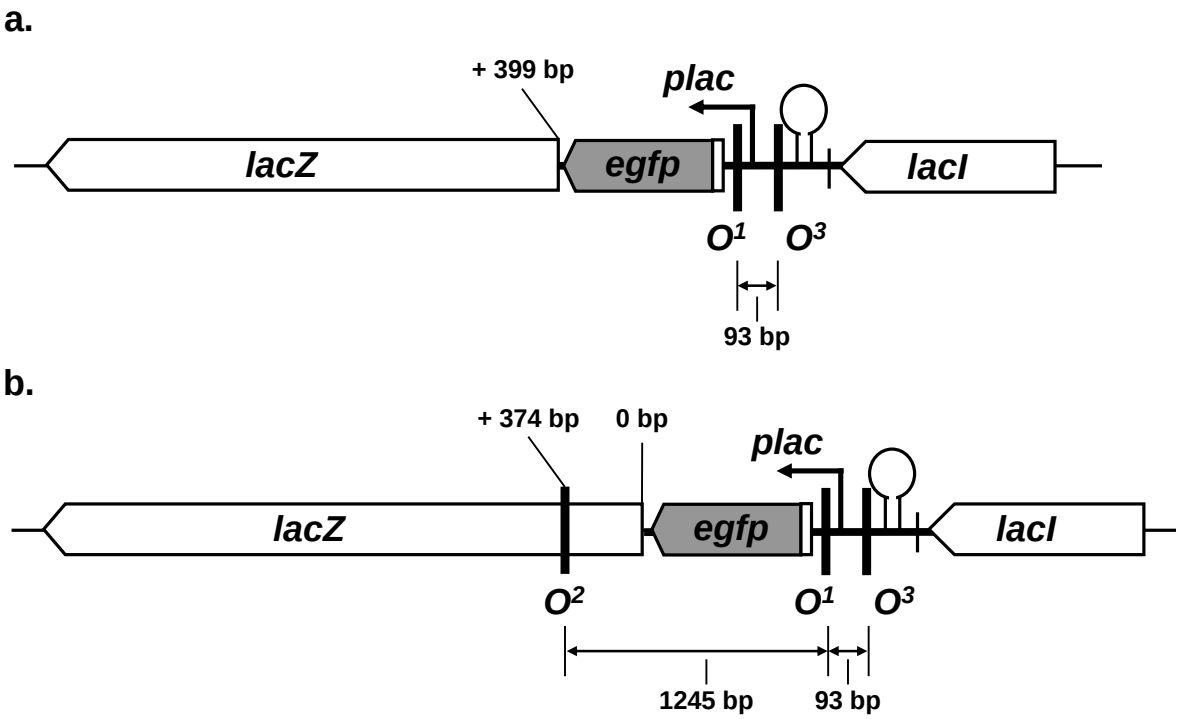

Figure 3.18: Schematic diagram of gene organisation at the *lac* locus in the BRY40 and BRY12 strains

White arrows represent genes, while black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. Direction of replication is from left to right. Thick black vertical lines represent the position of *lac* operator sites and are labelled. The arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. **a.** organisation of genes at the *lac* locus in strain BRY40 **b.** BRY12

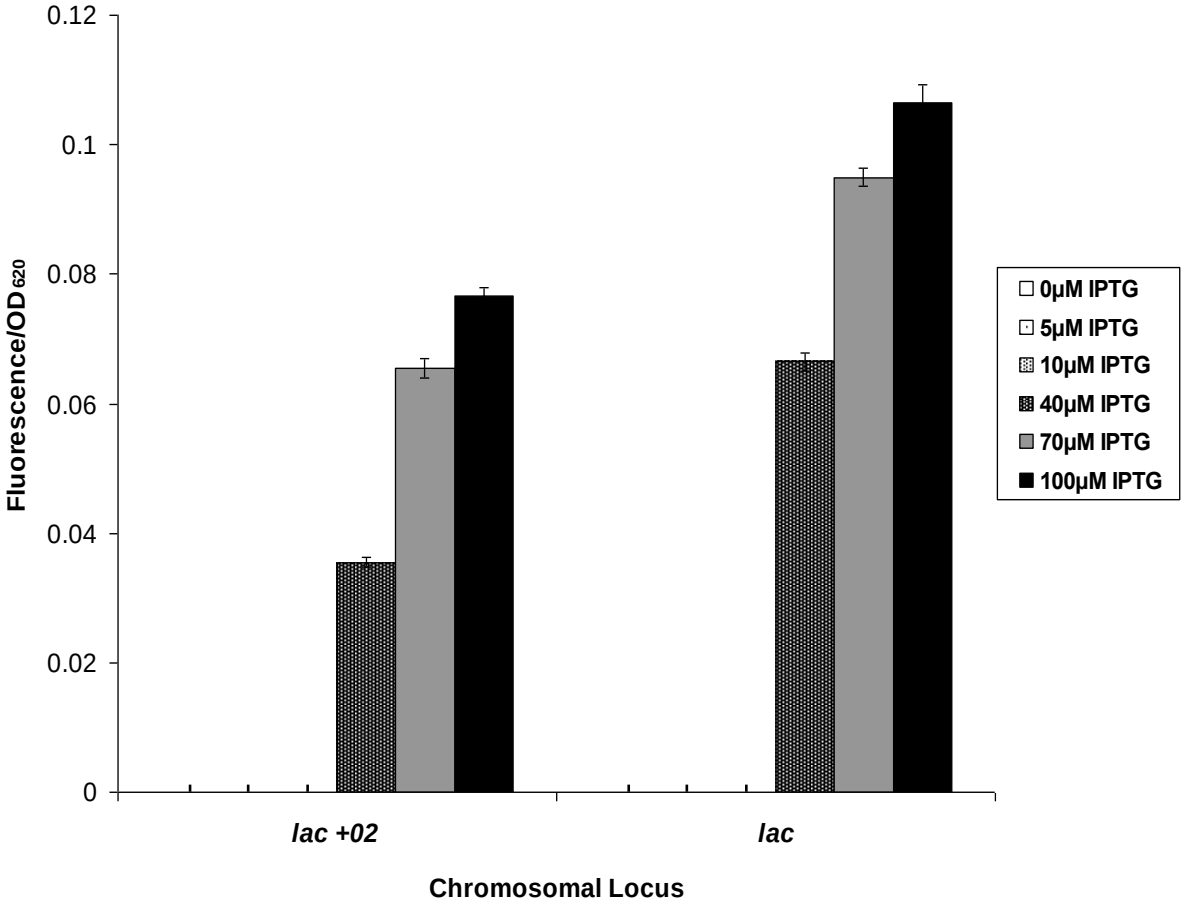

Figure 3.19: Effect of *lac* operator O^2 on *lac28* promoter activity at the *lac* locus

Fluorescence/OD_{620} measured in strains BRY12 (*lac* $+O^2$) and BRY40 (*lac*), each carrying *egfp* at the chromosomal *lac* locus. The *lac* operator O^2 is present where stated (BRY12). Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon IPTG. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean.

at longer distances, therefore the minimal effect of O^2 on the induction of *lac28* fragment promoter in strain BRY12 could be due to the extra 844 bp that separate O^1 and O^2 compared to the 401 bp in the starting strain (Oehler *et al*., 1994). These data suggest that local chromosomal context and the positioning of transcription factor binding sites in the vicinity of a promoter can have an effect on its total output.

3.7 Promoter activity during stationary phase

3.7.1 Effects of chromosome position on promoter activity

The results presented up to this point have only been concerned with expression in relatively fast-growing cultures. Therefore, the effect of position within the chromosome on promoter activity was also analysed during stationary phase, by measuring increase in fluorescence on induction in *E. coli* strains carrying the *lac28::egfp* fusion at the different chromosomal locations shown in figure 3.5. Promoter activity was measured during the stationary phase of growth on M9 minimal salts medium supplemented with 250 μM IPTG. The higher concentration of IPTG was found to be required for induction during stationary phase, potentially due to the increased cell density of the cultures compared to that of mid-late exponential growth. Overnight cultures were split in two and the inducer added to one half of the culture to allow a direct comparison between culture with and without inducer after 5 hours induction. As in section 3.5, expression of *egfp* was derived from the fluorescence at 510 nm divided by OD_{620} of at least three separate experiments to give the mean and standard deviation. The starting *E. coli* K-12 MG1655 strain was found to auto-fluoresce at the higher cell densities required by this experiment, as shown in figure 3.20, therefore activity of un-induced cultures were subtracted from that of induced cultures for each separate experiment, to give induction of reporter expression, before calculation of mean and standard

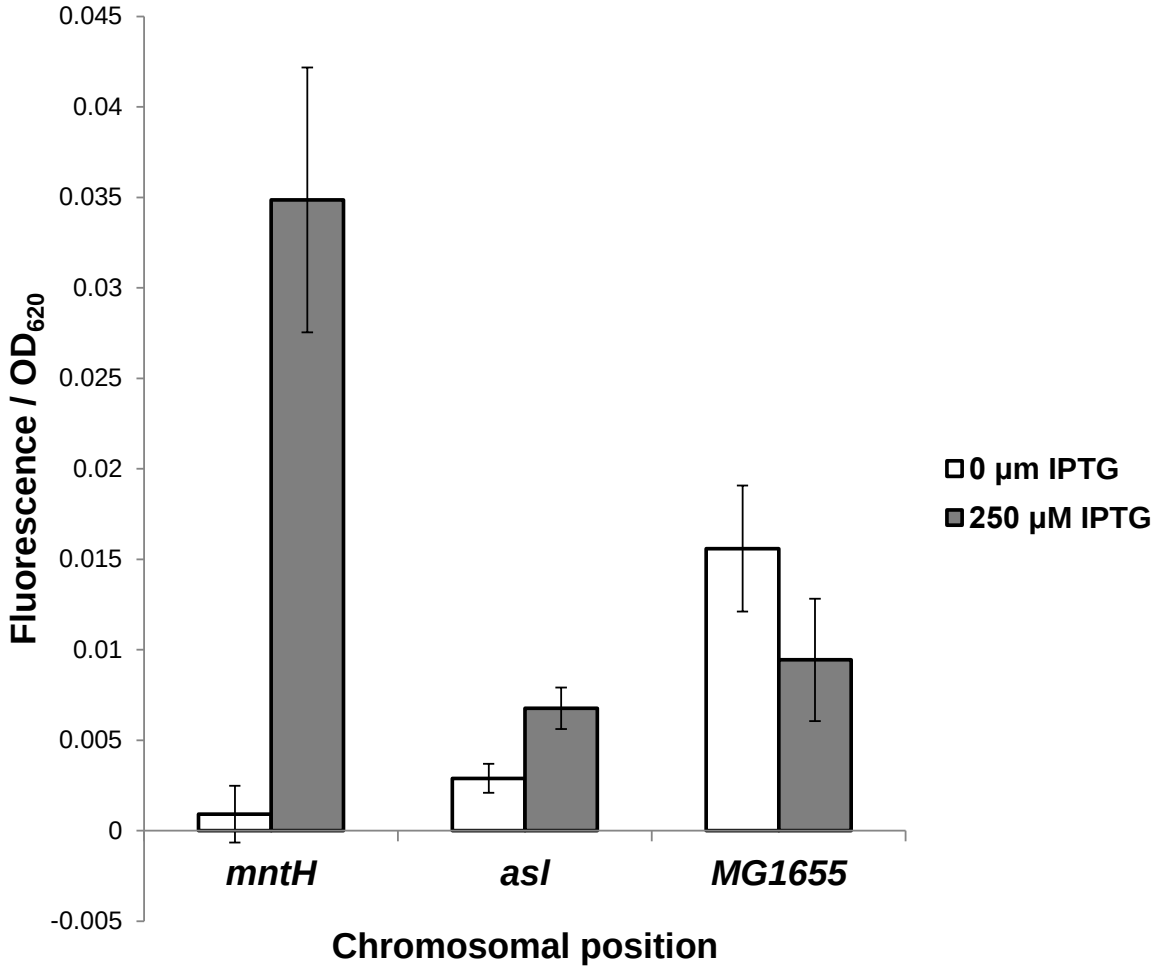

Figure 3.20: Promoter activity at the *mntH* and *asl* loci during stationary phase

Fluorescence/OD_{620} measured in strains BRY73 (*mntH*), BRY35 (*asl*), carrying *lac28::egfp* at the designated chromosomal locus, or the starting strain MG1655. Samples were taken during stationary phase (OD_{620} ~ 1.5) at 37 ˚C in M9 minimal medium supplemented with 250 µM IPTG, where stated. Overnight cultures were split into two aliquots prior to induction of one of these to allow direct comparison between induced and un-induced cultures. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Data show that fluorescence of the *mntH* and *asl* strains does vary slightly in the absence of inducer, but this is always lower than the starting MG1655 strain and is representative of all other strains used in this work. Data also show that fluorescence of the *mntH* and *asl* strains increased on addition of IPTG, however fluorescence of MG1655 decreased.

deviation. The autofluorescence of un-induced cultures did vary between strains, however this was always less than the starting MG1655 strain. Further to this, the fluorescence of strains carrying the *lac28::egfp* fusion always increased on average, whereas the fluorescence of the starting strain (MG1655) decreased on addition of IPTG (Figure 3.20). This suggests that variation of fluorescence in the un-induced culture was not due to leakage of the promoter (Figure 3.20). Subtraction of the background auto-fluorescence is the reason for the absolute level of fluorescence reported on the y-axis, of figure 3.21, being so much lower than that in figure 3.7 (exponential growth).

Expression of *egfp* during stationary phase was not only dependent upon regulation by the promoter region, but also upon position within the chromosome, as reported for exponential growth. Position-dependent modulation of promoter activity during stationary phase caused large variation of *egfp* expression by the *lac28* fragment promoter across the chromosomal positions tested (Figure 3.21). The total variation of promoter activity was much less than the 310-fold differences found during exponential growth or even the 11-fold variation between non-tsEPOD loci. Also, the level of expression at all loci was very low compared to that during exponential growth, which was most likely due to the promoter being σ^{70}-dependent (Figure 3.21; Figure 3.7).

The largest induction of fluorescence during stationary phase was at the *mntH* locus, which is positioned within the Left macrodomain. When positioned at the *mntH* locus, increase of fluorescence on inductrion was approximately 8-fold greater than that at the *tam* locus, at which the promoter activity was weakest (Figure 3.21; Figure 3.22). These results suggest that position within the *E. coli* chromosome does affect promoter activity during stationary phase, however the pattern of expression is significantly different compared to that during exponential growth and activity is generally much weaker.

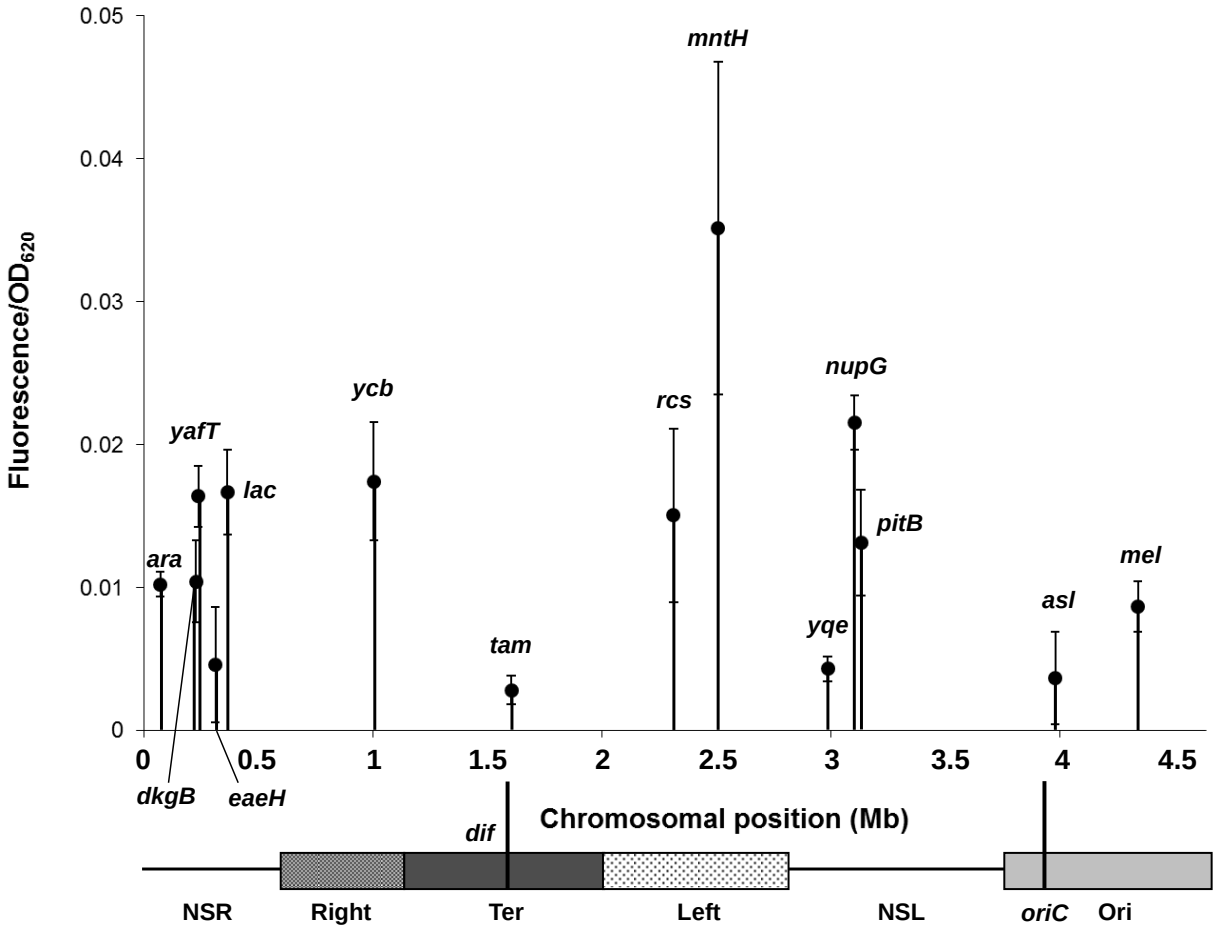

Figure 3.21: Promoter activity at different chromosomal positions during stationary phase

Expression of the *lac28*::*egfp* fusion during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 μM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the $OD_{620,}$ of the un-induced sample subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Position of each point on the x-axis indicates chromosomal position in Mb with respect to co-ordinate system origin and is to scale. The positions of the six macrodomains are represented by the grey shaded boxes with *oriC* and the *dif* indicated by vertical black lines.

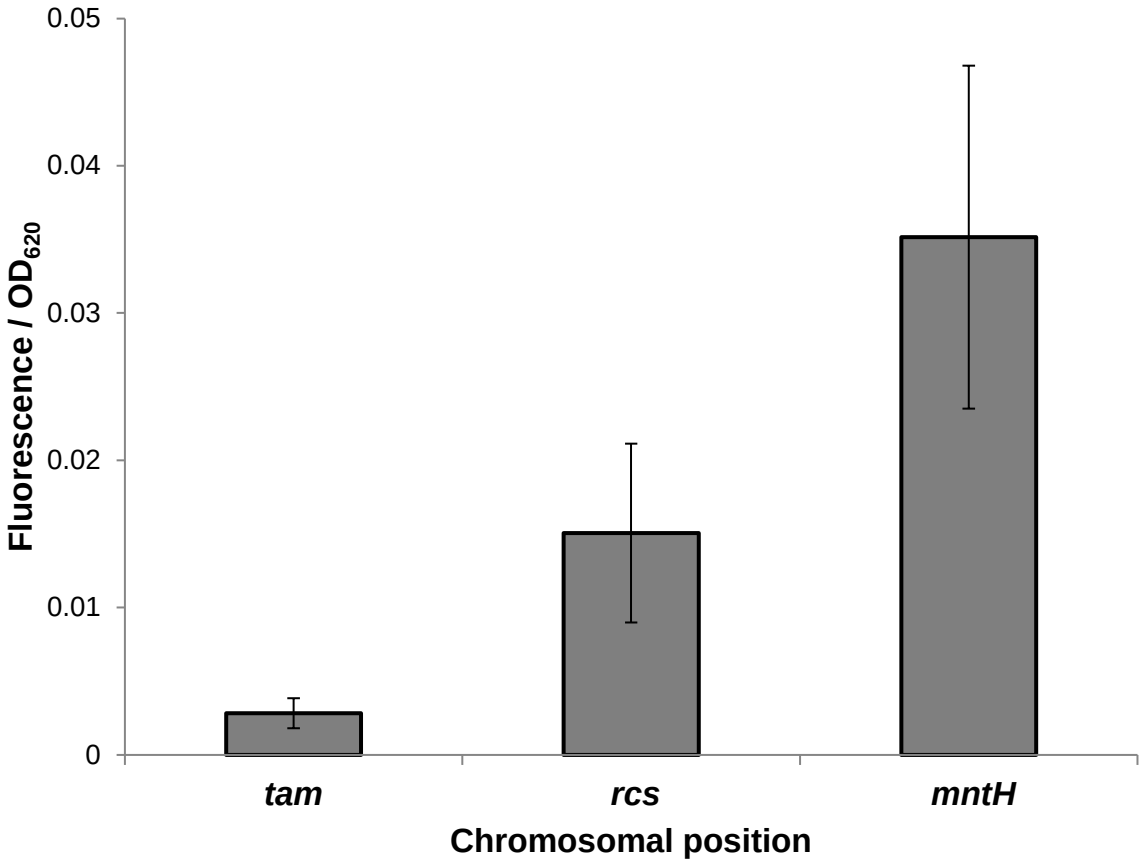

Figure 3.22: Promoter activity at the *tam, rcs* and *mntH* loci during stationary phase

Expression of *egfp* in the BRY33 (*tam*), BRY27 (*rcs*) and BRY73 (*mntH*) strains during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 μM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by $OD_{620,}$ of the un-induced sample subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Data show that promoter activity was strongest at the *mntH* locus, which was approximately 8-fold greater than the weakest activity at the *tam* locus. Expression of *egfp* at the *mntH* locus was also 2-fold stronger than that of the *rcs* locus, which also resides within the Left macrodomain

3.7.2 Promoter activity at the *asl* and *nupG* loci during stationary phase

Activity of the *lac28* fragment promoter at the *nupG* locus during stationary phase was one of the most highly expressed, giving approximately 75% of the activity reported for the *mntH* locus (Figure 3.23). Activity of the promoter at the *nupG* locus was approximately 6-fold greater than that at the *tam* locus, whereas, during exponential growth, the difference between activity at the two loci was approximately 10-fold (Figure 3.23; Figure 3.7). Therefore this result suggests that while the *nupG* locus is one of the most highly expressed loci during stationary phase, some difference in the nature of the *nupG* locus between the two growth phases may have caused a decrease in promoter activity. While expression of *egfp* at the *nupG* and *asl* loci was similar during exponential growth, increase in fluorescence on induction at the *asl* locus was significantly weaker than at the *nupG* locus during stationary phase (Figure 3.23; Figure 3.7). Induction of the *lac28* fragment promoter at the *asl* position, which is located near to the origin of replication, was similar to that at the *tam* locus, located near to the *dif* site (Figure 3.23). This suggests that the reasons for differences in *lac28::egfp* expression at the *tam* and *asl* loci during exponential phase, which caused an 11-fold variation in promoter activity between the two loci, are negated during stationary phase.

3.7.3 The effect of proximity to the origin or terminus of replication

The general pattern of expression during stationary phase was very different to that during exponential growth, in which the strongest promoter activity was seen with increasing proximity to the origin of replication (Figure 3.7; Figure 3.24). During stationary phase, increase in fluorescence on induction was smallest when positioned in the Ori and Ter macrodomains. Strength of promoter activity increased as proximity to these regions decreased. Thus, the *lac28* fragment promoter was strongest when positioned at the *mntH*

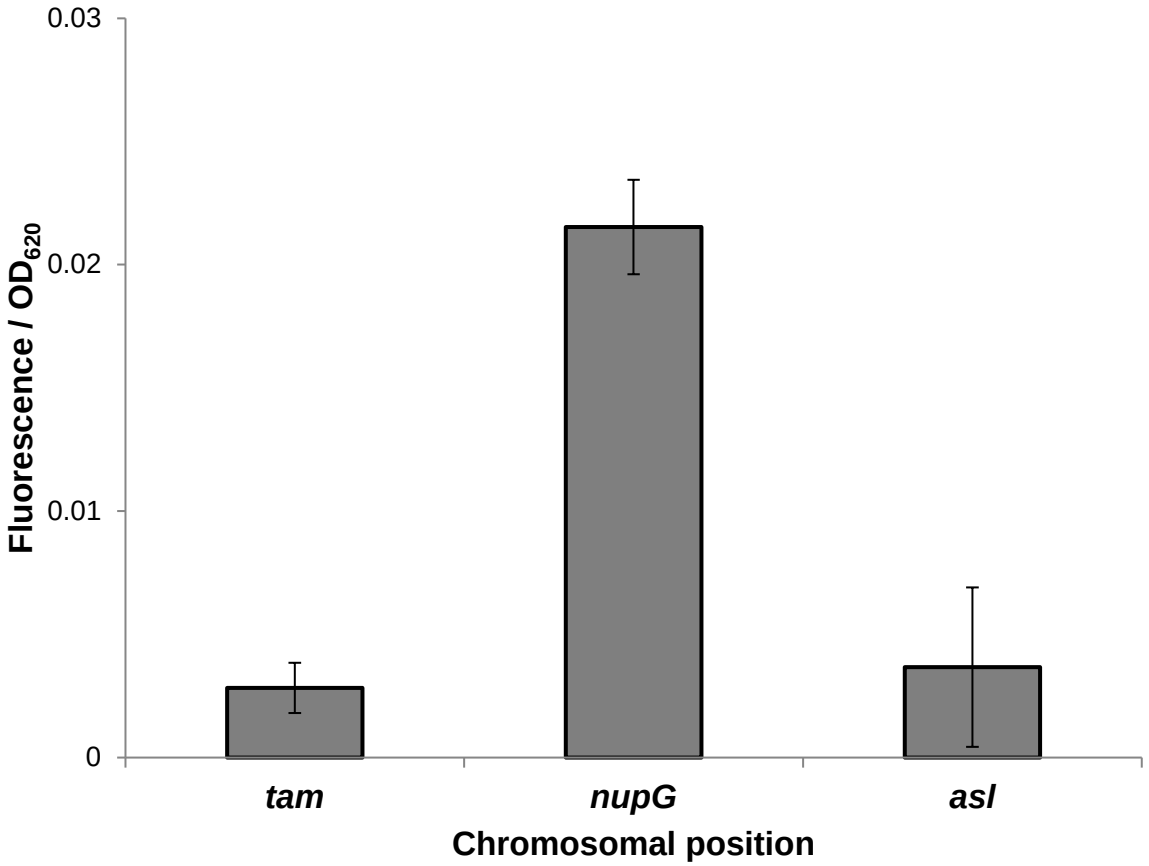

Figure 3.23: Promoter activity at the *tam*, *nupG* and *asl* loci during stationary phase

Expression of *egfp* in the BRY33 (*tam*), BRY34 (*nupG*) and BRY35 (*asl*) strains during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 µM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the un-induced sample, subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Data show that activity of the promoter at the *nupG* locus was approximately 6-fold greater than at the *tam* locus. Expression of *egfp* at the *asl* locus, which was positioned near to the origin of replication was the same as that of the *tam* locus, which was positioned near to the *dif* site.

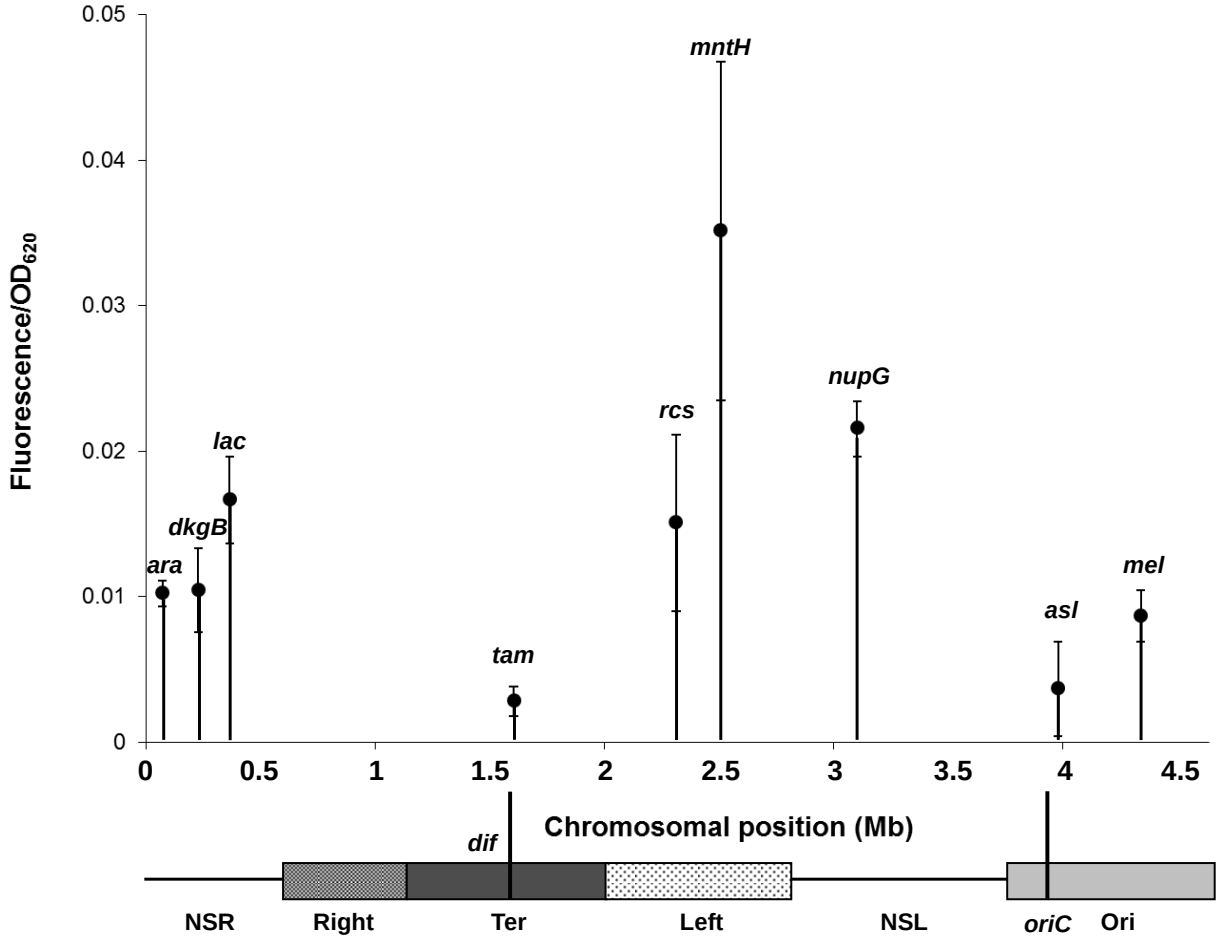

Figure 3.24: The effect of chromosomal position with respect to *oriC* and *dif* on promoter activity during stationary phase

Expression of the *lac28*::*egfp* fusion during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 μM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the un-induced sample subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. Position of each point on the x-axis indicates chromosomal position in Mb with respect to the coordinate system origin and is to scale. The positions of the six macrodomains are represented by the grey shaded boxes with *oriC* and the *dif* indicated by vertical black lines.

locus in the Left macrodomain, with the other well expressed loci being in a mid-replichore position, as opposed to being in the terminal macrodomains (Figure 3.24). This result suggests that the characteristics of the Ori macrodomain are conducive to strong *lac* promoter activity during exponential growth, but become poor for promoter activity during stationary phase.

3.8 Discussion

3.8.1 5' mRNA secondary structure improves *lac28::egfp* expression

The work described in this chapter sought to analyse the effects of position within the *E. coli* chromosome on promoter activity. This was achieved through use of the improved recombineering system, gene doctoring, and detailed knowledge of the sequence and annotation of the *E. coli* genome, allowing targeting of specific loci within the chromosome. Design of the *lac* promoter::*egfp* fusion revealed that direct fusion of the promoter region and ATG start codon to *egfp,* via a *Hin*dIII target site, produced very weak expression of the reporter gene (Figure 3.2). Data collected by Kudla *et al*. (2009) demonstrated that addition of a 28 codon leader sequence, with weak secondary mRNA structure, to a promoter::GFP fusion increased the fluorescent output of cells carrying the fusion. This result was shown to be due to improved translational efficiency. Hence, the first 28 codons of the *lacZ* gene were included in the *lac28* promoter::*egfp* fusion, improving the expression of *egfp* approximately 90-fold (Kudla *et al*., 2009; Figure 3.2; Figure 3.3).

Scharff *et al*. (2011) found that mRNAs lacking a Shine-Dalgarno sequence also lack mRNA secondary structure at or around the start codon. The lack of mRNA secondary structure was found to promote faithful translation initiation at the correct AUG in the absence of a Shine-Dalgarno sequence. Recent work has also suggested that evolution of weak mRNA secondary structure near the start codon is selected only in functional genes, whereas this pressure is not

observed in pseudogenes, which are not expressed, or poorly expressed genes of unknown function (Keller *et al*., 2012). The evidence presented by Keller *et al*. (2012) and Scharff *et al*. (2011) suggests that improved expression of the *lac28::egfp* fusion was most likely caused by changes in mRNA secondary structure near to the AUG start codon, therefore improving translational efficiency.

3.8.2 Experimental approach improves sensitivity to position-dependent effects

The large effects of chromosome position on promoter activity reported here are in disagreement with the results of previous studies and this could be due to differences in experimental approaches (Beckwith *et al*., 1966; Chandler and Pritchard, 1975; Schmid and Roth, 1987; Sousa *et al*., 1997; Block *et al*., 2012). As highlighted by Schmid and Roth, (1987) the experimental approaches of previous studies were similar in that they utilised a transposition vehicle of some variety to move the promoter::reporter probe to different loci. Beckwith *et al*. (1966) utilised a temperature sensitive $F_{TS}lac^{+}$ episome to insert the *lacZYA* operon at various chromosomal locations within *E. coli* to create 11 different *Hfr* strains. This also means that the *lac* promoter was contained within the large $F_{TS}lac^{+}$ episome, therefore separating the promoter from the genomic DNA (Beckwith *et al.*, 1966). The work of Schmid and Roth, (1987) utilised two *Tn10* insertion sequences flanking the *hisGDBCH'* operon, which is approximately 5 kb. Therefore the promoter could have been separated from local sequence by the ~ 5 kb *hisGDBCH'* genes and Tn*10* sequence on the downstream side or the Tn*10* insertion sequence upstream. As illustrated by Schmid and Roth, the large amount of flanking sequence could serve to shield, or insulate, the promoter from the effects of local genetic elements. This is also true of the probe used by Sousa *et al*. (1997) in which the *Pseudomonas putida sal* promoter is flanked by a 19-bp Tn*5* insertion terminus, an ~1.4 kb kanamycin resistance determinant and the ~800 bp *nahR* regulator upstream, and the ~3 kb *lacZ* reporter gene downstream.

The *lac28::egfp* probe used in this study was designed to keep the insert size to a minimum, therefore the kanamycin resistance gene was removed from the chromosome by FLP recombinase. This left only 157 bp of foreign DNA upstream of the promoter fragment, which included the single FRT site and the λ *oop* terminator to prevent read-through from neighbouring operons. Also, only 875 bp of foreign DNA was incorporated downstream of the transcription start site, which includes the *egfp* coding region (Figure 2.6). The minimal size of the *lac28::egfp* construct, compared to the promoter::reporter probes used in other studies, should leave the promoter better exposed to the local genetic context and leave the local chromosomal architecture relatively undisturbed. These differences in strain construction could explain why larger variation in gene expression levels were found in this work and different conclusions were made.

3.8.3 Insertion target site may improve sensitivity to position-dependent effects

Another major variation between this work and that of the previous studies is the target site of insertion. The work of Sousa *et al*. (1997) does not define the exact locus of insertion for the *psal*::*lacZ* insertion, however Beckwith *et al.* (1966) utilised the $F_{TS}lac^{+}$ episome to insert the *lacZYA* operon into the chromosome. Use of the episome leads to insertion within gene coding regions, due to integration of the episome being non-random and targeted to disrupt specific genes and phage integration sites (Beckwith *et al.*, 1966; Scaife, 1967). Schmid and Roth, (1987) utilised the Tn*10* transposon to create gene knockouts, which could then be targeted for homologous recombination by the Tn*10* flanked *hispGDCB* operons, therefore being inserted into the non-functional gene coding region. The work of Block *et al*. (2012) also inserted a promoter::reporter fusion within gene coding regions, however this was through choice as opposed to limitations of cloning technology. The promoter was inserted in the opposite orientation to that of the target gene regulatory element. In contrast, the work

presented in this chapter targeted non-coding regions between convergent genes, upstream of promoter regions, or, in the case of the *eaeH* locus, within pseudogenes. Intergenic regions between convergent genes were targeted because this would allow minimal disruption of the neighbouring genomic processes, such as transcription, translation and the local chromosomal architecture. The targeting of gene coding regions would cause premature termination of local transcription, therefore potentially having an effect on expression of the promoter::reporter probe through changes in local chromosomal architecture and structure, or cell physiology. The other difference was in the selection of the tsEPODs as target sites, which were demonstrated to be uniquely poor regions for promoter activity, potentially due to the large quantities of protein bound to the region. The silencing of transcription within tsEPODs was investigated further in section 4.5 of chapter 4.

3.8.4 Proximity to the origin of replication improves promoter activity through a mechanism other than gene dosage

Analysis of *egfp* expression during exponential growth, revealed that promoter activity varies by approximately 310-fold, dependent upon position of the promoter within the chromosome. Measurement of *egfp* mRNA levels showed a correlation with eGFP fluorescence, measured from three loci, suggesting that the variation was not due to changes in regulation of translation. The strongest promoter activities were found at the *asl* and *nupG* loci, which are in the Ori and NSL macrodomains, respectively. The *lac28::egfp* fusion at the *asl* locus resides between the *aslB* and *aslA* genes, the former of which is regulated by a σ^{54} promoter with a large σ^{70} common antigen biosynthesis pathway operon and some tRNA genes immediately upstream of the *aslB* gene (Figure 2.14). The generally low level of promoter activity in the neighbouring area could mean less competition for σ^{70} RNAP, therefore possibly explaining the elevated levels of expression during exponential growth. However, the

poorly expressed *tam* locus also had two stationary phase transcription units upstream of the *lac28* fragment promoter along with many operons dependent on alternative σ subunits in the vicinity. Therefore, if neighbouring gene transcription is a factor in position-dependent modulation of promoter activity, then it appears not to be a factor at these loci. The proximity of these two highly expressed loci to the origin of replication was the most apparent common feature.

The general pattern of increasing strength in activity, with proximity to the origin of replication, was found in the dataset presented in figure 3.10. and this type of pattern has been previously suggested to be due to gene dosage (Beckwith *et al*., 1966; Chandler and Pritchard, 1975; Schmid and Roth, 1987; Sousa *et al*., 1997; Block *et al*., 2012). The difference between promoter activity near *oriC* compared to the *dif* site was approximately 11-fold, therefore greatly exceeding the maximum 3-fold variation reported by Schmid and Roth (1987), the ~ 6-fold variation reported by Block *et al.* (2012) or the ~ 1.8 fold difference in gene copy number at the *ori* compared to the *ter* presented by Bremer and Dennis (1996).

Growth of *E. coli* in this study used M9 minimal salts medium supplemented with 0.3% fructose as the carbon source, which yields a slow doubling time of ~ 60 minutes and is much slower than that used by the previously mentioned studies. Hence, the gene dose ratio between the origin of replication and the terminus was expected to be small and this was confirmed, through the use of quantitative real-time PCR of genomic DNA, to be approximately 1.4 to 1.5 (Figure 3.13). This is in agreement with that presented by Bremer and Dennis (1996). Hence, the results presented in this chapter show that transcriptional activity, during exponential growth, does increase with proximity to the origin of replication. However, in contrast to previous work, this relationship is not solely dependent upon gene dosage, and alternative mechanisms are discussed in chapter 4.

3.8.5 Anomalies to the expected pattern of increasing activity with proximity to *oriC*

The results presented in section 3.6 also demonstrated that, while there was a general relationship between proximity of the promoter to the origin of replication and strength of activity, not all loci tested conformed to this general pattern. Previous studies have also detected "anomalies" from the pattern of gene expression expected from the gene dosage effect, such as the two presented by Beckwith *et al*. (1966). One of these "anomalies" was the EC-7 *Hfr* strain, which has the $F_{TS}lac^{+}$ episome inserted between *argB* and *thr*, and is one of the closest sites to *oriC*. The other anomaly was the EC-28 strain containing the $F_{TS}lac^{+}$ episome at the most distal position from *oriC*. The expression of *lacZ* from the EC-7 strain was shown to be ~ 80% of that from a similar position on the left replichore. The EC-28 strain contained the $F_{TS}lac^{+}$ episome at the most distal position from *oriC*, however *lacZ* expression in this strain was far greater than most others carrying the episome at positions proximal to *oriC*. The work of Block *et al.* (2012) reports that all sites tested follow the expected pattern of gene dosage-dependent expression, provided that transcription terminators flanking the insert are present. Three sites are found to fall dramatically outside of the predicted pattern in the absence of flanking terminators. However, due to the target sites being within coding regions, these effects are attributed to "local sequence effects", such as clashing of elongation complexes, reduced translation or increased mRNA degradation.

In this study, activity of the *lac28* fragment promoter at the *ara* and *mel* loci was shown to be less than expected from the general pattern. Activity of the promoter at the *mntH* and *dkgB* loci was also very low, considering the activity of the promoter at other loci in these areas and the location with respect to the origin of replication (Figure 3.11). The weak promoter activity demonstrated at the *dkgB* locus could potentially be due to the activity of neighbouring operons and this is discussed further in chapter 4. However the weak activity at the *mntH* locus was not expected, since this locus did not reside within a tsEPOD, and the neighbouring

operons did not show particularly strong expression. Strikingly, expression of *egfp* at the *mntH* locus did increase upon entry into stationary phase to become the best expressed locus (Figure 3.11; Figure 3.21). Data published by Grainger *et al.* (2006) shows that there is a small peak of FIS binding at the downstream neighbouring promoter region. Each cell contains approximately 60,000 molecules of the FIS protein during exponential growth, however upon entry into stationary phase the concentration drops to undetectable levels. Therefore binding in the neighbouring region could be having a negative effect on *lac28* promoter activity, which is relieved upon entry into stationary phase when quantities of FIS are reduced (Azam *et al*., 1999).

3.8.6 Hierarchies of promoter activity change dramatically upon entry into stationary phase

The effect of chromosomal position on promoter activity during stationary phase was analysed and demonstrates an 8-fold variation in activity across the positions tested (Figure 3.21). The general pattern of increasing activity with proximity to the origin of replication was not observed during stationary phase. Promoter activity was strongest when positioned within the NSL, NSR, Left and Right macrodomains, while expression from the Ori and Ter was relatively weak. The exceptionally high activity of the *lac28* fragment promoter at the *asl* locus, during exponential growth, was greatly reduced during stationary phase, as was that at the *nupG* locus, however not to the same degree (Figure 3.7; Figure 3.21). Due to the high activity not being associated with gene dosage, this change must be due to another factor, which could potentially be superhelical density. DNA superhelicity is more relaxed during stationary phase, which could have large effects on *lac28* fragment promoter activity, due to the sensitivity of the *lac* promoter to superhelical density (Sanzey, 1979; Balke and Gralla, 1987). This global change in superhelical density could explain levelling out of gene

expression patterns during stationary phase and the effect of superhelicity was investigated further in section 4.4 of chapter 4. The transition to stationary phase also appeared to have an effect on expression of the *lac28*::*egfp* construct at tsEPOD loci, which would be expected, due to the large changes seen in NAP concentrations within the cell and general restructuring of the nucleoid (Azam, 1999; Figure 3.21). These effects were analysed further in section 4.5 of chapter 4.

3.8.7 Position of target DNA sites affects repression by LacI

The efficiency of LacI-dependent repression was found to be sensitive to the position of the target *lac28* fragment promoter (Section 3.6.2). Montero Llopis *et al*. (2010) demonstrated that chromosomally expressed mRNA shows limited dispersion from the site of transcription. These data lead to the hypothesis that the Lac repressor protein has a higher probability of localising to the same region as the *lacI* gene, due to the coupling of transcription and translation. The increased local concentration of LacI repressor around the site of production would function in a similar way to the pseudo-operators of the *lac* operon, increasing repression of the promoter through increased LacI local concentration in proximity of the target (Oehler *et al*., 1994). Evidence for this hypothesis has recently been supplied by the work of Kuhlman and Cox, (2012) in which fluorescently labelled LacI shows limited dispersion and localises with the position of the *lacI* gene. Further evidence demonstrates that repression of the target promoter is increased with proximity to the *lacI* gene, however this effect is quite small (Kuhlman and Cox, 2012). This is expected, as the LacI repressor protein is found at low levels with 8-12 tetramers per cell, therefore cannot be present at all loci (Gilbert and Müller-Hill, 1966). Interestingly, genes encoding transcription factors are often found closer to the target binding site than would be expected if located at random (Kolesov *et al*., 2007; Janga *et al*., 2009).

The strains used in the experiments presented here carry the *lacI* gene at the wild type location under control of the wild type promoter, whilst the position of the target *lac28* fragment was varied. Repression of the *lac28* fragment promoter was analysed in each strain, by measuring induction of the promoter in the presence of varying IPTG concentrations, and was found to be affected by the chromosomal position of the target. The *lac28* fragment promoter was most tightly repressed at the *ycb* locus, which suggests that this locus is highly accessible to LacI. However, when positioned at this location, the promoter is very weak, so this result may be an artefact of the low levels of promoter activity. Other than the *ycb* locus promoter activity was most tightly repressed at the *ara* locus, which is the closest to the *lacI* gene on the circular map and resides in the same macrodomain (Figure 3.16). This result provides evidence that the repressive action of LacI is dependent upon proximity to the *lacI* gene, as suggested by Kuhlman and Cox (2012).

Repression of the *lac28* fragment promoter at the *mel* and *asl* locations was the weakest but these positions both reside within the Ori macrodomain (3.14; 3.15). This macrodomain is the neighbour of the NSR macrodomain on the circular map, in which the *lacI* gene resides, however this does not mean that these loci are proximal within the folded nucleoid. These data suggest that the Ori macrodomain may not reside in a similar location to the site of Lac repressor production. Alternatively, access to target promoters within the Ori macrodomain may be limited, potentially due to structuring, architecture or some topological characteristic of the domain.

Activity of the promoter is similar when fully induced at the *asl* and *nupG* loci, however at lower concentrations of IPTG, expression at the *asl* locus is greater, suggesting a difference in repression by LacI (Figure 3.16). The *nupG* locus is on the opposite side of the circular map to the *lacI* gene, however this would not be the case in the folded nucleoid as the non-structured macrodomains reside within similar home positions in the cell and exhibit

more freedom of movement than the other domains (Espeli *et al*., 2008). Therefore the *lac28* promoter at the *nupG* position would be expected to locate to a position that is relatively close to the *lacI* gene within the cell, compared to that of the *asl* locus in the Ori macrodomain. This could allow tighter repression if the Lac repressor demonstrates limited dispersion from the site of production, the *lacI* gene, as suggested by Kuhlman and Cox, (2012). These results demonstrate that the circular chromosome map may be of little use in attempting to predict position-dependent effects on transcription factor activity. These data also suggest that some loci may be more or less accessible to certain proteins, potentially affecting the regulation of gene expression.

3.8.8 Lac permease is involved in transport of IPTG at low concentrations

Expression of the *lacY* gene product was shown to play a role in the induction of *lac28* activity by IPTG at low concentrations. Elimination of *lacY* expression, in the strain carrying the *lac28* fragment promoter at the *rcs* locus, showed that induction of *egfp* expression by IPTG became similar to that at the *lac* locus, where *lacY* expression is already disrupted. The concentration of IPTG required to induce expression of *egfp* at the *rcs* locus was increased, as was the concentration required for full induction. This result suggests that Lac permease is involved in transport of IPTG into the cell at low concentrations, however diffusion across the membrane is sufficient at higher concentrations. These data are in agreement with that of Marbach and Bettenbrook (2012). Repression of *lac28* activity, by LacI, at the *rcs* locus was also shown to be weaker than at the *lac* locus. However, further work, in which the experiments performed in section 3.6.2 are repeated in LacY depleted cells, would allow direct comparison of *lac* promoter repression at all loci tested with that at the *lac* locus, the closest to the *lacI* gene.

Chapter 4:

Mechanisms of position-dependent modulation of promoter activity

4.1 Introduction

The effect of position within the bacterial chromosome on promoter activity has been investigated by several groups, including in this study (Chapter 3). To date, the effects of chromosomal position on expression have been attributed to gene dosage, which is the increase in gene expression correlating with proximity to the origin of replication. However, this study has demonstrated position-based modulation of promoter activity, which does not correlate with that expected from gene dosage alone (Beckwith *et al*., 1966; Schmid and Roth; 1987; Sousa *et al*., 1997; Block *et al*., 2012; Chapter 3.). Therefore the mechanisms behind position-based modulation of promoter activity required further investigation.

4.1.1 Orientation

There are many factors that could affect the activity of a promoter in different chromosomal contexts, one of which is orientation, and this issue was addressed by previous studies (Beckwith *et al*., 1966; Schmid and Roth; 1987; Block *et al*., 2012). Collisions between the DNA polymerase and RNAP are bound to occur due to the two highly processive enzyme complexes sharing the same template. Also, the DNA polymerase proceeds much faster along the template than the RNAP. Head-on-collisions between RNAP and the replisome cause inhibition of replication fork progress to a greater extent than co-directional collisions (French 1992; reviewed in Merrikh *et al*., 2012). Therefore, this leads to evolutionary selection of highly expressed and essential genes locating to the leading-strand of DNA replication, to allow the two processes to occur in the same direction (French, 1992; Price *et al*., 2005; Mirkin and Mirkin, 2005). Transcription is disrupted by collisions with the replisome, however the degree to which it is disrupted is greater in head-on-collisions. Therefore, the orientation of a gene is predicted to have an effect on its level of expression (French, 1992; Price *et al*., 2005).

4.1.2 Neighbouring gene expression

When moved to a different chromosomal position, promoters will have different neighbouring operons with varying levels of transcriptional activity. Based on the transcription factory hypothesis, put forward by Cook (2010) and described in section 1.4.3, or if local pools of RNA polymerase exist, then strong neighbouring transcription could subtract from the locally available pool of RNAP (Jin *et al*., 2006). Activity of transcription units can also have an effect on the distribution of superhelicity due to the fact that an actively transcribing RNAP elongation complex increases negative supercoiling behind the complex and decreases negative supercoiling, or introduces positive supercoils, ahead (Wu *et al*., 1988; Tsao *et al*., 1989). These changes can affect the activity of promoters, especially those which are sensitive to changes in negative superhelicity, such as the *lac* promoter (Sanzey, 1979; Brahms *et al*., 1985). Therefore the degree of neighbouring transcription could have an effect on target gene transcription through changes in local chromosomal architecture. These effects were analysed by assessment of the interaction between neighbouring inducible operons, or promoter activity in the vicinity of highly transcribed regions.

4.1.3 Distribution of superhelicity

The transition from exponential growth to stationary phase causes a large number of changes to the *E. coli* cell, one of which is the relaxation of the chromosomal superhelical density to produce a less negatively supercoiled molecule (Balke and Gralla, 1987). This change in the global superhelical density was suggested to be the mechanism behind shifts in transcription patterns presented in section 3.5. The superhelical density of the chromosome is not only changed during growth phase transitions, but is also changed and adapts in response to processes such as replication and transcription (Postow, 2004; Nöllmann, 2007). Further to

this, the chromosome is organised into many dynamic supercoiled domains, so the local superhelical density can vary across the genome (Postow, 2004). Variations in negative superhelicity across the genome are predicted to be one of the contributing factors to position-dependent modulation of *lac28* fragment promoter activity, due to the sensitivity of the *lac* promoter to superhelical density (Sanzey, 1979). These effects were tested by measuring promoter activity at different positions during perturbation of chromosomal superhelical density, by inhibition of DNA gyrase.

4.1.4 Transcriptionally silent extended protein occupancy domains

The work of Vora *et al*. (2009) sought to identify the protein occupancy landscape of the *E. coli* genome and identified tsEPODs, as discussed in section 1.4.2. The operons contained within tsEPODs could be transcriptionally silent due to the sequence of the promoter, or due to some silencing effect produced by the domain in which they reside. This has not been tested, hence the effect of tsEPODs, on activity of the *lac28* fragment promoter, was analysed by insertion of the reporter within tsEPODs and replacement of tsEPODs with the reporter.

4.2 Effects of orientation on promoter activity

4.2.1 Effects of orientation on promoter activity at the *lac, ara* and *mel* positions

The effect of orientation on expression of the *lac28* promoter::*egfp* construct was assessed by inserting the construct in the opposite orientation, to previously, at the *lac, ara* and *mel* loci to create the *inverse lac, inverse ara* and *inverse mel* strains, BRY41, BRY21 and BRY37, respectively (Figure 4.1). This was achieved by utilising PCR to amplify the homology regions used to create the *lac*, *ara* and *mel* gene doctoring donor plasmids; however, the *MfeI*

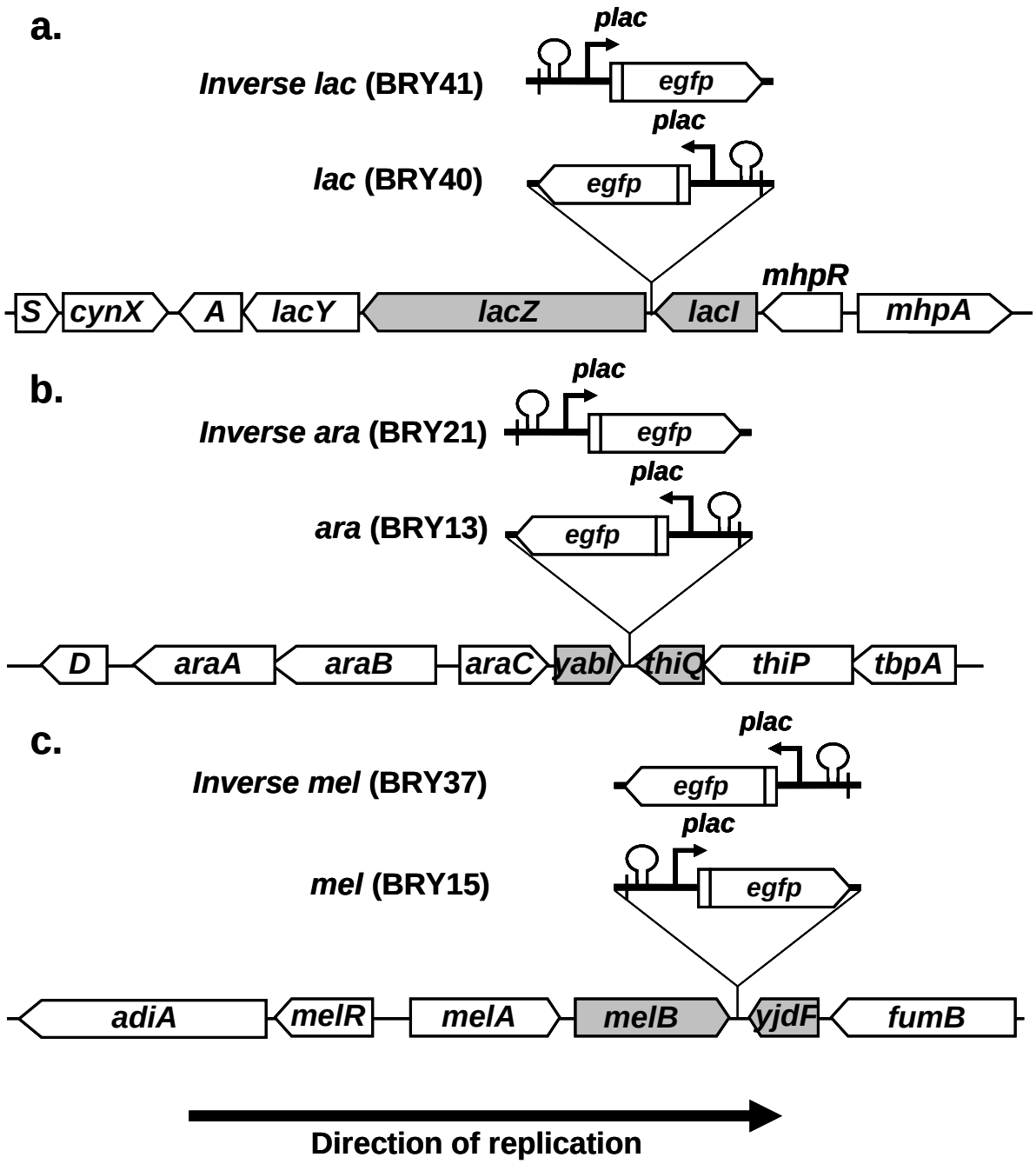

Figure 4.1: Schematic diagram of gene organisation at the *lac*, *ara* and *mel* loci in the BRY40, BRY41, BRY13, BRY21, BRY15 and BRY37 strains

White arrows represent genes and grey arrows represent target convergent genes, while black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. Direction of replication is from left to right in each panel. The arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. **a.** organisation of genes at the *lac* locus in strain BRY40 (*lac*) and BRY41 (*Inverse lac*) **b.** organisation of genes at the *ara* locus in strain BRY13 (*ara*) and BRY21 (*Inverse ara*) **c.** organisation of genes at the *mel* locus in strain BRY15 (*mel*) and BRY37 (*Inverse mel*).

and *XmaI* target sites in the oligonucleotide primers were replaced with *SacI* and *NheI* target sites, respectively, and vice versa (Figure 3.4). Homology regions were cloned into the gene doctoring donor plasmid pJB22, then resulting plasmids were used to insert the *lac28::egfp* fusion into the chromosome using the gene doctoring method, as described in section 2.8. The name "*Inverse*" has no bearing on the relationship between the direction of transcription and replication, but instead merely refers to the order in which the strains were made.

The effect of orientation on *egfp* expression was assayed by measuring fluorescence of strains, carrying the *lac28::egfp* fusion at the *lac, ara* and *mel* positions, in both orientations, during mid-late exponential growth on M9 minimal salts medium. As in section 3.5, activity of the promoter was derived from fluorescence at 510 nm divided by OD_{620} of at least three separate experiments to give the mean and standard deviation. The starting *E. coli* K-12 MG1655 strain was included as a control for auto-fluorescence, however the control samples and the un-induced (no IPTG) cultures registered no fluorescent signal, therefore indicating no auto-fluorescence of the culture at the OD_{620} measured and no un-induced expression of the construct. Therefore, orientation had no measurable effect on the un-induced basal activity of the promoter and no read-through from neighbouring operons was detected in either orientation at any of the loci.

Expression of *egfp* at the *lac* locus was weakly affected by orientation of the insert. Thus, activity of the promoter in the BRY41 *inverse lac* strain was approximately 80% of that in the BRY40 *lac* strain, indicating that activity of the promoter was weakly affected by orientation at the *lac* locus (Figure 4.2). The activity of the *lac28* fragment promoter at the *ara* locus was low, giving approximately 9% of that at the *lac* locus. However, upon inversion to the opposite orientation, in the BRY21 strain, activity was increased 10-fold, having similar activity to that of the promoter in the *inverse lac* strain (Figure 4.2). Expression of the *lac28::egfp* construct at the *mel* locus was approximately 4.8-fold stronger than that of

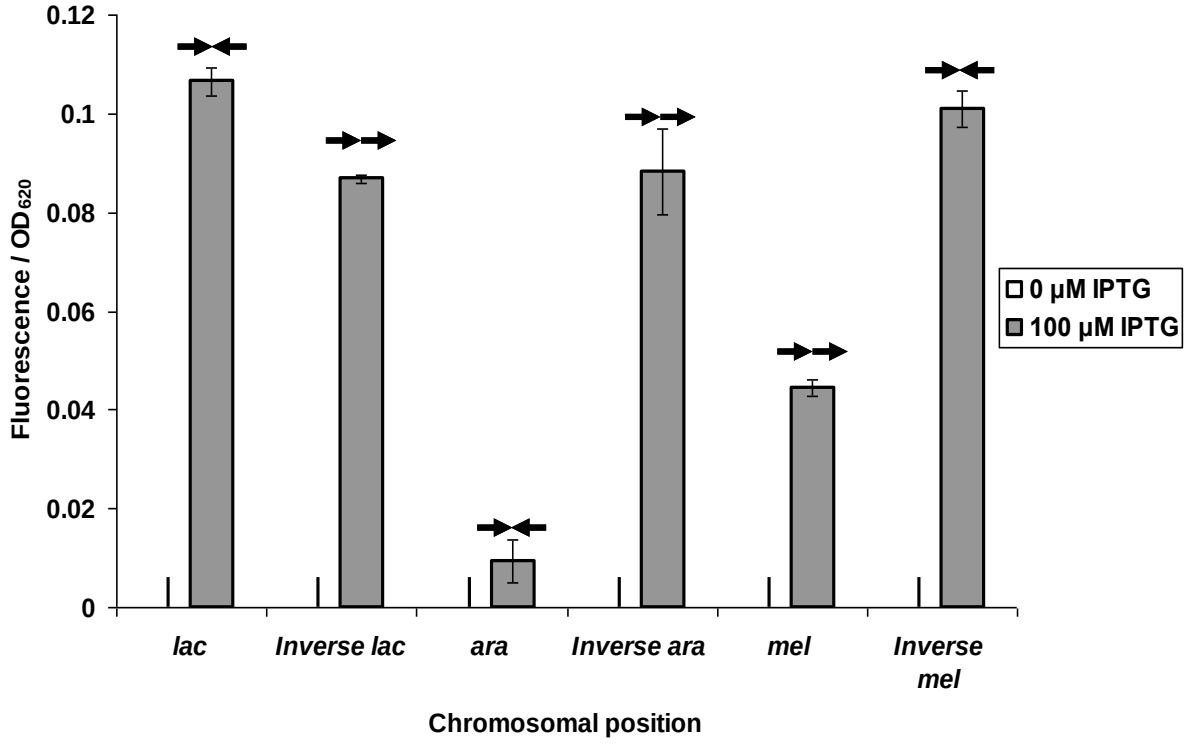

Figure 4.2: Effect of orientation on *lac28::egfp* activity at the *lac, ara* and *mel* loci

Fluorescence/OD_{620} measured in strains BRY40 (*lac*), BRY41 (*Inverse lac*), BRY13 (*ara*), BRY21 (*Inverse ara*), BRY15 (*mel*) or BRY37 (*Inverse mel*). Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with the inducer of the *lac* operon IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Arrows indicate whether replication and transcription are head to head (→←) or co-directional (→→). Orientation has a minimal effect at the *lac* locus, whereas the *Inverse ara* locus demonstrates a 10-fold increase in expression and the *Inverse mel* locus a 2.3-fold increase in expression compared to that of the original *ara* and *mel* locus strains.

the promoter at the *ara* locus. Inversion of the construct, like at the *ara* locus, increased expression levels, giving an activity that was approximately 2.3-fold greater than the promoter in the opposite orientation (Figure 4.2). Expression of the *lac28::egfp* construct was similar in the *lac, inverse lac, inverse ara* and *inverse mel* strains, whereas the promoter in the *ara* and *mel* strains demonstrated less activity (Figure 4.2).

The *lac, ara* and *mel* loci are all located on the right arm of replication. The *lac28* fragment promoter in the *lac* and *ara* strains was oriented in the opposite direction to that of replication, whereas the promoter in the *mel* strain was oriented co-directionally with replication (Figure 4.1). Insertion of the fusion in the opposite orientation at the *ara* locus caused the *egfp* gene to be transcribed co-directionally with replication, which significantly increased expression (Figure 4.1; Figure 4.2). However, the *lac28* fragment promoter in the *mel* strain was already oriented co-directionally with replication and insertion in the head to head orientation caused an increase in promoter activity (Figure 4.1; Figure 4.2). Therefore the effects of orientation at these three positions did not correlate with that expected due to collisions with replication. This could be due to the slow growth rate in minimal medium and reduced rounds of replication per cell cycle.

4.2.2 Analysis of local sequence context

Analysis of ChIP and microarray data presented by Grainger *et al*. (2006) demonstrates that there is no significant amount of transcription occuring in the vicinity of the *ara* and *mel* loci in the conditions of this experiment. These data also demonstrate that there is no significant enrichment in FIS, IHF or H-NS binding within 10 kb of the *ara* and *mel* loci. This suggests that the difference in expression is not due to the promoter being positioned closer to a repressive site due to any of these proteins.

The *ara* and *mel* insertion sites are between the convergent genes *yabI* and *thiQ*, and *melB* and *yjdF*, respectively. Analysis of these intergenic regions between these two sets of convergent genes revealed that they each contain two repetitive extragenic palindromic (REP) sequences each (Figure 4.3). The function of REP elements is not entirely understood, however these sites are thought to play a role as mRNA stabilisers at the 3' end of transcripts (Higgins *et al.*, 1988; Khemici and Carpousls, 2004). Further to this, these elements are also known to be target binding sites for DNA polymerase I, DNA gyrase and the NAP IHF, all of which are known to modify DNA structure (Gilson *et al*., 1990; Espeli and Boccard, 1997; Engelhorn *et al*., 1995).

The *ara* and *mel* loci insertion sites are both situated within one of the two REP sites between the respective convergent genes, therefore disrupting one of the REP elements and leaving the other intact (Figure 4.3). This situation meant that in the *ara* and *mel* strains, BRY13 and BRY15, there was an intact REP element upstream of the *lac28* fragment promoter. However, the inverse orientation strains, BRY21 and BRY37, have the intact REP element downstream of the *egfp* transcript (Figure 4.3). Expression of the *lac28::egfp* construct was greater in the inverse orientation strains, in which the intact REP element is downstream of the transcript. These data suggest that the increased expression could be due to greater mRNA stability, created by REP element-dependent protection of the 3' transcript end (Newbury *et al*., 1987a, b; Higgins *et al*., 1988; Khemici and Carpousls, 2004). Contrary to the fact that REP elements are known to act as binding sites for IHF (Engelhorn *et al*., 1995), data presented by Grainger *et al*. (2006) confirms that these REP elements are not bound by IHF during exponential growth. The effect of DNA polymerase I and DNA gyrase at these sites has yet to be determined, however further study on the action of DNA gyrase at these sites is presented in section 4.4. These data suggest that REP element position may be a contributing factor to orientation-based effects on promoter activity at these loci. However this effect could be due either to the action of DNA gyrase at these target sites, or stabilisation of mRNA transcripts.

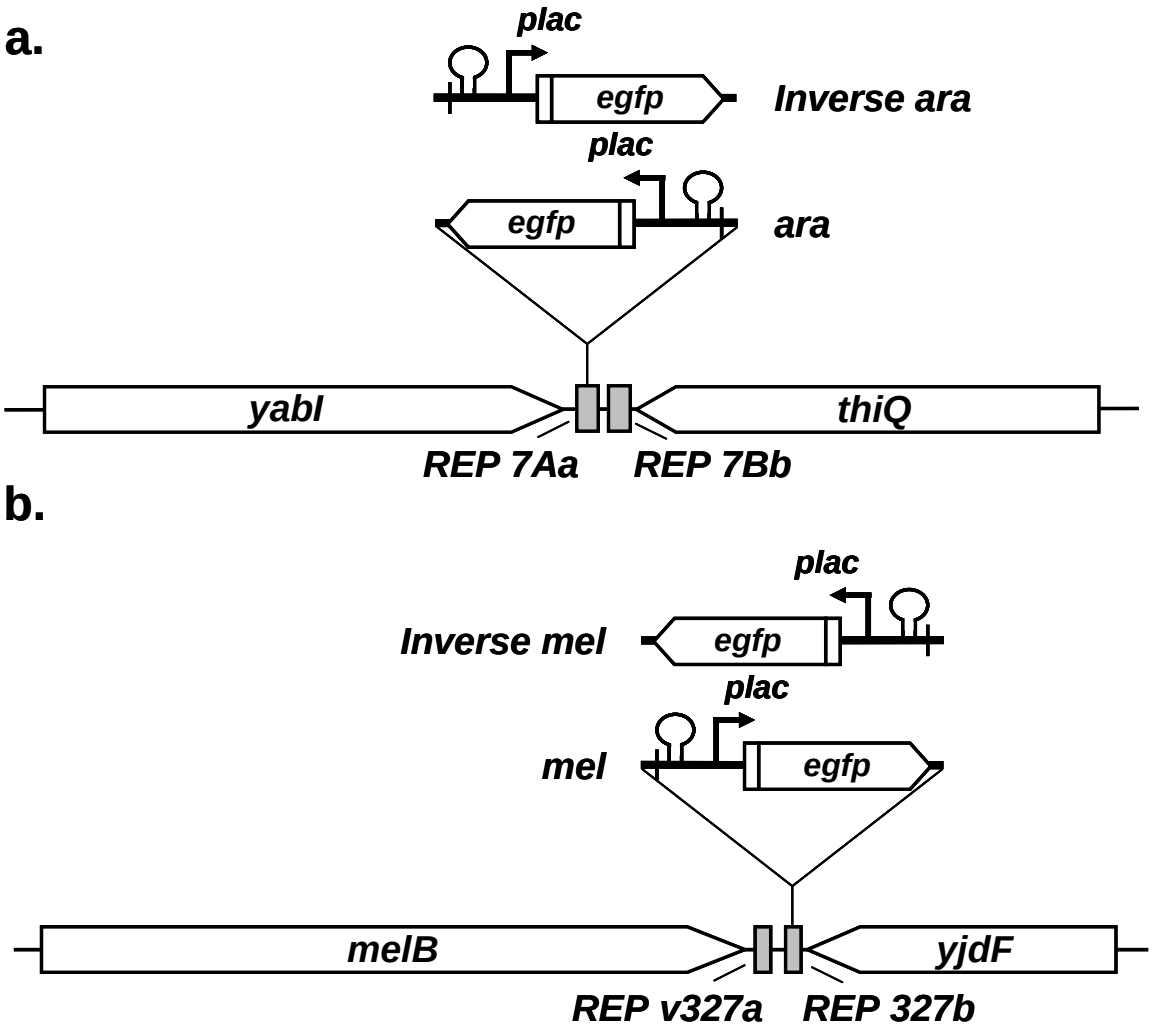

Figure 4.3: Schematic diagram of REP element disruption at the *ara* and *mel* loci

White arrows represent genes and grey boxes represent repetitive extragenic palindromic (REP) elements, while black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. Direction of replication is from left to right in each panel. The arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. **a.** organisation of genes at the *ara* locus in strain BRY13 (*ara*) and BRY21 (*Inverse ara*) **b.** organisation of genes at the *mel* locus in strain BRY15 (*mel*) and BRY37 (*Inverse mel*). Insertion of the *lac28::egfp* fusion caused disruption of the REP7Aa element at the *ara* locus (**a**) and the REP327b element at the *mel* locus (**b**)

4.2.3 Effects of orientation on promoter activity during stationary phase

Expression of the *lac28::egfp* fusion in both orientations at the *ara* and *mel* loci was assayed during stationary phase, as described in section 3.7. In stationary phase, activity of the promoter originally inserted at the *ara* and *mel* loci, in strains BRY13 and BRY15, was stronger than that of the promoter in the inverse orientation (Figure 4.1; Figure 4.4). This result was the opposite of that during exponential growth, during which the construct in the *inverse ara* and *inverse mel* strains were more highly expressed (Figure 4.2).

Expression of *egfp* was similar at the *ara* and *mel* loci, in strains BRY13 and BRY15, however the *inverse ara* strain, BRY21, gave a signal that was approximately 2.5-fold weaker (Figure 4.4). This result suggests that the previously mentioned difference in REP element-dependent transcript stability does not entirely explain the differences in expression caused by orientation. This is due to the fact that the proposed difference in transcript stability should have been maintained during stationary phase, as mRNA stability is known to be unaffected by growth transition (Kuzj *et al*., 1998). REP elements are also known to be targets for DNA gyrase, which is active during exponential growth but inhibited by GyrI on entry into stationary phase, as discussed in section 1.2.2 (Nakanishi *et al*., 1998; Baquero *et al*., 1995; Oh *et al*., 2001). Therefore these effects may be partially due to the action of DNA gyrase and this is investigated further in section 4.4.

4.3 Effects of neighbouring promoter activity

4.3.1 Effects of neighbour operon induction

Transcription of neighbouring operons is a potential factor affecting the expression of the *lac28::egfp* construct in the experiments presented in chapter 3. The insertion of *egfp* at the

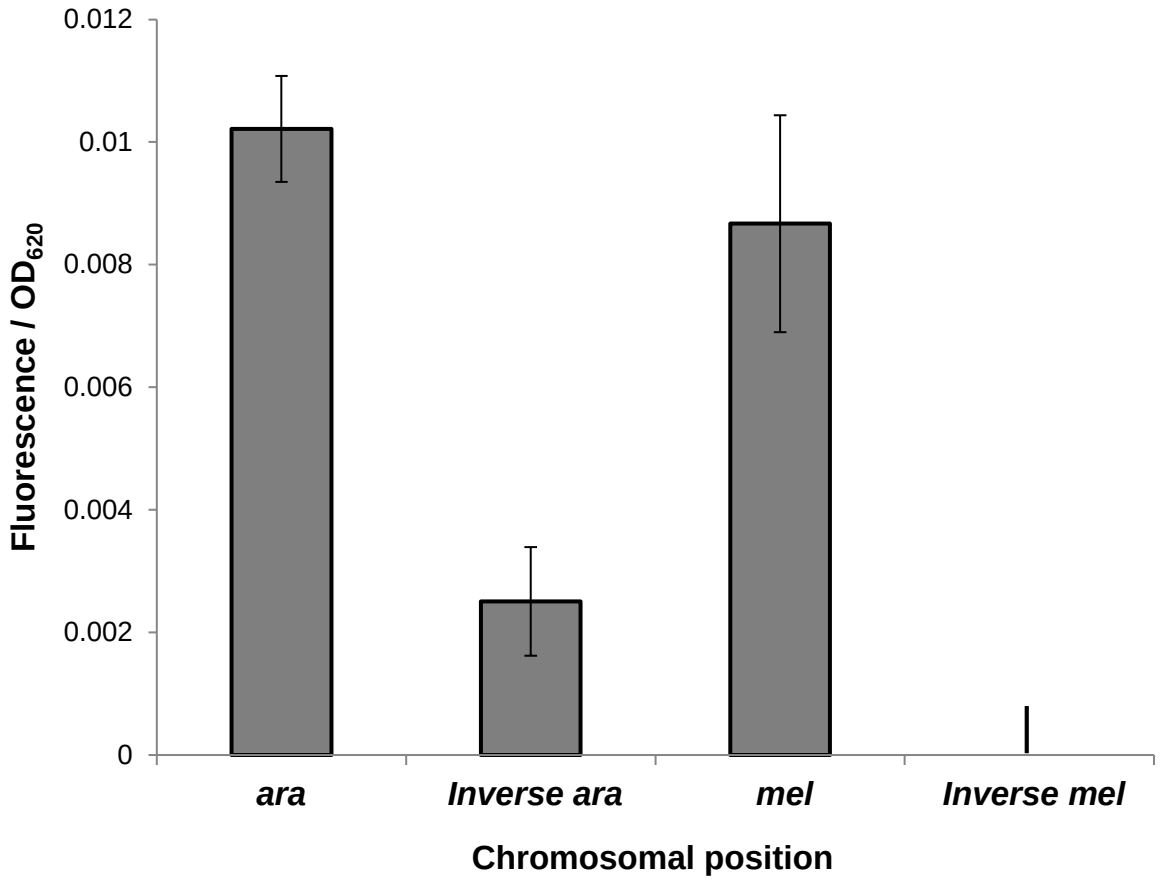

Figure 4.4: Effect of orientation on *lac28::egfp* activity at the *ara* and *mel* loci during stationary phase

Expression of the *lac28::egfp* fusion in the BRY13 (*ara*), BRY21 (*Inverse ara*), BRY15 (*mel*) and BRY37 (*Inverse mel*) strains during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 μM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the un-induced sample, subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. The signal obtained from the induced *inverse mel* strain (BRY37) was unchanged that of the un-induced culture, therefore is represented as zero. The activity of the construct at the *ara* and *mel* loci was similar and both were greater than in the opposite orientation.

mel position, in the BRY15 and BRY37 strains, allows the effect of neighbouring gene expression on *lac28::egfp* activity to be investigated. The *mel* insertion locus is situated immediately downstream of the melibiose inducible operon *melAB*, of which the *melA* gene encodes the α-galactosidase enzyme (Figure 4.5). The activity of α-galactosidase can be measured *in vivo* by quantifying the release of o-nitrophenyl during hydrolysis of o-nitrophenyl-α-D-galactopyranoside by the enzyme. At 37˚C, transport of the inducer, melibiose, into the cell requires the *lacY* gene product, due to the temperature sensitive nature of the melibiose transporter. However, at 30˚C, transport can be facilitated by either the *melB* or *lacY* gene products (Prestidge and Pardee, 1965; Kennedy, 1970; Tamai *et al*., 1998). Hence, the pKH3 gene gorging plasmid was used to knock out the wild type *lac* promoter to abolish any interference from the transport of melibiose into the cell by the *lacY* encoded Lac permease, as in section 3.6.3. The wild-type *lacZYA* regulatory region of the BRY15, BRY37 and BRY35 strains was knocked out to create the BRY75, BRY79 and BRY78 strains, carrying *egfp* at the *mel* locus in both orientations and at the *asl* locus. Therefore this system allowed the effect of *melAB* transcription on that of the *egfp* gene and the effect of *egfp* transcription on that of the *melAB* operon, to be analysed.

The effect of neighbouring *melAB* expression on *lac28::egfp* activity at the *mel* locus was assayed by measuring fluorescence of the BRY75, BRY79 and BRY78 strains during exponential growth at 30˚C on M9 minimal salts medium supplemented with 100 μM IPTG or 0.2% melibiose, where stated. The addition of 0.2% melibiose to M9 minimal salts medium is known to induce expression of the *melAB* operon (Webster *et al*., 1987) The starting *E. coli* K-12 MG1655 strain was included as a control for auto-fluorescence, however the control samples and the un-induced cultures (- IPTG) registered no fluorescent signal. This demonstrated that no read-through from the neighbouring *melAB* operon was occuring. The BRY78 strain, containing the *lac28::egfp* fusion at the *asl* position, which is completely un-connected to the *melAB* locus, was included as a control to ensure the addition of

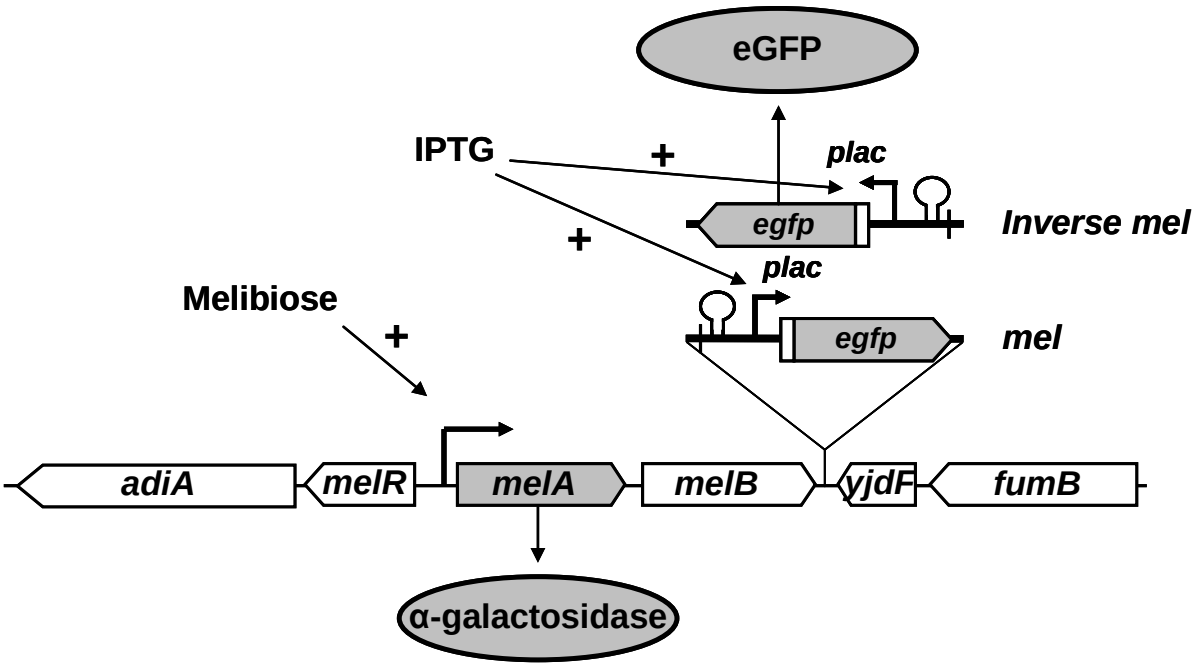

Figure 4.5: Schematic diagram of gene organisation in the BRY75 and BRY79 *mel* locus strains

White arrows represent genes and grey arrows represent reporter genes, while black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. The grey ovals indicate reporter gene products. Direction of replication is from left to right. The arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. Induction of the *melAB* operon promoter by melibiose, or induction of the *lac28* promoter by IPTG is indicated by a + symbol. The *mel* orientation is present in the BRY75 strain and the *inverse mel* orientation is present in the BRY79 strain. Both of these strains are $lacY^-$ due to disruption of the *lacZYA* regulatory region.

melibiose to the growth medium had no direct effect on the *lac28* fragment promoter activity. The activity of the *melA* gene product, α-galactosidase, was also assayed, as described in section 2.16, in the same cultures used for the fluorescence measurements, therefore making these results directly comparable.

The expression of *egfp* at the *asl* locus, in the BRY78 strain, was unaffected by the addition of melibiose to the growth medium, therefore indicating that the *lac28* fragment promoter was not directly affected by the presence of melibiose in the growth medium (Figure 4.6b). The α-galactosidase activity of the BRY78 strain culture was induced by the presence of 0.2% melibiose, however was unaffected by the presence of IPTG. These data show that the expression of the *melAB* operon was not directly affected by the presence of IPTG in the growth medium (Figure 4.6a). These results confirm that each promoter is only responsive to the respective inducer molecules with no cross-talk between the systems.

Activity of the *lac28* fragment promoter at the *mel* locus, in both orientations, was depressed by the induction of the upstream *melAB* operon (Figure 4.6b). Expression of *egfp* in the BRY75 *mel* strain, was reduced approximately 4-fold by the presence of melibiose in the growth medium, which induced expression of the *melAB* operon. Expression of the *lac28* promoter::*egfp* construct was also repressed in the BRY79 strain, which contains the fusion in the *inverse mel* orientation, and gave approximately 3-fold reduced activity in the presence of melibiose. This suggests that induction of upstream transcription represses *lac28* fragment promoter activity, regardless of orientation of the fusion (Figure 4.6b). Repression of *lac28* activity by active upstream transcription could possibly be due to positive supercoiled domains created ahead of the elongating transcription complex, to which the *lac* promoter is sensitive (Wu *et al*., 1988; Sanzey, 1979).

The α-galactosidase activity in the BRY75 *mel* strain, and BRY79 *inverse mel* strain, was significantly increased above that in the BRY78 strain, in which the *lac28::egfp* fusion is

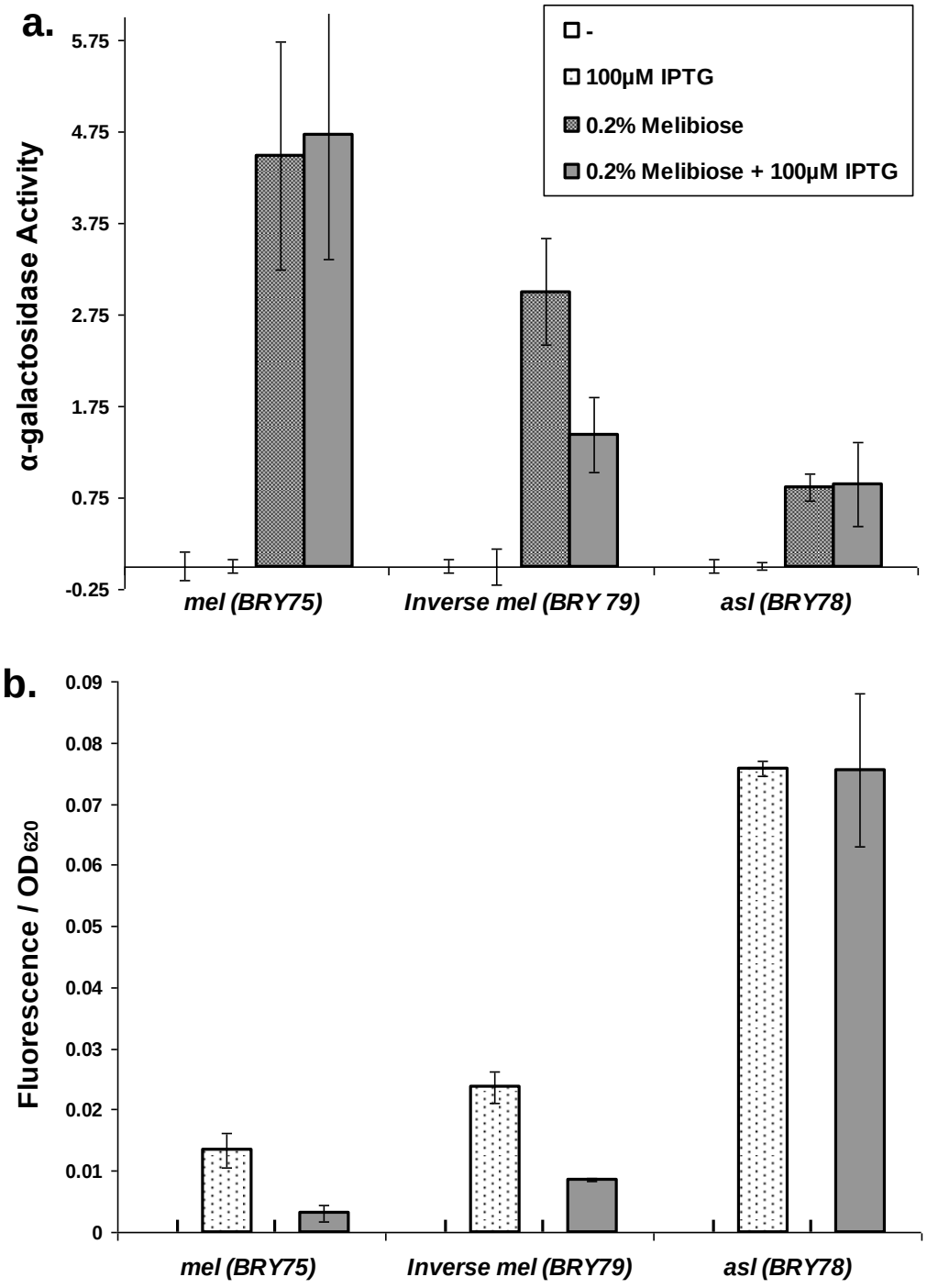

Figure 4.6: Effect of neighbouring transcription on *lac28::egfp* and the *melAB* operon
Strains BRY75 (*mel*), BRY79 (*Inverse mel*) and BRY78 (*asl*) were grown aerobically at 30°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with IPTG or the inducer of the *melAB* operon, melibiose, where stated. Data shown are average measurements from at least three independent experiments, and error bars show one standard deviation from the mean. **a.** α-galactosidase activity was measured *in vivo* and represented as a function of optical density as described in section 2.16. Data demonstrate that induction of the *lac28* fragment promoter is only detrimental to *melAB* activity in the BRY79 *inverse mel* strain. **b.** Fluorescence/OD_{620} was used to measure the *lac28* fragment promoter activity. Data demonstrate that induction of the upstream *melAB* operon is detrimental to *lac28* fragment activity regardless of orientation.

absent from the *mel* locus (Figure 4.6a). This increased expression could be due to the presence of the *lac28* fragment promoter causing a change to local chromosomal architecture. Alternatively, both the *lac* and *mel* promoters are activated by CRP, therefore the presence of an additional CRP binding site could serve to increase the local concentration of CRP.

The α-galactosidase activity in the BRY75 *mel* strain was not affected by induction of the co-oriented *lac28* fragment promoter (Figure 4.6a). However, α-galactosidase activity in the BRY79 *inverse mel* strain was reduced by 50% upon induction of the downstream, convergently oriented, *lac28* fragment promoter (Figure 4.6a). These results demonstrate that expression of the *melAB* operon is only negatively affected by activity of the *lac28::egfp* fusion when positioned downstream of the *lac28* fragment promoter. The data presented in this section shows that active transcription can have a significant negative effect on downstream promoter activity.

4.3.2 Effects of highly transcribed neighbouring genes

The *dkgB* locus is located immediately downstream of the *rrsHileValaVrrlHrrfH* and *aspU* highly expressed, ribosomal and tRNA operons, thereby allowing further analysis of the effects of neighbouring gene transcription (Figure 4.7). The high levels of transcription in the upstream region of the *dkgB* locus are shown by the abundant RNAP binding demonstrated by the ChIP-microarray data presented by Grainger *et al*. (2006) and re-interpreted in figure 4.7. However the large amount of rRNA transcription is known to be inhibited and maintained at low levels on entry into stationary phase (Jacobson and Gillespie, 1968; Paul, *et al*., 2004). Therefore the switch from very high activity during exponential growth to very low activity during stationary phase allowed the effect of highly expressed neighbouring transcription, on *lac28* fragment promoter activity at the *dkgB* locus, to be analysed.

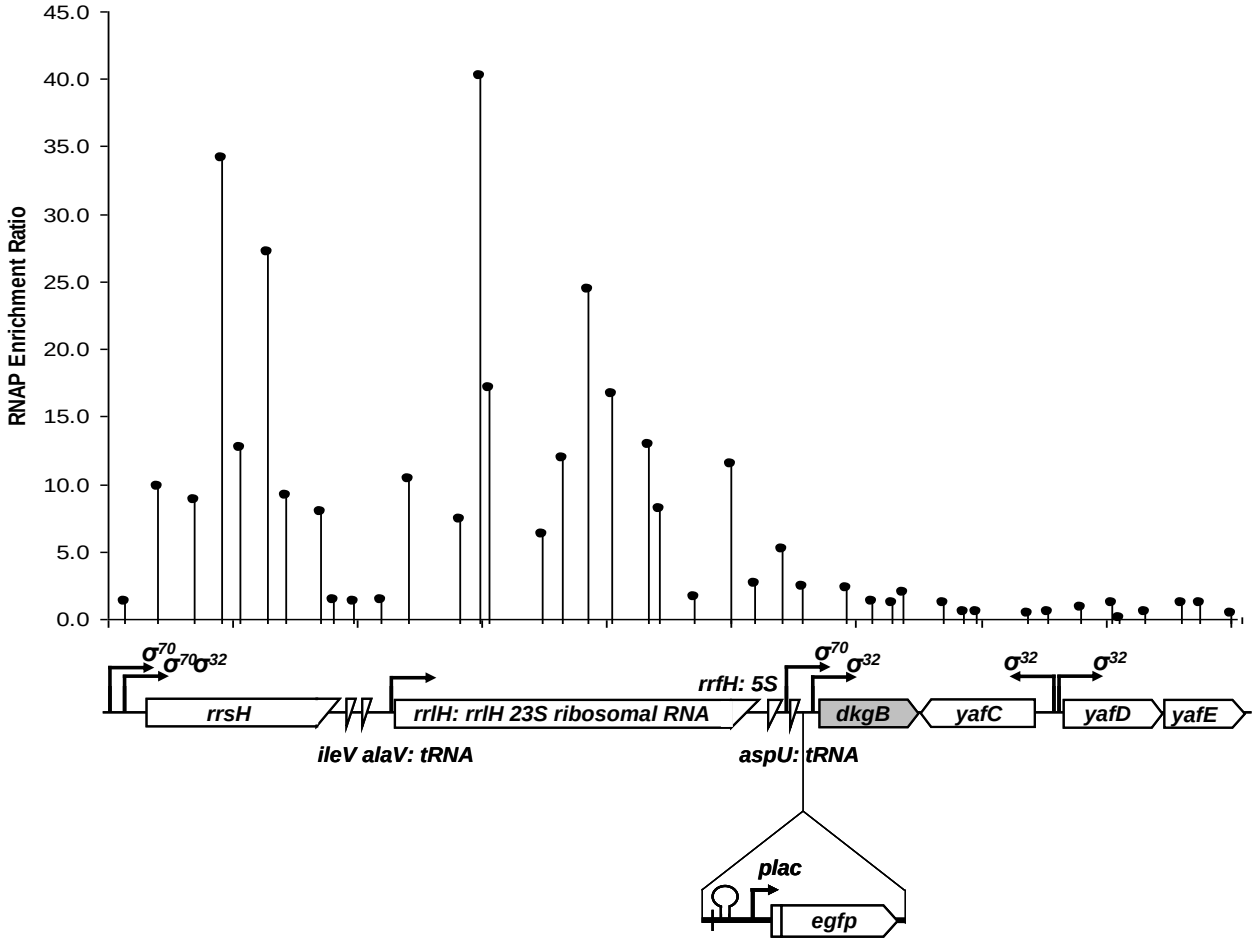

Figure 4.7: Schematic diagram of gene organisation in the BRY58 *dkgB* strain and enrichment of RNAP binding

White box arrows represent genes and the grey arrow illustrates the *dkgB* gene, while black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. Direction of replication is from left to right. The arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. Enrichment ratio of RNAP is shown on the y-axis, as measured by chromatin immunoprecipitation and microarray analysis. Data gathered from *E. coli* MG1655 during mid-exponential growth (OD_{650} 0.3-0.4) on M9 minimal salts medium supplemented with fructose. Data adapted from Grainger *et al*., (2006).

Expression of *egfp* at the *dkgB* position during exponential growth was analysed in section 3.5 and no measureable level of eGFP fluorescence could be detected (Figure 4.8). Therefore this result indicates that during exponential growth, when transcription of the *rrsHileValaVrrlHrrfH* and *aspU* operons is highly active, *lac28* fragment promoter activity was suppressed. However, when the activity of the *lac28* fragment promoter was measured during stationary phase it was shown to be 4-fold greater than that of the promoter at the *tam* locus, which was more active during exponential growth (Figure 3.21; Figure 4.8). Therefore this result suggests that high transcriptional activity of the upstream *rrsHileValaVrrlHrrfH* and *aspU* operons causes suppression of the *lac28* fragment promoter during exponential growth, but this is relieved upon entry into stationary phase (Jacobson and Gillespie, 1968; Paul, *et al*., 2004; Grainger *et al*., 2006). These results are in agreement with those presented in section 4.3.1. and suggest that active transcription can have a repressive effect on downstream promoters.

4.4 Effects of local superhelical density

4.4.1 Inhibition of DNA gyrase activity by novobiocin

The DNA gyrase of *E. coli* is solely responsible for introducing negative supercoils into the genome and contains two subunits, GyrA and GyrB, the latter of which is inhibited by the coumarin drug novobiocin (Nöllmann *et al*, 2007; Sugino *et al*, 1978). Therefore this allowed the general disruption of global chromosomal supercoiling, in all strains tested in section 3.5., by the addition of novobiocin to the growth medium. This enables analysis of effects due to changes in local superhelical density, by the action of DNA gyrase, on the activity of the *lac28* fragment promoter at the different locations.

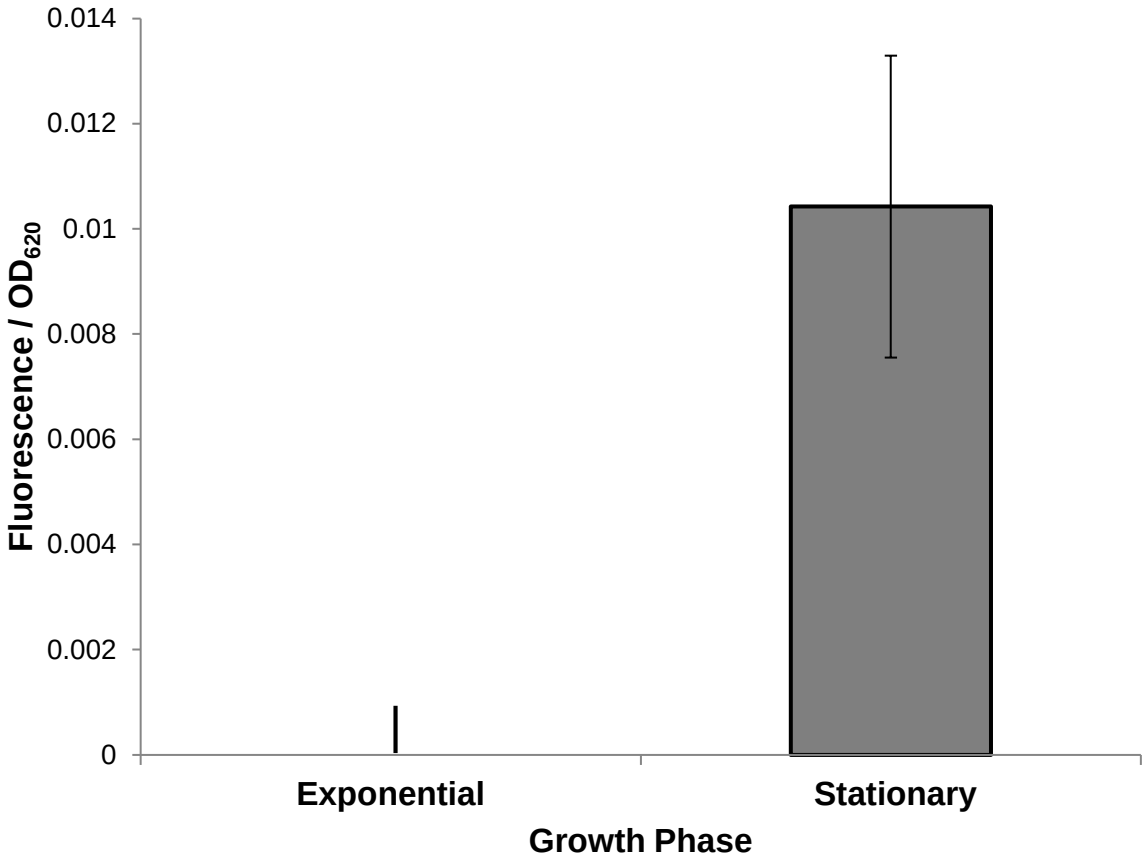

Figure 4.8: Effect of growth phase on promoter activity at the *dkgB* locus

Fluorescence/OD_{620} measured in the BRY58 (*dkgB*) strain. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) or stationary phase (OD_{620} ~ 1) in M9 salts medium supplemented with 100 μM IPTG during exponential growth or 250 μM IPTG during stationary phase, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Fluorescence/OD_{620} measurements for un-induced (- IPTG) cultures are subtracted from that for induced cultures. Data show that the *lac28* promoter::*egfp* construct at the dkgB locus is not inducible during exponential growth, however during stationary phase an increase in fluorescence is seen.

Plasmid pBR322 was used as a control to ensure that the addition of increasing concentrations of novobiocin affects the superhelical density of DNA within the cell. Plasmid was harvested from transformed BRY35 cells during mid-late exponential growth at 37˚C on M9 minimal salts medium supplemented with 80 μg/ml ampicillin and varying concentrations of novobiocin. Plasmid DNA was then visualised by 1% agarose gel electrophoresis supplemented with 2.5 μg/ml chloroquine, to separate the different supercoiled plasmid topoisomers, and stained with ethidium bromide to allow UV visualisation (Section 2.4.2). The addition of chloroquine causes DNA molecules to become more positively supercoiled, therefore increasing the migration speed. This accentuates differences in the supercoiled forms, aiding in separation. The intensity of fluorescence for each lane was plotted against the distance travelled through the gel as illustrated in Figure 4.9. The change in migration of pBR322 through the gel demonstrates that superhelical density is shifted to a more positively supercoiled state with increasing concentrations of novobiocin in the growth medium. This confirms that, in these conditions, the addition of novobiocin to the growth medium did cause a change to the global superhelical density of DNA within the cell, through the inhibition of GyrB (Figure 4.9).

4.4.2 Effects of DNA gyrase inhibition on promoter activity

The effect of novobiocin on *lac28::egfp* expression was investigated by measuring fluorescence of the BRY13, BRY15, BRY21, BRY22, BRY23, BRY27, BRY32, BRY33, BRY34, BRY35, BRY37 and BRY40 strains during mid-late exponential growth on M9 minimal salts medium supplemented with 100 μM IPTG to induce expression of the *lac28* fragment promoter. Novobiocin was added to the growth medium at 50 μg/ml, which is sub-inhibitory to growth. The starting *E. coli* K-12 MG1655 strain was included as a control for auto-fluorescence, however the control samples and the un-induced (no IPTG) cultures

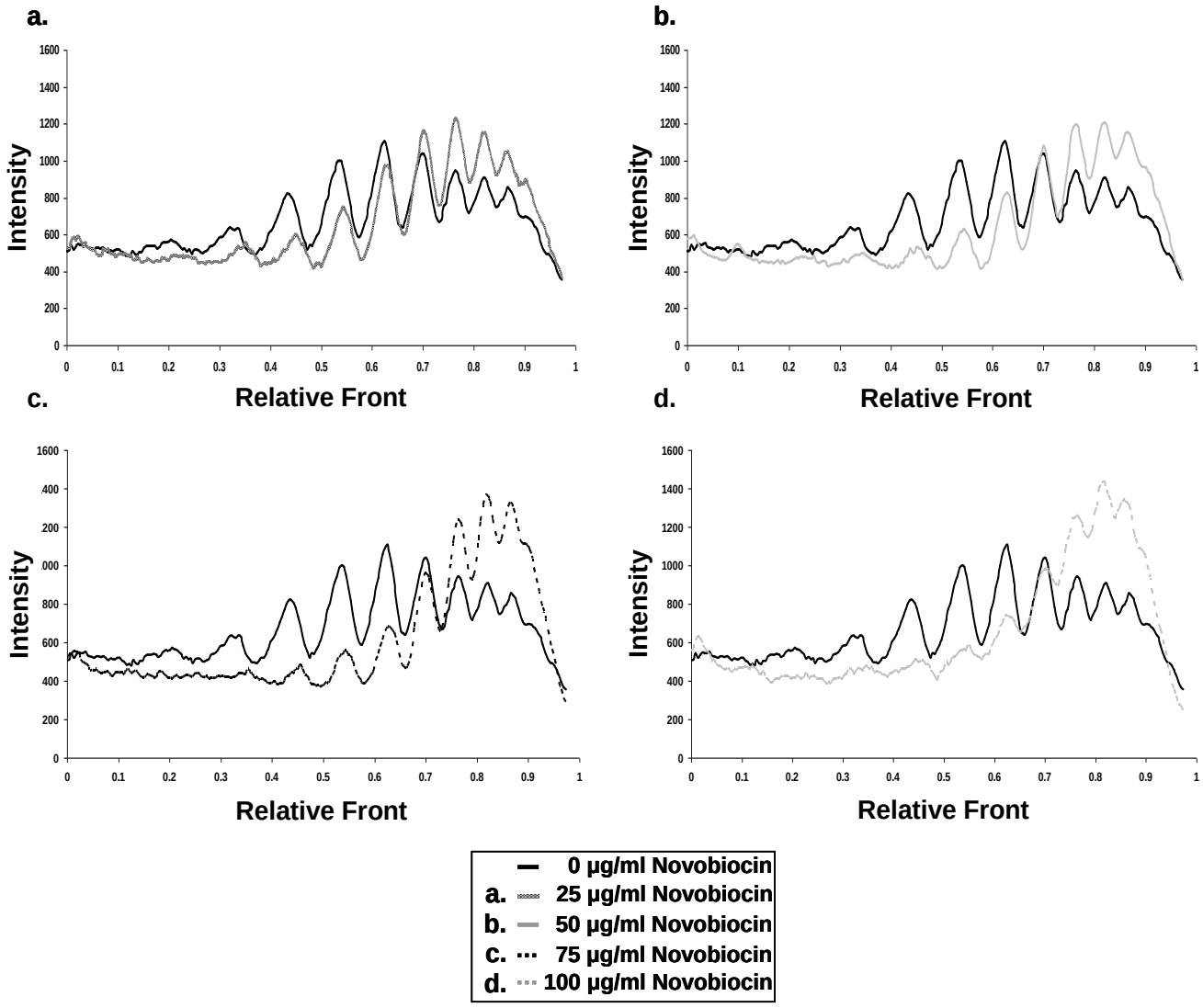

Figure 4.9: Effect of novobiocin inhibition of gyrB on pBR322 plasmid supercoiling

pBR322 plasmid DNA was harvested from BRY35 cells during mid-late exponential growth at 37˚C on M9 minimal salts medium supplemented with 80 µg/ml ampicillin and varying concentrations of novobiocin, where stated. Plasmid DNA was visualised by 1% agarose gel electrophoresis supplemented with 2.5 µg/ml chloroquine to separate the varying supercoiled plasmid topoisomers and stained with ethidium bromide to allow UV visualisation. The intensity of fluorescence for each lane was plotted against the distance travelled through the gel (relative front). **a.** Separation of supercoiled pBR322 in cells grown with no novobiocin or 25 µg/ml novobiocin. **b.** Separation of supercoiled pBR322 in cells grown with no novobiocin or 50 µg/ml novobiocin. **c.** Separation of supercoiled pBR322 in cells grown with no novobiocin or 75 µg/ml novobiocin. **d.** Separation of supercoiled pBR322 in cells grown with no novobiocin or 100 µg/ml novobiocin. Data presented demonstrate a shift in pBR322 supercoiling to a more positively supercoiled state in the presence of increasing concentrations of novobiocin, indicated by increased migration speed through the gel (increased relative front).

registered no fluorescent signal in the presence of novobiocin. Therefore this indicated no measurable effect of novobiocin on the un-induced, basal level, activity of the construct. Results demonstrated that *egfp* expression was unaffected by the addition of novobiocin to the growth medium in most of the strains tested. However, fluorescence of the BRY21 (*inverse ara*), BRY27 (*rcs*), BRY34 (*nupG*), BRY35 (*asl*) and BRY37 (*inverse mel*) strains was affected by GyrB inhibition at concentrations of novobiocin sub-inhibitory to growth.

The effect of GyrB inhibition, by novobiocin, on activity of the *lac28* fragment promoter in the *inverse ara, rcs, nupG, asl* and *inverse mel* strains was further characterised by the addition of increasing concentrations of novobiocin to the growth medium. The expression of *egfp* in the *inverse ara* strain was approximately 10-fold greater than that of the fusion in the opposite orientation in the BRY13 strain, as reported in section 4.2. However, upon addition of increasing concentrations of novobiocin the activity of the promoter was reduced incrementally by up to 3.4-fold (Figure 4.10). This argues that a large part of the increased activity in the *inverse ara* strain may be due to the effects of DNA gyrase on local superhelical density. This is in agreement with the reduction in activity in the *inverse ara* strain during stationary phase, during which DNA gyrase is inhibited and the chromosome becomes relaxed, as reported in section 4.2.3. (Figure 4.4).

The expression of the *lac28::egfp* construct at the *rcs* and *inverse mel* loci, in the BRY27 and BRY37 strains respectively, was only significantly affected by the presence of novobiocin at growth inhibitory concentrations of 75 μg/ml and 100 μg/ml (Figure 4.10). This suggests that DNA gyrase plays only a small part in the regulation of superhelical density at these locations. This implies that the increased expression in the BRY37 *inverse mel* strain, compared to that of the BRY15 strain reported in section 4.2, was not strongly affected by the action of DNA gyrase.

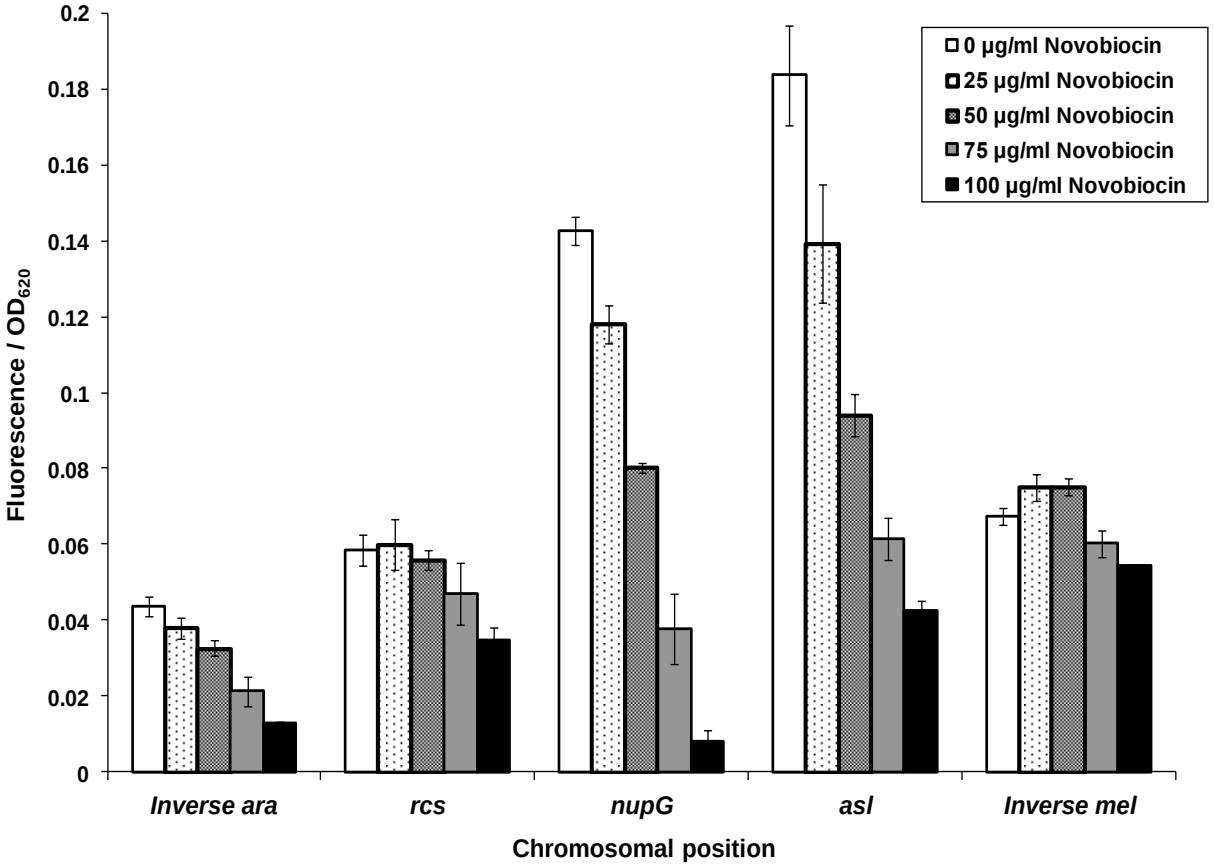

Figure 4.10: Effects of GyrB inhibition by novobiocin on promoter activity in the *inverse ara, rcs, nupG, asl* and *inverse mel* strains

Fluorescence/OD_{620} measured in strains BRY21 (*Inverse ara*), BRY27 (*rcs*), BRY34 (*nupG*), BRY35 (*asl*) or BRY37 (*Inverse mel*), each encoding the *lac28::egfp* fusion at the designated chromosomal position. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with IPTG and the GyrB inhibitor novobiocin, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Minimal effects were seen at the *rcs* and *Inverse mel* loci, at growth sub-inhibitory concentrations of novobiocin. The greatest effects of novobiocin on the *lac28* promoter::*egfp* construct were at the *nupG* and *asl* loci.

The expression of *egfp* was most affected by DNA gyrase inhibition in the BRY34 and BRY35 strains, which carry *egfp* at the *nupG* and *asl* loci respectively. During exponential growth, the *lac28::egfp* fusion was most strongly expressed at the *nupG* and *asl* positions, giving approximately 10 and 11-fold greater levels of expression than that at the *tam* locus, near to the *dif* site, as reported in section 3.5 (Figure 3.7; Figure 3.10). The inhibition of GyrB during exponential growth was shown to reduce the activity of the *lac28* fragment promoter at the *nupG* locus incrementally by up to 17-fold (Figure 4.10). However, activity of the *lac28* fragment promoter at the *nupG* locus in the presence of the sub-inhibitory concentration of 50 μg/ml was approximately 2-fold reduced to a level similar to that of the promoter at the *lac* locus (Figure 4.10). The high promoter activity at the *asl* locus was also shown to be reduced incrementally by up to 4-fold upon the addition of increasing concentrations of novobiocin (Figure 4.10). The addition of novobiocin at the sub-inhibitory to growth concentration of 50 μg/ml was also shown to reduce the expression of *egfp* at the *asl* locus by approximately 2-fold. This reduction made the activity similar to that at the *nupG* locus in the presence of 50 μg/ml novobiocin, or to that at the *inverse mel* or *lac* loci in the absence of novobiocin (Figure 4.10). These results demonstrate that the exceptionally strong *lac28* fragment promoter activity at the *nupG* and *asl* loci is due to the local action of DNA gyrase in maintaining locally available negative superhelicity, to which the *lac* promoter is known to be sensitive (Nöllmann *et al*, 2007; Sanzey, 1979).

4.5 Promoter activity within extended protein occupancy domains

4.5.1 Effects of tsEPOD replacement on promoter activity during exponential growth

The supercoiled domains of the *E. coli* chromosome are expected to contribute to the variation in *lac28* promoter::*egfp* expression, which could be due to variations in local superhelical

density, as presented in section 4.4 (Postow *et al*., 2004). The issue of how supercoils are constrained in these different domains is addressed by the work of Vora *et al*. (2009), who sought to identify the protein occupancy landscape of the *E. coli* genome. This work identified EPODs, which are large regions of the chromosome, between 1 and ~ 14 kb, bound by higher than average amounts of protein. The EPODs were classified as either highly transcribed (heEPODs) or transcriptionally silent (tsEPODs). Hence, the *lac28::egfp* fusion was inserted within a number of tsEPODs to ascertain whether the silent profile is due to the domain containing poor or repressed promoters, or due to intrinsic properties of the domain (Vora *et al*., 2009).

As described in section 3.3, the *lac28::egfp* fusion was, where possible, targeted to non-coding regions between convergent genes within tsEPODs. However, the non-expressed pseudogene *eaeH* was interrupted in one of the contructions. Firstly the *yafT, eaeH, yqe, pitB* and *ycb* loci were targeted to create the BRY22, BRY23, BRY54, BRY55 and BRY32 strains, respectively, as described in section 3.3. Homology regions were used that did not replace any of the existing chromosomal DNA of the starting *E. coli* K-12 MG1655 strain, as in section 3.3, therefore only disrupting the contiguous nature of the tsEPODs (Figures 2.20-2.24; Figure 4.11b). Homology regions were then designed to flank each of the tsEPODs, thereby replacing the tsEPOD with the *lac28::egfp* fusion, allowing expression from the locus to be analysed without the presence of EPOD DNA (Figures 2.20-2.24; Figure 4.11c). The strains in which the tsEPODs have been replaced were assigned BRY numbers and the locus referred to as *ΔtsEPOD* (e.g. *ΔyafT*). The *ΔyafT* strain, BRY59, was created to leave the 3' end of *dnaQ* and the downstream tRNA gene, which reside within the flank of the tsEPOD, present in order to minimise disruption to cell growth (Figure 2.20). Construction of the *ΔeaeH* strain, BRY60, was designed to replace the tsEPOD and the neighbouring *insF-1* insertion element, due to the requirement of a homology region that is unique within the chromosome (Figure 2.21). The *Δyqe* and *ΔpitB* strains, BRY68 and BRY69 respectively, were created as

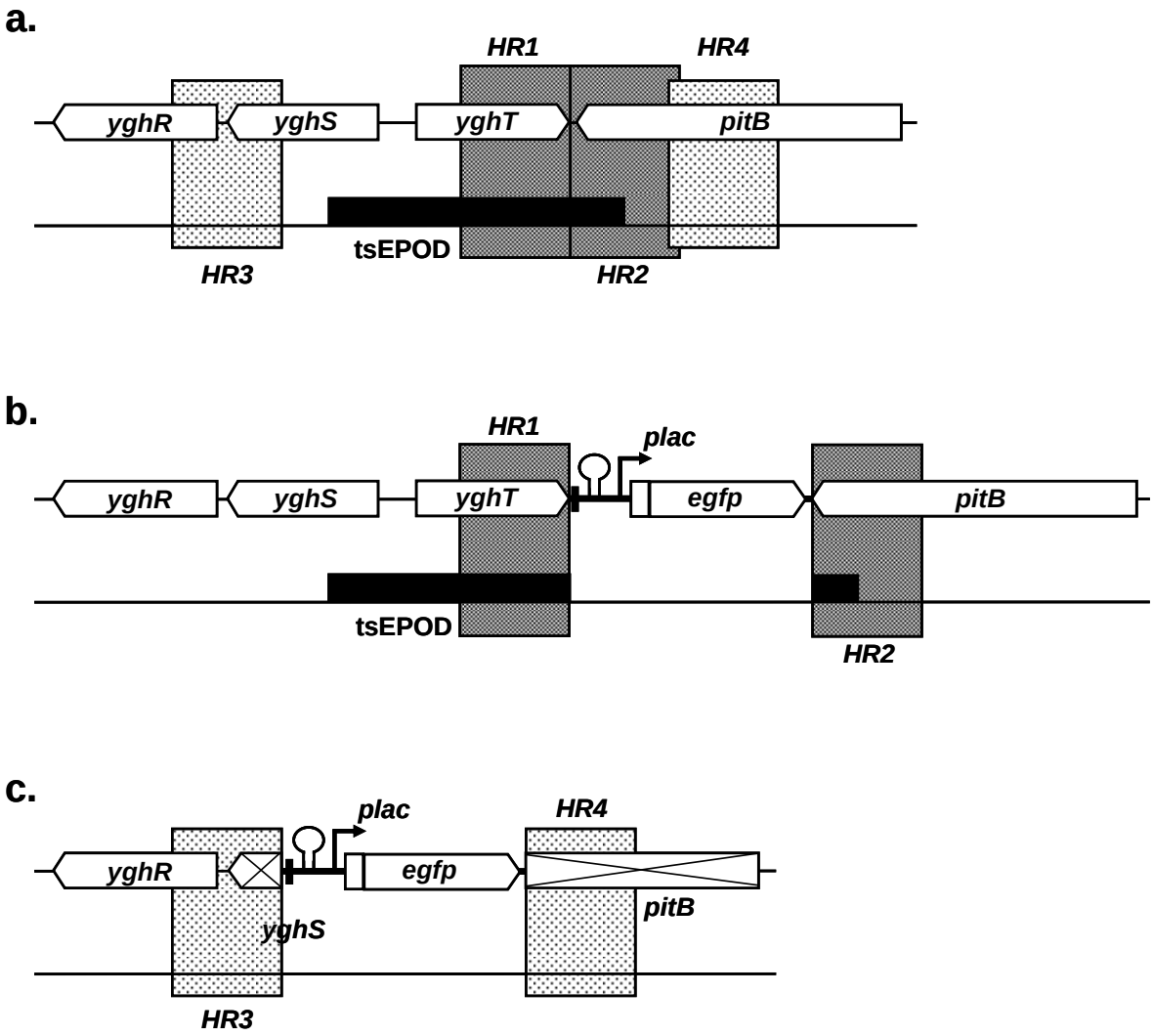

Figure 4.11: Schematic diagram of *lac28::egfp* insertion in the *pitB* and *ΔpitB* strains

White box arrows represent genes, whereas black lines denote intergenic regions and the thick black line represents the insert (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. Disrupted genes are represented by crossed boxes. Direction of replication is from left to right. The bent arrow represents the position and orientation of the *lac* promoter. The hairpin-loop structure represents the position of the λ *oop* transcription terminator. The position of tsEPODs is represented by the black box on the lower line and the positions of homology regions used to insert the *lac28* promoter::*egfp* fusion are indicated by labelled shaded boxes. **a.** Organisation of genes at the *pitB* locus in *E. coli* MG1655 with the homology regions to insert the *lac28* promoter::*egfp* fusion, and the position of the tsEPOD, indicated. **b.** Organisation of genes at the *pitB* locus in the BRY55 strain. The *lac28* promoter::*egfp* fusion was inserted using the *HR1* and *HR2* homology regions shown. **c.** Organisation of genes in the *ΔpitB* BRY69 strain. The tsEPOD region was replaced with the *lac28* promoter::*egfp* fusion using the *HR3* and *HR4* homology regions shown.

described in section 2.9 (Figures 2.23-2.24), however a *Δycb* locus recombinant was never successfully obtained.

Activity of the *lac28* fragment promoter at positions within tsEPODs was assayed by measuring fluorescence of the BRY22, BRY23, BRY32, BRY54, BRY55, BRY59, BRY60, BRY68 and BRY69 strains during mid-late exponential growth on M9 minimal salts medium. The starting *E. coli* K-12 MG1655 strain was included as a control for auto-fluorescence, however the control samples and the un-induced (no IPTG) cultures registered no fluorescent signal. The BRY22, BRY23 and BRY54 strains, carrying the *lac28::egfp* fusion at the *yafT, eaeH* and *yqe* tsEPOD loci, gave no measureable fluorescence signal upon addition of the inducer IPTG to the growth medium (Figure 4.12). This result suggests that the promoter was silenced at these positions. Also, very low levels of fluorescence were detected in some of the induced BRY32 strain cultures, in which the promoter was positioned at the *ycb* locus, therefore indicating that this locus is also repressive to *lac28* fragment promoter activity (Figure 4.12). However, activity of the *lac28* fragment promoter at the *pitB* tsEPOD locus, in strain BRY55, was stronger, giving approximately the same weak level of activity as the promoter at the *tam* locus, reported in section 3.5 (Figure 4.12; Figure 3.7).

Results presented in figure 4.12 show that deletion of the tsEPOD encoding DNA led to increased activity of the *lac28* fragment promoter. Expression of the fusion in the *ΔyafT* strain was greater than that of the promoter in the *yafT* strain, giving a signal that was approximately 80% of that seen from the *pitB* and *tam* loci (Figure 4.12). Promoter activity in the *ΔeaeH* strain was also greater than that of the tsEPOD version. However, the promoter activity was weaker than in the *ΔyafT* strain, giving an activity that was approximately 2-fold lower than that of the *tam* locus (Figure 4.12). Further to this, activity of the *lac28* fragment promoter in the *Δyqe* strain was also increased, compared to that in the *yqe* tsEPOD strain, and had an activity comparable to that in the *pitB* or *tam* strains (Figure 4.12). The activity of the *lac28*

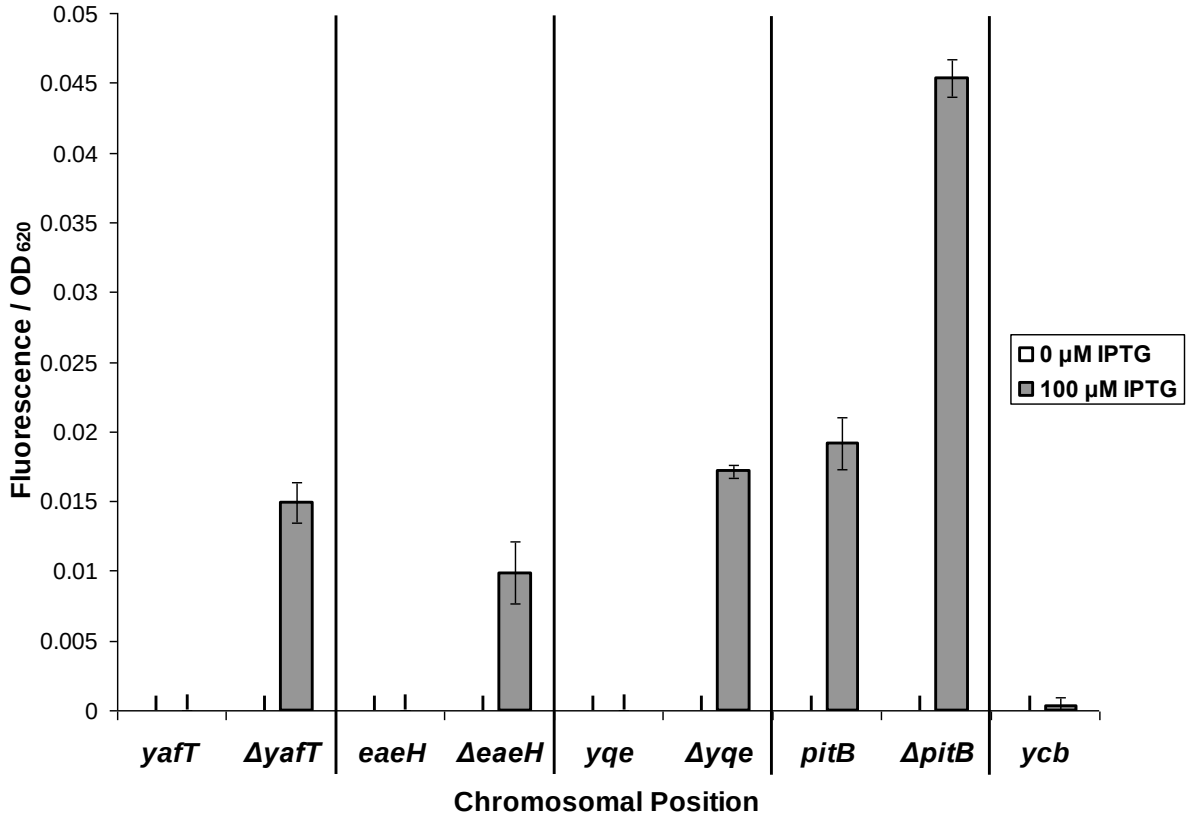

Figure 4.12: Effect of deletion of tsEPODs on *lac28* promoter activity during exponential growth

Fluorescence/OD_{620} measured in strains BRY22 (*yafT*), BRY59 (*ΔyafT*), BRY23 (*eaeH*), BRY60 (*ΔeaeH*), BRY54 (*yqe*), BRY68 (*Δyqe*), BRY55 (*pitB*), BRY69 (*ΔpitB*) or BRY32 (*ycb*). Cells were grown aerobically at 37°C to mid-exponential phase (OD_{620} 0.3-0.5) in M9 salts medium supplemented with IPTG, where stated. Data shown are averages of fluorescence/OD_{620} measurements from at least three independent experiments, and error bars show one standard deviation from the mean. Activity of the *lac28* fragment promoter is silenced at the tsEPOD loci, however upon replacement of the tsEPOD regions, repression is lifted. Silencing is weaker at the *pitB* locus.

fragment promoter was also increased approximately 2.3-fold by deletion of the tsEPOD encoding region at the *pitB* locus (Figure 4.12). Therefore these results suggest that the tsEPODs analysed in this study silenced promoter activity. However, the degree of silencing was less at the *pitB* tsEPOD locus, indicating that silencing by tsEPODs is not equal.

4.5.2 Effects of tsEPOD replacement on promoter activity during stationary phase

Promoter activity at tsEPOD loci was analysed during stationary phase, as described in section 3.7 and 2.10.2. (Balke and Gralla, 1987; Azam *et al*., 1999). Activity of the *lac28* promoter when positioned at the *yafT* tsEPOD locus was increased during stationary phase, giving a similar signal to that in the *ΔyafT* strain and at the *nupG* locus (Figure 3.21; Figure 4.13). This indicates that the silencing created by the *yafT* tsEPOD during exponential growth is relieved upon entry into stationary phase. The activity of the *lac28* promoter at the *pitB* locus was also unchanged by deletion of the tsEPOD encoding region, therefore indicating that either the silencing element of the tsEPOD was absent or that the expression of the *ΔpitB* locus was not improved (Figure 4.13). However, the activity of the *lac28* fragment promoter at the *eaeH* and *yqe* tsEPOD loci was approximately 3-fold less than that of the construct in the *ΔeaeH* and *Δyqe* strains, which was similar to that in the *ΔyafT* strain (Figure 4.13). These results indicate that the repressive element responsible for promoter silencing at the *eaeH* and *yqe* tsEPOD loci was reduced, however still present during stationary phase. The expression of *egfp* at the *ycb* locus was also similar to that of the other ΔtsEPOD strains, indicating that tsEPOD silencing at the *ycb* locus may also be absent during stationary phase (Figure 4.13). These results suggest that the tsEPOD landscape does have a silencing effect on promoter activity and that this landscape is dynamic and changes with growth phase.

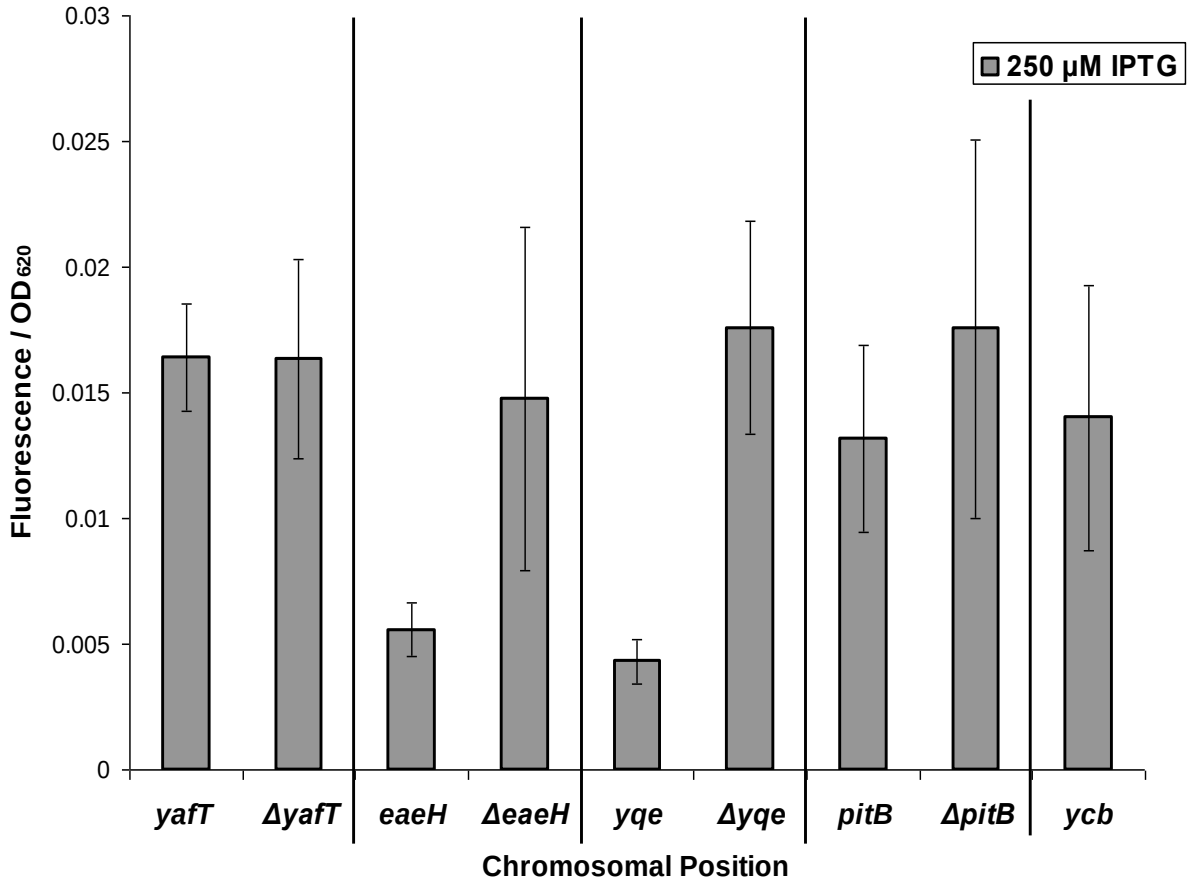

Figure 4.13: Effect of tsEPODs on *lac28* promoter activity during stationary phase

Activity of the *lac28* fragment promoter in the BRY22 (*yafT*), BRY59 (*ΔyafT*), BRY23 (*eaeH*), BRY60 (*ΔeaeH*), BRY54 (*yqe*), BRY68 (*Δyqe*), BRY55 (*pitB*), BRY69 (*ΔpitB*) or BRY32 (*ycb*) strains during stationary phase (OD_{620} ~ 1) at 37 ˚C in M9 minimal medium supplemented with 250 µM IPTG. Each point represents the fluorescence at 485 nm excitation, 510 nm emission, divided by the OD_{620} of the un-induced sample subtracted from the fluorescence/OD_{620} of the induced culture. Each point is the average of at least three independent experiments with the error bars showing one standard deviation from the mean. All loci give a signal above that of background. Activity of the *lac28* fragment promoter is the same in the *yaft, ΔyafT, ΔeaeH, Δyqe, pitB, ΔpitB* and *ycb* strains. Some silencing occurs at the *eaeH* and *yqe* tsEPOD loci indicating that some repressive element remains.

4.5.3 Analysis of nucleoid-associated protein binding of tsEPODs

The organising centres of the *E. coli* nucleoid are expected to be formed and maintained by the cooperative binding of NAPs, due to their high abundance, ability to increase further NAP binding, low sequence specificity and the bending, wrapping, bridging and clustering that can result upon binding to DNA (Azam and Ishihama, 1999; Luijsterbeurg *et al*., 2008; Browning *et al*., 2010). The work of Vora *et al*. (2009) showed that tsEPODs are enriched for NAP binding by position-weight matrix prediction and comparison with the Chip-chip data sets for the IHF, FIS and H-NS proteins presented by Grainger *et al*. (2006). Therefore the IHF, FIS and H-NS ChIP-chip data sets, presented by Grainger *et al*. (2006), were analysed at the tsEPOD loci used in this study to gain further insight into the observed promoter silencing. Further work from our lab identified one of the tsEPODs used in this study as a target for SeqA protein binding. SeqA is a regulator of DNA replication initiation that binds hemi-methylated GATC sequences after passage of the replication machinery (Sánchez-Romero *et al*., 2010). Therefore the data presented by Sánchez-Romero *et al*. (2010) were also analysed for the tsEPOD loci used in this study.

Analysis of SeqA ChIP-chip data for the *yafT* insertion site during exponential growth demonstrated that SeqA binding was not significantly enriched at that site (Sánchez-Romero *et al*., 2010). Further to this, no significant quantities of IHF binding could be detected, however enrichment ratios of approximately 12 for FIS and H-NS were found for the *yafT* locus (Grainger *et al*., 2006; Figure 4.14). Binding of FIS and H-NS in the vicinity of the *yafT* locus, correlates with silencing of expression of the *lac28::egfp* fusion during exponential growth. Therefore during stationary growth, when levels of FIS decrease to un-detectable levels and H-NS levels are reduced by approximately 60% (Azam *et al*., 1999), these proteins will be less likely to bind to the *yafT* target site and so less likely to interfere with *egfp* expression, as reported in Figure 4.13.

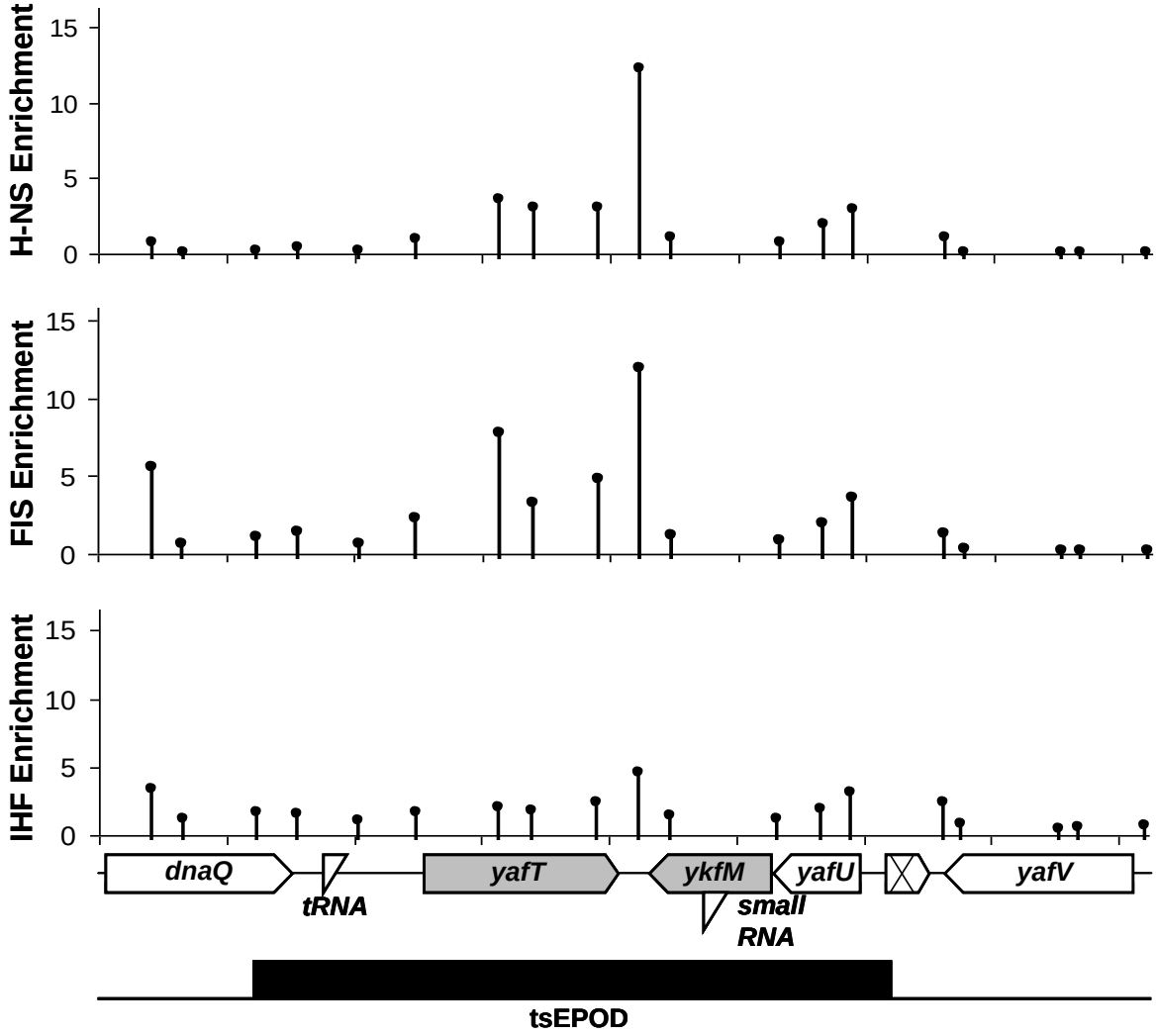

Figure 4.14: Association of IHF, FIS and H-NS with the *yafT* tsEPOD locus

Enrichment ratio of IHF, FIS and H-NS protein at the *yafT* tsEPOD locus, measured by chromatin immunoprecipitation and microarray analysis. Data gathered from *E. coli* MG1655 during mid-exponential growth (OD_{650} 0.3-0.4) on M9 minimal salts medium supplemented with fructose. Data adapted from Grainger *et al*. (2006). Data demonstrate enriched binding of FIS and H-NS at the *yafT* tsEPOD locus.

The *pitB* target tsEPOD was also found to be bound by IHF, FIS and H-NS, however in smaller quantities than that of the *yafT* tsEPOD. The enrichment ratio was approximately 6 for IHF, FIS and H-NS, which was approximately 50% as strong as the FIS and H-NS signals at the *yafT* tsEPOD (Grainger *et al*., 2006; Figure 4.14; Figure 4.15). Analysis of the SeqA Chip-chip data demonstrated that the *pitB* tsEPOD was not bound by SeqA protein (Sánchez-Romero *et al*., 2010). These results suggest that relatively weak promoter activity at the *pitB* tsEPOD position, during exponential growth, could also be due to NAP binding in the local area.

The *yqe* tsEPOD locus was found to be the most bound by NAPs and was also the largest tsEPOD which I targeted with the *lac28::egfp* fusion. Analysis of the ChIP-chip data demonstrated that the *yqe* locus tsEPOD contained one of the largest enrichment ratio peaks in the genome for FIS or H-NS, approximately 70 and 76, respectively (Figure 4.16). The IHF peak signal for the *yqe* tsEPOD had an enrichment ratio of approximately 6, which was similar to that of the *pitB* locus (Figure 4.16). These results demonstrate that the *yqe* tsEPOD region was strongly bound by FIS and H-NS, however silencing of the *lac28* fragment promoter, by the tsEPOD region, seemed at least partially functional during stationary phase, indicating the existence of further silencing mechanisms (Figure 4.13).

In contrast to the protein binding detected at the *yafT*, *pitB* and *yqe* loci, no FIS, IHF, H-NS or SeqA binding was detected at the *eaeH* tsEPOD (Grainger *et al*., 2006). This result suggests that silencing of *lac28* fragment promoter activity, at the *eaeH* tsEPOD locus, was not due to binding of FIS, IHF, H-NS or SeqA, therefore the tsEPOD must be formed of some other protein binding. Analysis of the ChIP-chip data for the *ycb* tsEPOD locus also confirmed that there was no significant FIS or H-NS binding to the tsEPOD and only weak levels of IHF with an enrichment ratio of 5 (Grainger *et al*., 2006). However, SeqA ChIP-chip data from Sánchez-Romero *et al*. (2010), confirmed that the *ycb* locus tsEPOD was formed of the SeqA

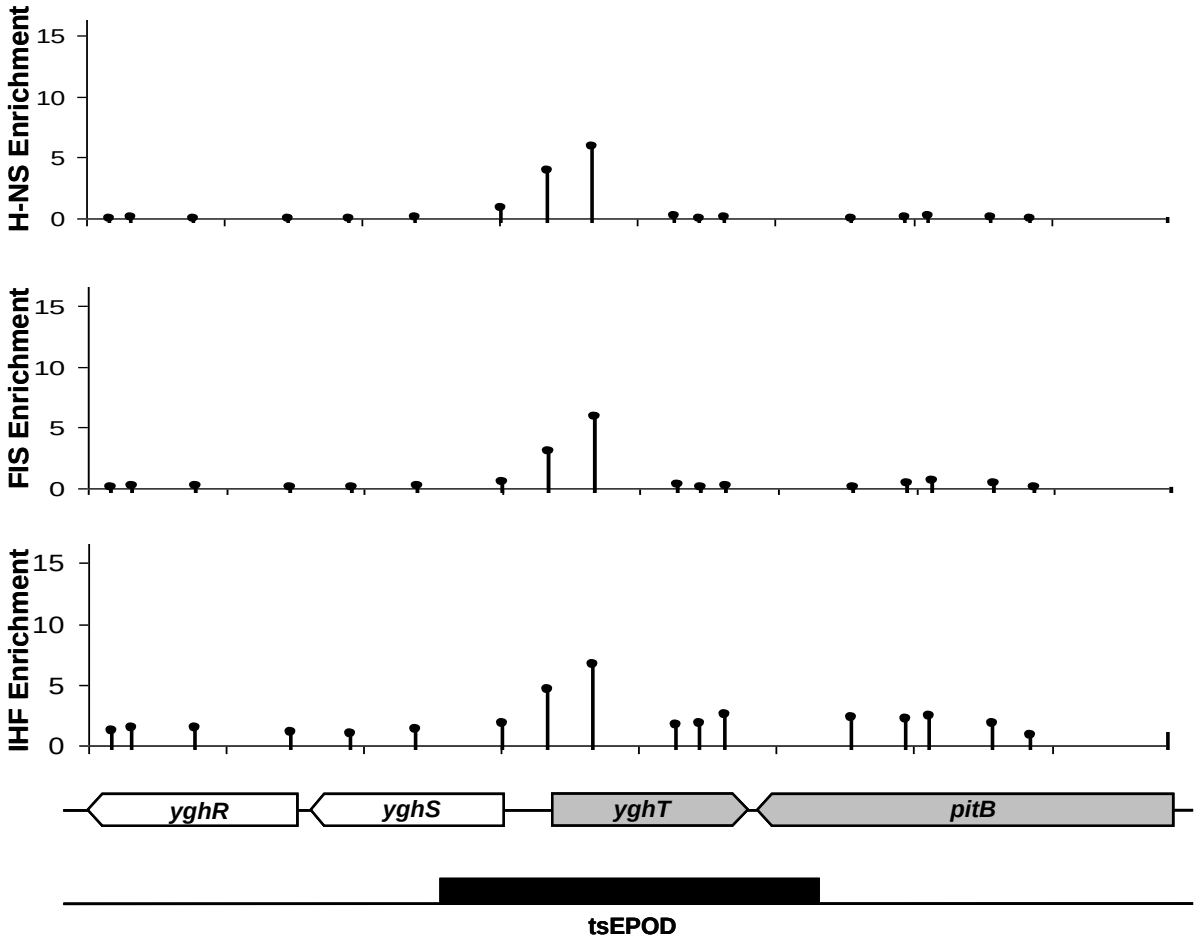

Figure 4.15: Association of IHF, FIS and H-NS with the *pitB* tsEPOD locus

Enrichment ratio of IHF, FIS and H-NS protein at the *pitB* tsEPOD locus, measured by chromatin immunoprecipitation and microarray analysis. Data gathered from *E. coli* MG1655 during mid-exponential growth (OD_{650} 0.3-0.4) on M9 minimal salts medium supplemented with fructose. Data adapted from Grainger *et al*. (2006). Data demonstrate a small signal, indicating low levels of IHF, FIS and H-NS binding to the *pitB* tsEPOD region.

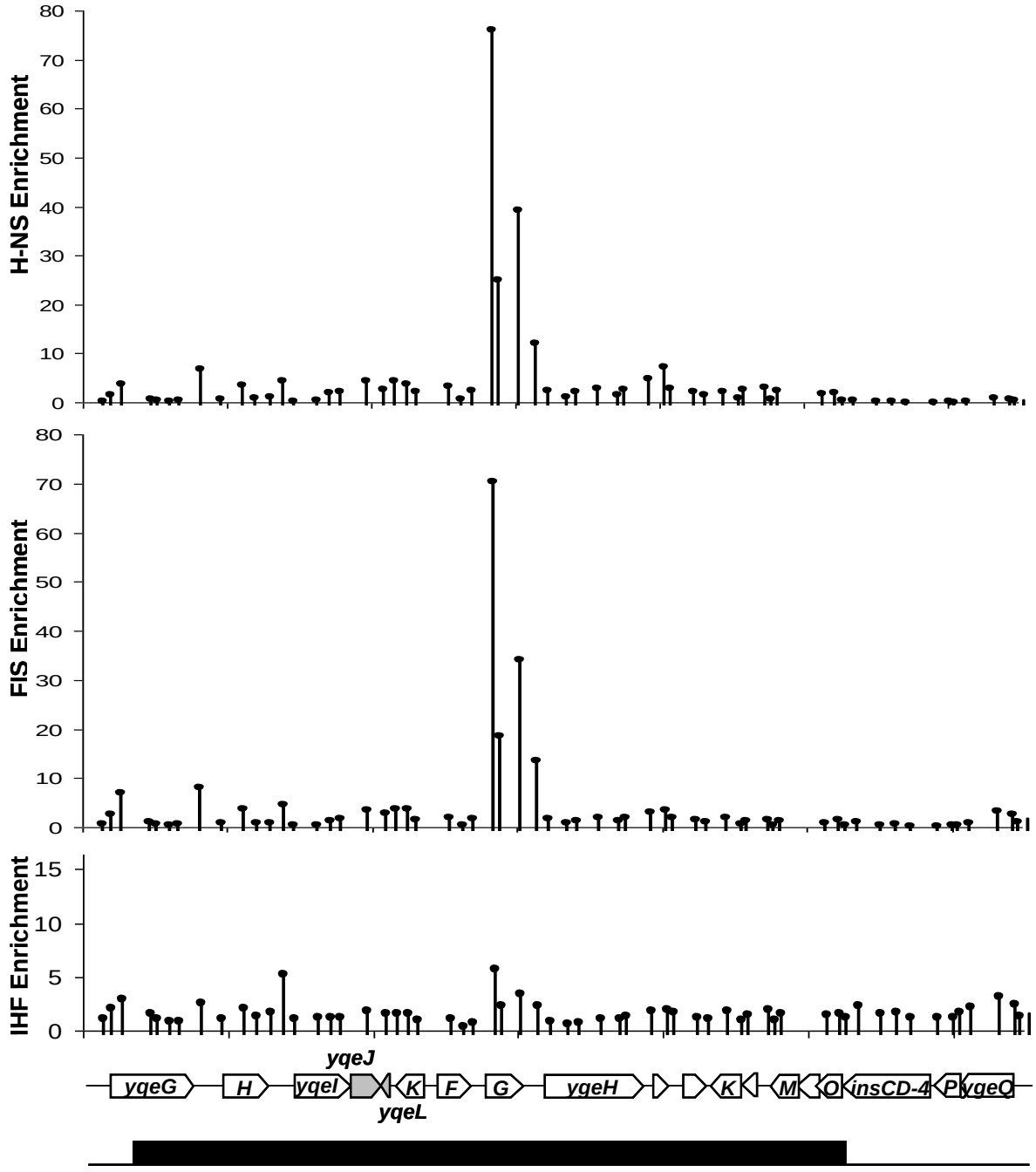

Figure 4.16: Association of IHF, FIS and H-NS with the *yqe* tsEPOD locus

Enrichment ratio of IHF, FIS and H-NS protein at the *yqe* tsEPOD locus, measured by chromatin immunoprecipitation and microarray analysis. Data gathered from *E. coli* MG1655 during mid-exponential growth (OD_{650} 0.3-0.4) on M9 minimal salts medium supplemented with fructose. Data adapted from Grainger *et al*. (2006). Data demonstrate a small IHF signal, however extremely large signals of FIS and H-NS binding are detected within the *yqe* tsEPOD.

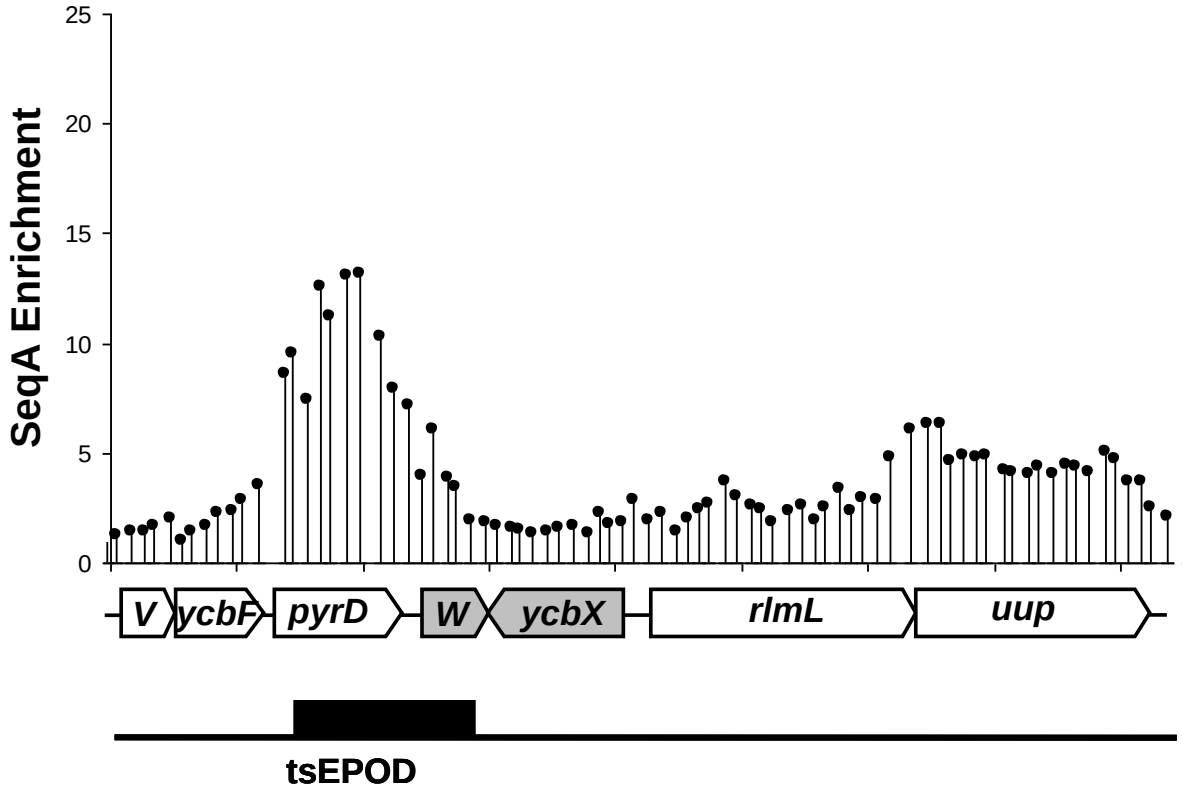

Figure 4.17: Association of SeqA with the *ycb* tsEPOD locus

Enrichment ratio of SeqA protein at the *ycb* tsEPOD locus, measured by chromatin immunoprecipitation and microarray analysis. Data gathered from *E. coli* CMT940 during mid-exponential growth (OD_{650} 0.3-0.4) on LB. Data adapted from Sánchez-Romero *et al*. (2010). Data demonstrate a SeqA binding peak that correlates with the *ycb* tsEPOD locus, indicating that this tsEPOD consists of bound SeqA.

protein. Therefore these results suggest that silencing of the *lac28* fragment promoter, reported during exponential growth, could be due to upstream SeqA binding (Figure 4.12; Figure 4.17). In summary, these results and data show that the protein content of the tsEPODs is not uniform and, whilst the observed silencing of the *lac28* promoter is likely due to tsEPOD-bound protein, there is probably no single mechanism operating.

4.6 Discussion

The work presented in chapter 3 demonstrates that position within the chromosome can affect gene expression levels and that this variation does not always correlate with the changes expected due to proximity to the chromosome replication origin. Previous studies have demonstrated position-mediated variation in gene expression, however this was always attributed to gene dosage or, more recently, "local sequence effects" created by local DNA processes (Beckwith *et al*., 1966; Schmid and Roth; 1987; Sousa *et al*., 1997; Block *et al*., 2012). Therefore the work of this chapter sought to investigate and evaluate the mechanisms behind the position-based effects on promoter activity described in chapter 3.

4.6.1 Orientation effects are unrelated to replication

The orientation of a promoter with respect to the direction of DNA replication is expected to affect gene expression due to disruptive collisions between DNA polymerase and RNAP. Collisions are bound to occur due to DNA polymerase moving much faster than RNAP and the two highly processive enzymes sharing the same template (McGlynn *et al*, 2012). The effect of collisions between elongation complexes and the replisome is dependent upon whether the collision is head-on or co-directional, and the former is far more disruptive to both complexes than the latter (French, 1992; Mirkin and Mirkin, 2005). Due to this

inequality, highly expressed and essential genes mostly locate to the leading strand of DNA replication to keep the majority of transcription co-directional with replication (Price *et al*., 2005; Merrikh *et al*., 2011; McGlynn *et al*., 2012). The effect of orientation on promoter activity was assessed at three loci, *lac ara* and *mel*, that are all on the right arm of replication. Orientation had little to no effect at the *lac* locus, however a 10-fold difference in activity was found at the *ara* locus and a 2.3-fold difference at the *mel* locus. There was no correlation between increased expression and direction of replication at these two loci. The lack of correlation between gene expression level and orientation relative to replication, is probably because of the slow growth rate used in these experiments, in which one doubling takes approximately 60 minutes and only one round of replication is expected to be initiated per cell cycle (Cooper and Helmstetter, 1968).

It was expected that the effects of orientation on expression at the *ara* and *mel* loci would be abolished during stationary phase if they were due to replication-transcription clashes (Chang *et al*., 2002). Surprisingly, analysis of *egfp* expression at the *ara* and *mel* loci, revealed that the expression patterns were reversed. Activity in the "*inverse*" orientation strains was stronger during exponential growth, but weaker during stationary phase (Figure 4.4). This underscores that the effects of orientation on *egfp* expression at the *ara* and *mel* loci are unrelated to DNA replication and highlights our ignorance of what is really going on.

4.6.2 Repetitive extragenic palindromic sequences play a role in orientation effects

Insertion of the *lac28::egfp* fusion at the *ara* and *mel* loci causes disruption of REP sites between the convergent target genes (Figure 4.3). REP elements are 35-40 nucleotide stem-loop structures, which are known to be binding targets for DNA polymerase I, and the NAP IHF, and provide cleavage sites for DNA gyrase (Gilson *et al*., 1990; Espeli and

Boccard, 1997; Engelhorn *et al*., 1995; Messing *et al*., 2012). REPs are also known to increase mRNA stability and terminate transcription when positioned at the 3' end of a transcript (Gilson *et al*., 1986; Higgins *et al.*, 1988; Khemici and Carpousls, 2004). Insertion of the *lac28::egfp* fusion at the *ara* and *mel* loci caused disruption of one of two REP elements between each set of convergent genes. The resulting *ara* and *mel* strains both have one intact REP element upstream of the *lac28* fragment promoter, whereas inversion of the fusion positioned the intact REP element downstream of the *egfp* transcript (Figure 4.3). The REP element at the 3' end of the *egfp* mRNA transcript could improve expression in the "*inverse*" orientation strains.

Data presented by Grainger *et al*. (2006), confirmed that the REP elements at the *ara* and *mel* target loci were not bound by IHF during exponential growth, contrary to the fact that they are known IHF binding sites (Engelhorn *et al*., 1995). Therefore IHF had no role in the orientation effects at these loci. However, inhibition of DNA gyrase, for which the REP elements are known targets, caused a dramatic reduction in the orientation-dependent difference in expression at the *ara* locus (Figure 4.10). Data presented in section 4.2.3 showed that the increased expression in the "*inverse*" orientation strain is lost. The action of DNA gyrase is inhibited upon entry into stationary phase and the chromosome becomes more relaxed, therefore creating a similar effect at the *ara* locus to that of DNA gyrase inhibition (Nakanishi *et al*., 1998; Baquero *et al*., 1995; Oh *et al*., 2001). Therefore, increased expression in the *inverse ara* strain, compared to that in the opposite orientation, was due partly to the action of DNA gyrase at the downstream REP element. Unlike the *ara* position, inhibition of DNA gyrase had little effect on *egfp* expression at the *mel* locus. Therefore orientation-dependent differences in expression at the *mel* locus are most likely due to enhanced *egfp* transcript stability, due to REP element-dependent protection of the 3' transcript end (Figure 4.3; Figure 4.10; Newbury *et al*., 1987a, b; Higgins *et al*., 1988; Khemici and Carpousls, 2004). These data show that position and nature of local sequence

elements, such as REPs, can have significant effects on promoter activity and gene expression.

4.6.3 DNA gyrase-dependent enhanced expression at the *asl* and *nupG* positions

The transition from exponential growth to stationary phase causes many changes to *E. coli* cells, one of which is the relaxation of chromosomal superhelical density to produce a less negatively supercoiled molecule (Balke and Gralla, 1987). This change was suggested to be the mechanism behind the shift in transcription patterns presented in section 3.7. The superhelical density of the chromosome is not only changed during growth phase transitions, but also changes in response to processes such as replication and transcription (Postow, 2004; Nöllmann *et al*, 2007). Sanzey (1979) showed that expression of the *lac* operon is negatively affected, by up to 75%, upon inhibition of DNA gyrase. Therefore, variations in negative superhelicity across the genome were predicted to be one of the contributing factors to the large variation in *lac28* fragment promoter activity reported in section 3.5.

Inhibition of GyrB by novobiocin had little to no effect on *lac28::egfp* expression at the majority of positions tested, however the exceptionally high activity at the *asl* and *nupG* positions was greatly reduced (Figure 4.10). In the presence of novobiocin concentrations sub-inhibitory to growth the expression was reduced to a similar level to that at the wild-type *lac* position, therefore proving that the exceptionally strong activity was due to negative superhelicity, maintained by the action of DNA gyrase. Analysis of the local sequence at the *asl* and *nupG* loci reveals that there are no REP elements present, therefore DNA gyrase is not acting at these sites as at the *ara* locus. However, DNA gyrase binding sites are found at a 5 to 10-fold higher average density near to the origin compared to the Ter proximal region, therefore suggesting a gradient of negative superhelicity from Ori to Ter (Sobetzko *et al*.,

2012). The density of gyrase binding sites also appears to be greater in the NSL macrodomain than in the NSR macrodomain, therefore potentially explaining the difference in expression at the *nupG* locus compared to similar positions on the right arm of replication (Sobetzko *et al*., 2012).

4.6.4 Transcription can have a negative effect on downstream promoter activity

Activity of the *lac28* fragment promoter at the *mel* position was found to be depressed by active transcription of the upstream *melAB* operon, regardless of *lac28::egfp* orientation (Figure 4.6). Also, activity of *melAB* was unaffected by induction of *lac28* when the promoters were oriented co-directionally, however *melAB* expression was reduced when downstream of the *lac28* fragment promoter (Figure 4.6). Further to this, when positioned at the *dkgB* locus, the *lac* promoter was only active during stationary phase, during which expression of the upstream rRNA operon is inhibited (Jacobson and Gillespie, 1968; Paul, *et al*., 2004; Grainger *et al*., 2006; Figure 4.8). These results could be due to the fact that an actively transcribing RNAP elongation complex increases negative superhelicity behind the complex and decreases it ahead (Wu *et al*., 1988; Tsao *et al*., 1989). Under-winding (negative supercoiling), but not over-winding (positive supercoiling), should assist in local DNA strand separation required at promoter transcription start sites for open complex formation and the completion of transcription initiation (Browning and Busby, 2004; Travers and Muskhelishvili, 2007). Also if the transcription factory hypothesis, as discussed in section 1.4.3., is correct or if local pools of RNA polymerase exist, then strong neighbouring transcription could subtract from the locally available pool of RNAP (Cook, 2010; Jin *et al*., 2006). However this is unlikely to be the case, as the *lac28* fragment promoter is closer in the co-directional orientation, yet has no negative effect on *melAB* expression upon induction (Figure 4.6).

4.6.5 tsEPODs silence transcription

Recent research by Vora *et al*. (2009), demonstrated that there are regions of the chromosome that are both transcriptionally silent and highly occupied by protein. Activity of the *lac28* fragment promoter from within a small subset of tsEPODs was analysed by insertion into the tsEPOD without replacement of any genetic material, therefore potentially leaving the protein binding as in the starting strain. Promoter activity at tsEPOD locations was suppressed, however replacement of the entire tsEPOD region with the *lac28::egfp* fusion caused increased levels of *lac28* fragment promoter activity at loci tested during exponential growth (Figure 4.12). Therefore, promoters in these regions were not necessarily transcriptionally silent due to the sequence of the promoter, but due to a silencing effect produced by the domain in which they reside.

Silencing of promoter activity was not maintained at all tsEPOD loci tested on entry into stationary phase. Analysis of ChIP-chip data for the tsEPOD loci revealed FIS, H-NS and IHF binding during exponential growth, however this binding was unequal at the different tsEPOD loci. During stationary phase, when levels of FIS decrease to un-detectable levels and H-NS levels are reduced by approximately 60% (Azam *et al*., 1999), tsEPODs may be less likely to form and interfere with promoter activity. Suppression of *egfp* expression at the *yqe* tsEPOD position was not relieved upon entry into stationary phase, however ChIP-chip data for this locus shows extremely large signals for FIS and H-NS binding during exponential growth (Grainger *et al*., 2006). Thus, as H-NS levels are only reduced by approximately 60% upon entry into stationary phase, some residual H-NS binding may still occur at this highly preferential binding site (Azam *et al.*, 1999). The repression of the *lac28* fragment promoter was also maintained at the *eaeH* tsEPOD locus during stationary phase, however no NAP binding could be detected. Therefore the suppression of *egfp* expression at this locus occurred by some other mechanism and the constituents of the *eaeH* tsEPOD remain unknown.

Analysis of ChIP-chip data for the *ycb* tsEPOD locus revealed that this locus was bound by the SeqA protein, which is a regulator of DNA replication that binds hemi-methylated GATC sequences after passage of the replication machinery (Sánchez-Romero *et al*., 2010). Therefore the *lac28::egfp* fusion was silenced during exponential growth when the chromosome was being replicated, however was active during stationary phase when replication of the chromosome is inhibited (Chang *et al*., 2002).

In summary these results demonstrated that tsEPODs are capable of repressing the activity of a strong promoter, therefore showing that position within the chromosome and local chromosomal architecture can play a major role in the regulation of transcription. However, the tsEPODs are dynamic and unequal in their protein constituent parts, with other protein components left unidentified. Therefore, further work is required to characterise the pattern of tsEPODs during other growth phases and to identify the core proteins involved in their formation and the exact mechanism behind repression of transcription initiation.

Chapter 5:

Comparison of plasmid-located and chromosome-located promoters in *Escherichia coli*

5.1 Introduction

5.1.1 Plasmid-based gene expression systems

Work reported in chapters 3 and 4 shows that bacterial gene expression is not only dependent upon promoter sequence, but also on position within the folded chromosome. However, many studies into promoter function utilise mobile extra-chromosomal plasmid-encoded reporter gene expression systems, such as the low copy number *lac* expression vector pRW50 (Lodge *et al*., 1992). While multi-copy plasmids have long been assumed to diffuse freely throughout the cell, they have also been shown to localise to positions that are on the periphery of the nucleoid at various positions, including the cell poles (Pogliano *et al*., 2001). Further to this, the work of Sánchez-Romero *et al*. (2012), demonstrated that the congregation of several partition-system deficient plasmids in one location was driven by high transcriptional activity. This suggested the presence of favoured locations for transcription, and that promoters may be moved to these specific locations as transcription occurs (Sánchez-Romero *et al*., 2012). Localisation to the space surrounding the ribosome-free nucleoid would facilitate coupled transcription and translation, therefore being positioned within a plasmid could affect the accessibility of the promoter to RNAP and ultimately the level of gene expression (Robinow and Kellenberger, 1994; Azam *et al*., 2000; Woldringh, 2002). Therefore the activity of several promoters, with a range of activities, was compared when the promoters were located either on the bacterial chromosome or a multi-copy, RK2-based, *lac* expression vector.

5.1.2 Construction of single copy promoter::*lacZ* fusions by gene gorging

The "gene gorging" method (Herring *et al*., 2003) was exploited by Hollands (2009) to allow the endogenous *lac* promoter, of the *E. coli* K-12 chromosome, to be replaced with any promoter of interest (Figure 5.1). The pKH3 master donor plasmid was developed to allow

any promoter to be cloned between regions of homology to the *lacI* and *lacZ* genes, which are flanked by recognition sites for the I-SceI yeast endonuclease in the plasmid (Figure 5.1b). The recipient strain is co-transformed with the donor plasmid and the pACBSR mutagenesis plasmid, which encodes the I-SceI endonuclease and the λ-Red homologous recombination machinery under control of an arabinose-inducible promoter (Figure 5.1b). Addition of arabinose to the growth medium allows the donor plasmid to be linearised by the I-SceI endonuclease to provide a linear donor molecule for λ-Red-mediated recombination of the homologous regions with the chromosome, therefore replacing the endogenous *lac* promoter with the promoter of interest (Figure 5.1).

To allow simple Lac^+ or Lac^-, red-white, screening, Hollands created two starting strains to be used as recipients for the gene gorging system. The region of the *lacI* gene encoding the DNA binding domain of the Lac repressor was deleted in *E. coli* K-12 MG1655, to create the constitutively Lac^+ KH000 strain. Gene gorging of weak promoters into this strain allows recombinants to be selected as white or pink in a red background by selection on MacConkey lactose agar. The starting strain used in this study was *E. coli* K-12 MG1655 KH001, in which gene gorging method was used to replace the *lacZYA* regulatory region with the empty pKH3 cloning site in strain KH000 (Hollands, 2009). Therefore KH001 has a lac^- phenotype, allowing recombinants carrying the promoter of interest on the chromosome, fused to *lacZ,* to be identified by red/white selection on MacConkey lactose agar.

5.1.3 The *hcp, lac, melR* and *CC* promoters

Previous chapters of this study utilised one promoter at many different positions to study the effects of location on gene expression. Hence a range of promoters, with varying activities was selected to compare gene expression in single copy on the *E. coli* chromosome compared

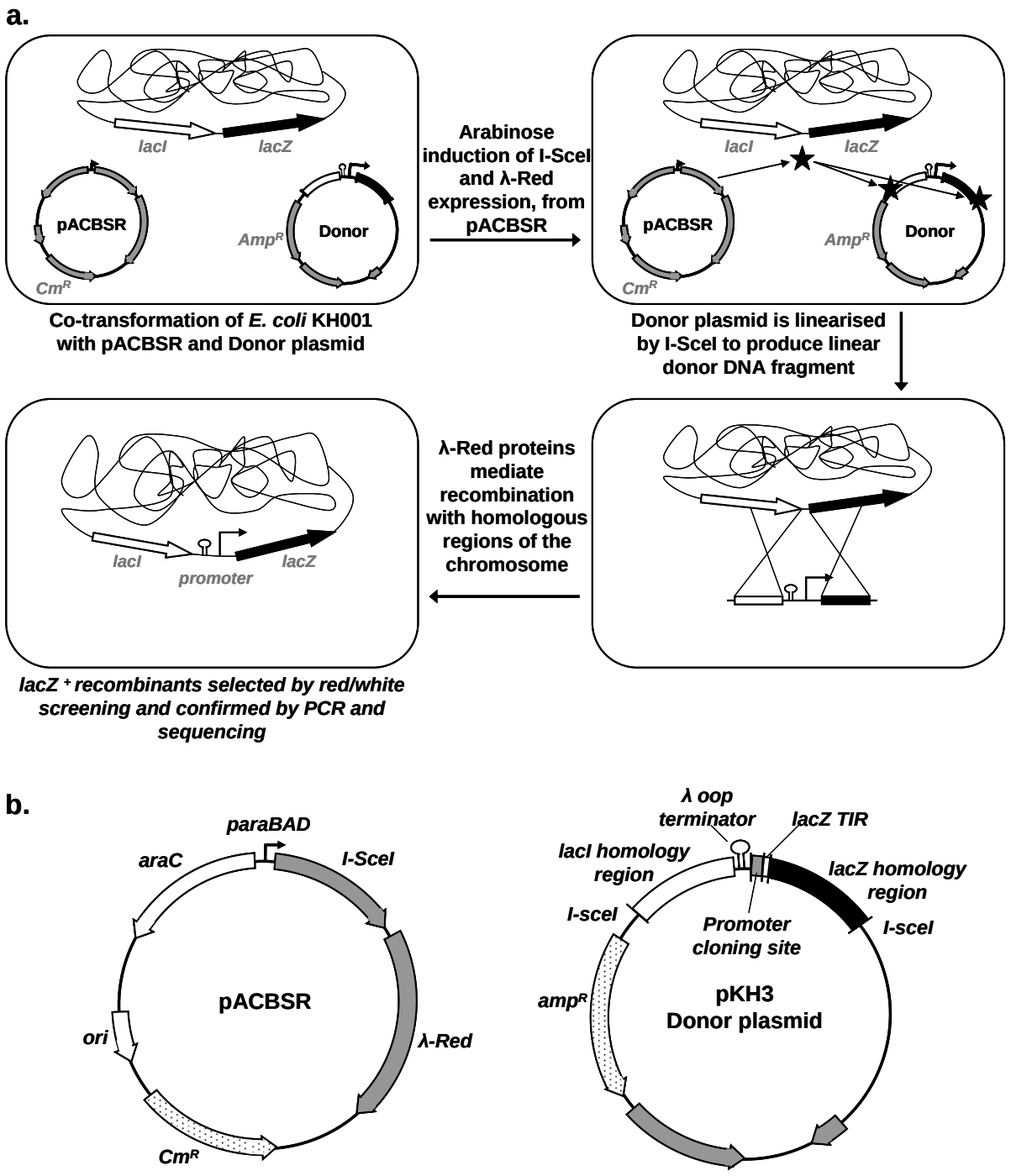

Figure 5.1: Diagram of chromosomal promoter::*lacZYA* fusion construction by gene gorging

a. Recipient *E. coli* MG1655 KH001 is co-transformed with the pACBSR recombineering plasmid and the pKH3 donor plasmid. Addition of arabinose to the growth medium induces expression of I-SceI endonuclease and the λ-Red recombination machinery from the pACBSR plasmid. The donor plasmid is linearised by the I-SceI endonuclease to allow λ-Red mediated recombination of the donor fragment with the homologous regions of the chromosome. Recombinants, carrying the promoter of interest fused to *lacZYA* on the chromosome, are identified by red/white selection on MacConkey lactose agar and verified by PCR amplification of genomic DNA and sequencing. **b.** Schematic diagram of pACBSR and pKH3 donor plasmids. Arrows represent genes and indicate the direction of transcription.

to that on a multi-copy plasmid. The *hcp* gene of *E. coli* encodes the Hybrid Cluster Protein (HCP), which is a cytoplasmic iron-sulphur protein that is thought to be involved in protection against reactive nitrogen species created as intermediates of nitrate metabolism (Filenko *et al*., 2005; Constantinidou *et al*., 2006; Filenko *et al*., 2007). The *hcp* promoter (*phcp*) is positively regulated, via a class I mechanism, by the FNR protein, which is a global transcription regulator of genes required for anaerobiosis. The promoter is also activated by the NarL and NarP proteins, which are the response regulators for the two-component regulatory systems required for response to nitrate and nitrite. The promoter is negatively regulated by the transcription factor NsrR, which is a nitrite sensing repressor that is also thought to respond to nitric oxide. Promoter *LacZ* fusion assays show that *phcp* has very high activity, therefore this promoter was used as one of the stronger promoters in this study and also to determine effects during anaerobic growth conditions (Rabin and Stewart, 1993; Filenko *et al*., 2005; Filenko *et al*., 2007; Chismon *et al*., 2011). The class I CRP-activated *lac* promoter also shows strong activity and is constitutively active in the *lacI*^{q} *E. coli* MG1655 KH000 strain, therefore this promoter was also used in this study (Hollands, 2009).

The *CC* promoters are synthetic derivatives of the *E. coli melR* gene promoter. Recall that *melR* encodes the MelR transcription factor responsible for regulation of the melibiose utilisation genes. The promoters are activated by CRP, acting at consensus binding sites located at positions -41.5 bp or -61.5 bp relative to the transcription start site. These promoters were chosen for this study because they are well charaterised, give a middle range of activity, were available in our lab, and allow direct comparison between effects on class I and class II activated promoters (Gaston *et al*., 1990). The wild-type *melR* promoter has much lower activity than the synthetically modified *CC* constructs and was also available from previous studies within our lab, therefore this was used to measure effects with weak promoters (Hollands, 2009). These promoters were inserted into the *E. coli* K-12 chromosome, at the *lac* site, in single copy to create promoter *lacZ* fusions by the gene

gorging method. Activities were then compared to that of the same fusions carried by a *lac* expression vector.

5.2 Construction of chromosome and plasmid-encoded promoter::*lacZ* fusions

5.2.1 Construction of chromosomal promoter::*lacZYA* fusions by gene gorging

To compare the activity of several promoters at two separate loci, a set of promoters with a range of activities, were first fused to *lacZYA* in single copy on the *E. coli* K-12 MG1655 KH001 chromosome by gene gorging (Figure 5.1). The promoters of interest can be fused to the *lacZ* homology region, encoded by the pKH3 plasmid, as a translation fusion, using the translation initiation region (TIR) of the promoter of interest. Alternatively fusions can be made as transcription fusions, using only the transcription initiation region of the promoter of interest and the TIR of the *lacZ* gene. All promoter fragments used were fused to *lacZYA* as transcription fusions to eliminate any potential complication due to variation in translation efficiency, except the *melR* promoter, which was used as a translation fusion due to the inactivity of the transcription fusion (Hollands, 2009).

The *phcp3831NS* fragment, carrying the regulatory region of the *hcp* gene, from position -125 bp to +11 bp relative to the transcription start site, was cloned into the *Eco*RI/*Bam*HI site of the pKH3 gene gorging vector as a transcription fusion (Figure 5.2). The NsrR DNA binding target site was inactivated in this promoter fragment, therefore abolishing repression by the NsrR transcriptional regulator (Chismon *et al*, 2011). A DNA fragment carrying the model class II CRP-activated *CC-41.5* promoter, from position -76 bp to +2 bp relative to the transcription start site, was cloned into the *Eco*RI/*Bam*HI site of pKH3, however the fragment was amplified to incorporate an upstream *Eco*RI target site and a downstream *Bgl*II target site,

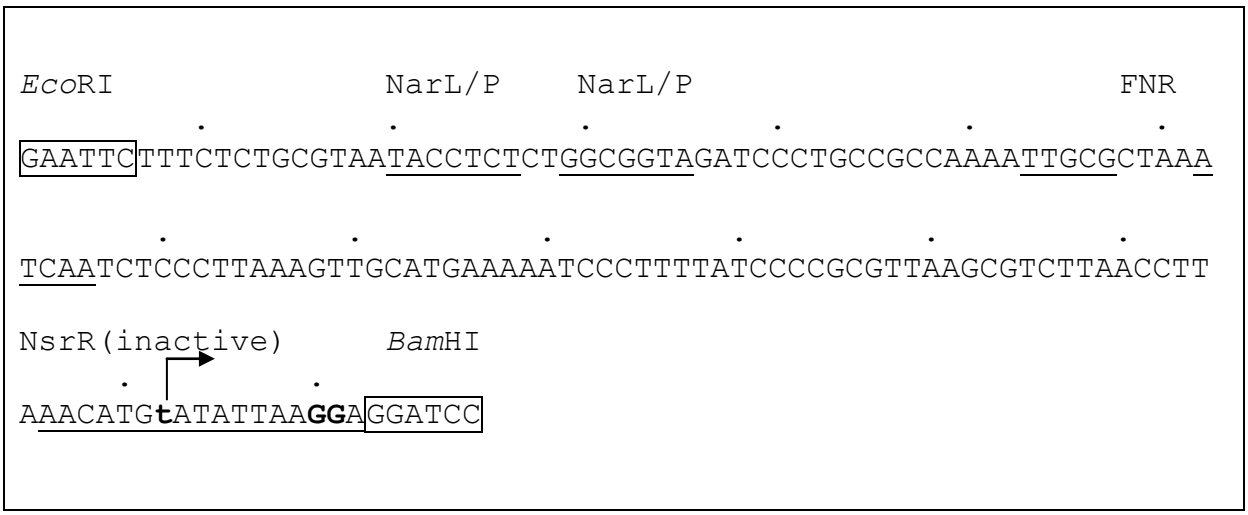

Figure 5.2: Sequence of the *phcp3831NS* promoter fragment

The *phcp3831NS* promoter fragment is an *Eco*RI-*Bam*HI fragment carrying the regulatory region of the *hcp* gene from position -125 bp to +11 bp relative to the transcription start site. Target sites for the restriction endonucleases *Eco*RI and *Hin*dIII are highlighted with borders and transcription factor DNA binding sites (NarL/P, FNR and inactivated NsrR) are underlined. The transcription start is indicated with a bold lower case t and an arrow. Mutations inactivating the NsrR site are bold.

which is compatible with *Bam*HI sticky ends. This approach had to be used due to the *CC* promoter fragments containing a *Bam*HI site downstream of the CRP target binding site (Figure 5.3). The same cloning procedure was utilised to insert the model class I CRP-activated *CC-61.5* promoter fragment into the *Eco*RI/*Bam*HI site of pKH3 as a transcription fusion (Figure 5.4). Finally the *TB10a* fragment, carrying the regulatory region for the *melR* gene from position -106 bp to +52 bp relative to the transcription start site, had already been cloned into the *Eco*RI/*Hin*dIII site of pKH3 as a translation fusion and used for gene gorging to produce the KH002 strain by Hollands (2009) (Figure 5.5). The translation fusion was used in place of the transcription fusion due to the transcription fusion having exceptionally weak activity, which was similar to that of the KH001 *lac*$^-$ starting strain (Hollands, 2009).

The pKH3 donor plasmids carrying the promoter::*lacZ* fusions were then used to insert the desired promoter into the chromosome of *E. coli* K-12 MG1655 KH001, to create single copy chromosome-encoded promoter::*lacZYA* fusions, by the gene gorging method (Figure 5.1; Section 2.12). Resulting strains were verified by PCR of genomic DNA and sequencing of the promoter region. Strains carrying the *phcp3831NS, CC-41.5, CC-61.5* and *TB10a* promoter fragments as transcription fusions to the *lacZYA* genes were named BRY01, BRY03, BRY05 and KH002 respectively (Table 2.1).

5.2.2 Construction of promoter::*lacZYA* fusions in the *lac* expression vector pRW500

To allow comparison between expression at the chromosomal *lac* locus and expression on a multi-copy plasmid, the chromosomally encoded promoter::*lacZYA* fusions were cloned into the pRW50 plasmid. However, the pRW50 *lac* expression vector encodes a *trpBA::lacZYA* fusion, therefore direct cloning of the amplified promoter fragments into this vector would not be a direct comparison, due to the *trpBA* fragment (Figure 2.1; Figure 5.6). The *lacZ* gene

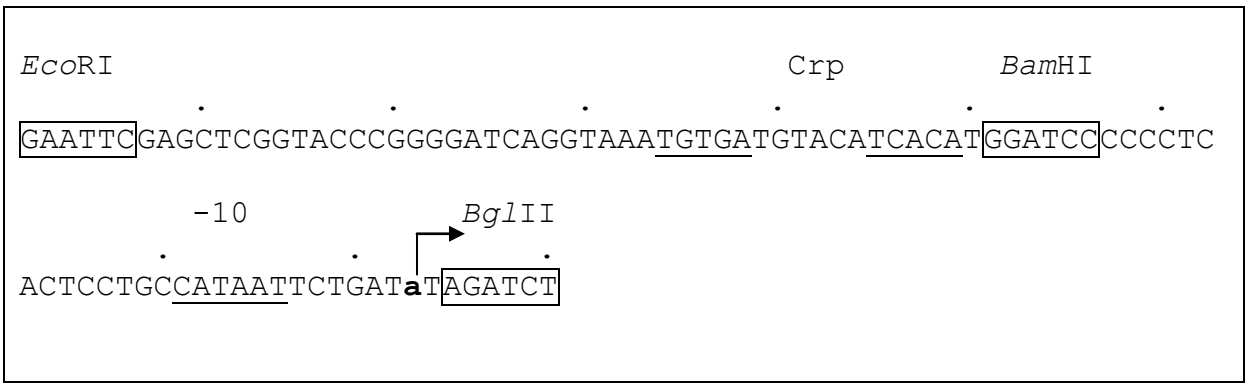

Figure 5.3: Sequence of the *CC-41.5* promoter fragment

The *CC-41.5* promoter fragment is an *Eco*RI-*Bgl*II fragment carrying the *CC-41.5* promoter, a synthetic model CRP regulated promoter based on the regulatory region of the *melR* gene. The CRP DNA binding site is underlined and centred at -41.5 bp relative to the transcription start site, which is indicated by a bold lower case a and an arrow. Restriction enzyme recognition sites are highlighted with borders.

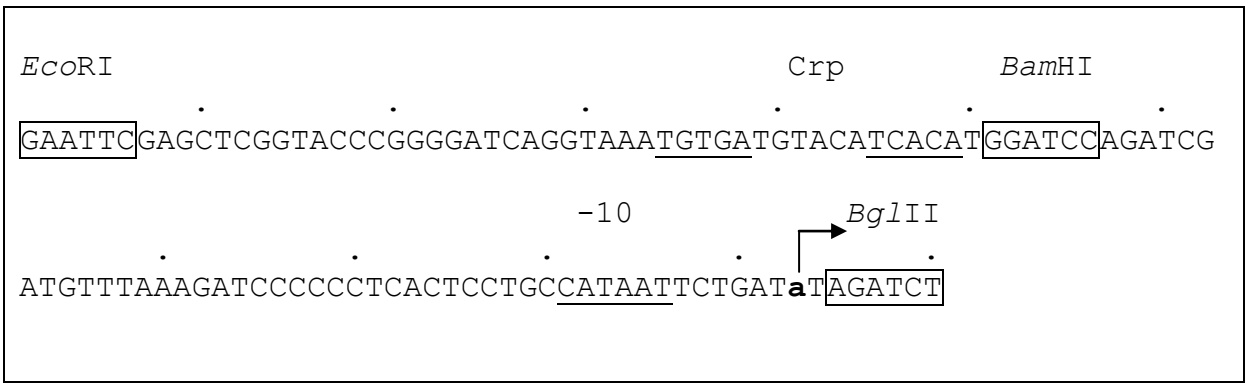

Figure 5.4: Sequence of the *CC-61.5* promoter fragment

The *CC-61.5* promoter fragment is an *Eco*RI-*Bgl*II fragment carrying the *CC-61.5* promoter, a synthetic model CRP regulated promoter based on the regulatory region of the *melR* gene. The CRP DNA binding site is underlined and centred at -61.5 bp relative to the transcription start site, which is indicated by a bold lower case a and an arrow. Restriction enzyme recognition sites are highlighted with borders.

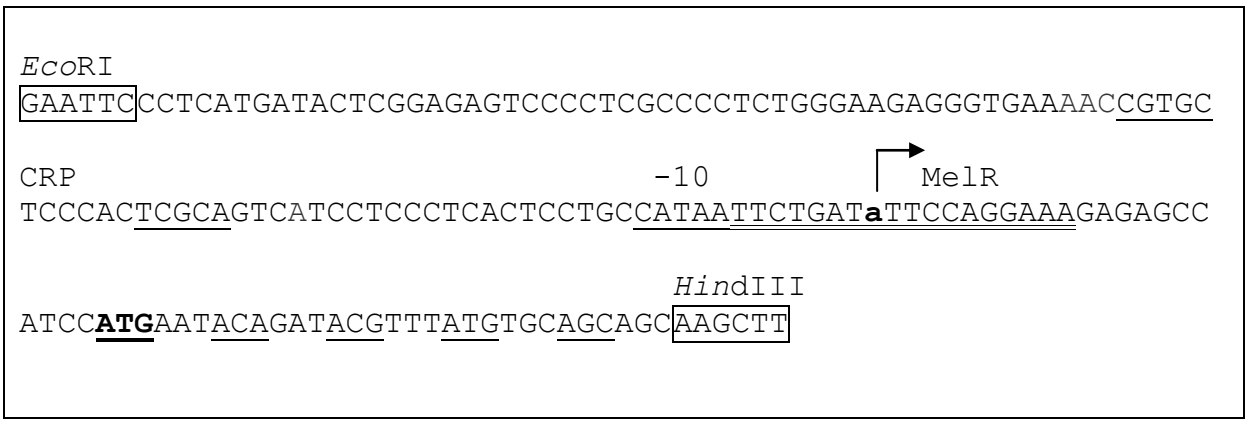

Figure 5.5: Sequence of the *TB10a* promoter fragment

The *TB10a* promoter fragment is an *Eco*RI-*Hin*dIII fragment carrying the regulatory region of the *melR* gene and was designed to create a translational fusion in pKH3. The CRP activator binding site is underlined and the MelR repressor binding site is double underlined. The promoter -10 is underlined and labelled, with the transcription start indicated by a lower case bold a and an arrow. The ATG start codon is in bold and underlined. Restriction enzyme recognition sites are highlighted by borders and are labelled.

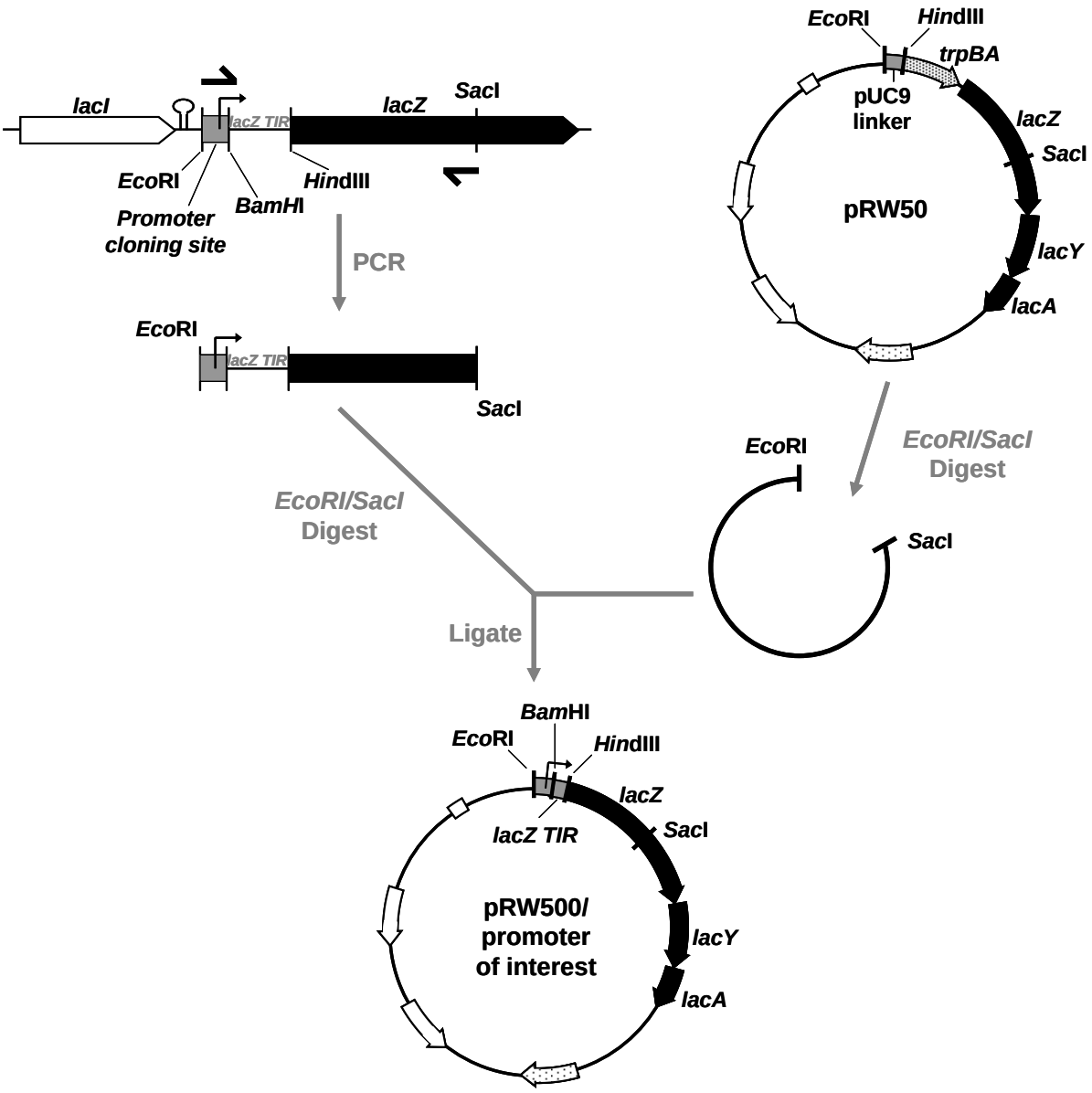

Figure 5.6: Schematic diagram of pRW500 plasmid and promoter::*lacZYA* fusion construction

Schematic diagram representing basic protocol for construction of the pRW500 plasmid or pRW500/promoter of interest fusions. Arrows represent genes, while black lines denote intergenic regions (approximately to scale). Direction of transcription of each gene is indicated by the arrow heads. DNA modifying enzyme target sites are represented by vertical black lines and are labelled. The *Eco*RI-*SacI* fragment at the *lac* locus on the chromosomes of strains MG1655, BRY01, BRY03, BRY05, KH002, and KH001 were amplified by PCR and cloned into the pRW50 plasmid. This cloning procedure removes the *trpBA* fragment of the pRW50 plasmid to create promoter::*lacZYA* fusions.

encodes a target site for the *SacI* restriction endonuclease, which is unique in the pRW50 plasmid. Therefore promoter fragments were amplified by PCR from BRY01, BRY03, BRY05 and KH002 genomic DNA with primers designed to incorporate the upstream *Eco*RI site and the *SacI* site within the *lacZ* gene. The resulting *Eco*RI-*Sac*I promoter fragments were cloned into the pRW50 *lac* expression vector to create plasmid encoded versions of the promoter::*lacZYA* fusions described in section 5.2.1 (Figure 5.6).The same procedure was used to clone the empty pKH3 cloning site of the KH001 strain or the wild-type *lacZYA* regulatory region into the pRW50 vector to create the empty pRW500 plasmid, to be used as a control, and pRW500 carrying the *lacZYA* operon (Table 2.2; Figure 5.6).

5.3 Comparison of activities of chromosome or plasmid-located promoters

The effect of position on gene expression was tested by comparison of promoter activities encoded on the *E. coli* K-12 MG1655 chromosome with that on the multi-copy *lac* expression vector pRW500. The *E. coli* KH001, KH002, BRY01, BRY03 and BRY05 strains, carrying the pKH3 empty cloning site, *TB10a, phcp3831NS, CC-41.5* or *CC-61.5* promoter fragments fused to the *lacZYA* operon were transformed with the empty pRW500 plasmid as a control. The *E. coli* KH001 strain was also transformed with pRW500 carrying the *TB10a, phcp3831NS, CC-41.5* or *CC-61.5* promoter fragments fused to the *lacZYA* operon encoded by the plasmid. Cells were then grown to mid-late exponential growth phase in LB broth supplemented with tetracycline for plasmid maintenance and promoter activities were derived from measured β-galactosidase activities, as outlined in section 2.15.

Expression of *lacZ* by the *phcp3831NS* fragment promoter was approximately 9-fold greater when encoded by the pRW500 plasmid when compared to the chromosomally encoded copy of the fusion (Figure 5.7; Table 5.1). Activity of the class I, CRP activated, *CC-61.5* fragment

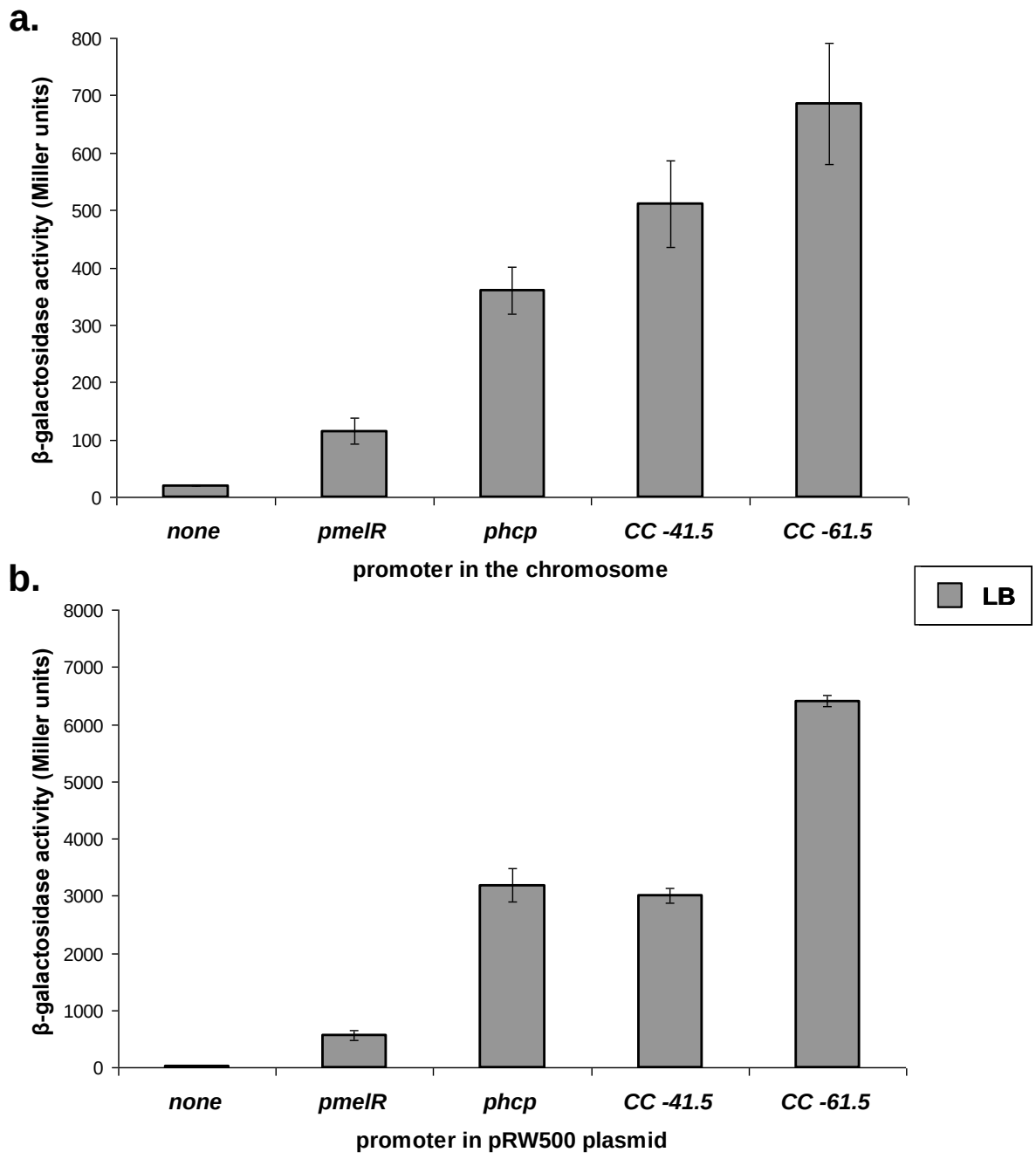

Figure 5.7: Measured expression driven by different promoters at the chromosomal *lac* locus and on the pRW500 plasmid

The figure shows β-galactosidase activities measured in (**a.**) strain KH001 carrying different promoter::*lacZ* fusions cloned in pRW500 or (**b.**) the KH002 (*pmelR*), BRY01 *(phcp3831NS*), BRY03 (*CC-41.5*) and BRY05 (*CC-61.5*) strains encoding the same promoter::*lacZYA* fusions on the chromosome and carrying the empty pRW500 plasmid as a control. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{650} 0.3-0.5) in LB broth supplemented with tetracycline. Data shown are averages of β-galactosidase activity measurements from at least three independent experiments, and error bars show one standard deviation from the mean. The KH001 strain (no promoter at the chromosomal *lac* locus) transformed with the empty pRW500 plasmid was included as a control ("none").

Table 5.1: Summary of promoter activities at the chromosomal *lac* locus or the pRW500 plasmid

The table lists β-galactosidase activities (in Miller units) measured in strain KH001 carrying different promoter::*lacZ* fusions cloned in pRW500 or the KH002 (*pmelR*), BRY01 *(phcp3831NS*), BRY03 (*CC-41.5*) and BRY05 (*CC-61.5*) strains encoding the same promoter::*lacZYA* fusions on the chromosome and carrying the empty pRW500 plasmid as a control. Cells were grown aerobically in LB medium at 37°C to mid-exponential phase (OD_{650} 0.3-0.5). Data listed are averages from at least three independent experiments, shown ± one standard deviation. The KH001 strain (no promoter at the *lac* locus on the chromosome) transformed with the empty pRW500 plasmid was included as a control ("none").

		β-galactosidase activity (Miller units ± one standard deviation)			
Growth media	**Promoter**	**Chromosome**	**Plasmid**	**Ratio**	**Promoter Class**
LB	None	20 ± 0.8	20 ± 0.8	-	-
	phcp (phcp3831NS)	360 ± 40	3187 ± 290	8.9 ± 0.02	I
	CC-61.5tx	685 ± 105	6410 ± 93	9.4 ± 0.02	I
	CC-41.5tx	511 ± 75	2999 ± 128	5.9 ± 0.03	II
	pmelR (TB10a)	115 ± 22	570 ± 89	5 ± 0.05	II

promoter was also approximately 9-fold greater when the fusion was encoded by the pRW500 multi-copy plasmid compared to that at the *lac* locus on the chromosome (Figure 5.7 Table 5.1). However, activity of the class II CRP-activated *CC-41.5* fragment promoter encoded by the pRW500 plasmid was approximately 6-fold greater than at the *lac* locus on the chromosome.

Activity of the *CC-41.5* promoter was approximately 1.5-fold greater than the *phcp3831NS* promoter fragment when positioned on the chromosome, however activities of the two promoters were similar when encoded by the plasmid (Figure 5.7 Table 5.1). Expression of *lacZ* by the class II, CRP-activated, *TB10a* fragment *melR* promoter was approximately 5-fold increased compared to that of the chromosomally encoded *TB10a::lacZYA* fusion. Activity of chromosomally encoded *TB10a* was approximately 3-fold weaker than that of the *phcp3831NS* promoter fragment, whereas the plasmid-encoded *TB10a* promoter was approximately 6-fold weaker than the *phcp3831NS* promoter (Figure 5.7; Table 5.1).

Comparison of the chromosomally encoded class II *CC-41.5* and the class I *CC-61.5* promoters revealed that the class I activated promoter was approximately 34% stronger than the class II, whereas the plasmid encoded class I promoter was approximately twice as strong as the class II activated promoter (Figure 5.7; Table 5.1). The pRW500 plasmid is derived from the broad host range RK2 plasmid and is expected to be present at approximately six to eight copies per cell (Figurski *et al*., 1979; Kues and Stahl, 1989; Lodge *et al*., 1992). Therefore these results suggest that activity of the class I activated *phcp3831NS* and *CC-61.5* promoter fragments increased by the expected fold change. In contrast, the fold increase in expression between chromosomally encoded class II CRP-activated *TB10a* and *CC-41.5* promoter fragments, compared with that encoded by the pRW500 plasmid, was slightly less than expected (Figure 5.7; Table 5.1). This suggests that the class II CRP-activated promoters either have improved expression when

encoded by the chromosome at the *lac* locus or decreased expression when encoded by the multi-copy pRW500 plasmid.

5.4 Effect of position of *phcp* activity under optimal conditions

The *phcp3831NS* promoter is activated by the FNR transcription factor, consequently activity of the promoter is enhanced in response to anaerobiosis, allowing comparison of the chromosome and plasmid-located promoter at different levels of induction (Chismon *et al*., 2011). Therefore activity of the *phcp3831NS* fragment promoter, encoded either by the BRY01 strain chromosome or the pRW500 plasmid, was tested during aerobic or anaerobic growth. The BRY01 strain was transformed with the empty pRW500 plasmid and the KH001 strain transformed with pRW500 carrying the *phcp3831NS* fragment fused to the *lacZYA* operon. Cells were grown to mid-late exponential phase (OD_{650} 0.3-0.5) in minimal salts medium supplemented with 10% LB, glycerol, sodium fumarate and tetracycline under aerobic or anaerobic conditions at 37˚C.

Activity of the *phcp3831NS* fragment promoter was induced approximately 4.7-fold under anaerobic conditions, compared to the promoter activity during aerobic growth, when encoded on the chromosome and approximately 4.3-fold when carried by the pRW500 multi-copy plasmid (Figure 5.8; Table 5.2). This result suggests that activation of the promoter by FNR is unaffected by the change in location between chromosome and plasmid. Expression of *lacZ* by the *phcp3831NS* fragment promoter, carried by the pRW500 plasmid, was approximately 8.2-fold greater than that of the chromosomally encoded fusion during aerobic growth and approximately 7.5-fold greater during anaerobic growth (Figure 5.8; Table 5.2). The ratio between chromosomally encoded *lacZ* expression and that of the plasmid was similar to that of the promoter::*lacZ* fusion during aerobic growth in LB and was within the expected change

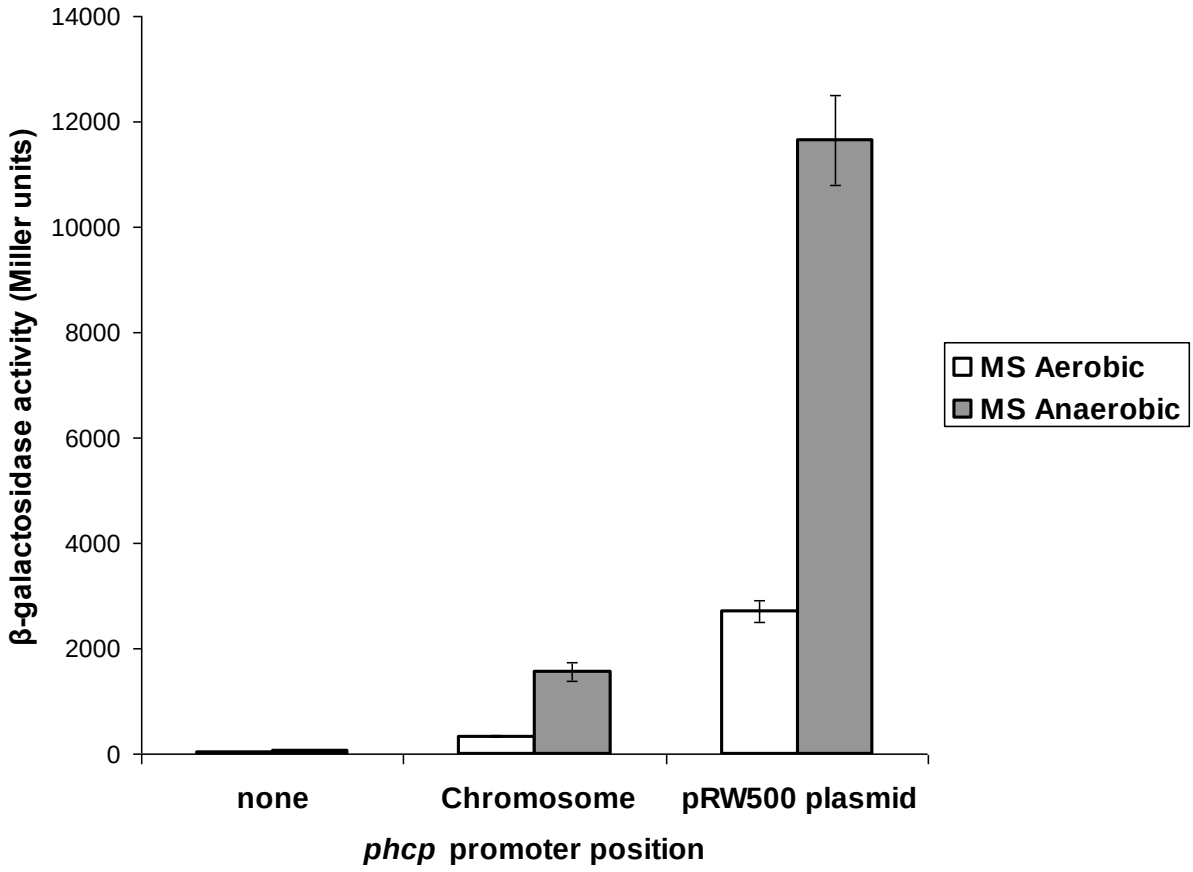

Figure 5.8: Activity of the *hcp3831NS* promoter encoded by the chromosome or pRW500 plasmid

The figure shows β-galactosidase activities measured in strain BRY01 transformed with empty pRW500, as a control, or strain KH001 transformed with pRW500 carrying the *phcp3831NS* fragment fused to *lacZ*. Cells were grown aerobically or anaerobically at 37°C to mid-exponential phase (OD_{650} 0.3-0.5) in minimal salts media supplemented with 10% LB, glycerol, sodium fumarate and tetracycline for maintenance of the pRW500 plasmid. Data shown are averages of β-galactosidase activity measurements from at least three independent experiments, and error bars show one standard deviation from the mean. The KH001 strain (no promoter at the *lac* locus on the chromosome) transformed with the empty pRW500 plasmid was included as a control ("none").

Table 5.2: Summary of *hcp3831NS* promoter activity encoded at the chromosomal *lac* locus or by the pRW500 plasmid

The table lists β-galactosidase activities (in Miller units) measured in strain BRY01 transformed with empty pRW500, as a control, or strain KH001 transformed with pRW500 carrying the *phcp3831NS* fragment fused to *lacZ*. Cells were grown aerobically or anaerobically at 37°C to mid-exponential phase (OD_{650} 0.3-0.5) in minimal salts media supplemented with 10% LB, glycerol, sodium fumarate and tetracycline for maintenance of the pRW500 plasmid. Data listed are averages from at least three independent experiments, shown ± one standard deviation. The KH001 strain (no promoter at the *lac* locus on the chromosome) transformed with the empty pRW500 plasmid was included as a control ("none").

		β-galactosidase activity (Miller units ± one standard deviation)		
Growth media	**Promoter**	**Chromosome**	**Plasmid**	**Ratio**
MS + 10% LB Aerobic	*phcp3831NS*	329 ± 14	2708 ± 205	8.2 ± 0.01
MS + 10% LB Anaerobic	*phcp3831NS*	1556 ± 167	11647 ± 863	7.5 ± 0.02
Ratio	-	4.7 ± 0.02	4.3 ± 0.03	-

based on plasmid copy number (Figure 5.7; Figure 5.8; Table 5.2; Figurski *et al*., 1979; Kues and Stahl, 1989; Lodge *et al*., 1992). These results suggest that regulation of the class I activated *hcp* promoter is unaffected by change in location from the chromosome to the multi-copy plasmid.

5.5 Comparison of chromosome or plasmid-located *lac* promoter activity

The *lac* promoter is activated by CRP acting by a class I activation mechanism and this activation is repressed by the presence of glucose in the medium (Hollands, 2009). Hence, growth on LB broth or M9 minimal medium supplemented with fructose allowed analysis of *lac* promoter activity at different levels of induction. Therefore KH000 was transformed with the empty pRW500 plasmid and the Lac$^-$ KH001 strain was transformed with pRW500 carrying the *lacZYA* operon including the regulatory region from -93 bp relative to the transcription start site.

Activity of the *lac* promoter, encoded either by the chromosome or the multi-copy plasmid, was measured during mid-late exponential growth phase (OD_{650} 0.3-0.5) in LB broth or M9 minimal media supplemented with tetracycline. Activity of the *lac* promoter encoded by the chromosome of strain KH000, during growth on M9 minimal media, was increased approximately 3.2-fold compared to that during growth on LB broth (Figure 5.9; Table 5.3). However, activity of the promoter carried by the pRW500 plasmid was only increased approximately 2.6-fold during growth on M9 minimal media, compared to that during growth on LB broth (Figure 5.9; Table 5.3). This difference in glucose repression of the *lac* promoter at the chromosomal and plasmid loci is minimal and could potentially be due to inaccuracies caused by the extremely high levels of activity achieved when expressed from the plasmid

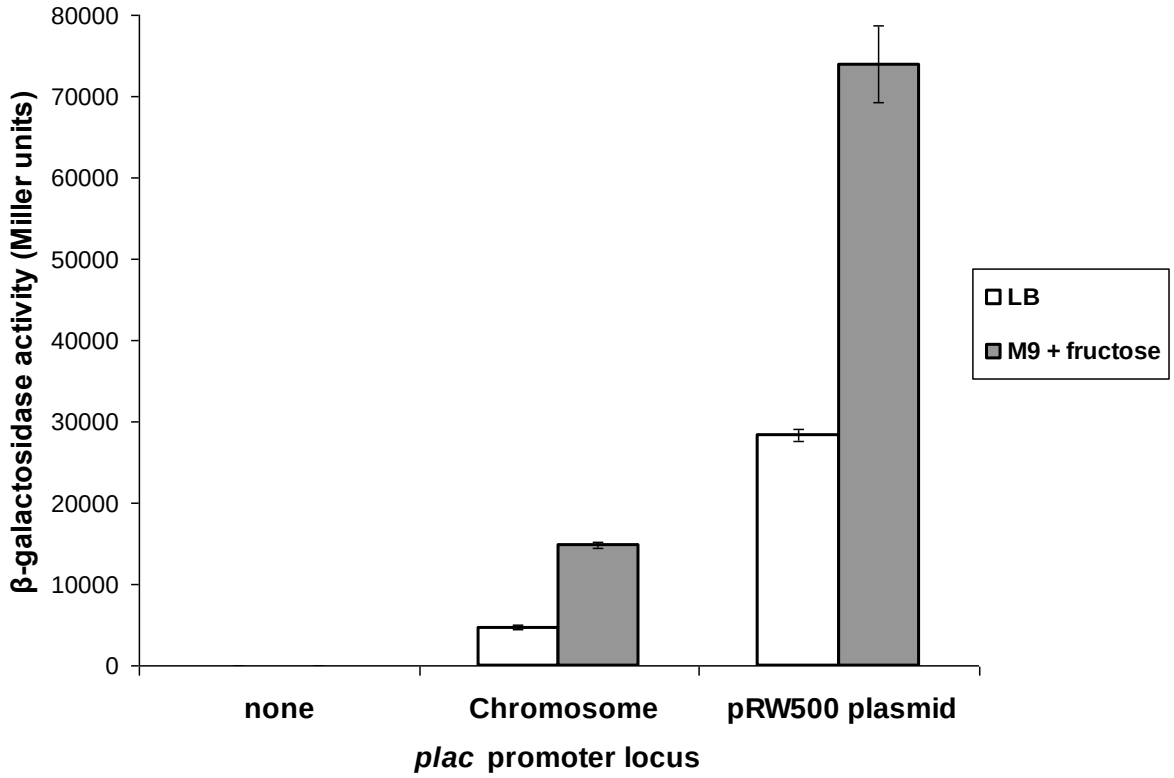

Figure 5.9 Activity of the *lac* promoter encoded by the chromosome or pRW500 plasmid

The figure shows β-galactosidase activities measured in strain KH000 transformed with empty pRW500, as a control, or strain KH001 transformed with pRW500 carrying the *lacZYA* operon including the *lac* promoter region. Cells were grown aerobically at 37°C to mid-exponential phase (OD_{650} 0.3-0.5) in LB broth or M9 minimal media supplemented with fructose and tetracycline for maintenance of the pRW500 plasmid. Data shown are averages of β-galactosidase activity measurements from at least three independent experiments, and error bars show one standard deviation from the mean. The KH001 strain (no promoter at the *lac* locus on the chromosome) transformed with the empty pRW500 plasmid was included as a control ("none").

Table 5.3: Summary of *lac* promoter activity encoded at the chromosomal *lac* locus or by the pRW500 plasmid

The table lists β-galactosidase activities (in Miller units) measured in strain KH000 transformed with empty pRW500, as a control, or strain KH001 transformed with pRW500 carrying the *lacZYA* operon including the *lac* promoter region. Cells were grown aerobically in LB broth or M9 minimal medium supplemented with fructose and tetracycline at 37°C to mid-exponential phase (OD_{650} 0.3-0.5). Data listed are averages from at least three independent experiments, shown ± one standard deviation. The KH001 strain (no promoter at the *lac* locus on the chromosome) transformed with the empty pRW500 plasmid was included as a control ("none").

		β-galactosidase activity (Miller units ± one standard deviation)		
Growth media	**Promoter**	**Chromosome**	**Plasmid**	**Ratio**
LB	*plac*	4662 ± 310	28375 ± 698	6.1 ± 0.01
M9 + 0.3% fructose	*plac*	14805 ± 311	73955 ± 4729	5 ± 0.01
Ratio	-	3.2 ± 0.02	2.6 ± 0.03	-

locus during growth on M9 minimal media. Growth in this situation was slower than un-induced cells and β-galactosidase activity was detected in the supernatant.

Expression of *lacZYA*, during growth on LB broth, was approximately 6-fold stronger when encoded by the pRW500 plasmid compared to that at the chromosomally encoded locus (Figure 5.9; Table 5.3). The ratio of *lacZ* expression between the chromosome and plasmid was at the lower end of the expected range during growth on LB broth, but below during growth on minimal media. Slower growth and the presence of β-galactosidase in the supernatant suggests that the cells grown on minimal media and carrying the *lacZYA* operon on the plasmid were stressed and potentially undergoing lysis, due to the massive production of the *lacZYA* gene products. The exceptionally strong activity of the *lac* promoter may have exceeded the limit of that capable of being measured through β-galactosidase assay, therefore potentially skewing the ratios measured.

5.6 Discussion

Promoter activity can be affected by position within the folded chromosome of *E. coli* through mechanisms such as variation of DNA superhelicity, neighbouring gene expression, local chromosomal architecture and neighbouring sequence elements (Chapter 4). Therefore the work presented here investigated the effect of positioning promoters on an extra-chromosomal piece of DNA, an RK2-based *lac* expression vector, on promoter activity. The expression of *lacZ* by the class I activated promoters, *hcp* and *CC-61.5*, was increased by the expected amount due to the copy number of the pRW500 plasmid being approximately six to eight per genome (Figurski *et al*., 1979; Kues and Stahl, 1989; Lodge *et al*., 1992). However the fold difference in class II-activated promoter activity between the chromosome and plasmid loci was much lower than that of the class I-activated promoters. These results suggest that the class II-activated

promoters either have increased activity when encoded by the chromosome or decreased activity when encoded by the multi-copy pRW500 plasmid.

Differences in *lac* promoter activity between the chromosomally encoded fusion and the plasmid fusion were smaller, giving a change of approximately 5-fold during growth on M9 minimal medium and approximately 6-fold during growth on LB broth. The difference in *hcp* promoter activity, between the chromosomally encoded fusion and the plasmid encoded copy, was also less during growth on minimal medium compared to that in LB broth. Growth on minimal medium is slower than in the relatively rich LB broth and could cause there to be less copies of the plasmid per cell (Kolatka *et al*., 2010). Therefore this could account for the lower fold-difference in *lac* and *hcp* promoter activities during growth on minimal medium compared to that on LB broth. The ratio of activity between the chromosome and plasmid-located copies of the *lac* promoter was smaller than that of the model class I, CRP-activated, *CC-61.5* promoter and similar to that of the *CC-41.5* and *TB10a* fragment promoters. However this lower ratio could well be due to the strength of the *lac* promoter exceeding that of the maximum possible promoter activity measureable through β-galactosidase activity assay. This was shown by poor growth and cell lysis, indicated by the presence of β-galactosidase in the culture supernatant.

Differences in the effects of position on different promoters are unlikely to be due to variation in plasmid copy number, as the strains used were all of the same genetic background and the growth conditions were the same. Hence these variations must be due to some other factor, such as plasmid mobility, accessibility by RNAP to the promoter or a change of superhelical density. The *lacZYA* locus on the chromosome is tethered to the rest of the chromosome, which is known to be highly structured and compacted with specific macrodomains locating to specific home positions within the cell (Espeli *et al*., 2008). However, the pRW500 plasmid should be relatively mobile within the cell, which is evidenced by the fact that fluorescently

tagged pRW50 derivative plasmids in *E. coli* do not form specific fluorescent foci, suggesting that they are diffuse (Sánchez-Romero *et al*., 2012). Further to this, active transcription of a highly expressed plasmid can drive plasmids to co-localise to positions on the periphery of the *E. coli* nucleoid, suggesting that there are preferred locations within the cell for transcription to take place (Sánchez-Romero *et al*., 2010). Therefore this increased mobility and localisation to the edge of the nucleoid could account for variation, if accessibility to the promoter on the chromosome is affecting the promoters differently. However this is unlikely as the *CC* promoters are regulated by the same transcription factors, yet the effect of changing the position from chromosome to plasmid is different between class I and class II CRP-activated promoters.

Activation of the *CC* promoters by CRP may respond differently to variation in negative superhelicity due to the fact that the mechanisms of class I and class II activation are different. Class I activation at the *CC-61.5* promoter acts via simple recruitment of RNAP through a single protein-protein interaction to yield the RNAP-promoter closed complex. Whereas, class II CRP-dependent activation, at the *CC-41.5* promoter, acts via binding of CRP to a DNA site overlapping the -35 element. Bound CRP acts to recruit RNAP, as with class I activation, but also facilitates post-recruitment promoter open complex formation (Lawson *et al*., 2004). Further to this, to produce the model for class II CRP-dependent activation, large changes in DNA geometry within the downstream half-site of the CRP DNA site were required to allow the interactions that facilitate isomerisation of the closed complex (Lawson *et al*., 2004). These differences in the class I and class II CRP-dependent activation mechanisms could cause them to be differentially responsive to variation in DNA superhelical density. Therefore any variation between superhelicity at the chromosomal *lac* locus and the promoter cloning site of pRW500, could differentially affect activity of the different promoters used in this study.

Chapter 6:

Final discussion

To date, study of transcriptional regulation in bacteria has focussed on individual promoters and the complex mechanisms of interaction between different transcription factors. These studies have yielded vast quantities of information about the mechanisms by which internal and external signals are integrated into the transcriptional response at individual promoters. However, only a handful of studies have attempted to study the regulation of promoters in their natural context within the chromosome. Further to this, little has been done to quantify global effects on transcription and to assess the effects of position within the chromosome on transcriptional output. The extreme compaction of the bacterial chromosome has been expected to have an effect on the process of gene expression since the first studies of the bacterial nucleoid in the 1970's. However, until recently, only 3 studies had directly assessed the effect of position within the chromosome on gene expression, therefore the work described in this study sought to characterise these effects and any potential mechanisms behind them.

6.1 Promoter activity is dependent upon position within the chromosome

Previous studies by Beckwith *et al*. (1966), Schmid and Roth, (1987) and Sousa *et al*. (1997) attempted to address the question of whether position within the chromosome can affect gene expression (reviewed in chapter 1). These studies reported no effect of location other than that expected due to increased copy number of gene loci in proximity to the actively replicating origin, *oriC*. In contrast, the results presented in this thesis report a 310-fold variation in promoter activity, which is dependent upon position within the chromosome and unrelated to the gene dosage effect. This raises the question as to why this level of position-dependent modulation of promoter activity was not previously reported.

Details of the experimental differences between the study presented here and previous work are discussed in section 3.8.2 and 3.8.3. The major difference appeared to be exposure of the promoter sequence, used in this study, to neighbouring local sequences. The gene doctoring homologous recombination system, used here, required homology to any desired region of the chromosome, unlike previous studies which used transposition systems. This allowed minimal amounts of flanking sequence to be used in this study, therefore giving greater exposure to the local sequence elements. As was reported in chapter 4, features of the local sequence at each of the different locations can affect promoter activity. For example, locally available DNA superhelicity played a major role in the high activity seen at the *asl* and *nupG* loci, whereas this effect may have been blocked if insulating sequence were included.

DNA superhelicity appeared to be the causative agent of a large portion of the variation observed, either through the natural differences across the genome or due to the changes induced by neighbouring transcription. The study of Schmid and Roth, (1987) utilised transposition of a reporter operon in *S. typhimurium* to characterise position-dependent effects on gene expression, however the only measurable effects were due to gene dosage. Champion and Higgins, (2007) demonstrated that the average supercoil density of DNA isolated from *S. typhimurium* was approximately 13% lower than that of *E. coli*. This data suggests that the *S. typhimurium* chromosome is generally more relaxed than that of *E. coli*. Therefore, there may be less variation in superhelical density, which could account for the smaller differences in the results of Schmid and Roth (1987) and those presented here.

During production of this thesis, a study was published by Block *et al*. (2012), in which a promoter::*yfp* fusion was inserted at a total of 8 different chromosomal loci within *E. coli*. The results of Block *et al*. (2012) demonstrated no significant effect of position on promoter activity outside of that expected by gene dosage. However, the fusion used was flanked by transcriptional terminators and the experiment was repeated without these. Removal of the

insulating terminators saw a change in the level of expression at three of the loci tested, however these effects were attributed to "local sequence effects", with no further investigation. It was suggested that the effects were potentially due to antisense transcription of the reporter from neighbouring operons, which would interfere with transcription of the target gene and reduce its expression. Due to an oversight in design of the probe, the system reported here did not incorporate a downstream transcription terminator beyond the reporter gene, however neighbouring gene terminators were preserved, where possible (Chapter 2). This could be a shortcoming of this study, however there are several lines of evidence which argue to the contrary. Firstly, no change in fluorescence or *egfp* mRNA level was detected in the absence of IPTG and the levels of *egfp* mRNA were approximately the same in the absence of inducer at several loci. This data demonstrated that the *egfp* reporter gene was not being transcribed by read-through from neighbouring operons. Further to this, no un-induced activity is reported during the investigation of neighbouring promoter activity effects in section 4.3 of chapter 4. During this experiment, the activity of the *lac28::egfp* fusion and the *melAB* operon was analysed with the promoter::reporter fusion placed immediately downstream of the *melAB* operon. Activity of both reporters was only detected in the presence of the corresponding inducer, but not in the control experiments, therefore indicating no read-through from either promoter (Figure 4.6).

The gene regulatory element used by Block *et al*. (2012) was a variant of the exceptionally strong P_L promoter of phage lambda, which was modified to be regulated by the Lac repressor, therefore allowing induction of promoter activity by the addition of IPTG to the medium (Lutz and Bujard, 1997). The P_L promoter is strong in the absence of any activator and will bind RNAP efficiently. This is unlike the *lac* promoter of *E. coli*, used in this study, which requires the assistance of CRP to activate transcription by recruitment of RNAP for binding to the promoter (Lutz and Bujard, 1997). Supposing that some locations within the cell are more or less conducive to transcription, due to variations in locally available RNAP,

the P_L promoter would be less sensitive to these effects than the *lac* promoter. This is because the P_L promoter does not require the simultaneous presence of both RNAP and an activator, like the *lac* promoter, for transcription initiation. This difference in the rate limiting steps of promoter activation could explain why position-dependent effects are not observed by Block *et al*. (2012), but were detected in this study. Further to this, phage DNA can be inserted into the host chromosome at secondary integration sites during the phage lysogenic pathway, as opposed to the primary integration site *attB* (Nash, 1981). Therefore the phage promoters must be more resistant to position-dependent modulation of promoter activity, due to their natural role in phage DNA transcription at different chromosomal integration sites.

6.2 DNA supercoiling: an important factor in position-dependent modulation of promoter activity

The large variation of promoter activity between different chromosomal loci, reported in chapter 3, was greater than can be explained by the gene dosage model of Cooper and Helmstetter (1968). Further to this, the difference in copy number between the origin and terminus during growth on minimal media, as measured by qRT-PCR, did not exceed a 1.5-fold difference (Figure 3.13). However, a general trend was observed in which promoter activity was found to increase with proximity to the origin of replication, despite this effect being unrelated to gene dosage (Section 3.5.4). When the promoter was positioned at either of two ori-proximal loci, *asl* and *nupG*, activity was found to be approximately 11-fold greater than activity of the promoter positioned near to the *dif* site (Figure 3.7). Further investigation into the mechanism behind this high level of promoter activity discovered that it was reliant upon the activity of DNA gyrase, with the activity being reduced by gyrase inhibition (Section 4.4). These results suggest that the ori-proximal increase of promoter activity is due to

variation of locally available DNA superhelicity by the action of DNA gyrase. Further support for an *oriC* to *dif* gradient of superhelical density comes from the observation made by Sobetzko *et al*. (2012), which highlighted an increase in the density of DNA gyrase binding sites with increasing proximity to the origin of replication. This gradient of DNA gyrase binding sites is expected to reflect a gradient of superhelical density from the origin of replication to the terminus. The work described in this study has demonstrated the importance of this gradient of DNA supercoiling for position-dependent modulation of promoter activity.

Section 3.7 of this study describes experiments in which induction of promoter activity at different chromosomal positions was analysed during stationary phase. The patterns of position-dependent activity were found to be drastically different to those during exponential growth, with no ori-proximal increase in promoter activity. The 11-fold difference in activity between the promoter at the *oriC*-proximal *asl* locus and the *dif*-proximal loci was lost during stationary phase (Figure 3.21). Loss of the *dif* to *oriC* increase in promoter activity is likely to be due to changes in DNA supercoiling, heralded by entry into stationary phase. DNA gyrase is known to be inhibited upon entry into stationary phase and DNA isolated from cells in stationary phase is more relaxed than during exponential growth (Balke and Gralla, 1987; Nakanishi *et al*., 1998; Baquero *et al*., 1995; Oh *et al*., 2001). The loss of the *oriC*-proximal increase in promoter activity, upon entry into stationary phase, is most likely due to inhibition of the action of DNA gyrase, by the gyrase inhibitor SbmC, and general relaxation of DNA superhelicity. This result provides yet more evidence that DNA supercoiling plays an exceptionally important role in position-dependent modulation of promoter activity and that this effect is coordinated with growth-phase.

The results presented in this study demonstrate that position of a promoter within the chromosome is important for its activity and that DNA supercoiling plays a major role in this effect. This phenomenon may be relevant to the natural organisation of the *E. coli* genome,

for example, promoters requiring increased levels of negative superhelicity may be located nearer to *oriC* than to the *dif* site. The *E. coli* genome encodes seven rRNA operons, four of which are clustered in proximity to the origin of replication, with the other three venturing no further from *oriC* than the boundaries of the less-structured macrodomains. This positioning means that during rapid growth there would be increased copies of the rRNA operons, due to gene dosage, therefore increasing their transcriptional output in line with physiological need. However, activity of the rRNA operons is dependent upon negative superhelicity, *in vitro* and *in vivo*, which is sensed by the GC-rich "discriminator" sequence, positioned between the -10 element and transcription start site (Travers, 1980; Glaser *et al*., 1983; Ohlsen and Gralla, 1992; Rochman *et al*., 2002). The "discriminator" sequence acts as a structural barrier to isomerisation of closed complexes to the promoter open complex, however this can be overcome by negative supercoiling of the promoter DNA (Pemberton *et al*., 2000). Therefore positioning near to the origin would be advantageous for promoter activity, due to the increased negative supercoiling in this region. Another excellent example of an *oriC*-proximal, supercoiling-sensitive promoter is that of the gene encoding the NAP, FIS. The *fis* promoter is also activated by high negative superhelicity, contains a "discriminator" sequence, and is positioned midway between the *nupG* and *asl* loci used in this study (Travers *et al*., 2001). Therefore the positioning of the supercoiling-sensitive *fis* and rRNA operon promoters, near to the origin of replication, may also be influenced by the apparent gradient of negative superhelicity from *oriC* to the *dif* site, reflected in the results presented in chapters 3 and 4.

The position-dependent modulation of promoter activity by variation in DNA supercoiling, described in this study, could be an important feature of bacterial gene regulation on a global scale. The sexually-transmitted disease-causing bacterium, *Mycoplasma genitalium*, has the smallest known self-replicating genome, which encodes a small number of genes, no conventional transcription factors, and only one sigma factor (reviewed in Dorman, 2011). This lack of conventional transcriptional regulatory mechanisms raises the question of how

this bacterium regulates its transcriptional output in response to environmental stimuli. Zhang and Baseman, (2011) found that transcription of an osmolarity-responsive gene was affected by inhibition of DNA gyrase, through the addition of novobiocin. This data prompted Dorman, (2011) to propose that variation of DNA supercoiling composes a major part of the regulatory repertoire for such a small genome. The regulation of transcription by DNA supercoiling in *M. genitalium* is likely to manifest a position-dependent modulation of promoter activity within the chromosome. Therefore the mechanisms of position-dependent modulation of promoter activity, characterised in chapter 4 of this thesis, could be of use in determining the mechanisms by which *M. genitalium* regulates its transcriptional response.

The effect of differential DNA superhelicity on transcription initiation is difficult to characterise, as the effect is expected to vary between different promoters. The work presented in chapter 5 analysed the activity of several different promoters when fused to *lacZ* on the chromosome of *E. coli* or encoded by a multi-copy *lac* expression vector. The difference between the chromosomal and plasmid locus was found to affect class I activated promoters differently to class II activated promoters. This result is likely due to the difference in physical nature between the chromosomal locus and the plasmid. Mobility of the chromosomal locus would be highly restricted compared to the relatively small plasmid, which are known to move freely within the cytoplasm and to congregate in one position when carrying a strong, active promoter (Sánchez-Romero *et al*., 2012). This may affect the two classes of promoter differently, however this is unlikely due to the promoters utilising the same transcription factor and being derived from the same sequence.

The other difference in the physical characteristics of the chromosomal locus and the plasmid is the superhelical density of the template. Due to the plasmid being an extra-chromosomal DNA fragment, the superhelical density would not necessarily be the same as that of the chromosomal *lac* locus, used for promoter::*lacZ* fusions. Therefore, template DNA

superhelical density is likely to affect promoters differently, based on the mechanism of activation, due to changes in the alignment of promoter elements brought about by variation in DNA twist. The majority of promoter characterisation studies to date have used multi-copy reporter gene expression plasmids, therefore studying the promoter out of context and on a different physical template. The work presented here demonstrates that supercoiling of the bacterial chromosome plays a major role in regulation of transcription from the promoters encoded by it. Rovinsky *et al*. (2012) report a "supercoiling sensor" module, which is 50-fold more sensitive to changes in superhelical density than a *gyrB::lacZ* promoter fusion. Development of this technology would allow locally available supercoiling to be measured at each of the chromosomal loci tested in this study, during different growth phases and conditions. Hence I am confident that rapid progress will soon be made in characterisation of global supercoiling effects and understanding of the mechanism by which they occur.

6.3 Regulation by organisation of genes within the chromosome

Changing the chromosomal locus of a promoter not only moves it to a different physical space within the cell, but also gives it different neighbouring operons. The promoter may naturally be located in a transcriptionally quiet region of the chromosome, which is only transcribed in the presence of a specific sugar. Therefore, if this promoter were moved to a different location, it may be positioned next to a highly transcribed locus. This could possibly have an effect on the optimum transcription level of the promoter of interest. Work described in section 4.3 describes the effects of neighbouring promoter activity on the *lac* promoter at the *mel* locus and the effect of *lac* promoter induction on *melAB* activity. Active transcription was found to have a negative effect on downstream promoters, but no measurable effect on upstream promoters.

For a single RNAP elongation complex to proceed with the transcription process, either the elongation complex, including its associated mRNA transcript and RNA associated proteins, must rotate around the DNA template, or the DNA must rotate around its own axis (Wu *et al*., 1988). The former motion is unlikely, due to the fact that the nascent transcript would become entwined around the helical DNA template as RNAP tracks around it, therefore requiring an unwinding mechanism for which there is currently no explanation (Cook, 2010). Rotation of the helical DNA template around its own axis would lead to an increase in negative supercoils behind the elongation complex and a loss of negative supercoils, or gain of positive supercoiling, ahead of the complex. This effect is referred to as the twin-supercoiled-domain model and was first described by Liu and Wang, (1987). Reduction in negative superhelicity, or increase in positive supercoiling, ahead of the actively transcribing elongation complex is likely to have a negative effect on downstream promoters, which will also become positively supercoiled. Over-winding of the promoter DNA, positive supercoiling, would oppose the unwinding of DNA around the transcription start site to form the promoter open complex and initiation of transcription. Further to this, changes in superhelicity of the promoter could misalign the different promoter elements, therefore leading to suppression of transcription initiation, much like a promoter activated by conformational change, as described in section 1.3.3 (Figure 1.8). Therefore the results presented in this study suggest that transcription from separate promoters can be topologically coupled by transcription-induced supercoiling, which diffuses between neighbouring operons and affects promoter activity.

Supercoiling-mediated topological coupling of promoters has been suggested to be the mechanism behind the supercoiling-sensitive *leu-500* promoter mutation, which leads to leucine autotrophy in *S. typhimurium* (Lilley and Higgins, 1991). The work of el Hanafi and Bossi, (2000) demonstrated that active transcription of inducible *tet* or *cat* gene cassettes, inserted upstream of the chromosomal *leu* locus, caused inhibition of *leu* promoter activity. This result is very similar to that reported here, however the activity of both partners of the

topological coupling is reported by this study, demonstrating that both promoters are affected by each others activity. The evidence presented here, and that of the *leu-500* system, indicate that organisation of promoters and genes within the chromosome may be influenced by interaction between topologically coupled transcription units. This may be based on the signals to which they are responsive and the appropriate integration of different responses. Organisation at this complex level would lead to co-operation between divergently oriented promoters and antagonism between convergently oriented operons.

Further evidence of transcriptional regulation by neighbouring promoter activity was presented in section 4.3.2, in which the activity of the *lac28::egfp* fusion was shown to be suppressed when positioned downstream of the *rrsHileValaVrrlHrrfH* highly transcribed rRNA operon. Silencing of *lac28* promoter activity was shown to be relieved upon entry into stationary phase when the transcription of the rRNA operons is inhibited (Jacobson and Gillespie, 1968; Paul, *et al*., 2004; Grainger *et al*., 2006; Figure 4.7; Figure 4.8). This result is most likely due to the positively supercoiled domain that would be created immediately downstream of the *rrsHileValaVrrlHrrfH* operon. This could be tested by measurement of the change in superhelical density at the *dkgB* locus during growth-phase transition by utilisation of the "supercoil sensor", described by Rovinskiy *et al*. (2012). The *dkgB, yafC* and *yafD* genes all reside immediately downstream of the *rrsHileValaVrrlHrrfH* and *aspU* highly transcribed operons and are all transcribed from promoters reliant upon the stationary phase sigma factor, σ^{32} (Figure 4.7). Transcription of these genes would only be required during stationary phase when the conditions for transcription at this locus are optimal, suggesting that this higher level chromosome organisation may have naturally evolved. However further bioinformatic studies into the organisation of differentially regulated genes would be required to test this hypothesis. The work of Peter *et al*. (2004) identified genes responsive to supercoiling perturbations. Therefore a bioinformatic study of the orientation and transcriptional regulation of genes neighbouring supercoiling-sensitive operons could be

analysed to test whether genes are regulated by organisation. This would be a fascinating study and would make an excellent contribution, however this work was beyond the scope of this thesis.

The topological coupling of neighbouring promoters could potentially explain why genes encoding repressor proteins are often found directly upstream of the target promoter, such as is the case at the *lac* operon. However, a more likely explanation for this organisation is that put forward by Janga and colleagues (2009) who utilised a bioinformatics approach to show that transcription factors are positioned differently, with respect to their targets on the genome, depending on the size of their regulons. Janga *et al*. (2009) find that the gene for transcription factors controlling small regulons are located close to the target genes on the chromosome, therefore, due to the coupling of transcription and translation, the protein product should be located near to the target DNA site. Data presented in section 3.6 of this study provides support for this hypothesis through the analysis of *lac28* promoter induction with varying distance of the promoter from the *lacI* gene on the linear chromosomal map. The promoter carried by the *lac28* fragment was found to be activated in the presence of lower inducer concentrations at certain loci. Therefore this work demonstrates that position of the target with respect to the gene encoding a transcription factor can have an effect on its efficacy.

Further evidence for this hypothesis has recently been supplied by the work of Kuhlman and Cox, (2012) in which visualisation of fluorescently labelled LacI shows limited dispersion and distribution based on the position of the *lacI* gene. Further evidence demonstrates that repression of the target promoter is increased with proximity to the *lacI* gene (Kuhlman and Cox, 2012). However these data are in disagreement with the work presented by Block *et al*. (2012), in which no effect of changing the position of the target promoter with respect to the gene encoding the transcription factor was measured. Therefore, further work utilising the

promoter reporter probe inserted at more loci, and super-resolution imaging of the physical location of each locus with respect to the *lacI* gene locus, would provide a definitive answer as to whether this level of organisation is important to the regulation of transcription.

As previously discussed, changing the location of a promoter also alters its neighbours, however other local sequences may also play a part in position-dependent modulation of promoter activity. Experiments described in section 4.2 were initially designed to assess the effect of orientation with respect to replication, however no correlation could be found. This result is in agreement with that found by Beckwith *et al*. (1966), Schmid and Roth, (1987) and Block *et al*. (2012). Analysis of neighbouring promoter sequences at the *ara* and *mel* loci identified two REP elements at each of the target loci, with one of them being disrupted by the insertion. The remaining REP element at the *ara* locus was found to be a target for DNA gyrase, with the orientation-dependent differences in promoter activity being reliant upon the action of DNA gyrase. However, orientation-dependent effects at the *mel* locus were found to be unresponsive to gyrase inhibition, therefore these effects were attributed to mRNA transcript stabilisation by positioning of the REP element at the 3' end (Section 4.2). The *E. coli* genome encodes hundreds of REP elements positioned throughout the chromosome in extragenic space, therefore these local sequence elements are likely to be involved in position-dependent variation of promoter activity. I am confident that greater understanding of the effects produced by organisation of REP elements with respect to transcription units will advance the understanding of transcription regulation.

Finally, the results presented in section 4.3.1 demonstrate that activity of the *melAB* promoter is increased by introduction of the *lac* promoter downstream, even when the *lac* promoter is not activated. These results also show that this effect is greater when the promoters are closer on the linear DNA molecule. The increased activity is weakened when the orientation of the *lac28::egfp* reporter is changed and the inter-promoter distance is increased. Both the *melAB*

and *lac* promoters are activated by the same transcription factor, CRP, therefore introducing an additional, *lac* promoter encoded, CRP DNA binding site downstream of the *melAB* promoter. The additional CRP binding site may act in the same way as the Lac repressor pseudo-operators at the *lac* operon, therefore increasing locally available CRP and increasing the probability of transcription initiation. Currently, no other hypothesis can be supplied for these results, therefore this would require further investigation.

6.4 Transcription factories and the uneven distribution of RNAP

Section 1.4.3 introduced the transcription factory model for organisation of the chromosome by clustering of actively transcribing RNAP complexes (Cook, 2010). Cook proposes a model for chromosomal organisation which is also suggested to play a role in the regulation of transcription and how neighbouring promoters interact. Due to the static nature of RNAP transcription factories, the DNA template would be pulled into the factory by active transcription. Hence, the probability of transcription initiation at downstream neighbouring promoters is increased. However, this would potentially be a negative effect for convergently oriented promoters, due to the template being pulled in opposite directions by tethered elongation complexes. Unfortunately, this hypothesis does not fit with the results presented in this thesis (Section 4.3.1), in which downstream promoters are negatively affected by active transcription, regardless of their orientation. The twin-domain effect, in which active transcription affects local DNA topology, is the more likely explanation for the results presented here, however integration of the two hypotheses may be possible in the future.

One particularly interesting suggestion is that different transcription factories could contain specific transcription factors. This puts forward an alternative hypothesis to explain the increased activity of the *melAB* promoter, by the proximity of the *lac* promoter. If

transcription factories do contain specific transcription factors, or at least a bias towards certain types, then activation of either of the CRP activated promoters, by movement to a CRP factory, would bring the other promoter into more frequent contact with CRP. Recent evidence supplied by Wang *et al*. (2011) demonstrates that the H-NS transcriptional regulator and NAP is localised into specific foci within *E. coli*, similar to that expected due to the presence of transcription factories. The H-NS transcriptional regulator is mainly involved in repression and is also a major architectural protein, therefore the clustering may not reflect any role played in transcription factory formation. Therefore, visualisation of a CRP::GFP fusion, expressed at wild type levels from the natural promoter on the chromosome, by super-resolution microscopy, would be an appropriate experiment to determine whether CRP transcription factories exist. This experiment was beyond the scope of this thesis, but may be possible in the near future. While specific transcription factories in bacteria are an appealing suggestion, little evidence is provided for their existence. Therefore the more likely explanation for these results is that the additional CRP DNA binding site, introduced downstream of the *melAB* promoter, acts like a pseudo-operator, to increase the local CRP concentration and probability of transcription activation.

While transcription factories are unlikely to exist in bacteria in the same way that is evident in eukaryotic systems, spatial distribution of RNAP within the cell, and its access to different chromosomal regions, is certainly not uniform. The work of Vora *et al*. (2009), identified large regions of the chromosome, tsEPODs, which are bound by higher than average levels of protein and encode transcriptionally silent genes. The work described in section 4.5 demonstrated that positioning within tsEPODs silences transcription of the strong *lac* promoter and that this can be relieved by deletion of the protein bound domain. These results confirm that the transcriptionally silent profile of tsEPODs is not merely due to the region containing poor promoters, but because of the intrinsic properties of the domain.

The tsEPODs have previously been suggested to be the organising centres of the *E. coli* nucleoid, and would be expected to be buried in the centre of the ribosome- and RNAP-free nucleoid space (Bakshi *et al*. 2012; Vora *et al*., 2009). This would mean that promoters within tsEPODs would be less accessible to RNAP, therefore providing an explanation for the silencing of *lac* promoter activity measured in section 4.5.1. The H-NS protein is a main candidate for organisation of the chromosome by tsEPODs, due to the observed clustering of the protein into two foci, presented in the fluorescence microscopy studies of Wang *et al*. (2011). This study lends support to the hypothesis that the *E. coli* chromosome is organised and structured with tsEPODs positioned at the centre of the nucleoid and supercoiled loops radiating from this central domain.

The data presented and discussed in this thesis demonstrate that chromosomal architecture and the compaction of the nucleoid plays a major role in transcriptional regulation. Data presented in section 3.7 demonstrates that the profile of position-dependent effects is shifted on entry into stationary phase, therefore suggesting a link between this global level of regulation and growth-phase transitions or stress responses. This effect is likely to be coordinated by shifting populations of NAPs to form tsEPODs, which silence transcription through occlusion of RNAP. Data presented in chapter 4 identifies variation in DNA superhelicity as a key mechanism behind position-dependent modulation of promoter activity. Further evidence presented in chapter 4 demonstrates that while DNA supercoiling can affect promoter activity, active transcription can cause variations in local superhelicity. These results suggest that organisation of transcription units within the genome could have an affect on their regulation with neighbouring operons being topologically coupled. Finally position of the target promoter, with respect to the gene encoding the Lac repressor, was shown to have an effect on its regulation. This suggests that organisation of genes with respect to their interaction targets can play a role in regulation of the transcriptional output. This thesis, as a

whole, demonstrates the importance of promoter position within the chromosome and illustrates the complexity of transcriptional regulation in bacteria.

References

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

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

Auner, H., Buckle, M., Deufel, A., Kutateladze, T., Lazarus, L., Mavathur, R., *et al*. (2003) Mechanism of transcriptional activation by FIS: Role of core promoter structure and DNA topology. *J Mol Biol.* **331:** 331-344.

Azam, T.A., Hiraga, S. and Ishihama, A. (2000) Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells.* **5:** 613–626.

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

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

Bachmann, B.J. (1996) Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12. In *Escherichia coli and Salmonella*. Neidhardt F.C. (ed.). Washington, DC: ASM Press, pp. 2460–2488.

Bakshi, S., Siryaporn, A., Goulian, M., and Weisshaar, J.C. (2012) Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Mol Microbiol.* **85:** 21-38.

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

Barne, K.A., Bown, J.A., Busby, S.J., 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.

Baquero, M.R., Bouzon, M., Varea, J., and Moreno, F. (1995) *sbmC*, a stationary-phase induced SOS *Escherichia coli* gene, whose product protects cells from the DNA replication inhibitor microcin B17. *Mol Microbiol.* **18:** 301-311.

Beckwith, J.R., Signer, E.R. and Epstein, W. (1966) Transposition of the *lac* region of *E. coli*. *Cold Spring Harb Symp Quant Biol*. **31:** 393-401.

Benham, C.J., and Mielke, S.P. (2005) DNA mechanics. *Annu Rev Biomed Eng.* **7:** 21.53.

Blake, T., Barnard, A., Busby, S.J.W., and Green, J. (2002) Transcription activation by FNR: Evidence for a functional activating region 2. *J Bacteriol.* **184:** 5855-5861.

Blattner, F.R., Plunkett, G.III., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., *et al*., (1997) The complete genome sequence of *Escherichia coli* K-12. *Science*. **277:** 1453-1462.

Block, D.H.S., Hussein, R. Liang, L.W. and Lim, H.N. (2012) Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.* doi: 10.1093/nar/gks694.

Blot, N., Mavathur, R., Geertz, M., Travers, A. and Muskhelishvili, G. (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO reports.* **7:** 710-715.

Brahms, J.G., Dargouge, O., Brahms, S., Ohara, Y. and Vagner, V. (1985) Activation and inhibition of transcription by supercoiling. *J Mol Biol.* **181:** 455-465.

Bremer, H. and Dennis, P. P. (1996). Modulation of chemical composition and other parameters of the cell growth rate. In *Escherichia col*i *and Salmonella*: *Cellular and Molecular Biology*, 2nd edn. Neidhardt, F. C. and others (eds). Washington, DC: American Society for Microbiology, pp. 1553–1568.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., and Hobman, J.L. (2003) The MerR family of transcriptional regulators. *FEMS Microbiol Rev.* **27:** 145–63.

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

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.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969) Factor stimulating transcription by RNA polymerase. *Nature.* **221**: 43-46.

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., and Ebright, R.H. (1997) Transcription activation at class II CAP-dependent promoters. *Mol Microbiol.* **23:** 853–859.

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

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

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.

Campbell, N.A., and Reece, J.B. (2005) *Biology* (7th ed.), San Francisco: Pearson Benjamin Cummings.

Calos, M.P. (1978) DNA sequence for a low-level promoter of the *lac* repressor gene and an "up" promoter mutation. *Nature*. **274:** 762-765.

Caro, L.G., and Berg, C.M. (1968) Chromosome replication in some strains of *Escherichia coli* K12. *Cold Spring Harbor Symp Quant Biol*. **33:** 559-573.

Ceci, P., Cellai, S., Falvo, E., Rivetti, C., Rossi, G.L., and Chiancone, E. (2004) DNA condensation and self-aggregation of *Escherichia coli* Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Res*. **32:** 5935–5944.

Champion, K., and Higgins, N.P. (2007) Growth rate toxicity phenotypes and homeostatic supercoil control differentiate *Escherichia coli* from *Salmonella enterica* serovar Typhimurium. *J bacteriol.* **189:** 5839-5849.

Chandler, M.G. and Pritchard, R.H. (1975) The effect of gene concentration and relative gene dosage on gene output in *Escherichia coli*. *Mol Gen Genet.* **138:** 127-141.

Chang, D.E., Smalley, D.J. and Conway, T. (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Mol Microbiol.* **45:** 289-306.

Cherepanov, P.P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene*. **158:** 9-14.

Chismon, D.L. (2011) *Architecture of Escherichia coli promoters that respond to reactive nitrogen species.* Ph.D. thesis, School of Biosciences, University of Birmingham.

Chismon, D.L., Browning, D.F., Farrant, G.K. and Busby, S.J.W. (2011) Unusual organization, complexity and redundancy at the *Escherichia coli hcp-hcr* operon promoter. *Biochem J.* **430:** 61-68.

Constantinidou, C., Hobman, J.L., Griffiths, L., Patel, M.D., Penn, C.W., Cole, J.A. and Overton, T.W. (2006) A reassessment of the FNR regulon and transcriptomic analysis of the effects of nitrate, nitrite, NarXL, and NarQP as *Escherichia coli* K12 adapts from aerobic to anaerobic growth. *J Biol Chem.* **281:** 4802-4815.

Cook, P.R. (2002) Predicting three-dimensional genome structure from transcriptional activity. *Nat Genet.* **32:** 347-352.

Cook, P.R. (2010) A model for all genomes: The role of transcription factories. *J Mol Biol.* **395:** 1-10.

Cooper, S. and Helmstetter, C.E. (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol.* **31:** 519-540.

Dame, R.T., Espéli, O., Grainger, D.C., and Wiggins, P.A. (2012) Multidisciplinary perspectives on bacterial genome organisation and dynamics. *Mol Microbiol*. **86:** 1023-1030.

Dame, R.T., Kalmykowa, O.J. , and Grainger, D.C. (2011) Chromosomal macrodomains and associated proteins: Implications for DNA organisation and replication in Gram negative bacteria. *PLOS Genet.* **7:** Epub e1002123.

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.

Dame, R.T., Wyman, C., and Goosen, N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res*. **28:** 3504–3510.
Darst, S.A., Polyakov, A., Richter, C., and Zhang, G.Y. (1998) Insights into *Escherichia coli* RNA Polymerase Structure from a Combination of X-Ray and Electron Crystallography. *J Struc Biol* **124:** 115-122.

Dickson, R.C., Abelson, J., Barnes, W.M., and Reznikoff, W.S. (1975) Genetic regulation: the Lac control region. *Science*. **187:** 27-35.

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

Dorman, C.J. (2011) Regulation of transcription by DNA supercoiling in *Mycoplasma genitalium*: global control in the smallest known self-replicating genome. *Mol Microbiol*. **81:** 302-304.

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.

Dupaigne, P., Tonthat, N.K., Espeli, O., Witfill, T., Boccard, F., and Schumacher, M.A. (2012) Molecular basis for a protein-mediated DNA-bridging mechanism that functions in condensation of the *E. coli* chrommosome. *Mol Cell*. **48:** 1-12.

Dworsky, P., and Schaechter, M. (1973) Membrane attachment of the nucleoid of *Escherichia coli*. *J Bacteriol.* **116:** 1364-1374.

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

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

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

el Hanafi, D., and Bossi, L. (2000) Activation and silencing of *leu-500* promoter by transcription-induced DNA supercoiling in the *Salmonella* chromosome. *Mol Microbiol*. **37:** 583-594.

Engelhorn, M., Boccard, F., Murtin, C., Prentkin, P., and Geiselmann, J. (1995) *In vivo* interaction of the *Escherichia coli* integration host factor with its specific binding sites. *Nucleic Acids Res.* **23:** 2959-2965.

Espeli, O. and Boccard, F. (1997) *In vivo* cleavage of *Escherichia coli* BIME-2 repeats by DNA gyrase: genetic characterisation of the target and identification of the cut site. *Mol Microbiol.* **26:** 767-777.

Espeli, O., Borne, R., Dupaigne, P., Thiel, A., Gigant, E., Mercier, R., and Boccard, F. (2012) A MatP-divisome interaction coordinates chromosome segregation with cell division in *E. coli*. *EMBO J.* **31:** 3198-3211.

Espeli, O., Mercier, R., and Boccard, F. (2008) DNA dynamics vary according to macrodomain topography in the *E. coli* chromosome. *Mol Microbiol.* **68:** 1418-1427.

Falkow, S. (1996) The evolution of pathogenicity in *Escherichia, Shigella* and *Salmonella*. In *Escherichia coli and Salmonella*. Neidhardt F.C. (ed.). Washington, DC: ASM Press, pp. *2723*–2729.

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.

Figurski, D.H., Meyer, R.J. and Helinski, D.R. (1979) Suppression of ColEl replication properties by the IncP-1 plasmid RK2 in hybrid plasmids constructed *in vitro. J Mol Biol.* **133:** 295-318.

Filenko, N.A., Browning, D.F. and Cole, J.A. (2005) Transcriptional regulation of a hybrid cluster (prismane) protein. *Biochem Soc Trans.* **33:** 195-197.

Filenko, N.A., Spiro, S., Browning, D.F., Squire, D., Overton, T.W., Cole, J. and Constantinidou, C. (2007) The NsrR regulon of *Escherichia coli* K-12 includes genes encoding the hybrid cluster protein and the periplasmic, respiratory nitrite reductase. *J Bacteriol.* **189:** 4410-4417.

French, S. (1992) Consequences of replication fork movement through transcription units *in vivo*. *Science*. **258:** 1362-1365.

Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimoni, E., Wolf, S.G., and Minsky, A. (2004) Nucleoid restructuring in stationary-state bacteria. *Mol Microbiol*. **51:** 395–405.

Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G.J., Thompson, N.E., Burgess, R.R., David, P.R., and Kornberg, R.D. (1999) Yeast RNA Polymerase II at 5 Å Resolution. *Cell.* **98:** 799-810.

Fukushima, M., Kakinuma, K, and Kawaguchi, R. (2002) Phylogenetic analysis of *Salmonella, Shigella* and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *J Clin Microbiol*. **40:** 2779-2785.

Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H., and Gourse, R.L. (1996) DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. *Genes Dev.* **10:** 16-26.

Gaston, K., Bell, A., Kolb, A., Buc, H., and Busby, S. (1990) Stringent spacing requirements for transcription activation by CRP. *Cell.* **62:** 733-740.

Gellert, M., Mizuuchi, K., O'Dea, M.H., and Nash, H.A. (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proc Natl Acad Sci USA.* **73:** 3872-3876.

Gilbert, W. and Müller-Hill, B. (1966) Isolation of the *lac* repressor. *Proc Natl Acad Sci USA.* **56:** 1891-1898.

Gilson, E., Rousset, J.P., Clement, J.M. and Hofnung, M. (1986) A subfamily of *E. coli* palindromic units implicated in transcription termination? *Ann Inst Pasteur Microbiol.* **137B:** 259–270.

Gilson, E., Perrin, D., and Hofnung, M. (1990) DNA polymerase I and a protein complex bind specifically to *E. coli* palindromic unit highly repetitive DNA: implications for bacterial chromosome organization. *Nucleic Acids Res.* **18:** 3941-3952.

Glaser, G., Sarmientos, P., and Cashel, M. (1983) Functional interrelationship between two tandem *E. coli* ribosomal RNA promoters. *Nature*. **302:** 74-76.

Gottesman, S. (1984) Bacterial regulation: Global regulatory networks. *Annu Rev Genet.* **18:** 415-441.

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., Harrison, M., Holdstock, J. and Busby, S.J. (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA.* **102:** 17693-17698

Grainger, D., Hurd, D., Goldberg, M.D., and Busby, S.J.W. (2006) Association of nucleoid proteins with coding and non-coding segments of the *Escherichia coli* genome. *Nucleic Acids Res.* **34(16):** 4642-4652.

Gribskov, M., and Burgess, R.R. (1986) Sigma factors from E. *coli, B. subtilis*, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res*. **14:** 6745–6763.

Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J., and Young, B. (1998) The Functional and Regulatory Roles of Sigma Factors in Transcription. *Cold Spring Harbor Symposia on Quantitative Biology* **63:** 141-155.

Gruber, T.M., and Gross, C.M. (2003) Multiple Sigma Subunits and The Partitioning of Bacterial Transcription Space. *Annu Rev Microbiol* **57:** 441-466.

Helmstetter, C.E. and Cooper, S. (1968) DNA synthesis during the division cycle of rapidly growing *Escherichia coli* B/r. *J Mol Biol.* **31:** 507-518.

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.

Higgins, C.F., McLaren, R.S., and Newbury, S.F. (1988) Repetitive extragenic palindromic sequences, mRNA stability and gene expression: evolution by gene conversion? – a review. *Gene*. **72:** 3-14.

Higgins, N.P., Yang, X., Fu, Q., and Roth, J.R. (1996) Surveying a supercoil domain by using the gamma delta resolution system in *Salmonella typhimurium*. *J Bacteriol.* **178:** 2825–2835.

Hochschild, A., and Dove, S.L. (1998) Protein-protein contacts that activate and repress prokaryotic transcription. *Cell.* **92**: 597-600.

Hohlbein, J., Gryte, K., Heilemann, M., and Kapanidis, A.N. (2010) Surfing on a new wave of single-molecule fluorescence methods. *Phys Biol.* **7:** 031001.

Hollands, K. (2009) *Post-genomic studies on the Escherichia coli cyclic AMP receptor protein*. Ph.D. Thesis, School of Biosciences, University of Birmingham.

Holmes, V.F., and Cozzarelli, N.R. (2000) Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proc Natl Acad Sci USA.* **97:** 1322-1324.

Ishihama, A. (2000) Functional Modulation of *Escherichia coli* RNA Polymerase. *Annu Rev Microbiol.* **54:** 499-518.

Jacob, F., and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol.* **3:** 318-356.

Jacobson A, Gillespie D. (1968) Metabolic events occurring during recovery from prolonged glucose starvation in *Escherichia coli*. *J Bacteriol.* **95:**1030–39.

Janga, S.C., Salgado, H. and Martinez-Antonio, A. (2009) Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res.* **37:** 3680-3688.

Jin, D.J., and Cabrera, J.E. (2006) Coupling the distribution of RNA polymerase to global gene regulation and the dynamic structure of the bacterial nucleoid in *Escherichia coli*. *J Struct Biol*. **156:** 284-291.

Jobe, A., and Bourgeois, S. (1972) *lac* repressor-operator interaction. VI. The natural inducer of the *lac* operon. *J Mol Biol*. **69:** 397–408.

Kavenoff, R., and Bowen, B.C. (1976) Electron microscopy of membrane-free folded chromosomes from *Escherichia coli*. *Chromosoma.* **59:** 89-101.

Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., *et al.* (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nuc Acids Res.* **39:** D583–90.

Keller, T., Mis, S.D., Jia, K.E., and Wilke, C.O. (2012) Reduced mRNA secondary-structure stability near the start codon indicates functional genes in prokaryotes. *Genome Biol Evol.* **4:** 80-88.

Kennedy, E. P. (1970) The Lactose Operon (Beckwith, J. R. and Zipser, D., eds) pp. 49–92, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Keseler, I.M., Collado-Vides, J., Santos-Zavaleta, A., Peralta-Gil, M., Gama-Castro, S., Muniz-Rascado, L., *et al.* (2011) EcoCyc: a comprehensive database of *Escherichia coli* biology. *Nuc Acids Res.* **39:** D583–90.

Khemici, V. and Carpousls, A.J. (2004) The RNA degradosome and poly(A) polymerase of *Escherichia coli* are required *in vivo* for the degradation of small mRNA decay intermediates containing REP-stabilizers. *Mol Microbiol.* **51:** 777-790.

Kleppe, K., Övrebö, S., and Lossius, I. (1979) The bacterial nucleoid. *J Gen Microbiol*. **112:** 1-13.

Kolatka, K., Kubik, S., Rajewska, M. and Konieczny, I. (2010) Replication and partitioning of the broad-host-range plasmid RK2. *Plasmid*. **64:** 119-134.

Kolesov, G., Wunderlich, Z., Laikova, O.N., Gelfand, M.S. and Mirny, L.A. (2007) How gene order is influenced by the biophysics of transcription regulation. *Proc Natl Acad Sci USA.* **104:** 13948–13953.

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

Kues, U. and Stahl, U. (1989) Replication of plasmids in Gram-negative bacteria. *Microbiol Rev.* **53:** 491–516.

Kuhlman, T.E. and Cox, E.C. (2012) Gene location and DNA density determine transcription factor distributions in *Escherichia coli*. *Mol Syst Biol.* **8:** 610.

Kusano, S., Ding, Q., Fujita, N., and Ishihama, A. (1996) Promoter selectivity of *Escherichia coli* RNA polymerase $E\sigma^{70}$ and $E\sigma^{38}$ holoenzymes: Effect of DNA supercoiling. *J Biol Chem*. **271:** 1998-2004.

Kuzj, A.E.S., Medberry, P.S. and Schottel, J.L. (1998) Stationary phase, amino acid limitation and recovery from stationary phase modulate the stability and translation of chloramphenicol acetyltransferase mRNA and total mRNA in *Escherichia coli*. *Microbiol.* **144:** 739-750.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, G.O., *et al*. (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.

Lange, R. and Hengge-Aronis, R. (1991) Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol Microbiol.* **5:** 49-59.

Lawson, C.L., Swigon, D., Murakami, K.S., Darst, S.A., Berman, H.M. and Ebright, R.H. (2004) Catabolite activator protein (CAP): DNA binding and transcription activation. *Curr Opin Struct Biol.* **14:** 10-20.

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, D.J, Minchin, S.D., and Busby, S.J.W. (2012) Activating transcription in bacteria. *Annu Rev Microbiol*. **66:** 125-152.

Lilley, D.M., and Higgins, C.F. (1991) Local DNA topology and gene expression: the case of the *leu-500* promoter. *Mol Microbiol*. **5:** 779-783.

Lim, C.J., Lee, S.Y., Kenney, L.J., and Yan, J. (2012) Nucleoprotein filament formation is the structural basis for bacterial protein H-NS gene silencing. *Sci Rep.* **2:** 509 Epub DOI: 10.1038/srep00509.

Livak, K. and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}$T method. *Methods*. **25:** 402-408.
Liu, L.F., and Wang, J.C. (1987) Supercoiling of the DNA template during transcription. *Proc Natl Acad Sci USA.* **84:** 7024-7027.

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.

Loewen, P.C., Hu, B., Strutinsky, J., and Sparling, R. (1998). Regulation in the *rpoS* regulon of *Escherichia coli*. *Can J Microbiol.* **44:** 707–17.

Lonetto, M., Gribskov, M., and Gross, C.A. (1992) The σ^{70} family: sequence conservation and evolutionary relationships. *J Bacteriol.* **174:** 3843–3849.

Louarn, J., Cornet, F., Francois, V., Patte, J., and Louarn, J.M. (1994) Hyperrecombination in the terminus region of the *Escherichia coli* chromosome: Possible relation to nucleoid organization. *J Bacteriol.* **176:** 7524-7531.

Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L., and Dame, R.Th. (2006) The architectural role of nuceloid-associated proteins in the organisation of bacterial chromatin: A molecular perspective. *J Struct Biol*. **156:** 262-272.

Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res*. **25:** 1203-1210.

Madigan, M.T. and Martinko, J.M. (2006) *Biology of Microorganisms* (11th ed.), Upper Saddle River, NJ: Pearson Prentice Hall.

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

Marbach, A. and Bettenbrock, K. (2012) *lac* operon induction in *Escherichia coli*: systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol.* **157:** 82-88.

Martinez-Antonio, A., and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol.* **6:** 482-489.

McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem.* **54:** 171-204.

McGlynn, P., Savery, N.J. and Dillingham, M.S. (2012) The conflict between DNA replication and transcription. *Mol Microbiol.* **85:** 12-20.

Meiklejohn, A.L., and Gralla, J.D. (1985) Entry of RNA polymerase at the *lac* promoter. *Cell.* **43:** 769-776.

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

Mercier, R., Petit, M.A., Schbath, S., Robin, S., Karoui, M.E., Boccard, F., and Espeli, O. (2008) The MatP/*mat*S site-specific system organizes the terminus region of the *E. coli* chromsome into a macrodomain. *Cell.* **135:** 475-485.

Merrikh, H., Machón, C., Grainger, W.H., Grossman, A.D. and Soultanas P. (2011) Co-directional replication-transcription conflicts lead to replication restart. *Nature*. **470:** 554-558.

Merrikh, H., Zhang, Y., Grossman, A. D., and Wang, J. D. (2012) Replication-transcription conflicts in bacteria. *Nat Rev Microbiol*. **10:** 449-458.

Messing, S.A.J., Ton-Hoang, B., Hickman, A.B., McCubbin, A.J., Peaslee, G.F., Ghirlando, R. *et al*. (2012) The processing of repetitive extragenic palindromes: the structure of a repetitive extragenic palindrome bound to its associated nuclease. *Nucleic Acids Res.* **1-16** doi:10.1093/nar/gks741

Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci USA.* **98**: 892-897.

Mirkin E.V. and Mirkin, S.M. (2005) Mechanisms of transcription-replication collisions in bacteria. *Mol. Cell Biol*. **25:** 888-895.

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.

Montero Llopis, P., Jackson, A.F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T., Jacobs-Wagner, C. (2010) Spatial organization of the flow of genetic information in bacteria. *Nature*. **466:** 77-81.

Mukherjee, K., and Chatterji, D. (1997) Studies on the omega subunit of *Escherichia coli* RNA polymerase--its role in the recovery of denatured enzyme activity. *Eur J Biochem.* **247**: 884-889.

Murakami, K., Fujita, N., and Ishihama, A. (1996) Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. *Embo J.* **15**: 4358-4367.

Murakami, K.S., Masuda, S., and Darst, S.A. (2002a) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å 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.

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

Murray, P.R., Rosenthal, K.S., and Pfaller, M.A. (2009) *Medicial Microbiology* (6th ed.), Philadelphia: Mosby Elsevier.

Nakanishi, A., Oshida, T., Matsushita, T., Imajoh-Ohmi, S., and Ohnuki, T. (1998) Identification of DNA gyrase inihibitor (GyrI) in *Escherichia coli*. *J Biol Chem*. **273:** 1933-1938.

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.

Nash, H.A. (1981) Integration and excision of bacteriophage λ: The mechanism of conservative site specific recombination. *Annu Rev Genet*. **15:** 143-167.

Newbury, SF., Smith, N.H., Robinson, E.C., HiIes, I.D. and Higgins, C.F. (1987a) Stabilization of translationally active mRNA by prokaryotic REP sequences. *Cell.* **48:** 297-310.

Newbury, S.F., Smith, N.H. and Higgins, CF. (1987b) Differential mRNA stability controls relative gene expression within a polycistronic operon. *Cell.* **5:** 113 l-l 143.

Nöllmann, M., Crisona, N.J. and Arimondo, P.B. (2007) Thirty years of *Escherichia coli* DNA gyrase: From *in vivo* function to single-molecule mechanism. *Biochimie.* **89:** 490-499.

Navarro-Llorens, J.M., Tormo, A. and Martinez-Garcia, E. (2010) Stationary phase in gram-negative bacteria. *FEMS Microbiol Rev.* **34:** 476-495.

Oehler, S., Eismann, E.R., Krämer, H., and Müller-Hill, B. (1990) The three operators of the *lac* operon cooperate in repression. *EMBO J.* **9:** 973-979.

Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B., and Müller-Hill B. (1994) Quality and position of the three lac operators of *E. coli* define efficiency of repression. *EMBO J.* **13:** 3348-3355.

Oh, T.J., Jung, I.L., and Kim, I.G. (2001) The *Escherichia coli* SOS gene *sbmC* is regulated by H-NS and RpoS during the SOS induction and stationary growth phase. *Biochem Biophys Res Commun.* **288:** 1052-1058.

Ohlsen, K.L., and Gralla, J.D. (1992) Interrelated effects of DNA supercoiling, ppGpp, and low salt on melting within the *Escherichia coli rrnB* P1 promoter. *Mol Microbiol*. **6:** 2243-2251.

Pastan, I., and Perlman, R. (1970) Cyclic adenosine monophosphate in bacteria. *Science*. **169:** 339-344.

Paul, B.J., Ross, W., Gaal, T. and Gourse, R.L. (2004) rRNA transcription in *Escherichia coli*. *Annu Rev Genet.* **38:** 749-770.

Pedersen, S., Bloch, P., Reeh, S., and Neidhardt, F. (1978). Patterns of protein synthesis in *E. coli:* a catalogue of the amount of 140 individual proteins at different growth rates. *Cell.* **14:**179–190.

Pemberton, I.K., Muskhelishvili, G., Travers, A.A., and Buckle, M. (2000) The G+C rich discriminator region of the *tyrT* promoter antagonises the formation of stable preinitiation complexes. *J Mol Biol*. **299:** 859-864.

Pèrez-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.

Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O. and Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*. *Genome Biol*. **5:** R87.

Pogliano, J., Ho, T.Q., Zhong, Z. and Helinski, D.R. (2001) Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc Natl Acad Sci USA.* **98:** 4486-4491.

Postow, L., Hardy, C.D., Arsuaga, J. and Cozzarelli, N.R. (2004) Topological domain structure of the *Escherichia coli* chromosome. *Genes Dev.* **18:** 1766-1779.

Prestidge, L. S., and Pardee, A. B. (1965) A second permease for methyl-thio-beta-D-galactoside in *Escherichia coli. Biochim Biophys Acta.* **100:** 591–593.

Price, M.N., Alm, E.J., and Arkin, A.P. (2005) Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. *Nucleic Acids Res.* **33:** 3224-3234.

Pruss, G.J., Manes, S.H., and Drlica, K. (1982) *Escherichia coli* DNA topoisomerase I mutants: Increased supercoiling is corrected by mutations near gyrase genes. *Cell*. **31:** 35-42.

Rabin, R.S. and Stewart, V. (1993) Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate and nitrite-regulated gene expression in *Escherichia coli* K-12. *J Bacteriol.* **175:** 3259-3268.

Reznikoff, W.S. (1992) The lactose operon-controlling elements: a complex paradigm. *Mol Microbiol*. **6:** 2419-2422.

Reznikoff, W.S., Winter, R.B., and Hurley, C.K. (1974) The location of the repressor binding sites in the *lac* operon. *Proc Nat Acad Sci USA.* **71:** 2314-2318.

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

Robinow, C.F. (1956) The chromatin bodies of bacteria. *Bacteriol Re*v. **20**: 207–242.

Robinow, C., and Kellenberger, E. (1994) The bacterial nucleoid revisited. *Microbiol Rev.* **58:** 211-232.

Rochman, M., Aviv, M., Glaser, G., and Muskhelishvili, G. (2002) Promoter protection by a transcription factor acting as a local topological homeostat. *EMBO reports*. **3:** 335-360.

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

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.

Rouvière-Yaniv, J., and Gros, F. (1975) Characterisation of a novel, low-loecular-weight DNA-binding protein from *Escherichia coli*. *Proc Natl Acad Sci USA.* **72:** 3428-3432.

Rovinskiy, N., Agbleke, A.A., Chesnokova, O., Pang, Z. and Higgins, N.P. (2012) Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome. *PLOS Genet*. **8:** e1002845. doi:10.1371/journal.pgen.1002845

Ryter, A., and Chang, A. (1975) Localisation of transcribing genes in the bacterial cell by means of high resolution autoradiography. *J Mol Biol*. **98:** 797-810.

Sánchez-Romero, M.A., Busby, S.J.W., Dyer, N.P., Ott, S., Millard, A.D., and Grainger D.C. (2010) Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12. *mBio*. **1:** e00012-10. doi:10.1128/mBio.00012-10.

Sánchez-Romero, M.A., Lee, D.J., Sánchez-Morán, E. and Busby, S.J.W. (2012) Location and dynamics of an active promoter in *Escherichia coli* K-12. *Biochem J.* **441:** 481-485.

Sanzey, B. (1979). Modulation of gene expression by drugs affecting deoxyribonucleic acid gyrase. *J Bacteriol.* **138:** 40-47.

Scaife, J (1967) Episomes. *Annu Rev Microbiol.* **21:** 601-638.

Scharff, L.B., Childs, L., Walther D. and Bock, R. (2011) Local absence of secondary structure permits translation of mRNAs that lack ribosome-binding sites. *PLOS Genet.* **7:** e1002155.

Schmid, M.B. and Roth, J.R. (1987) Gene Location Affects Expression Level in *Salmonella typhimurium. J Bact.* **169:6:** 2872-2875.

Schmitz, A. (1981) Cyclic AMP receptor protein interacts with lactose operator DNA. *Nucleic Acids Res*. **9:** 277-292.

Shaner, N.C., Steinbach, P.A. and Tsien, R.Y. (2005) A guide to choosing fluorescent proteins. *Nat methods.* **2:** 905-909.

Sheridan, S.D., Benham, C.J. and Hatfield, G.W. (1998) Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *J Biol Chem.* **273:** 21298-21308.

Shimomura, O., Johnson, F.H. and Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol.* **59:** 223-239.

Simpson, R.B. (1980) Interaction of the cAMP receptor protein with the *lac* promoter. *Nucleic Acids Res.* **8:** 759-766.

Sobetzko, P., Travers, A. and Muskhelishvili, G. (2012) Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc Natl Acad Sci USA.* **109:** E42-E50.

Sousa, C., de Lorenzo, V. and Cebolla, A. (1997) Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiol*. **143:** 2071-2078.

Stonington, O.G., and Pettijohn, D.E. (1971) The folded genome of *Escherichia coli* isolated in a protein-DNA-RNA complex. *Proc Nat Acad Sci USA.* **68:** 6-9.

Stragier, P., Parsot, C., and Bouvier, J. (1985) Two functional domains conserved in major and alternate bacterial sigma factors. *FEBS Lett*. **187:** 11–15.

Sugino, A., Higgins, N.P., Brown, P.O., Peebles, C.L. and Cozzarelli, N.R. (1978) Energy coupling in DNA gyrase and the mechanism of action of novobiocin. *Proc Natl Acad Sci USA.* **75:** 4838-4842.

Tamai, E., Shimamoto, T., Tsuda, M., Mizushima, T. and Tsuchiya, T. (1998) Conversion of temperature-sensitive to –resistant gene expression due to mutations in the promoter region of the melibiose operon in *Escherichia coli*. *J Biol Chem.* **273:** 16860-16864.

Thanbichler, M., Wang, S.C., and Shapiro, L. (2005) The bacterial nucleoid: A highly organised and dynamic structure. *J Cell Biochem.* **96:** 506-521.

Thiel, A., Valens, M., Vallet-Gely, I., Espeli, O., and Boccard, F. (2012) Long-range chromosome organisation in *E. coli*: A site-specific system isolates the Ter macrodomain. *PLOS Genet*. **8:** Epub: e1002672.

Travers, A.A. (1980) Promoter sequence for the stringent control of bacterial ribonucleic acid synthesis. *J Bacteriol*. **141:** 973-976.

Travers, A., Schneider, R., and Muskhelishvili, G. (2001) DNA supercoiling and transcription in *Escherichia coli*: The FIS connection. *Biochemie*. **83:** 213-217.

Travers, A., and Muskhelishvili, G. (2005) Bacterial chromatin. *Curr Opin Genet Dev*. **15:** 507-514.

Travers A. and Muskhelishvili, G. (2007) A common topology for bacterial and eukaryotic transcription initiation? *EMBO reports.* **8:** 147-151.

Tsao, Y. P., Wu, H. Y., and Liu, L. F. (1989) Transcription-driven supercoiling of DNA: Direct biochemical evidence from in vitro studies. *Cell*. **56:** 111-118.

Tse-Dinh, Y.C., and Beran, R.K. (1988) Multiple promoters for transcription of the *Escherichia coli* DNA topoisomerase I gene and their regulation by DNA supercoiling. *J Mol Biol*. **202:** 735–742.

Tsien, R.Y. (1998) The green fluorescent protein. *Annu Rev Biochem*. **67:** 509-544.

Valens, M., Penaud, S., Rossignol, M., Cornet, F. and Boccard, F. (2004) Macrodomain organization of the *Escherichia coli* chromosome. *EMBO J.* **23:** 4330-4341.

van Noort, J., Verbrugge, S., Goosen, N., Dekker, C., and Dame, R.T. (2004) Dual architectural roles of HU: formation of flexible hinges and rigid filaments. *Proc Natl Acad Sci USA*. **101:** 6969–6974.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature*. **417**: 712-719.

von Freiesleben, U., Rasmussen, K.V., and Schaechter, M. (1994) SeqA limits DnaA activity in replication from *oriC* in *Escherichia coli*. *Mol Microbiol.* **14:** 763–772.

Vora, T., Hottes, A.K. and Tavazoie, S. (2009) Protein occupancy landscape of a bacterial genome. *Mol Cell.* **35:** 247-253.

Waldminghaus, T., and Skarstad, K. (2009) The *Escherichia coli* SeqA protein. *Plasmid*. **61:** 141-150.

Wang, J.C. (1971) Interaction between DNA and an *Escherichia coli* protein. *J Mol Biol*. **55:** 523-533.

Wang, W., Li, G-W., Chen, C., Xie, S., and Zhuang, X. (2011) Chromosome organisation by a nucleoid-associated protein in live bacteria. *Science*. **333:** 1445-1449.

Watson, J.D., and Crick, F.H.C. (1953) Molecular structure of nucleic acids. *Nature*. **171:** 737-738.

Webster, C., Kempsell, K., Booth, I. and Busby, S. (1987) Organisation of the regulatory region of the *Escherichia coli* melibiose operon. *Gene*. **59:** 253-263.

Wilson, C.J., Zhan, H., Swint-Kruse, L., and Matthews, K.S. (2007) The lactose repressor system: Paradigms for regulation, allosteric behaviour and protein folding. *Cell Mol Life Sci.* **64:** 2-16.

Woldringh, C.L. (2002) The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol Microbiol.* **45:** 17-29.

Wolf, B., Pato, M.L., Ward, C.B., and Glaser, D.A. (1968) On the origin and direction of replication of the *E. coli* chromosome. *Cold Spring Harbor Symp Quant Biol*. **33:** 575-584.

Worcel, A., and Burgi, E. (1972) On the structure of the folded chromosome of *Escherichia coli*. *J mol Biol*. **71:** 127-147.

Wu, H.Y., Shyy, S., Wang, J.C., and Liu, L.F. (1988) Transcription generates positively and negatively supercoiled domains in the template. *Cell*. **53:** 433-440.

Zaritsky, A. and Woldringh, C.L. (1978) Chromosome replication rate and cell shape in *Escherichia coli*: Lack of coupling. *J Bacteriol.* **135:** 581-587.

Zhang, W., and Baseman, J.B. (2011) Transcriptional regulation of MG_149, an osmoinducible lipoprotein gene from *Mycoplasma genitalium. Mol Microbiol*. **81:** 327–339.

Zhang, W., and Baseman, J.B. (2011) Transcriptional response of *Mycoplasma genitalium* to osmotic stress. *Microbiology*. **157:** 548–556.

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 Å Resolution. *Cell.* **98:** 811-824.