UNDERSTANDING INTRAGENIC TRANSCRIPTION AND ITS REGULATION

by

LISA ELLEVERA LAMBERTE

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

June 2016

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

June 2016

ABSTRACT

Transcription is the first step in gene expression. Thus, RNA polymerase copies information from template DNA to generate the mRNA template for protein production. Transcription is divided into three steps; initiation, elongation, and termination. Each step provides a platform for regulation of gene expression. For example, transcription initiation is dependent on promoter strength and the presence or absence of transcription factors. This work contributes to understanding novel aspects of regulatory processes.

For example, complex regulation of expression of the nucleoid-associated protein *cbpA* was found to be dependent not only on sigma factor specificity, but also on binding of the growth phase transcription factor Fis (factor for inversion stimulation) upstream of the *cbpA* start codon. Binding of Fis to the *cbpA* regulatory region prevents the action of a strong σ^{70}-dependent promoter found within the coding region of a neighbouring gene, *yccE*. This work carefully dissects the sequence specificity and orientation specificity of DNA sequences found upstream of the *cbpA* coding region that allows for the binding of Fis. Binding of Fis to the *cbpA* regulatory region is a redundant process, whereby in the absence of Fis another factor binds to this region. Furthermore, the regulation of *cbpA* was also found to be dependent on the interplay between the strong σ^{70}-dependent promoter (i.e. *cbpA* P6) and a weak but convergent σ^{32}-dependent promoter (i.e. P*yccE*).

Finally, horizontally acquired DNA, which is AT-rich in nature, has been known to be regulated by the transcription factor, H-NS (histone-like nucleoid structuring protein). However, the field only has limited knowledge on the mechanisms behind this regulation. It was previously thought that AT-rich genes are subject to canonical regulation by H-NS (i.e. H-NS silences activity

coming from the genuine promoter). However, this work demonstrates the phenomenon of "pseudo-regulation". Here, H-NS silences activity coming from intragenic promoters, rather than the genuine promoter. This phenomenon is likely widespread, and is demonstrated by this study in a number of AT-rich genes. In addition, the transcription terminator Rho provides an additional mechanism for regulation by terminating intragenic transcripts coming from the H-NS-targeted AT-rich genes.

ACKNOWLEDGEMENTS

The first page of this work contains only one name, that of mine. However, underneath that name is a whole series of names; the names of a few dear people who have contributed greatly to the production of this work and who I am deeply grateful for. The first and foremost of these people is my supervisor, Dr. David Grainger, who has given me an extraordinary opportunity to join his laboratory and become his student. I am thankful for his continued patience, persistence, and encouragement, which allowed me to grow intellectually as a molecular microbiologist. I would also like to thank my second supervisor, Dr. David Lee, who has always given me the utmost support and advice both while in the laboratory and during the writing of my reports. I am very thankful to the wonderful postdocs of the Grainger Lab, past and present—Dr. Kiran Chintakayala, Dr. Shivani Singh, and Dr. James Haycocks—for their patience in teaching me the protocols and methods of the Grainger Lab, especially during my first year. I am also very grateful to Gabriele Baniulyte who, aside from being a brilliant MSci student working with me on my project, is also a wonderful friend. Of course, my utmost thanks to all the members of the Grainger Lab, past and present; Prateek Sharma, Rachel Kettles, and Stephanie MacDonald for putting so much fun into the lab, and especially to Jainaba Manneh, Laura Sellars, and Emily Perry for being such awesome friends at work. Indeed, I am very grateful to be a part of such a vibrant and productive laboratory, working under a brilliant and talented supervisor.

I am very thankful to also have such amazing support outside of the laboratory, in the form of my awesome friends Ann Shapiro, Lucia Scotti, and Magnus Anderson. Sharing a house with

you has made you become like family to me, supporting me through the ups and downs of life since my arrival here in Birmingham.

I am deeply thankful to the people who I fondly call my "Core Support System". These are the people who have always been there for me and who I know will always be there for me, no matter what happens. To my magnificent friend and housemate, Supriya Malik, you have always provided invaluable hugs and support, especially during these last few months of writing. Words cannot express how I appreciative I am of your friendship (and how it has kept me sane!). To my amazing family—Tatay, Nanay, and Albert—your love and continued support and encouragement have been always been vital not only during my PhD, but life in general. I always have been, and always will be very, very fortunate and blessed to have such a loving and superb family. And to my wonderful and brilliant partner, Colin Jones, who has painstakingly been with me on this journey until the end, although I would always joke around that your belief in me sounds just like a cheesy anime, I would like you to know that it is that same belief that has always encouraged me to take the next step…and the next one…and the next one. You are truly a blessing, Mahal Ko.

Finally, my sincerest thanks goes to you, my dear reader, for taking the time and patience to read three long years worth of blood, sweat and tears. It has been a fantastic journey.

"It is good to have an end to journey toward; but it is the journey that matters, in the end."

-Ernest Hemingway-

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

A	Adenine
A/Amp	Ampere
Amp^R	Ampicillin resistance
ArcA	Anaerobic respiratory control A
AR1	Activating region 1
AR2	Activating region 2
AR3	Activating region 3
APS	Ammonium persulphate
BCM	Bicyclomycin
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
cAMP	Cyclic AMP
CbpA	Curved DNA binding protein A
c-di-GMP	Cyclic di-GMP
ChIP	Chromatin immunoprecipitation
ChIP-chip	ChIP combined with DNA microarray
ChIP-seq	ChIP combined with sequencing
Ci	Curie
CIP	Alkaline phosphatase, calf intestine
CTD	C-terminal domain

CRP	cAMP receptor protein
°C	Celsius (degree)
D	Aspartate
ddH_2O	Double distilled water
DEPC	Diethylpyrocarbonate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate
dsDNA	Double stranded DNA
DTT	Dithiothreitol
E	*Escherichia coli* holoenzyme
E. coli	*Escherichia coli*
EDTA	Diaminoethanetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
EtBr	Ethidium bromide
F	Farad
Fis	Factor for inversion stimulation protein
FNR	Fumarate and nitrate reductase protein
G	Guanine
GDP	Guanosine diphosphate
GTP	Guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfphonic acid

H-NS	Histone-like nucleoid structuring protein
HTH	Helix-turn-helix
IHF	Integration host factor
Kan^R	Kanamycin resistance
kb	Kilobase
LB	Lennox broth
mRNA	Messenger RNA
MeOH	Methanol
Mg	Magnesium
NAP	Nucleoid associated protein
n/nt	Nucleotide
NCR	Non-conserved region
NTD	N-terminal domain
NTP	Nucleoside triphosphate
OD	Optical density
Oligo	Oligonucleotide
ONPG	Ortho-nitrophenyl-β-galactosidase
ORF	Open reading frame
Ω	Ohm
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein data bank

ppGpp	Guanosine tetraphosphate
pppGpp	Guanosine pentaphosphate
psi	Pounds per square inch
RNA	Ribonucleic acid
RNA seq	RNA sequencing
RNAP	RNA polymerase
RNase	Ribonuclease
RP_O	Open promoter complex
SDS	Sodium dodecyl sulphate
sRNA	Small RNA
T	Thymine
TEMED	Tetramethylethylenediamine
Tet^R	Tetracycline resistance
TF	Transcription factor
TNSC (buffer)	Transcription Buffer
T4 PNK	T4 polynucleotide kinase
Tris	Tris (hydroxymethyl) aminoethane
U	Uracil
UP element	Upstream promoter element
V	Volts
W	Watts

Introduction

Chapter 1

1.1. Regulation of gene expression in *Escherichia coli* (*E. coli)*

Bacteria are exposed a range of environmental conditions, which may either be opportunities or threats to the cell's growth and survival. As such, bacteria are able to adapt to either of these conditions through the regulation of gene expression. Such regulation enables bacteria to control their own cellular structures and functions within a given environment. One of the mechanisms that bacteria employ to regulate gene expression is transcriptional control.

Transcription is the first step in decoding the information found in DNA. Hence, RNA products are derived from the DNA within a gene (Figure 1.1). The RNA products are then used as templates for the synthesis of proteins. The enzyme responsible for all transcription in bacteria is the RNA polymerase (RNAP). Transcription initiation starts with the formation of a closed complex, when the RNA polymerase docks at promoter DNA. This step can be controlled by both promoter DNA sequences and DNA binding proteins that can enhance or repress binding of RNAP to the DNA. Once bound to the promoter an open complex is formed by RNAP (deHaseth *et al.*, 1998). This complex contains an unwound region close to the transcription start site. It is within this open complex that the initial RNA chain is formed. Transcription then moves into the elongation phase. Here, RNA polymerase proceeds along the DNA template, transcribing one DNA strand, and the RNA chain is extended.

Finally, the process ends with either one of two pathways: (1) transcription is terminated when RNA polymerase encounters a "terminator" sequence that causes the nascent RNA chain to dissociate from the enzyme, or (2) transcription is terminated by a protein called Rho.

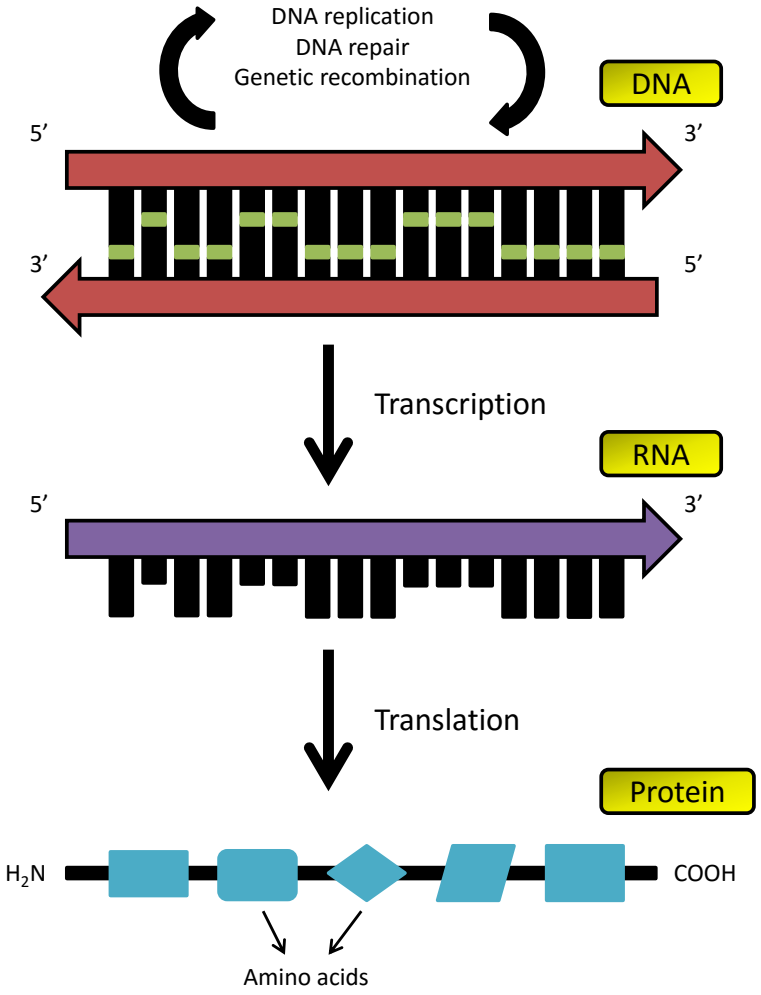


Figure 1.1. The pathway of gene expression.

Schematic diagram of the pathway of gene expression. Generally, genetic information flows from DNA to RNA to protein. DNA serves as a template for the synthesis of RNA (transcription), after which the resulting RNA molecules also serve as the template for the synthesis of polypeptides—the building blocks of proteins (translation). DNA is shown as red block arrows, bases are shown as black blocks, RNA is shown as a purple block arrow, the protein chain is shown as a black line, and amino acids are shown as light blue block shapes (Alberts *et al.*, 2002).

1.2. RNA polymerase (RNAP) core enzyme

The *Escherichia coli* RNA polymerase core enzyme is composed of five subunits; β, β', two α, and ω (Figure 1.2). The core enzyme is responsible for the generation RNA chains, in the 5' → 3' direction, via the hydrolysis of pyrophosphates from nucleoside triphosphates (NTPs). Each subunit of the core enzyme plays a role in this process (reviewed in Browning and Busby, 2004). Thus, the β and β' subunits contain the active site for RNA synthesis, and the sites of interaction between the DNA template and RNA product. The two α subunits (each made up of two independent domains connected by a flexible linker) are required for the first step of RNAP assembly. Thus, the assembly of RNA polymerase core enzyme initiates with the formation of α-subunit dimers. This is followed by association of the β and β' to form a claw structure with Mg^{2+} at the catalytic active site (Murakami and Darst, 2003). The smaller ω subunit facilitates the assembly of the RNA polymerase by assisting the folding of the β' subunit (Mathew and Chatterji, 2006).

1.3. Sigma factors (σ)

In order for transcription to initiate at specific sites, the RNA polymerase core enzyme requires a dissociable sigma factor (σ). The σ-associated RNA polymerase is called the holoenzyme. The σ factor plays an important role in transcription initiation because it establishes proper recognition of promoter sequences and aids open complex formation adjacent to the transcription start site (Browning and Busby, 2004). Many bacteria encode

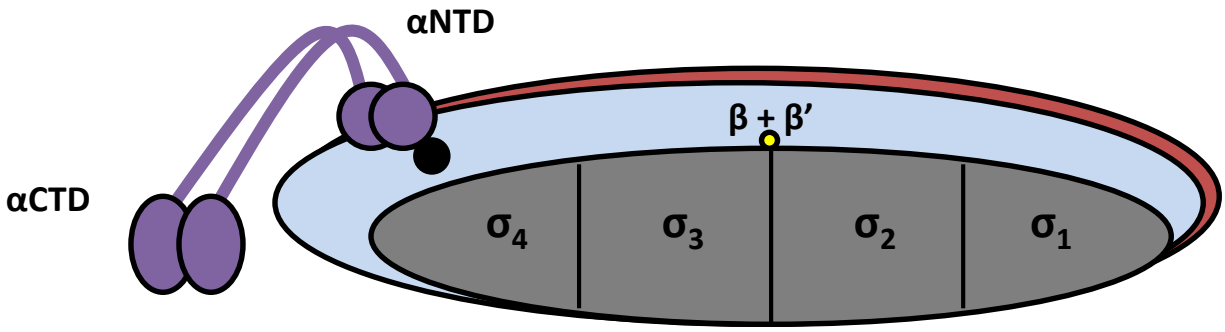


Figure 1.2. The RNA polymerase holoenzyme.

Schematic diagram of the RNA polymerase holoenzyme. The β and β' subunits are shown in coloured light blue and red oblongs, respectively. The α-subunits (αCTD and αNTD) are shown in purple, and the sigma factor (σ) is shown in grey. The Mg^{2+} is shown as a yellow circle, and is indicative of the RNA polymerase catalytic active site. The ω subunit is shown as a black circle (Browning and Busby, 2004)

multiple sigma factors. For example, *E. coli* contains seven derivatives of the σ subunit, each recognising a specific set of promoters (Lonetto *et al.*, 1992, Ishihama, 2000, Gruber and Gross, 2003, Paget and Helmann, 2003). The intracellular levels of these seven subunits vary according to environmental conditions; each subunit competes for a limited amount of core enzyme. For example, the intracellular concentrations of the housekeeping factor σ^{70} are at high levels during exponential phase. Thus, during exponential phase, σ^{70} is able to associate more frequently with RNAP. However, during altered physiological states (i.e. heat shock or stationary phase), the levels of the alternative sigma factors (e.g. σ^{38} or σ^{32}) vary considerably. Thus, during these altered physiological states, the alternative sigma factors are more likely to associate more frequently with RNAP than σ^{70}. Furthermore, after promoter escape, σ dissociates from RNAP and is only able to bind once more to RNAP after transcription termination. Thus, in some ways, σ acts as a molecular switch for the genome-wide pattern of transcription (Paget, 2015). Structurally, all sigma factors, except for those belonging to the σ^{54} family, have similar features. Each subunit is composed of four different domains joined by linkers, which bind across one face of the RNA polymerase (Figure 1.3). In the holoenzyme, the DNA binding determinants of σ are exposed to the -35 and -10 elements of the template DNA. Thus, the holoenzyme and promoter can interact. Domain 1 of σ is negatively-charged, and is only found in primary σ factors (collectively termed Group 1). Because of this negative charge, Domain 1 acts as a DNA mimic and occupies the RNAP active site that will subsequently be occupied by DNA in the open complex (Mekler *et al.*, 2002, Bae *et al.*, 2013, Murakami, 2013). Therefore, Domain 1 acts as a "gatekeeper" and ensures that σ recognises the promoter only when it is part of the holoenzyme. Domains 2, 3,

and 4 are involved in promoter recognition (Murakami *et al.*, 2002a, Murakami *et al.*, 2002b, Vassylyev *et al.*, 2002). The highly conserved section of Domain 2 (specifically, subdomain 2.4) recognises the -10 hexamer. Domain 4 (specifically, subdomain 4.2) recognises the -35 hexamer. The second domain of σ (specifically, subdomain 2.3) is also responsible for isomerisation; the binding of the non-template DNA strand of the open complex (Tomsic *et al.*, 2001).

The housekeeping σ factor, σ^{70}, is encoded by *rpoD* and is required to initiate transcription of most genes during normal growth (Lonetto *et al.*, 1992, Ishihama, 2000, Gruber and Gross, 2003, Paget and Helmann, 2003). This version of the holoenzyme ($E\sigma^{70}$) directs transcription by binding to a consensus sequence at target promoters. Through the years, experiments, using molecular through to genomic approaches, have been done to determine this consensus (Pribnow, 1975, Siebenlist *et al.*, 1980, Busby and Ebright, 1994, Gralla, 1996, Herring *et al.*, 2005, Cho *et al.*, 2009, Mooney *et al.*, 2009, Raghavan *et al.*, 2011, Giannoukos *et al.*, 2012). This consensus binding sequence contains two sets of hexanucleotides: (1) 5'-TTGACA-3' and (2) 5'-TATAAT-3', 35 bp and 10 bp upstream of the transcription start sites respectively. Between these two elements is a spacer region, 17 bp in length, which can also influence promoter activity. Around one fourth of $E\sigma^{70}$-dependent promoters are constitutive (Shimada *et al.*, 2014). Constitutive promoters are defined as sequences recognised *in vitro* by the RNAP holoenzyme in the absence of transcription factors. Hence, most promoters are classified as "inducible", meaning these promoters require the presence of transcription factors (TFs) for activity.

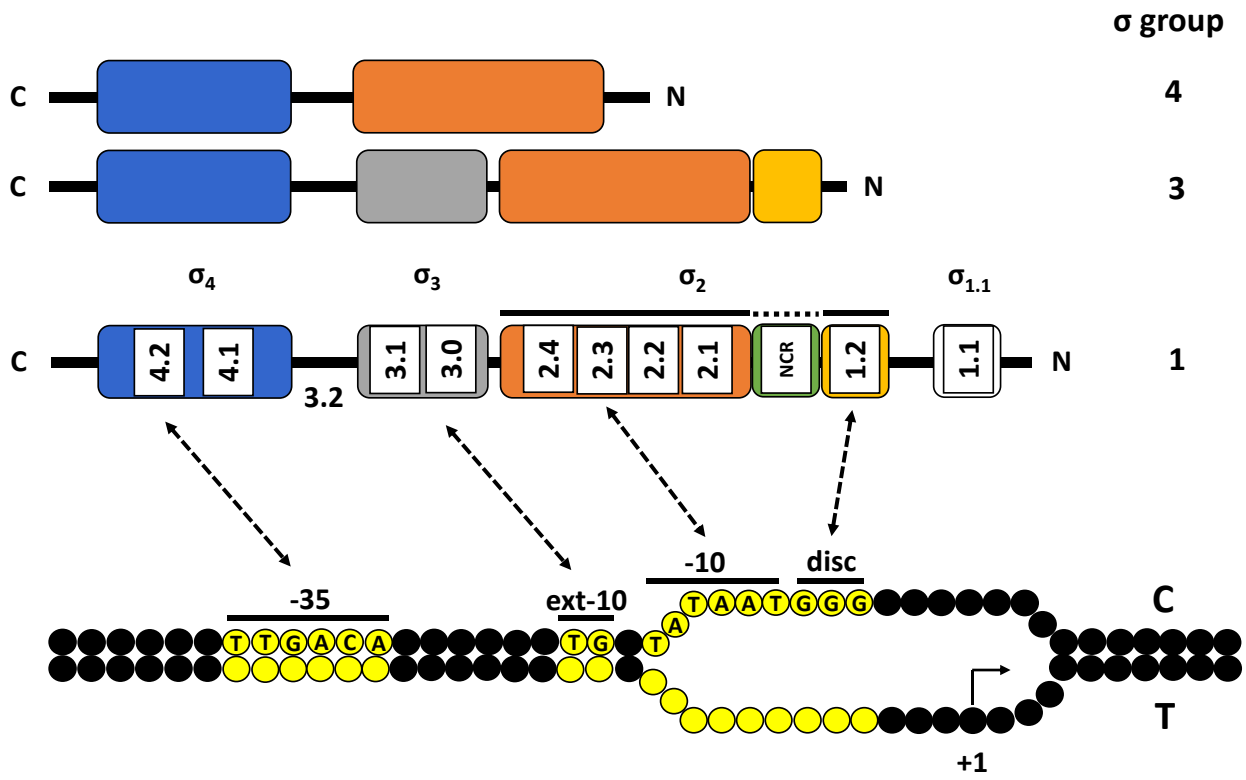


Figure 1.3. The organisation of σ^{70} domains.

Schematic diagram of the domain organization of Groups 1, 3, and 4 of the σ^{70} family. Structural domains are shown in coloured blocks. Conserved regions are shown in the same block colours. DNA is shown in solid black circles: the upper strand labelled "C" is the coding strand, the lower strand labelled "T" is the template strand. Promoter consensus sequences, which interact with σ, are shown in yellow block circles. These key promoter sequences are labelled "-35" (the -35 element), "ext-10" is (extended -10 element), "-10" (the -10 element), and "disc" (the discriminator). The transcription start site is labelled "+1". The coloured block labelled "NCR" indicates the non-conserved region (Paget, 2015).

Six genes encoding alternative σ factors are distributed across the *E. coli* genome (Wosten, 1998). The stationary phase σ factor, σ^{38}, is encoded by the *rpoS* gene and is required for the general stress response (Hengge-Aronis, 1996). Hence, a rise in the levels of intracellular σ^{38} is associated with restriction of nutrient availability. Numerous experiments have shown that the consensus promoter sequences for $E\sigma^{38}$ and $E\sigma^{70}$ are almost identical (Tanaka *et al.*, 1993, Espinosa-Urgel *et al.*, 1996, Gaal *et al.*, 2001). Because of this, many promoters bound by $E\sigma^{38}$ can also be bound by $E\sigma^{70}$ (Keseler *et al.*, 2005). Landini and colleagues (2014) have attributed the specificity of $E\sigma^{38}$ to specific determinants within the σ^{38} subunit and promoter. For example, site directed mutagenesis experiments reveal that different determinants in σ^{70} and σ^{38} are responsible for interactions within the -10 region and open complex formation. Specific nucleotide differences are found within the promoter elements in the DNA template, particularly within the extended -10 hexamer and the "discriminator" (region between the -10 element and the transcription start site). Notably, σ^{38} is more tolerant to a C nucleotide at position -12, compared to the more stringent requirement of a T nucleotide at the same position for σ^{70} (Figure 1.4) (Lacour *et al.*, 2003, Lacour *et al.*, 2004). Furthermore, σ^{38} specifically recognises a C nucleotide at position -13, thus extending the -10 element (Becker and Hengge-Aronis, 2001, Checroun *et al.*, 2004). Transcription factors also play an important role (Landini *et al.*, 2014). For example, at the *E. coli dps* promoter, the transcription factor Fis selectively represses transcription initiation by σ^{70}-associated RNAP, but not σ^{38}-associated RNAP. Fis represses by trapping the σ^{70}-associated RNAP at the *dps* promoter (Grainger *et al.*, 2008). Taken together, these experiments suggest

that the promoter selectivity for $E\sigma^{38}$ shows more tolerance in terms of specificity than that of $E\sigma^{70}$.

The heat shock σ^{32} factor is encoded by the *rpoH* gene and is responsible for the transcriptional response to heat stress (Grossman *et al.*, 1987, Segal and Ron, 1998). Thus, in contrast to $E\sigma^{70}$ and $E\sigma^{38}$, the RpoH holoenzyme, $E\sigma^{32}$, recognises a completely distinct set of promoter elements. By binding these promoters RNAP is directed to genes involved in the heat-shock response. The elements recognised by σ^{32} are (1) CTTGAAA and (2) CCCCATNT, located 35 bp and 10 bp upstream of the transcription start site respectively. The optimum spacer between these elements is 15 bp (Wang and deHaseth, 2003). Previous studies, using traditional biochemical and genetic methods, have identified a small number of σ^{32}-dependent genes in *E. coli* (Lemaux *et al.*, 1978, Zhao *et al.*, 2005). More recently, Wade and colleagues (2006), using chromatin immunoprecipitation coupled with microarray technology (ChIP-chip), were able to identify additional σ^{32} targets. Surprisingly, this work demonstrated that there is a functional overlap between $E\sigma^{32}$ and $E\sigma^{70}$, meaning that promoters recognised by $E\sigma^{32}$ often overlap promoters bound by $E\sigma^{70}$. The authors also proposed that such overlapping promoters often result in transcription initiating from the same nucleotide (Wade *et al.*, 2006).

As discussed above, σ factors establish promoter preferences for the RNA polymerase holoenzyme and subsequently guide the enzyme towards appropriate genes. Hence, when the cell responds to specific signals, the housekeeping σ factor (σ^{70}) is supplemented with an alternative σ factor (also known as "sigma switching") (Figure 1.5). This subsequently modifies the promoter preference of RNA polymerase within the cell (Gruber and Gross, 2003).

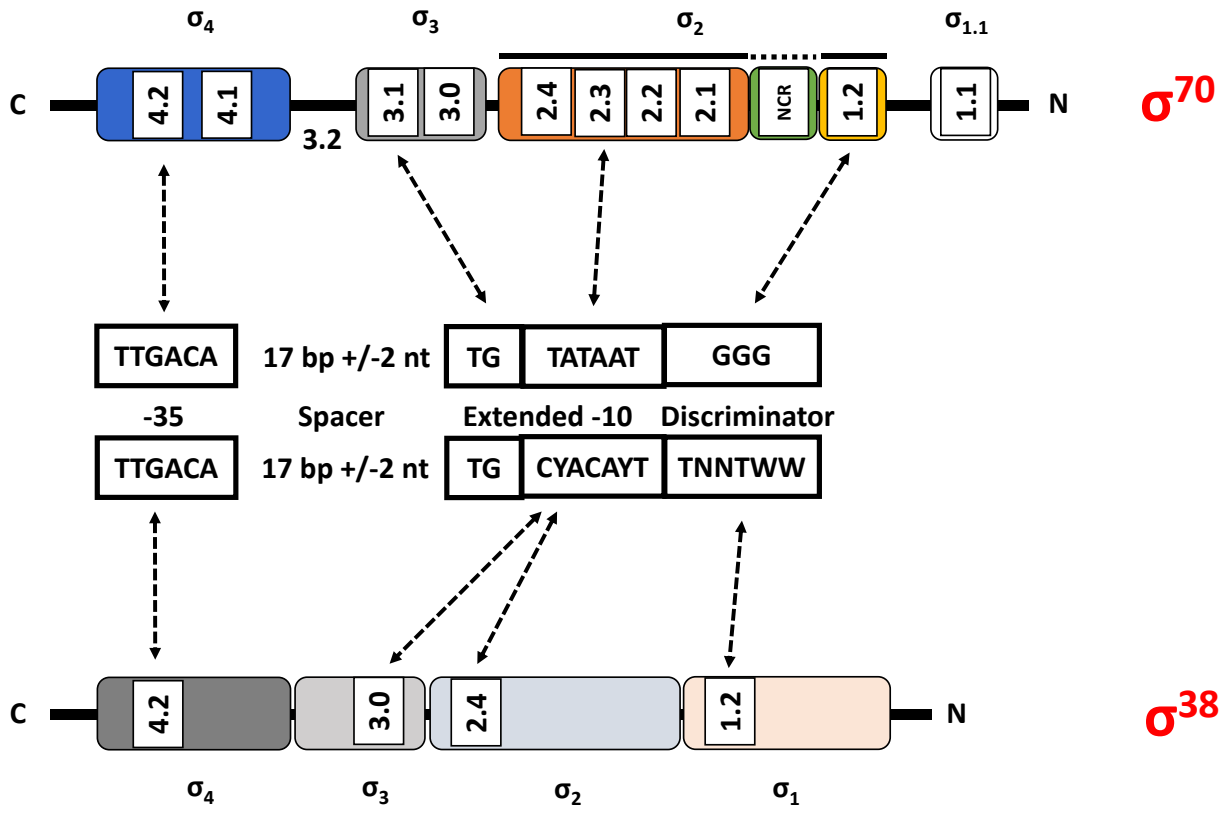


Figure 1.4. Comparison of specific determinants for σ^{70} and σ^{38}.

Schematic diagram comparing the different determinants in σ^{70} and σ^{38} responsible for interactions within the promoter region. Structural domains are shown in coloured blocks. Conserved regions are shown in the same block colours. Consensus sequences of the template DNA are shown in black outlined box with text. Y = C or T. W = A or T.

1.4. Promoter recognition

Promoter recognition by the RNA polymerase holoenzyme is a crucial step in transcription initiation, and, as such, promoter sequences have been studied extensively (Busby and Ebright, 1994, Gross, 1998). An optimal promoter is made up of four elements: the -35 and -10 hexamers, the extended -10 element, and the UP element. As outlined in Section 1.3, the -35 and -10 hexamers are located 35 and 10 base pairs upstream of the transcription start site. The extended -10 element has the sequence TGn and is located immediately upstream of the -10 hexamer (Bown *et al.*, 1997, Murakami *et al.*, 2002a, Sanderson *et al.*, 2003). UP elements are AT-rich, around 20 bp in length, and are located upstream of the -35 hexamer. With the exception of the UP element, which interacts with the C-terminal domain of the two RNA polymerase α subunits (αCTD), most promoter-RNAP interactions are mediated by the σ subunit (Gourse *et al.*, 2000, Ross *et al.*, 2003). Few promoters have a full complement of DNA sequence elements and those present need not be perfect. These imperfections, as well as competition between promoters for a limited supply of RNA polymerase (Ishihama, 2000, Maeda *et al.*, 2000), enables the cells to regulate the transcription of genes according to environmental conditions.

Schematic diagrams depict interactions between RNA polymerase, and the promoter DNA template, in Figures 1.3, 1.4, and 1.6. In many cases the two α subunits of RNA polymerase, along with the helix-turn-helix of σ Domain 4, first interact with the promoter via the UP element and the -35 element respectively (Buckle *et al.*, 1999, Sclavi *et al.*, 2005, Davis *et al.*, 2007, Saecker *et al.*, 2011). This may be facilitated by transcriptional activators

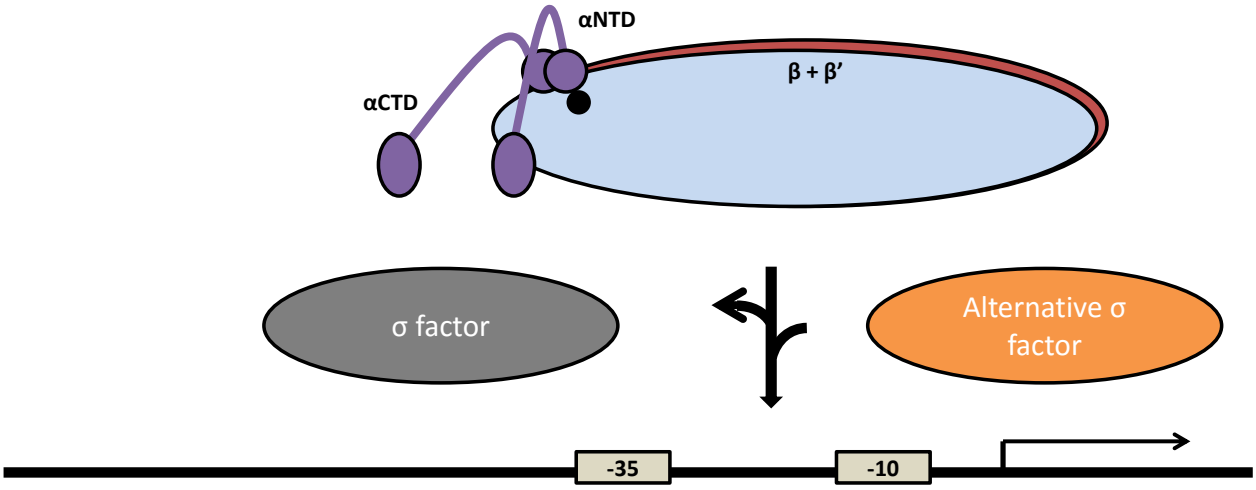


Figure 1.5. Sigma switching.

Schematic diagram of σ switching, or σ factor appropriation. An alternative σ factor alternates the promoter preference of RNA polymerase. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, and the alternative σ subunit is shown as an orange oval, the promoter regions are shown as white blocks, the extended -10 region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle (Lee *et al.*, 2012).

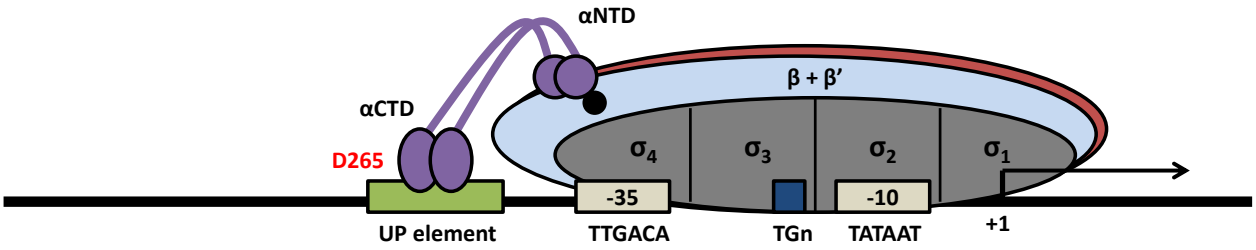


Figure 1.6. RNA polymerase and its promoter interactions.

The *E. coli* RNA polymerase is made up of a core enzyme (β, β', α_2, and ω) which is responsible for the actual transcription mechanism and a dissociable sigma subunit which directs the core enzyme to recognize appropriate promoters. Promoter sequences are typically made up of two sets of consensus sequences, TTGACA (-35) and TATAAT (-10). The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, the promoter regions are shown as white blocks, the extended -10 element (TGn) is shown as a dark blue box, the UP element is shown as the light green box, the α subunits are drawn in purple, and the DNA template is shown as an black line. Residue D265, which is responsible for the interactions between the αCTD and the UP element, is shown in red text. The ω subunit is shown as a black circle (Browning and Busby, 2004, Lee *et al.*, 2012).

bound upstream of the -35 element. Interactions involving the -10 region follow and result in DNA melting. Thus, the highly conserved σ Domain 2 interacts with the -10 element of the single-stranded non-template DNA (coding strand). This process traps the DNA and stabilises the open promoter complex (RP_o). In particular, the A_{-11} and the T_{-7} bases of the -10 element ($T_{-12}A_{-11}T_{-10}A_{-9}A_{-8}T_{-7}$) are flipped out of the base stack and buried in specific pockets of σ Domain 2 (Feklistov and Darst, 2011, Zhang *et al.*, 2012). Further downstream, subdomain 1.2 of σ interacts with the discriminator sequence (5'-GGG-3') (Zhang *et al.*, 2012). At promoters with an extended -10 element, the compact three-helix bundle of σ Domain 3 interacts with the major groove of the double helix (Mitchell *et al.*, 2003). Such interactions are particularly beneficial at promoters where the -35 promoter element is not present. Remarkably, σ Domain 1.1 does not interact with a particular sequence motif in the template DNA. Rather, this negatively-charged domain inhibits the RNA polymerase active site channel by blocking the site normally occupied by double-stranded DNA (Mekler *et al.*, 2002, Bae *et al.*, 2013, Murakami, 2013). Thus, σ Domain 1.1 acts as a DNA mimic and can either stimulate or inhibit promoter isomerisation (Vuthoori *et al.*, 2001, Hook-Barnard and Hinton, 2009). In some Group 1 σ factors, a non-conserved region (NCR) can be found. This domain has been implicated in core enzyme binding and promoter escape (Leibman and Hochschild, 2007). The various steps that occur between promoter binding and promoter escape are shown in Figure 1.7. These events are all subject to regulation by small ligands, transcription factors, and chromosome structure. Hence, cells fine tune gene expression according changes in the environment.

1.5. Regulation by small ligands

There are a number of small ligands that promote a fast and efficient response to environmental factors. In bacteria, these are cyclic AMP, cyclic di-GMP (c-di-GMP), and guanosine tetraphosphate (ppGpp). Of these, ppGpp can control transcription by directly interacting with RNA polymerase. Thus, ppGpp and guanosine pentaphosphate (pppGpp), collectively known as (p)ppGpp, accumulate in amino acid-deprived cells (Cashel *et al.*, 1996). Synthesis of (p)ppGpp is controlled by the synthestase RelA (Potrykus and Cashel, 2008). RelA associates with ribosomes. During amino acid depletion, uncharged tRNAs accumulate and are sensed by the ribosomes. This event triggers the synthetase activity of RelA. Formation of (p)ppGpp takes place when a pyrophosphate moiety is added to the 3' position of GDP (for ppGpp) and GTP (for pppGpp). The production of these small molecules is associated with growth arrest and the activation of adaptive responses (Cashel *et al.*, 1996, Takahashi *et al.*, 2004). This (p)ppGpp-induced change is commonly known as the "stringent" response and involves inhibition of RNA synthesis. Recent studies have shown that ppGpp interacts at a binding site between the RNAP β' subunit and the ω subunit (Ross *et al.*, 2013, Zuo *et al.*, 2013). This particular interaction restrains the movement of RNAP, which in turn, slows down the open complex formation. Thus, the ligand (p)ppGpp affects the production of translation machinery through the destabilisation of open complexes at rRNA promoters (Barker *et al.*, 2001b, Barker *et al.*, 2001a, Ross *et al.*, 2013, Zuo *et al.*, 2013). The cell's capacity to generate these small ligands greatly affect a number of key cellular processes such as replication, transcription, and translation, as well as other important adaptive responses such as the induction of virulence, differentiation, and persistence (Srivatsan and Wang, 2008).

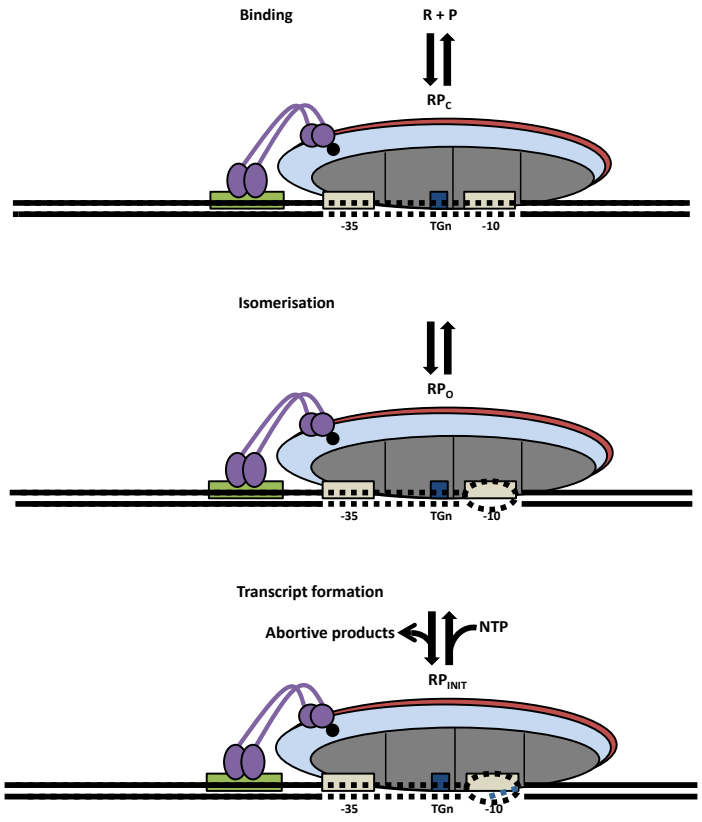


Figure 1.7. Transcription initiation at bacterial promoters.

Schematic diagram of transcription initiation at bacterial promoters. Transcription initiates when RNA polymerase binds to the promoter (binding). After binding, isomerisation to the open complex takes place when the duplex DNA is being unwound around the transcription start site (isomerisation). Finally, RNA transcripts are formed with the addition of NTPs (transcript formation). Stochastically, RNA polymerase can proceed either towards the pathway to promoter escape and elongation, or towards the release of the abortive products and, consequently, returns to the open complex formation. The β and β' subunits are shown in coloured light blue and red oblongs, respectively. The α-subunits (αCTD and αNTD) are shown in purple, and the sigma factor (σ) is shown in grey. The ω subunit is shown as a black circle. R – RNA polymerase. P

– promoter. RP_c – closed promoter complex. RP_o – open promoter complex. RP_{INIT} – initiating complex. NTP – nucleoside triphosphates. Double-stranded DNA is shown in black lines. Unwound DNA complex is shown as a bubble. The RNA chain is shown as blue dashed and solid lines. The UP element is shown as a green box (Browning and Busby, 2004, Lee *et al.*, 2012).

1.6. Regulation by transcription factors

Transcription factors can up- or down-regulate gene expression (Perez-Rueda and Collado-Vides, 2000, Madan Babu and Teichmann, 2003). Some transcription factors are specific to one or two promoters, whilst others have a global role at many promoters. Furthermore, certain regulators can act as an activator or a repressor depending on local promoter organisation (Perez-Rueda and Collado-Vides, 2000). In *E. coli,* global regulators include the cyclic AMP receptor protein (CRP), the factor for induction of fumarate reductase and nitrite reductase (FNR) and the anaerobic respiratory control A protein (ArcA) (Martinez-Antonio and Collado-Vides, 2003). These and other transcription factors can be controlled by small ligands, covalent modification, the concentration of the transcription factors themselves, and blocking by a regulatory protein to which it binds (Darwin and Stewart, 1996, Demple, 1996, Muller-Hill, 1996, Stock *et al.*, 2000, Plumbridge, 2002). Thus, these transcription factors are a direct link between gene expression and environmental conditions.

1.7. Transcription Activation

Transcription factors that increase promoter activity are called activators. Perhaps the most straightforward mode of transcription activation is the simultaneous binding of a single activator to a promoter and RNA polymerase. This process allows the activator to either stabilise the RNA polymerase-promoter complex or to accelerate the transition from the closed to the open complex. Traditionally, activator-dependent promoters have been grouped into three classes (I, II, and III) depending on the positions of the activator binding sites. However, in their recent review, Lee and

colleagues (2012) emphasised that a simple classification cannot be applied to most activator-dependent promoters. Thus, I will discuss transcription activators as either having a promoter-centric, RNA polymerase-centric, or indirect mode of action.

Promoter-centric mechanism: Some promoters are organised so that RNAP binding is unfavourable. To overcome this, an activator can bind at, or very near to, the core promoter elements and alter promoter conformation (Figure 1.8). The best example for this mechanism is transcription activation by the MerR family of regulators. Such proteins bind the sub-optimal spacer between the -10 and -35 regions and alter the juxtaposition of these elements (Heldwein and Brennan, 2001, Brown *et al.*, 2003, Reyes-Caballero *et al.*, 2011). Thus, at the *merP* promoter, non-optimal spacing between the -10 and -35 hexamers occurs because the -10 element is misplaced. This hinders subsequent DNA melting and interaction of the -10 element with σ Domain 2. This problem is negated by MerR or BmtR. These factors cause a base pair distortion upon DNA binding that results in realignment of the -10 and -35 elements. Thus, transcription is activated (Heldwein and Brennan, 2001).

RNA polymerase-centric mechanism: Whilst promoter-centric activators modulate DNA topology, RNAP-centric activators target RNAP directly. For example, an activator may bind immediately upstream of, or overlapping, the promoter -35 element and may contact σ Domain 4 (Figure 1.9) (Busby and Ebright, 1994, Hochschild and Dove, 1998, Perez-Rueda *et al.*, 1998, Dove *et al.*, 2003). However, contacts with the activator are not restricted to σ. Hence, in some cases, the activator interacts with αNTD or αCTD (Busby and Ebright, 1997). Unlike promoter-centric systems, the positioning of RNAP-centric activator binding sites is less flexible. This is due to structural constraints required for the activator:RNAP interaction. A well-studied RNAP-centric

mechanism involves the bacteriophage λ P_{RM} promoter. The promoter is activated by the bacteriophage λ cI protein (described as "molecular velcro"). Hence, a single cI dimer binds to a position next to the boundary of the -35 element. This binding process allows the λ cI protein to interact with σ Domain 4 (Nickels *et al.*, 2002). This process results in the acceleration of open complex formation, thereby activating transcription (Jain *et al.*, 2004). In some cases, binding of the transcription activator to the site overlapping the -35 element is triggered by a ligand (Martin and Rosner, 2001). For instance, the ligand arabinose acts on the transcription activator AraC that directs the binding of RNA polymerase to the *araBAD* promoter (Zhang *et al.*, 1996, Schleif, 2010). Activators may also bind further upstream of the -35 element. Here, the α subunits of RNA polymerase play a key role by functioning as "molecular antennas" (contact points for the activator protein). Interactions between the activator CRP and αCTD are particularly well defined (Figure 1.10). For example, at the *lac* promoter, the homodimeric CRP binds to a site upstream of the -35 element and interacts with one or both of the RNAP αCTDs through a surface-exposed patch known as activating region 1 (AR1) (Figure 1.10A) (Busby and Ebright, 1999, Lawson *et al.*, 2004). The AR1 interaction with αCTD is also important when CRP binds sites overlapping the -35 element. Thus, activation proceeds accordingly. First, the bound CRP homodimer prevents the αCTD from interacting with Domain 4 of the sigma subunit. This, in turn, places the αCTD in a position to bind upstream of the activator (Zhou *et al.*, 1994, Belyaeva *et al.*, 1996). Second, the bound CRP homodimer is now in a position to interact with three contact points in the RNA polymerase holoenzyme: activating region 1 (AR1) interacts with a target site on the surface of the αCTD (Savery *et al.*, 1998), activating region 2 (AR2) interacts with a site on the surface of the αNTD (Niu *et al.*, 1996), and activating region 3 (AR3) interacts with a target site on the surface of Domain 4 of the sigma subunit

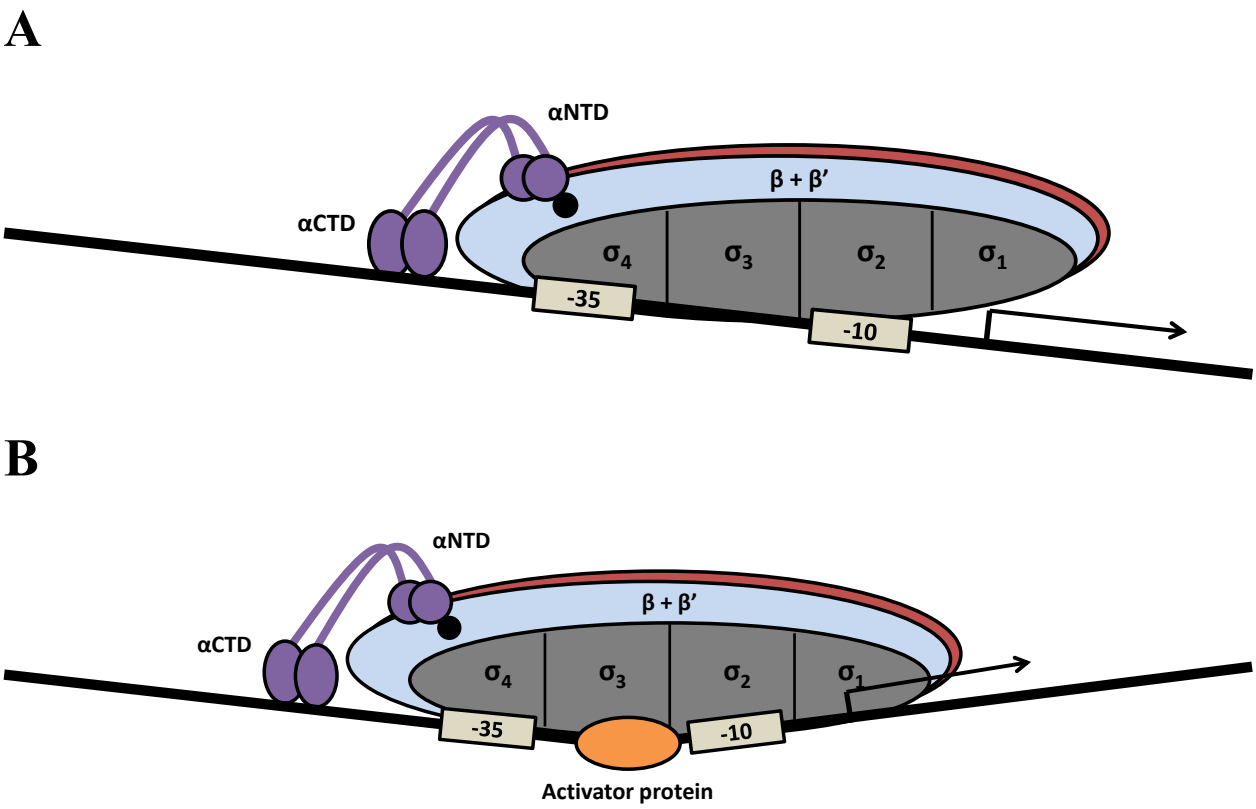


Figure 1.8. Simple activation of transcription (promoter-centric mechanism).

Schematic diagram of the promoter-centric mechanism of transcription activation. (A) The -10 element of the DNA template is misaligned with Domain 2 of the sigma subunit. (B) An activator proteins binds within the immediate area of the promoter regions and readjusts their conformation to increase the binding affinity of RNAP to the promoter. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, the promoter regions are shown as white blocks, the extended -10 region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle (Lee *et al.*, 2012).

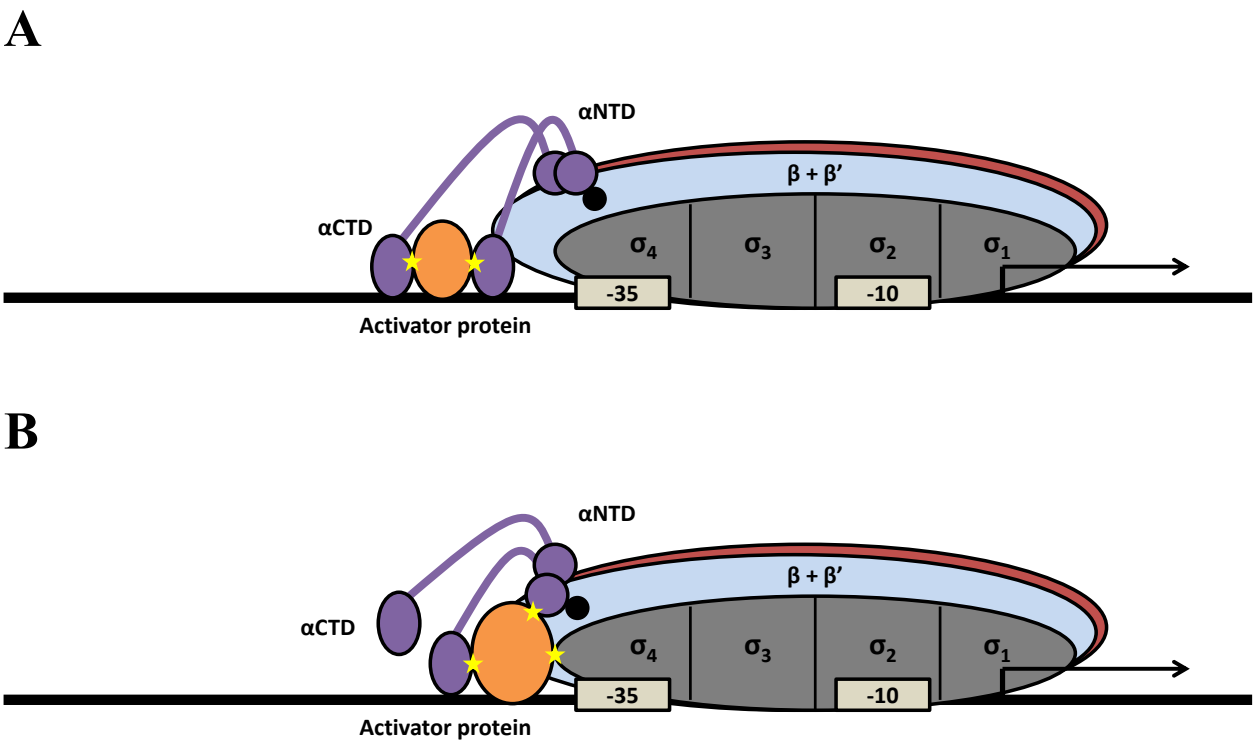


Figure 1.9. Simple activation of transcription (RNA polymerase-centric mechanism).

Schematic diagram of the RNA polymerase-centric mechanism of the transcription activation. (A) The activator protein binds upstream of the -35 element and interacts with one or both of αCTDs of RNA polymerase. (B) The activator binds to a site immediately adjacent to the -35 region and interacts with σ_4. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, the promoter regions are shown as white blocks, the extended -10 region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle. Interactions are illustrated by yellow stars (Lee *et al.*, 2012).

(Rhodius and Busby, 2000a, Rhodius and Busby, 2000b). Thus, the whole homodimer (i.e. both subunits) interacts with the RNA polymerase holoenzyme (i.e. core enzyme and the sigma subunit) in order to activate transcription (Figure 1.10B). CRP may also activate transcription by binding tandem targets. In these scenarios CRP can interact with the two αCTDs and/or other RNA polymerase subunits (Figures 1.10C and 1.10D) (Joung *et al.*, 1993, Murakami *et al.*, 1997, Belyaeva *et al.*, 1998, Tebbutt *et al.*, 2002, Beatty *et al.*, 2003).

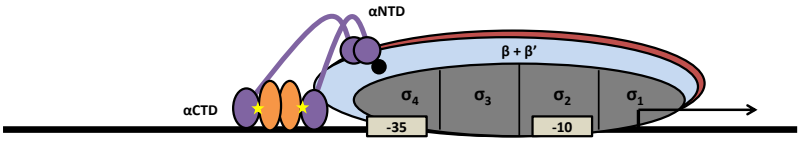
Indirect mechanism: In addition to direct (i.e. promoter-centric and RNA polymerase-centric) mechanisms, transcription can also be activated indirectly. For instance, "antirepressors" bind to DNA and displace transcriptional repressors (Dillon and Dorman, 2010, Frederix *et al.*, 2011). A schematic diagram of antirepression is shown in Figure 1.11. This mechanism is well-illustrated by the Ler and VirB proteins of enteropathogenic *E. coli* and *Shigella flexneri*, respectively. These proteins activate transcription of virulence genes by relieving repression due to the nucleoid-associated H-NS (histone-like nucleoid structuring) protein (Stoebel *et al.*, 2008).

1.8. Transcription repression

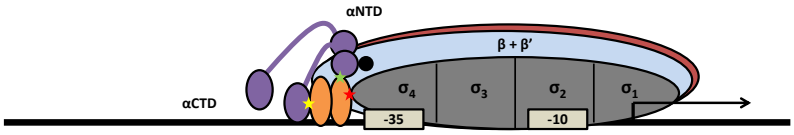
Transcription factors that reduce the affinity for RNAP for a target promoter are called repressors. A schematic diagram of transcriptional repression is illustrated in Figure 1.12. Repression mechanisms can be grouped into three classes:

Repression by steric hindrance: The most straightforward mechanism is steric hindrance; the repressor binds a site overlapping the promoter and prevents RNAP binding (Figure 1.12A).

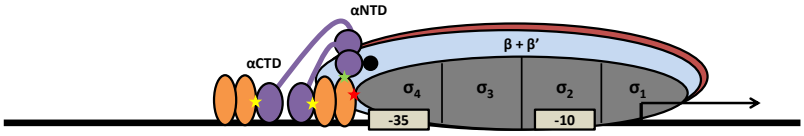
A



B



C



D

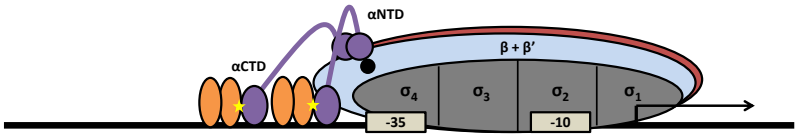


Figure 1.10. Transcription activation by CRP.

Schematic diagram of transcription activation by CRP. (A) The CRP homodimer binds upstream of the RNA polymerase holoenzyme and interacts with the αCTD through the AR1. (B) The CRP homodimer binds upstream adjacent to the RNA polymerase and interacts with different RNA polymerase subunits through the AR1 and AR2, and σ Domain 4 through AR3. (C) Tandem bound CRP homodimers contacting through AR1, AR2, and AR3. (D) Tandem bound CRP homodimers contacting through AR1. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, the promoter regions are shown as white blocks, the extended -10

region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle. Interactions are illustrated by yellow (AR1), green (AR2) and red (AR3) stars (Lee *et al.*, 2012).

Repression by steric hindrance is observed at the *lac* promoter. Here, the *lac* repressor binds to the *lac* operator sequences overlapping the RNAP binding site. Interactions between LacI and DNA are determined by the presence or absence of lactose. Hence, the operon is repressed in the absence of lactose, and is likewise de-repressed in the presence of lactose (Bell and Lewis, 2001).

Repression by DNA looping: There are instances when repressors interfere with post-recruitment steps of transcription initiation (Muller-Hill, 1998). For example, such a repressor mechanism is repression by DNA looping. Here, the repressor binds to sites further away from the promoter. Interactions between repressors cause the DNA to loop, rendering RNAP either unable to make contact with, or trapped at, the promoter (Figure 1.12B). This mechanism is illustrated when the transcription factor GalR represses the *gal* promoter (Choy and Adhya, 1996).

Repression by modulation of an activator: A further complex mechanism is repression by anti-activation. Thus, the repressor binds adjacent to an activator and prevents the activator from performing its function (Figure 1.12C). A well-characterised example of this mechanism is repression by CytR. Thus, CytR interacts with CRP to prevent CRP-dependent activation of CytR-repressed promoters, such as the *E. coli deo*P2 promoter (Valentin-Hansen *et al.*, 1996, Shin *et al.*, 2001).

1.9. Spurious transcription

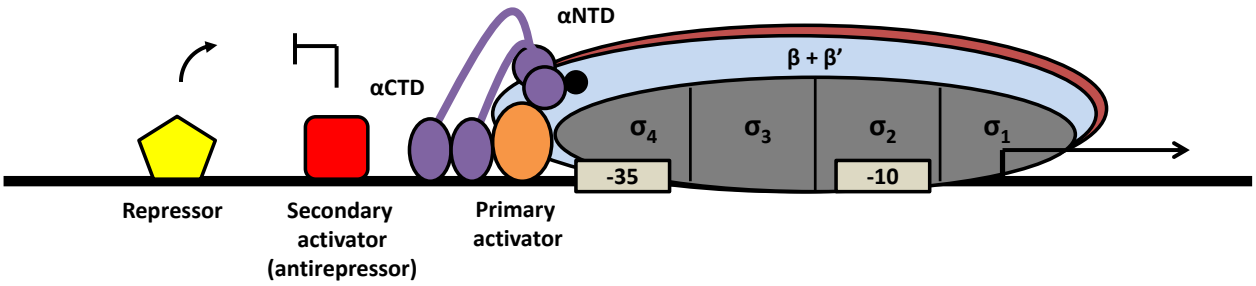
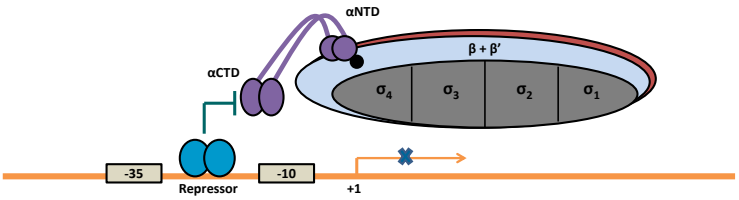


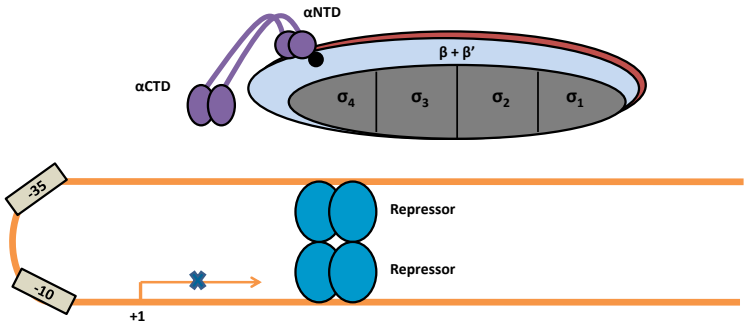
Figure 1.11. Transcription activation by indirect mechanisms (antirepression).

Schematic diagram of antirepression. Activation by the primary activator is maximised when the secondary activator binds to the DNA template and disrupts the repression caused by the repressor. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, and the alternative σ subunit is shown as an orange oval, the promoter regions are shown as white blocks, the extended -10 region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle. Protein factors are illustrated as an orange circle (primary activator), red square (antirepressor), and yellow pentagon (repressor) (Lee *et al.*, 2012).

A



B



C

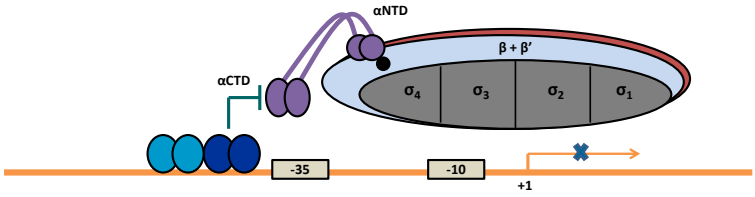


Figure 1.12. Simple repression of transcription.

Schematic diagram of the simple repression of transcription. Simple repression involves action by one repressor. There are three classes: (A) Repression by steric hindrance, where the repressor binds to a site overlapping the promoter regions thereby blocking promoter recognition by RNAP. (B) Repression by looping, where the repressors bind to sites distal to the promoter and causes the DNA to loop, rendering the promoter regions inaccessible to RNAP. (C) Repression by modulation of an activator, where the repressor interacts with the activator and prevents the activator from

performing its function. The β and β' subunits are shown as red and black ovals, the σ subunit is shown as the grey oval, and the alternative σ subunit is shown as an orange oval, the promoter regions are shown as white blocks, the extended -10 region is shown as a dark blue box, α subunits are drawn in purple, and the DNA template is shown as an orange line. The ω subunit is shown as a black circle (Browning and Busby, 2004).

Whilst transcription resulting in the formation of mRNA or rRNA is an essential process, and carefully regulated, there are instances when spurious RNA transcripts are made. The generation of these noncanonical transcripts can be due to the chance occurrence of promoters in genes or inefficient transcription termination (Peters *et al.*, 2009, Peters *et al.*, 2012, Singh *et al.*, 2014). Remarkably, intragenic promoters appear to be widespread in bacteria. For example, in *Neisseria gonorrhoeae*, non-coding RNAs are generated from sites within the majority of genes in both the sense and antisense orientations (Wachter *et al.*, 2015). Similarly, Singh and colleagues (2014) have shown that the AT-rich *ehxCABD* operon of *E.coli* 0157:H7 contains many intragenic promoters, both in the forward and reverse orientations. Such spurious transcription is potentially detrimental to the cell. Spurious transcripts may disturb the synthesis of mRNA or participate in the formation of DNA-RNA hybrids (R loops), thus destabilizing genome integrity (Gowrishankar and Harinarayanan, 2004, Dornenburg *et al.*, 2010). As such, it is essential that bacteria have control measures to suppress this phenomenon. In the recent years, a number of studies have investigated such mechanisms of suppression. For example, the transcription terminator Rho and the cofactor NusG were found to prevent the production of spurious transcripts occurring downstream of gene boundaries and from antisense promoters found within genes (Peters *et al.*, 2012). Furthermore, at the *E. coli* 0157:H7 *ehxCABD* operon, H-NS directly suppresses intragenic promoters in both sense and antisense orientations (Singh *et al.*, 2014).

1.10. Nucleoid associated proteins

Nucleoid-associated proteins (NAPs) are low molecular mass polypeptides that influence bacterial chromosome structure. NAPs are highly abundant in the cell but levels may fluctuate according to growth conditions. Binding of NAPs to DNA alters DNA shape (locally or genome-wide) and consequently NAPs can affect many DNA based processes including transcription, replication, transposition, and recombination (Makris *et al.*, 1990, Freundlich *et al.*, 1992, Haykinson and Johnson, 1993, Swingle *et al.*, 2004, Chodavarapu *et al.*, 2008, Browning *et al.*, 2010, Dillon and Dorman, 2010, Liu *et al.*, 2011). Thus, nucleoid architecture and the global regulation of gene expression are tightly linked.

1.11. Fis (Factor for inversion stimulation)

The factor for inversion stimulation, or Fis, is a highly abundant protein that is prolifically expressed during rapid growth (Azam and Ishihama, 1999). More specifically, Fis reaches peak concentrations during early growth phase, as the cells exit lag phase (Ball *et al.*, 1992, Keane and Dorman, 2003). Fis concentrations then rapidly regress until the protein becomes undetectable during stationary phase. The expression of *fis* however, can still be maintained throughout stationary phase under micro-aerobic conditions (O Croinin and Dorman, 2007, Cameron *et al.*, 2013). When present, Fis binds across the genome and greatly impacts the major processes of the cells, such as replication, transcription, translation, transposition, and site-specific recombination (Dorman, 2014). In particular, Fis influences the global transcription profile of the cell by acting either as an activator or a repressor at many promoters when necessary. For example, in the P2 promoter, Fis and CRP work in tandem albeit independently, as activators (Belyaeva *et al.*, 1998,

McLeod *et al.*, 2002). Additionally, at ribosomal RNA operon promoters, Fis interacts with RNAP directly to activate transcription (Hirvonen *et al.*, 2001, Aiyar *et al.*, 2002). Fis can also function as a transcriptional repressor at the *nir* and *nrf* promoters (Browning *et al.*, 2002, Browning *et al.*, 2000). Thus, the binding of Fis, along with IHF, represses transcription activation by FNR.

The X-ray crystal structure of Fis shows Fis to be a homodimer, with each 98-amino acid monomer comprising four α helices (A, B, C and D) (Kostrewa *et al.*, 1991, Yuan *et al.*, 1991). Fis binds to DNA through the C (residues 74-81) and D (residues 85-94) helices that form a helix-turn-helix (HTH) DNA binding motif. Specifically, the C-terminal D helix (the recognition helix of the HTH) is inserted into adjacent major grooves of the concave surface of the DNA (Figure 1.13) (Stella *et al.*, 2010). This results in the DNA being bent by ~50-90°. High affinity Fis binding sites are 15-bp AT-rich sequences bounded by a G and a C at the 5' and 3' ends respectively. The AT-content of the binding site allows compression of the minor groove on Fis binding (Shao *et al.*, 2008, Stella *et al.*, 2010). Additionally, most well defined Fis sites have a pyrimidine-purine (Y-R) step at positions ±(3-4). However, this feature of the Fis binding site is less apparent in more recent genome-wide studies (Hancock *et al.*, 2016). Structural studies of the Fis-DNA complex show that Fis contacts the DNA backbone over a total of 21 bp (Stella *et al.*, 2010). Residue R85, which is the first residue of the D helix, protrudes into the major groove and directly contacts the conserved guanines at ±7.

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

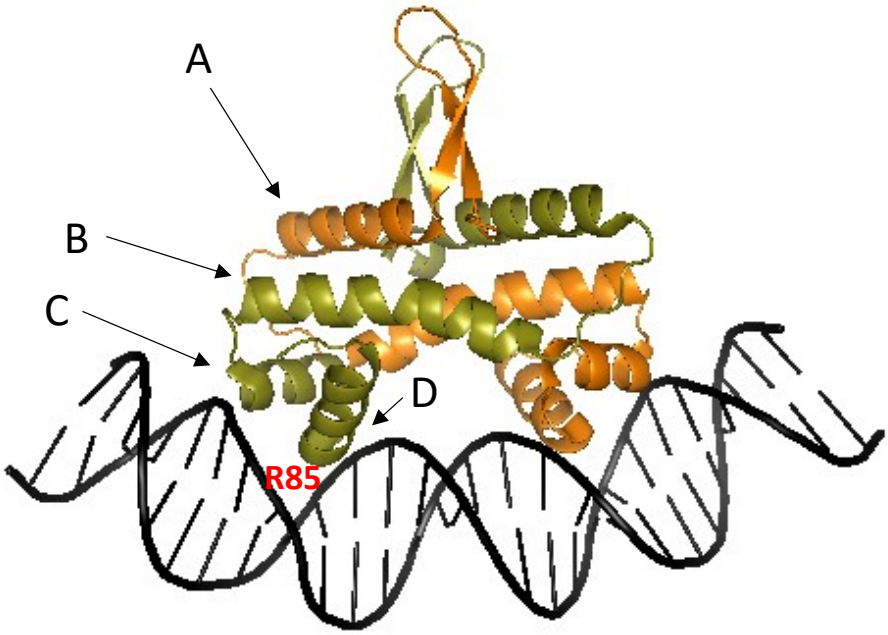


Figure 1.13. Structure of a Fis-DNA complex.

The figure shows the structure of a Fis homodimer-DNA complex (PDB ID: 3JRA). The four α helices of each Fis monomer are shown in similar colour schemes (e.g. orange and green) and are labelled as "A", "B", "C", and "D". Note that helices "C" and "D" are the HTH DNA binding domains. The C-terminal domain of helix "D" is the recognition helix, with residue R85 (shown in red text) contacting the DNA (shown in black).

Perhaps the most well-studied NAP is H-NS; a small (15.5 kDa) DNA folding protein highly abundant in bacteria. H-NS is expressed in all phases of growth and, importantly, exerts a global influence on the structure and architecture of the nucleoid as well as acting as a transcriptional silencer (Dorman, 2014). H-NS is composed of 137 amino acids, and is organised in at least two structural domains joined via an unstructured flexible linker (Figure 1.14). The N-terminal domain is involved in oligomerisation and the C-terminal domain controls DNA binding (Arold *et al.*, 2010, Renault *et al.*, 2013). H-NS can organise itself into DNA-protein complexes in either bridged or filamentous forms (Dame *et al.*, 2006, Liu *et al.*, 2010). Global studies show that H-NS binds many targets across the *E. coli* chromosome, notably AT-rich regions that are horizontally acquired (Grainger *et al.*, 2006, Lucchini *et al.*, 2006, Navarre *et al.*, 2006, Oshima *et al.*, 2006, Kahramanoglou *et al.*, 2011). Thus, it has been proposed that H-NS silences these genes and consequentially aids genome evolution. As a nucleoid-associated protein, H-NS can impact the structure and architecture of the bacterial nucleoid. This may result from the ability of H-NS to form bridges within and between distal DNA segments (Dame *et al.*, 2000, Dame *et al.*, 2006, Maurer *et al.*, 2009). These bridging functions result in the formation of DNA loops and may stabilise the 10- to 15-kb microdomains apparent in the folded bacterial chromosome (Hardy and Cozzarelli, 2005, Wang *et al.*, 2011). Importantly, bridging by H-NS can also confine RNA polymerase. In these instances, H-NS-DNA bridges form a looped structure that consequently traps RNA polymerase between these bridges (Dorman, 2014). The trapping of RNA polymerase provides a means for transcriptional silencing. Transcriptional silencing may also result from the formation of extended nucleoprotein structures (Schnetz, 1995, Jordi and Higgins, 2000, Petersen *et al.*, 2002, Dillon *et al.*, 2010). Examples of H-NS mediated silencing are found at the *proU* and *bgl* promoters.

H-NS also works in cooperation with other NAPs such as Fis and IHF, such as in the case of the *nir* regulatory region. Here, these proteins work together to control the organisation of DNA into a nucleoprotein complex that results in transcriptional repression (Browning *et al.*, 2000).

1.13. CbpA (Curved DNA-binding protein A)

The CbpA protein is a NAP specifically up-regulated during stationary phase (Azam and Ishihama, 1999). First isolated from crude cell extracts, *in vitro* studies have found that CbpA preferentially binds to curved DNA (hence curved DNA-binding protein) (Ueguchi *et al.*, 1994). The CbpA protein comprises four regions: an N-terminal J-domain, a flexible linker, and two C-terminal domains (Figure 1.15) (Cosgriff *et al.*, 2010). The C-terminal domain I (CTDI) and the flexible linker are responsible for DNA binding whilst the C-terminal domain II (CTDII) is responsible for dimerisation (Bird *et al.*, 2006, Cosgriff *et al.*, 2010). The J-domain is required for co-chaperone activity of CbpA. This co-chaperone activity entails stimulating the activity of the DnaK, a protein belonging to the Hsp70 family of proteins involved in the prevention of protein misfolding and aggregation (Gur *et al.*, 2004). Both co-chaperone and DNA binding are inhibited when the J-domain interacts with a modulator protein called CbpM (Chae *et al.*, 2004, Sarraf *et al.*, 2010, Chintakayala and Grainger, 2011).

Little is known about the function of CbpA in starved cells although a recent study has shown that CbpA contributes to changes in the supercoiled state of DNA during stationary phase (Chintakayala *et al.*, 2013, Moore *et al.*, 2015). In addition, CbpA has been shown to be required

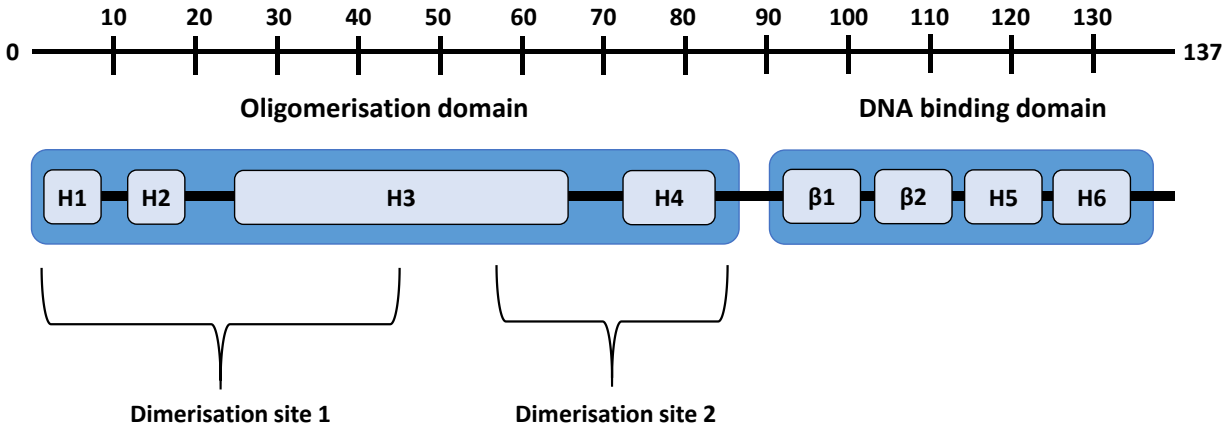


Figure 1.14. The domain organisation of H-NS.

Schematic diagram of the domain organisation of H-NS. The protein comprising 137 amino acids is divided into two domains (shown in blue boxes), the N-terminal (oligomerisation) domain and the C-terminal (DNA binding) domain. The ruled bar indicates the number of amino acids. Boxes labelled "H" (e.g. H1, H2, H4, H5, H6) indicate α helices and boxes labelled "β" (e.g. β1 and β2) indicate β sheets. The two observed dimerisation sites are indicated (Winardhi *et al.*, 2015).

for mutagenic break repair (MBR) by promoting F' plasmid maintenance (Moore *et al.*, 2015). Because CbpA is preferentially expressed in a specific phase of growth, further investigation of CbpA expression and regulation may provide a clearer picture of its function.

1.14. *yccE*

The *yccE* coding region is located adjacent to the *cbpA* coding region and is divergently orientated. While a study has implicated *yccE* in antibiotic resistance (Duo *et al.*, 2008), very little is known about the function of this gene. Importantly, however, this gene has been found to influence the expression of *cbpA* due to an intragenic promoter (i.e. *cbpA* P6) located within the coding region of the gene (Chintakayala *et al.*, 2013). The *cbpA* P6 promoter, located near the 5' end of the gene, is a strong, σ^{70}-dependent promoter. However, during growth phase, the nucleoid-associated protein Fis, binds to the intergenic region between *yccE* and *cbpA*, and subsequently insulates the activity coming from the *cbpA* P6 promoter.

1.15. Objective of this work

Previously, it was shown that *yccE* contains an internal antisense promoter that can influence *cbpA* transcription. Thus, this study seeks to understand i) how *cbpA* can be protected from these transcriptional effects and ii) the extent of intragenic promoters inside *yccE*.

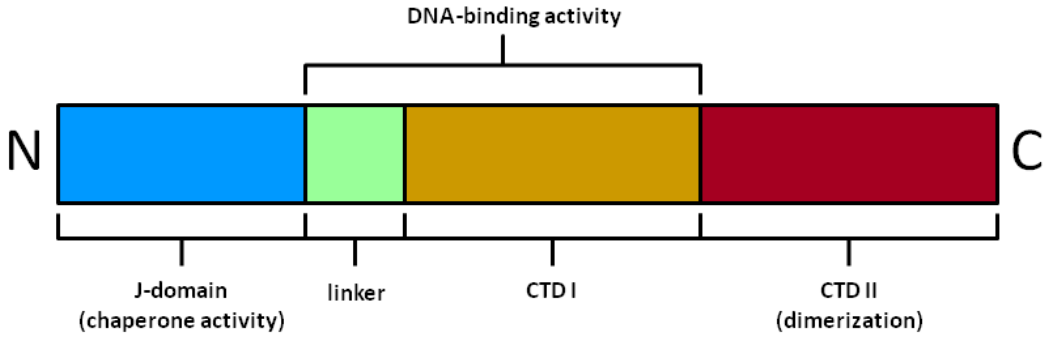


Figure 1.15. Schematic representation of the curved DNA-binding protein A (CbpA).

The CbpA protein is a nucleoid associated protein that is specifically upregulated during stationary phase. The protein consists of an N-terminal J-domain and two C-terminal domains joined by a linker. The J-domain is shown as a blue block. The linker is shown as a light green block. CTDI is shown as a light brown block, and CTDII is shown as a dark red block (Cosgriff *et al.*, 2010).

Materials and Methods

Chapter 2

2.1. General reagents, buffers and solutions

All solutions used in this work were dissolved in distilled deioinised water (ddH_2O) unless otherwise stated. Solutions were sterilised by autoclaving at 121°C for 15 mins at 5 psi. All reagents were obtained from Sigma-Aldrich Co. Ltd., Fisher Scientific, VWR, Bioline, and Life Technologies unless otherwise stated. Radioactive nucleotides were obtained from MP Biochemicals and Perkin Elmer.

Phenol/chloroform extraction:

- Phenol/chloroform/isoamyl alcohol pH 8.0 (25:24:1)

Ethanol precipitation:

- 100% (w/v) ethanol (EtOH)
- 70% (w/v) ethanol (EtOH)
- 3 M sodium acetate (CH_3COONa) pH 5.2
- 20 mg/ml glycogen

Competent cell preparation:

- 100 mM calcium chloride ($CaCl_2$)
- 50% (w/v) glycerol

Polyacrylamide gel electrophoresis (PAGE) and denaturing sequencing gels:

- 5X TBE (0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA). When in use, this solution was diluted to 1X.

- Ammonium persulphate [$(NH_4)_2S_2O_8$]
- Tetramethylethylenediamine, or TEMED ($C_6H_{16}N_2$)
- 30% (w/v) acrylamide: 0.8% (w/v) bisacrylamide mix. Obtained from Geneflow.
- Sequagel ureagel system. Obtained from Geneflow.

Agarose gel electrophoresis:

- 6x loading dye: 10 mM Tris pH 7.5, 1 mM EDTA, 20% (w/v) glycerol, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF
- 5X TBE (0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA). When in use, this solution was diluted to 1X.
- Agarose powder

β-galactosidase assays:

- Z-buffer: 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4 \cdot 2H_2O$, 0.75 KCl, 0.25 g $MgSO_4$ dissolved in 1 L of ddH_2O. Sterilised by autocloaving before use.
- 13 mM (8 mg/ml) 2-nitrophenyl-β-D-galactopyranoside, or ONPG ($C_{12}H_{15}NO_8$) dissolved in 1 L Z-buffer
- β-mercaptoethanol (C_2H_6OS). Before use, 271 µl β-mercaptoethanol was added per 100 ml of Z-buffer.
- 1% (w/v) sodium deoxycholate ($C_{24}H_{39}O_4Na$)
- 100% (w/v) toluene (C_7H_8)
- 1 M sodium carbonate (Na_2CO_3)

***In vitro* transcription:**

- 10X transcription buffer, or TNSC buffer (400 mM Tris acetate pH 7.9, 10 mM $MgCl_2$, 1 M KCl, 10 mM DTT)
- STOP solution [97.5% (w/v) deioinised formamide (CH_3NO), 10 mM EDTA, 0.3% (w/v) bromophenol blue/xylene cyanol FF]
- P^{32} labelled α-UTP

Radiolabelling of DNA fragments:

- G-50 sephadex beads, resuspended in a 12% (w/v) slurry with TE (Tris EDTA)
- T4 polynucleotide kinase, used with 10X buffer provided by the supplier.
- P^{32} labelled γ-dATP
- TE buffer or Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA pH 8.0)

Electrophoretic mobility shift assays (EMSA):

- Fis binding buffer (40 mM Tris-HCl pH 8.0, 10 mM $MgSO_4$, 1 mM $CaCl_2$, 100 mM NaCl, 1 mM DTT and 100 µg/ml BSA)

M13 sequencing reactions:

- 2 M sodium hydroxide (NaOH)
- Annealing buffer (1 M Tris-HCl pH 7.5, 100 mM $MgCl_2$, 160 mM DTT)
- 'A' mix short (840 each µM dCTP, dGTP, and dTTP, 93.5 µM dATP, 14 µM ddATP, 40 mM Tris-HCl pH 7.5, 50 mM NaCl)

- ‘C’ mix short (840 each µM dATP, dGTP, and dTTP, 93.5 µM dCTP, 17 µM ddCTP, 40 mM Tris-HCl pH 7.5, 50 mM NaCl)
- ‘G’ mix short (840 each µM dATP, dCTP, and dTTP, 93.5 µM dGTP, 17 µM ddGTP, 40 mM Tris-HCl pH 7.5, 50 mM NaCl)
- ‘T’ mix short (840 each µM dATP, dCTP, and dGTP, 93.5 µM dTTP, 17 µM ddTTP, 40 mM Tris-HCl pH 7.5, 50 mM NaCl)
- T7 polymerase dilution buffer (25 mM Tris-HCl pH 7.5, 5 mM DTT, 100 µg/ml BSA, 5% (w/v) glycerol)
- Label mix ‘A’ (1.375 mM dCTP, dGTP, and dTTP each, 333.5 mM NaCl).
- STOP solution [0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF, 10 mM EDTA pH 7.5, 97.5% (w/v) formamide]
- P^{32} labelled α-dATP

Primer extension:

- 1X hybridisation buffer [20 mM HEPES ($C_8H_{18}N_2O_4S$), 0.4 M NaCl, 80% (w/v) formamide]
- 3 M sodium acetate pH 5.2
- 3 M sodium acetate pH 7.0
- 5X reverse transcriptase buffer [50 mM Tris-HCl pH 8.3 (at 25°C), 50 mM KCl, 10 mM $MgCl_2$, 0.5 mM spermidine]
- STOP solution [97.5% (w/v) deioinised formamide (CH_3NO), 10 mM EDTA, 0.3% (w/v) bromophenol blue/xylene cyanol FF]

All solutions above were treated with DEPC and sterilised by autoclaving.

- 100% (w/v) ethanol
- 70% (w/v) ethanol, made up with DEPC-treated ddH_2O

G + A ladder:

- DNase I blue: 5 M urea (CH_4N_2O), 20 mM NaOH, 1 mM EDTA, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF
- 10 M piperidine ($C_5H_{11}N$), diluted with ddH_2O before use.
- 100% (w/v) formic acid (CH_2O_2)

DNase I footprinting:

- Recombinant DNase I. Obtained from Roche.
- DNase I blue: 5 M urea (CH_4N_2O), 20 mM NaOH, 1 mM EDTA, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol FF
- 10X transcription buffer, or TNSC buffer (400 mM Tris acetate pH 7.9, 10 mM $MgCl_2$, 1 M KCl, 10 mM DTT
- DNase I STOP solution: 0.3 M sodium acetate, 10 mM EDTA

Chromatin immunoprecipitation followed by PCR (ChIP-PCR):

- 1X TBS: 20 mM Tris-HCl pH 7.4, 0.9% (w/v) NaCl
- FA lysis buffer “150 mM NaCl”: 50 mM HEPES-KOH pH 7.0, 150 mM NaCl, 1 mM EDTA, 1% (w/v) Triton-X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS.

- FA lysis buffer "500 mM NaCl": 50 mM HEPES-KOH pH 7.0, 500 mM NaCl, 1 mM EDTA, 1% (w/v) Triton-X-100, 0.1% (w/v) sodium deoxycholate, 0.1% (w/v) SDS.
- ChIP wash buffer: 10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% (w/v) Nonidet-P40, 0.5% sodium deoxycholate
- ChIP elution buffer [50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1% (w/v) SDS]
- 1X TE buffer: 10 mM Tris-HCl pH 8.0, 1 mM EDTA
- Monoclonal anti-H-NS antibody (kindly donated by Jay Hinton).
- Protein A sepharose beads, washed with ddH_2O and stored in a 50% (v/v) slurry with 1X TBS. Obtained from GE Healthcare, GE17-0780-01.
- Spin-X columns

2.2. Antibiotics

Stock solutions of the following antibiotics were made and stored at -20°C.

Ampicillin: 100 mg/ml in ddH_2O

Tetracycline: 35 mg/ml in methanol (MeOH)

Kanamycin: 50 mg/ml in ddH_2O

Bicyclomycin: 100 mg/ml in methanol (MeOH)

When needed, antibiotics were added to autoclaved liquid or solid media to the following concentrations: 100 μg/ml (ampicillin), 35 μg/ml (tetracycline), and 50 μg/ml (kanamycin). Final concentrations used for bicyclomycin are stated in the Results and Discussion sections.

2.3. Enzymes

All enzymes were obtained from New England Biolabs, Inc. unless otherwise stated. Enzymes were then stored at -20°C and used following the manufacturer's instructions.

2.4. General media

All general media were obtained from Sigma Aldrich Co. Ltd and Oxoid Ltd. Media batches were made up to 1 L ddH_2O and were sterilised by autoclaving.

Lennox Broth (LB) agar: 35 g dissolved in 1 L ddH_2O

MacConkey lactose agar: 52 g dissolved in 1 L ddH_2O

Lennox Broth (LB): 20 g dissolved in 1 L ddH_2O

2.5. Phenol/chloroform extraction-Ethanol precipitation

To the DNA sample, an equal volume of the phenol-chloroform mix was added, after which the mixture was vortexed for 15 seconds. The mixture was then centrifuged at 17,900 x g for 3 minutes. The aqueous phase was then transferred to a new microcentrifuged tube, into which 2.5X volume ice-cold absolute ethanol and 0.1X volume 3M sodium acetate (CH_3COONa) pH 5.2 was added. Additionally, 1 µl 20 mg/ml glycogen was added to the mixture if the size of the DNA is <100bp. The mixture was then placed at -80°C for 30 minutes, after which precipitated DNA was then collected via centrifugation at 17,900 x g. The resulting DNA pellet was washed in ice-cold 70% (v/v) ethanol, and was subsequently dried under a vacuum. Typically, DNA was resuspended in 100 µl double distilled water.

2.6. Radioactive end-labelling of DNA fragments

DNA fragments to be used for EMSAs and DNase I footprinting assays were initially cloned into pSR and purified in large scale to generate large amounts of plasmid DNA. Plasmid DNA was first digested with *Hin*dIII, after which it was treated with calf alkaline phospatase (CIP). Resulting DNA was then extracted using phenol chloroform extraction. Next, DNA was digested with *Aat*II, and was subsequently purified using the QIAquick Gel Extraction Kit (Qiagen) as per manufacturer's instructions. The resulting fragments were radiolabelled at both ends using T4 polynucleotide kinase (T4 PNK) as per the manufacturer's instructions (New England Biolabs). Briefly, each reaction contains the following: (1) *Aat*II and *Hin*dIII-digested DNA fragment (up to 50 pmol); (2) 5 µl 10X T4 PNK Reaction Buffer; (3) 1 µl of P^{32} labelled γ-dATP (10 µCi/µl); (4) 2 µl of T4 PNK; and made up to 50 µl with double distilled water. Each reaction was incubated at

37°C for 30 minutes, after which enzymes were heat inactivated by incubating at 65°C for 20 minutes.

Unincorporated nucleotides were removed by running the reaction twice into 200 µl G-50 Sephadex columns.

2.7. Preparation of competent cells

Competent cells were prepared using the calcium chloride method. Overnight cultures were grown in 5-ml LB broth with the appropriate antibiotic, and 1 ml of the culture was added to 50 ml of LB-antibiotic broth. This culture was incubated at 37°C with shaking for 1.5-2.0 hours to an OD_{650} of 0.4-0.5. Cells on growth phase were placed on ice for 10 minutes, after which were pelleted at 4°C at 1,600 x g for 5 minutes. Cells were resuspended in 25 ml ice-cold $CaCl_2$ then incubated on ice for 20 minutes. They were then centrifuged at 4°C at 1,600 x g for 5 minutes, resuspended once more in 3.3 ml $CaCl_2$, and left overnight for 24 hours on ice at 4°C. The following day, 1.2 ml of 50% glycerol was added to the competent cells, and stored in -80°C until use.

2.8. Transformation of competent cells

Calcium chloride competent cells were transformed by heat shock. Plasmid DNA (approximately 0.2 µg) was incubated in 1.5 ml eppendorf tubes with 100 µl of competent cells on ice for 1 hour. The tubes were placed on a 42°C heat block for 2 minutes, after which 750 µl LB

broth was added to tubes. The cell suspensions were subsequently incubated for 30-60 minutes at 37°C with shaking. Cells were spun down at 2,400 x g for 2 minutes, and were resuspended in 100 µl LB broth. The suspensions were then spread on LB (with antibiotic) agar plates and incubated overnight at 37°C.

2.9. Strains and plasmid vectors

Bacterial strains used in this work are listed in Table 2.1. Strains M182 and JCB387 are highly competent *lac*$^-$ strains that were used as wildtype strains in assays and experiments. When in use, bacterial strains were streaked on solid media plates and stored at 4ºC. Overnight cultures were prepared by inoculating the required liquid culture with a fresh colony from the agar plates and subsequently incubated at 37ºC with shaking. Optical densities were measured using a Jenway 6300 Spectrophotometer (Thermo Fisher Scientific, Inc.). For long-term storage, glycerol stocks of the bacterial strains were prepared and stored in -80ºC.

Plasmid vectors used in this study are listed in Table 2.2. Plasmid vectors were purified using the QIAquick PCR Purification kit (Qiagen) according to manufacturer's instructions and were subsequently stored at -20ºC until use. Schematic diagrams of the plasmid vectors used are shown in Figures 2.1 and 2.2.

2.10. Polymerase chain reaction (PCR)

A typical PCR reaction is composed of Phusion DNA polymerase (New England Biolabs) or Velocity DNA polymerase (Bioline), template DNA (plasmid or genomic DNA), appropriate upstream and downstream oligonucleotides at a final concentration of 1 µM, deoxynucleotide triphosphates (dNTPs) (Bioline) at a final concentration of 2 mM, and sterile distilled water.

Cycling parameters for PCR were used according to the needs of the experiment. The cycling parameters for a typical PCR are listed in Table 2.3.

2.11. Quikchange II-E site directed mutagenesis (Agilent Technologies)

Site directed mutagenesis was performed using the Quikchange II-E Site Directed Mutagenesis Kit (Agilent Technologies). Briefly, sample reactions were prepared using the following components: 5 µl of 10x reaction buffer, 50 ng of dsDNA template, 125 ng of forward primer, 125 ng of reverse primer, 1 µl of dNTP mix, ddH_20 to a final volume of 50 µl, and 1 µl *PfuUltra* HF DNA polymerase (2.5 U/µl). Each reaction was cycled using the cycling parameters specified in Table 2.4.

After temperature cycling, all reactions were placed on ice and allowed to cool. DNA amplification was checked by agarose gel electrophoresis, after which 1 µl of the *DpnI* restriction enzyme (10 U/µl) was added directly to the reaction in order to digest the parental (i.e, nonmutated) supercoiled dsDNA. Each digestion reaction was incubated at 37°C for 1 hour.

Table 2.1. Strains used in this study

Strain	Genotype	Source
M182	Δ*lac galK galU strA*	(Busby *et al.*, 1983)
M182Δ*hns* (JRG4864)	M182Δ*hns* Cm^R	(Wyborn *et al.*, 2004)
JCB387	Δ*nir* Δ*lac*	(Typas and Hengge, 2006)
JCB3871Δ*fis*	JCB387 *fis985* (*str/spcR*)	(Wu *et al.*, 1998)

Table 2.2. Plasmids used in this study

Plasmid	Description	Source
pRW50	Broad host-range *lac* fusion vector used for cloning promoter fragments with *Eco*RI and *Hin*dIII restriction sites. Contains the RK2 origin of replication and encodes Tet^R. Low copy number.	(Lodge *et al.*, 1992)
pSR	pBR322-derived plasmid. Contains *Eco*RI-*Hin*dIII fragment upstream of the *λoop* transcription terminator and encodes Amp^R. High copy number.	(Kolb *et al.*, 1995)
pJ204	pUC-derived plasmid. Encodes Amp^R	(DNA 2.0)
pMK-RQ	ColE1 origin. Encodes Kan^R	(Life Technologies)

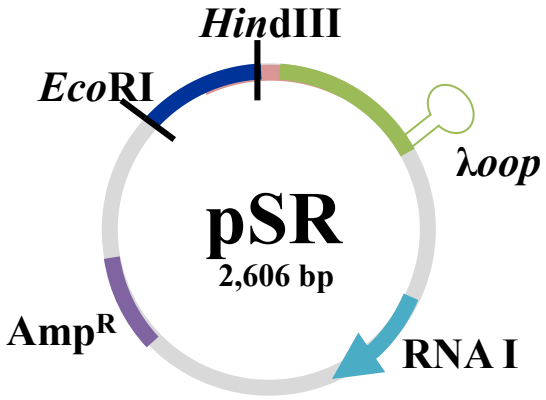


Figure 2.1. The pSR plasmid.

A schematic diagram of the pSR plasmid, a 2.606 kb high copy number plasmid used for *in vitro* transcription assays. A representative DNA fragment with *Eco*RI-*Hin*dIII restriction sites are shown in dark blue. The *λoop* transcription terminator is shown in light green. RNAI, which functions as an internal control, is shown as a light blue arrow. This plasmid encodes resistance to ampicillin (shown in purple).

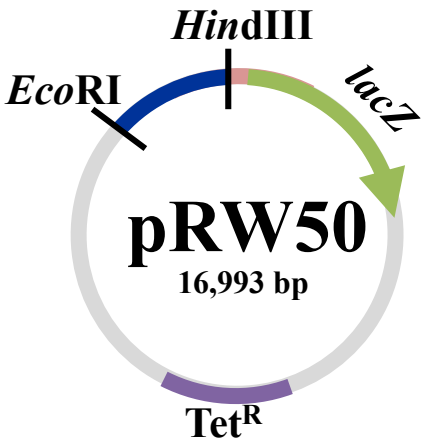


Figure 2.2. The pRW50 plasmid.

A schematic diagram of the pRW50 plasmid, a 16.993 kb low copy number plasmid used for *in vivo* reporter assays. A representative DNA fragment with *Eco*RI-*Hin*dIII restriction sites is shown in dark blue. The *lacZ* gene is shown as a light green arrow. This plasmid encodes resistance to tetracycline (shown in purple).

Mutant DNA was subsequently transformed into XL1-Blue Electroporation Competent Cells using the electroporation method. Briefly, 7 µl of the StrataClean slurry was added into the mutagenesis reaction. This mixture was vortexed for 15 seconds, and then incubated at room temperature for 1 minute. The resin was then pelleted by centrifugation at >2,000 x g for 1 minute. After centrifuging, the supernatant carrying the DNA was carefully pipetted out and placed into a fresh microcentrifuge tube. Then, into a pre-chilled microcentrifuge tube containing 40 µl of the electroporation-competent cells, 2 µl of resin-purified mutagenesis reaction was added. The reaction was mixed thoroughly and placed into a chilled electroporation cuvette (0.1-cm gap). The cuvette was then placed into the chilled electroporation chamber, pulsed once, and quickly removed. The following settings are used for electroporation:

Applied volts (0.1 cm gap): 1700 V

Field strength (0.1 cm gap): 17 kV/cm

Resistance: 200 Ω

Capacitance: 25 µF

After pulsing, 960 µl of sterile LB medium (held at 37°C) was added into the cuvette. The cell suspension was then transferred into a sterile microcentrifuge tube and incubated at 37°C for 1 hour while shaking. Then 100 µl of the transformation reaction was plated into MacConkey agar plates containing 35 mg/ml of the appropriate antibiotic. Subsequent colonies were then picked and sent for sequencing to confirm the mutagenesis.

Table 2.3. Cycling parameters for a typical PCR reaction.

SEGMENT	DESCRIPTION	CYCLES	TEMPERATURE	TIME
1	Initial melting of DNA	1	94°C	1 min
2	Melting of DNA	35	94°C	30 sec
	Primer annealing		50°C	30 sec
	Elongation		72°C	35 sec
3	Final elongation	1	72°C	10 min

Table 2.4. Cycling parameters used for Quikchange II-E site directed mutagenesis.

SEGMENT	CYCLES	TEMPERATURE	TIME
1	1	95°C	30 sec
2	12-15	95°C	30 sec
		55°C	1 min
		68°C	1 min/kb plasmid length

2.12. Plasmid DNA purification (Miniprep)

Plasmid DNA was purified using the Miniprep Kit (Qiagen) as per the manufacturer's instructions. Briefly, a 5-ml overnight culture was placed into 15-ml falcon tubes and centrifuged at 1,600 x g for 10 minutes. The resulting bacterial pellet was resuspended in 250 µl Buffer P1 and subsequently transferred into a microcentrifuge tube. Then, 250 µl Buffer P2 was added, after which the mixture was gently inverted. The solution was incubated at room temperature for not more than 5 minutes. Next, 350 µl Buffer N3 was added and the microcentrifuge tube was again gently inverted a few times until the solution became cloudy. The solution was then centrifuged for 10 minutes at 17,900. The resulting supernatant was applied to the QIAprep spin column and subsequently centrifuged for 1 minute at 17,900 x g. After spinning, the flow-through was immediately discarded and 750 µl Buffer PE was added to the spin column to wash. The spin column was centrifuged for 1 minute at 17,900 x g, after which the flow-through was again discarded. The spin column was subjected to an additional centrifuge step in order to remove residual wash buffer. Finally, the spin column was transferred to a new microcentrifuge tube and eluted with 30 µl double distilled water.

2.13. Plasmid DNA purification (Maxiprep)

Plasmid DNA was purified using the Maxiprep Kit (Qiagen) as per the manufacturer's instructions. Briefly, a 100 ml (for high-copy number plasmids) or 400 ml (for low-copy number plasmids) overnight culture was centrifuged at 1,600 x g for 15 minutes to harvest bacterial cells. The resulting pellet was resuspended in 10 ml Buffer P1, after which the suspension was transferred

into a 15-ml falcon tube. Next, 10 ml Buffer P2 was added to the suspension and the solution was gently inverted a few times to mix. The solution was incubated at room temperature for not more than 5 minutes, after which 10 ml Buffer P3 was added. Again, the suspension was gently inverted a few times until it became cloudy, after which it was centrifuged briefly for 3 minutes at 1,600 x g to clear the debris. The supernatant was then applied to a QIAfilter Maxi Cartridge and left at room temperature for 10 minutes to allow separation of debris. Meanwhile, the QIAGEN-tip 500 column was equilibrated using 10 ml Buffer QBT and allowed to empty using gravity flow. Then, the supernatant was applied to the QIAGEN-tip 500 column using the QIAfilter Maxi Cartridge, and was allowed to enter the resin using gravity flow. The column was then washed twice with Buffer QC, also allowing the buffer to move via gravity flow. Afterwards, DNA was eluted from the column with 15 ml Buffer QF into a 50-ml falcon tube containing 10.5 ml isopropanol in order to precipitate DNA. The solution was then divided into 10-ml aliquots and placed into 10-ml OakRidge Centrifuge Tubes. The isopropanol mixture was then centrifuged at 17,900 x g at 4°C for at least 45 minutes to pellet the DNA. The resulting DNA pellet was resuspended with the same isopropanol mixture and was distributed evenly into microcentrifuge tubes. The mixture was then centrifuged at 17,900 x g for 10 minutes to pellet the DNA. The resulting DNA pellet was washed with 750 µl 70% ethanol and centrifuged once more at 17,900 x g for 10 minutes. The supernatant was then removed and the pellet was allowed to dry under vacuum for not more than 5 minutes. Pellet DNA was then resuspended in 100 µl double distilled water ready for use.

2.14. RNA extraction

RNA was purified using the RNeasy Kit (Qiagen) as per the manufacturer's instructions. Briefly, a 5 ml culture (with the appropriate antibiotic) was grown overnight at 37°C. The next day, 200 µl of the overnight culture was used to inoculate 10 ml LB plus antibiotic. This culture was grown to an OD_{650} of 0.4-0.5. At the appropriate OD, the culture was centrifuged at 1,600 x g for 20 mins at 20°C to collect cells. The supernatant was discarded, and the cell pellets were used for the RNA extraction immediately or, storage at -70°C. Cell pellets were resuspended in 200 µl TE buffer + 40 µg/ml lysozyme, and incubated for 15 mins at room temperature. Then, 700 µl of buffer RLT (with 10 µl β-mercaptoethanol per 1 ml buffer) was added to the sample. After vortexing, 500 µl ethanol was added and each sample was split between two RNeasy mini columns and centrifuged for 15 s at 8,000 x g. The flow-through was discarded and 700 µl RW1 was added to the column. Again, the sample was centrifuged at 8,000 x g for 15 s. The flow-through was discarded and the column was placed into a new 2 ml collection tube. To the column, 500 µl buffer RPE was added, and the sample was centrifuged at 8,000 x g for 15 s. Then, a further 500 µl buffer RPE was added and the sample was subsequently centrifuged at 8,000 x g for 2 mins. Then, the column was placed into a new collection tube and again centrifuged at full speed for 1 min to remove residual ethanol. Next, the column was transferred into a new microfuge tube. RNA was eluted in 30 µl RNase-free water. To remove DNA contamination, 3.5 µl 10X Turbo DNase Buffer and 1 µl Turbo DNase (Ambion) was added to the sample, after which it was incubated at 37°C for 30 mins. Next, 6.1 µl DNase inactivating reagent was added to the sample, and after which the sample was incubated at room temperature for 2 mins. The sample was centrifuged at 8,000 x g for 1.5 mins, and the resulting supernatant was transferred to a clean tube for freezing at -70°C.

2.15. Agarose gel electrophoresis

Agarose gel electrophoresis was done using 1% (w/v) agarose gel. Briefly, 1 g of agarose (Bioline) was dissolved in 100 ml 1X TBE by microwaving on a high setting. The solution was allowed to cool, after which 1% (v/v) ethidium bromide was added before the gel solidified. Typically, gels were run at 100V for 30 minutes in 1X TBE running buffer.

2.16. Extraction of DNA fragments (gel extraction)

DNA fragments were extracted from agarose gels using the QIAquick Gel Extraction Kit (Qiagen). Briefly, DNA fragments were excised from the agarose gel using a scalpel, after which the gel slices were weighed in a microcentrifuge tube. Next, 3 volumes Buffer QG was added to the microfuge tube and was then incubated at 50°C for 10 minutes until the gel slice was completely dissolved. To ensure that the solution had the correct pH, 10 µl 3M sodium acetate pH 5.2 was added. One volume of isopropanol was added if the DNA fragment to be recovered ranged from 70 bp to 10 kb long. Next, the sample was applied to the QIAquick spin column and was centrifuged at 17,900 x g for 1 minute. The flow-through was discarded, and 500 µl Buffer QG was added to the spin column and was subsequently centrifuged at 17,900 x g for 1 minute. Again, the flow-through was discarded, and this time 750 µl Buffer PE was added. The spin column was then centrifuged twice at 17,900 x g for 1 minute. After centrifugation, the column was transferred into a new 1.5-ml microcentrifuge tube. DNA fragments were eluted in 50 µl double distilled water.

2.17. Polyacrylamide gel electrophoresis (PAGE)

Polyacrylamide gel electrophoresis (PAGE) was done using Protogel (Geneflow). Protogel contains 30% (w/v) acrylamide/methylene bisacrylamide (37.5:1 ratio). A solution of 18.8 ml Protogel and 20 ml 5X TBE was made up to 100 ml with ddH_2O. Then, 100 µl of 10% APS and 40 µl TEMED was added in order for the gel to polymerise. Samples for the electrophoresis were loaded into each well with a 1:1 ratio (DNA sample:loading dye). The gel tank was then filled with 1X TBE. Next, the gels were run for 25 minutes at 30 mV. Gels were visualised using a UV transilluminator after staining with ethidium bromide (EtBr).

2.18. Purification of PCR products

PCR products were purified using the QIAquick Purification Kit (Qiagen). Briefly, 5 volumes of Buffer PB was added to 1 volume of the PCR reaction, after which the reaction was mixed by vortexing. The sample was then applied to the QIAquick spin column provided and the column was then centrifuged at 17,900 x g for 1 minute. The flow-through was discarded, and 750 µl Buffer PE was added. The spin column was then centrifuged twice at 17,900 x g for 1 minute. After centrifugation, the column was transferred into a new 1.5-ml microcentrifuge tube. DNA fragments were eluted in 50 µl ddH_2O.

After purification, the PCR products were digested with the required restriction enzymes. The digested reactions were then subjected to further purification, and the resulting DNA fragments were used for the ligation reactions.

2.19. Digestion of DNA with restriction enzymes

Restriction enzymes were obtained from New England Biolabs and used according to the manufacturer's instructions. Reactions contained 10 units of each restriction enzyme, 1 µg DNA, and 5 µl 10X buffer, made up to a total reaction of 50 µl with ddH_2O. The reaction was then incubated for 2-3 hours at 37°C.

2.20. Ligation reactions

Before ligation, plasmids were digested with appropriate restriction enzymes. Then, the digested plasmid was treated with alkaline phosphatase in order to cleave the 5' hydroxyl groups. Each treatment reaction contained 1 pmol of plasmid DNA, 2 µl 10X buffer, and 1 unit of calf intestinal alkaline phosphatase (CIP), made up to a total reaction volume of 20 µl with ddH_2O. These reactions were then incubated at 37°C for 30 mins. Reactions containing the plasmid vectors were subjected to purification. Plasmid vectors which have a size of less than 10 kb (e.g. pSR) were purified using the gel extraction protocol described above. Vectors which have a size of more than 10 kb (e.g. pRW50) were purified using phenol/chloroform extraction-ethanol precipitation protocol. Briefly, to the DNA sample, an equal volume of the phenol-chloroform mix was added, after which the mixture was vortexed for 15 seconds. The mixture was then centrifuged at 13,300 rpm for 3 minutes. The aqueous phase was then transferred to a new microcentrifuged tube, into which 2.5X volume ice-cold absolute ethanol and 0.1X volume 3M sodium acetate (CH_3COONa) pH 5.2 was added. Additionally, 1 µl 20 mg/ml glycogen was added to the mixture if the size of the DNA is <100bp. The mixture was then placed at -80°C for 30 minutes, after which precipitated DNA was then collected via centrifugation at 13,000 rpm. The resulting DNA pellet was washed

in ice-cold 70% (v/v) ethanol, and was subsequently dried under a vacuum. Typically, DNA was resuspended in 100 µl double distilled water. Each 20 µl ligation reaction is composed of: 1-3 µl appropriate plasmid vector, 1-5 µl digested and purified DNA fragment, 2 µl 10X buffer supplied by the manufacturer, and 2 µl T4 DNA ligase.

2.21. Plasmid construct sequencing

Sanger sequencing of plasmid constructs were done by the Functional Genomics, Proteomics and Metabolomics Facility at the School of Biosciences, University of Birmingham. Each sequencing reaction contained 200 ng of DNA and primers were used at a concentration of 10 mM.

2.22. Oligonucleotides and synthesised DNA fragments

All oligonucleotides used in this study were synthesised by Invitrogen or Alta Biosciences. Oligonucleotide sequences are listed in Table 2.5. Oligonucleotides were resuspended in sterile distilled water according to manufacturer's instructions before use.

All DNA fragments used in this study are synthesised by Invitrogen or DNA 2.0. These are listed in the Appendices. DNA fragments were resuspended in sterile distilled water according to manufacturer's instructions before use.

2.23. Electrophoretic mobility shift assays (EMSA)

Reactions contained radiolabelled DNA (10-40 nM), Fis binding buffer, 50% glycerol, and ddH_2O. Purified proteins were added to each reaction at indicated concentrations and reactions were incubated at 37°C for 20 minutes. Each reaction was then loaded into 5% polyacrylamide gel. The gel was run at 250V for 1-2 hours, and was subsequently dried and exposed to a phosphorscreen.

2.24. DNase I footprinting experiments

Purified protein was mixed with radiolabelled 10-40 nM of DNA and 12.5 µg/ml Herring sperm DNA (a nonspecific competitor) in TNSC buffer, made up to a total reaction volume of 20 µl. This reaction was incubated at 37°C for 20 minutes. Then to each reaction, 2 µl DNase I was added, and the reaction was incubated at room temperature for 40 seconds. Reactions were then subsequently stopped with the addition of 200 µl STOP solution.

Reactions containing the digested DNA were purified using phenol/chloroform extraction, followed by ethanol precipitation. The resulting DNA pellet was resuspended in 8 µl gel loading

Table 2.5. Oligonucleotides used in this study.

Note: Bold text are oligonucleotide uses. Underlined text are restriction sites.

Number (F – Forward; R – Reverse)	**Name**	**Sequence (5' → 3')**
Oligos used for site-directed mutagenesis of the *cbpA* regulatory region		
LL0912-1F	-126C Forward	CTGGTTCAACTATCAAAAATCCCTCACCCTTTTTCACCT
LL0912-1R	-126C Reverse	AGGTGAAAAAGGGTGAGGGATTTTTGATAGTTGAACCAG
LL0912-2F	-127G Forward	AGTGCCTGGTTCAACTATCAAAAATGGCTCACCCTTTTT
LL0912-2R	-127G Reverse	AAAAAGGGTGAGCCATTTTTGATAGTTGAACCAGGCACT
LL0912-3F	-144C Forward	CTAGTTAAACTTAAAGTGCCTGCTTCAACTATCAAAAACG CTCA
LL0912-3R	-144C Reverse	TGAGCGATTTTTGATAGTTGAAGCAGGCACTTTAAGTTTA ACTAG
LL0912-4F	-145C Forward	CCTAGTTAAACTTAAAGTGCCTCGTTCAACTATCAAAAAT CGCTC
LL0912-4R	-145C Reverse	GAGCGATTTTTGATAGTTGAACGAGGCACTTTAAGTTTAA CTAGG
LL0113-1F	-144C145C F	CCCTAGTTAAACTTAAAGTGCCTCCTTCAACTATCAAAAA TCGCTCAC
LL0113-1R	-144C145C R	GTGAGCGATTTTTGATAGTTGAAGGAGGCACTTTAAGTTT AACTAGGG
LL0113-2F	-126C144C 145C F	CCCTAGTTAAACTTAAAGTGCCTCCTTCAACTATCAAAAA TCCCTCAC
LL0113-2R	-126C144C 145C R	GTGAGGGATTTTTGATAGTTGAAGGAGGCACTTTAAGTTT AACTAGGG
LL1013-1F	-126C127G F	TAAAGTGCCTGGTTCAACTATCAAAAATGCCTCACCCTTT TTCAC

LL1013-1R	-126C127G R	GTGAAAAAGGGTGAGGCATTTTTGATAGTTGAACCAGGCACTTTA

Oligos used for the *cbpAΔ203.1.Reversed* deletion analysis

LL1013-2F	*cbpA*D203.1 RD10F	GGCTGCGAATTCGCAACTAATTCCATGTATATTAC
LL1013-3F	*cbpA*D203.1 RD20F	GGCTGCGAATTCCCATGTATATTACTACCC
LL1013-4F	*cbpA*D203.1 RD30F	GGCTGCGAATTCTACTACCCATATATAGCG
LL1013-5F	*cbpA*D203.1 RD40F	GGCTGCGAATTCATATAGCGTCTATAAAATTTAATAAATAATG
LL1013-6F	*cbpA*D203.1 RD50F	GGCTGCGAATTCTATAAAATTTAATAAATAATG
Ll1013-7F	*cbpA*D203.1 RD60F	GGCTGCGAATTCAATAAATAATGACGCCCCTAG
LL1013-8F	*cbpA*D203.1 RD70F	GGCTGCGAATTCGACGCCCTAGTTAAACTTAAAG
LL1013-9F	*cbpA*D203.1 RD80F	GGCTGCGAATTCTTAAACTTAAAGTGCCTG
LL1013-10F	*cbpA*D203.1 RD90F	GGCTGCGAATTCAGTGCCTGGTTGAACATATTTTAAAC
LL1013-11F	*cbpA*D203.1 RD100F	GGCTGCGAATTCTGAACATATTTTAAACAG
LL1013-12F	*cbpA*D203.1 RD110F	GGCTGCGAATTCTTAAACAGGTGAAAAAGGG
LL1013-13F	*cbpA*D203.1 RD120	GGCTGCGAATTCGAAAAAGGGTGAGCGATTTTTG
LL1013-14F	*cbpA*D203.1	GGCTGCGAATTCAGCGATTTTTGATAGTTG

	RD130	
LL1013-2R	*cbpA*D203.1 RD10 Rev.	CGGGCGAAGCTTCATAGCGTTATCTCGCGTAAATC

Oligos used for mutations in the *cbpAΔ203.1* and its derivatives

LL0713-1F	*cbpA*D203.1.Rev -217G-216G F	GGCTGCGAATTCTATTTGCAGTGCAACTAATTCCATGGGTATTACTAC
LL0114-1R	*cbpA*D203.1.Rev -217G-216G R	CGGGCGAAGCTTCATAGCGTTATCTCGCGTAAATCAACACAAATTGAAG

Oligos used for the *yccE* intergenic region

	yccE upstream1	GGCTGCGAATTCCATAGCGTTATCTCGCGTAAATC
	yccE downstream	CGCCCGAAGCTTCATATATAGCGTCTATAAAATTTAATAAAT

Oligos used for the *yccE* intergenic region deletion analysis

LL1013-15F	*yccE*D10	GGCTGCGAATTCTAAATCAACACAAATTGAAG
LL1013-16F	*yccE*D20	GGCTGCGAATTCCAAATTGAAGGAACCCCTG
LL1013-17F	*yccE*D30	GGCTGCGAATTCGAACCCCTGTAAGGTAACTC
LL1013-18F	*yccE*D40	GGCTGCGAATTCAAGGTAACTCCTATAAGTG
LL1013-19F	*yccE*D50	GGCTGCGAATTCCTATAAGTGTAGGGTAATCC
LL1013-20F	*yccE*D60	GGCTGCGAATTCAGGGTAATCCTCAAAATTTC
LL1013-21F	*yccE*D70	GGCTGCGAATTCTCAAAATTTCATATGCCAAC
LL1013-22F	*yccE*D80	GGCTGCGAATTCATATGCCAACACAGAATATG
LL1013-23F	*yccE*D90	GGCTGCGAATTCACAGAATATGTTATTGAAATC
LL1013-24F	*yccE*D100	GGCTGCGAATTCTTATTGAAATCATCGCGG
LL1013-25F	*yccE*D110	GGCTGCGAATTCCATCGCGGAGAGGAGGTC
LL1013-26F	*yccE*D120	GGCTGCGAATTCAGGAGGTCGCCATCAAGATG
LL1013-27F	*yccE*D130	GGCTGCGAATTCCATCAAGATGGGTTGCTG
LL1013-28F	*yccE*D140	GGCTGCGAATTCGGTTGCTGAACATATTTTAAAC
LL1013-29F	*yccE*D150	GGCTGCGAATTCCATATTTTAAACAGGTGAAAAAG

LL1013-30F	*yccE*D160	GGCTGCGAATTCACAGGTGAAAAAGGGTGAG
LL1013-31F	*yccE*D170	GGCTGCGAATTCAAGGGTGAGCGATTTTTG
LL1013-32F	*yccE*D180	GGCTGCGAATTCGATTTTTGATAGTTGAAC
LL1013-33F	*yccE*D190	GGCTGCGAATTCAGTTGAACCAGGCACTTTAAG
LL1013-34F	*yccE*D200	GGCTGCGAATTCGCACTTTAAGTTTAACTAG
LL1013-35F	*yccE*D210	GGCTGCGAATTCTTTAACTAGGGCGTC
LL1013-36F	*yccE*D220	GGCTGCGAATTCCGTCATTATTTATTAAATTTTATAG
LL1013-37F	*yccE*D230	GGCTGCGAATTCATTAAATTTTATAGACGC
LL1013-15R	*yccE*D10 R	CGGGCGAAGCTTCATATATAGCGTC

Oligos used to make cbpA full pyccEΔ20 and cbpAΔ203.1 pyccEΔ20 with P6 and Fis binding site mutations

LL0314-1F	*cbpA* full P*yccE*Δ20-217G-216G F	CATGGGTATTACTACCCATATATAG
LL0314-1R	cbpA full P*yccE*Δ20-217G-216G R	CATTATATGGGTAGTAATACCCATG
LL0314-2F	*cbpA* full P*yccE*Δ20-127G-126C-108C-94G F	CAAAAATGCCTCACCCTTTTTCACCTCTTTAAAATATGTTGAG
LL0314-2R	*cbpA* full P*yccE*Δ20-127G-126C-108C-94G R	CTCAACATATTTTAAAGAGGTGAAAAAGGGTGAGGCATTTTTG
LL0314-3F	*cbpA*Δ203.1-127G-126C-108C-94G F	CAAAAATGCCTCACCCTTTTTCACCTCTTTAAAATATGTTGATAG
LL0314-3R	*cbpA*Δ203.1-127G-126C-108C-94G R	CTATCAACATATTTTAAAGAGGTGAAAAAGGGTGAGGCATTTTTG

Oligos used for the *yccE* gene in sense and antisense orientation

	yccE sense Fwd	TATGATGAATTCATGGGTAGTAATATACATGGAATTAGTTGC
	yccE sense Rev	CTCATAAAGCTTAATTGCATAACGATTTAAATAGTCGAGTTGCGC
	yccE antisense Fwd	TATGATGAATTCAATTGCATAACGATTTAAATAGTCGAGTTGCGC
	yccE antisense Rev	CTCATAAAGCTTATGGGTAGTAATATACATGGAATTAGTTGC
	yccE sense Fwd MfeI	TATGATCAATTGATGGGTAGTAATATACATGGAATTAGTTGC
	yccE antisense Fwd MfeI	TATGATCAATTGAATTGCATAACGATTTAAATAGTCGAGTTGCGC
Oligos used for the *yccE* gene in sense and antisense orientation and its internal -10 promoters		
	yccE gene AS no P6 R	CGGGCGAAGCTTCATGGAATTAGTTGCACT
Oligos used for the mutations in the *yccE* promoter (yccEΔ200)		
LL0814-1F	*yccE*D200-29G-28G-26G-25G	GGCTGCGAATTCGCACGGTGGGTTTAACTAGGGCGTC
LL0814-1R	*yccE*D200-5G-4G	CGGGCGAAGCTTCATATATAGCGTCTATAAAATTTAATAACCAATGACGCCCTAG
Oligos used for the *yccE* gene in sense orientation with its promoter region (yccEΔ200)		
LL1214-1F	Forward oligo for the *yccE* gene with its promoter region	GGCTGCCAATTGGCACTTTAAGTTTAACTAGGG
LL1214-1.1F	Forward oligo for the *yccE* gene with its promoter	GGCTGCCAATTGGCACTTTAAGTTTAACTAGGGC

	region (1bp extended)	

Oligos used to disrupt the internal promoters within the *yccE* gene

LL1214-2F	Forward oligo for the disrupted M10 *yccE* 1.2 fragment	GGCTGCGAATTCTAATTATTTGCAGTGCAACTAATTCCATGGGTATT
LL1214-2R	Reverse oligo for the disrupted M10 *yccE* 1.2 fragment	CGGGCGAAGCTTCCTCCTCTATATATGGGTAGTAATACCCATGGAATTAGTTG
LL1214-3F	Forward oligo for the disrupted M10 *yccE* 4 fragment	GGCTGCGAATTCATTTGAGGAAATGCAGAATGATAATGATCGGTCAT
LL1214-3R	Reverse oligo for the disrupted M10 *yccE* 4 fragment	CGGGCGAAGCTTCCTCCTATATTCTCATAAATAATGACCGATCATTATCAT
LL1214-4F	Forward oligo for the disrupted M10 *yccE* 6 fragment	GGCTGCGAATTCTTTATTCAGGGGTTTCAGGAGGTATTACAGGTACT
LL1214-4R	Forward oligo for the disrupted M10 *yccE* 1.2 fragment	CGGGCGAAGCTTCCTCCTTCGAACATATTGTATAGTACCTGTAATACCTCCTG
LL1214-5F	Reverse oligo for the disrupted M10 *yccE* 1.2 fragment	GGCTGCGAATTCCTAACAATATCAATATCTCGAACATATTGGGTAGT
LL1214-5R	Forward oligo for the disrupted M10 *yccE* 4 fragment	CGGGCGAAGCTTCCTCCTGGAGGTATTACATATACTACCCAATATGTTCGAGA
LL1214-6F	Reverse oligo for the disrupted M10 *yccE* 4 fragment	GGCTGCGAATTCAGGGGTTTCAGGAGGTATTACATATACTAGGCAAT
LL1214-6R	Forward oligo for the disrupted M10 *yccE* 6 fragment	CGGGCGAAGCTTCCTCCTCAATATCTCGAACATATTGCCTAGTATATGTAATA

LL12-14-7F	Forward oligo for the disrupted M10 *yccE* 1.2 fragment	GGCTGCGAATTCCTCCATAAAGTTATACCAATAATAACCTTGGAAAT
LL1214-7R	Reverse oligo for the disrupted M10 *yccE* 1.2 fragment	CGGGCGAAGCTTCCTCCTCAGAGTCTATCACAGATTTCCAAGGTTATTATTGG
LL1214-8F	Forward oligo for the disrupted M10 *yccE* 4 fragment	GGCTGCGAATTCCTTCAGAGTCTATCACAGATTTTAAAGGTGGTTAT
LL1214-8R	Reverse oligo for the disrupted M10 *yccE* 10 fragment	CGGGCGAAGCTTCCTCCTCATAAAGTTATACCAATAACCACCTTTAAAATCTG
LL1214-9F	Forward oligo for the disrupted M10 *yccE* 12 fragment	GGCTGCGAATTCAAGGTTATTATTGGTATAACTTTATGGAGGGTATT
LL1214-9R	Reverse oligo for the disrupted M10 *yccE* 10 fragment	CGGGCGAAGCTTCCTCCTCGCATTAATGTTTTCAATACCCTCCATAAAGTTAT
LL1214-10F	Forward oligo for the disrupted M10 *yccE* 14 fragment	GGCTGCGAATTCTCAGCGTGAAACTAACCAGGCATTAGGATGGAAAT
LL1214-10R	Reverse oligo for the disrupted M10 *yccE* 14 fragment	CGGGCGAAGCTTCCTCCTCATCTACAGGAGCATATTTCCATCCTAATG CCTGG
LL1214-11F	Forward oligo for the disrupted M10 *yccE* 15 fragment	GCCTGCGAATTCGATGGCGATGGATATTGTCTGTTAAGAGCGGTACT
LL1214-11R	Reverse oligo for the disrupted M10 *yccE* 15 fragment	CGGGCGAAGCTTCCTCCTTGTTGTTTTAAAACCAGTACCGCTCTTAACAGACA
LL1214-12F	Forward oligo for the disrupted M10 *yccE* 18 fragment	GGCTGCGAATTCCGATCGAGGCTCTTGTTGATACGGCATTCGGTAAT

LL1214-12R	Reverse oligo for the disrupted M10 *yccE* 18 fragment	CGGGCGAAGCTTCCTCCTTACATCTTCCCTGAGATTACCG AATGCCGTATCAA
	Oligos used to amplify other genes similar to *yccE*	
LL0115-1F	Forward oligo to amplify *fepE* wildtype	GGCTGCGAATTCATGTCATCACTGAATATTAAAC
LL0115-1R	Reverse oligo to amplify *fepE* wildtype	CGGGCAAGCTTTTAAACTAAGTGGTCTGCCATC
LL0115-2F	Forward oligo to amplify *fepE* with disrupted promoters	GGCTGCGAATTCATGTCATCACTGAATATCCAAC
LL0115-2R	Reverse oligo to amplify *fepE* with disrupted promoters	CGGGCAAGCTTTTAAACCCAGTGGTCTGCCATC
LL0115-3F	Forward oligo to amplify *yfdF* wildtype	GGCTGCGAATTCATGCTACCATCTATTTCAATCAAC
LL0115-3R	Reverse oligo to amplify *yfdF* wildtype	CGGGCAAGCTTTTAAATATCATCAAAATTAACAC
LL0115-4F	Forward oligo to amplify *yfdF* with disrupted promoters	GGCTGCGAATTCATGCGGCCATCTATTTCAATCAAC
LL0115-4R	Reverse oligo to amplify *yfdF* with disrupted promoters	CGGGCAAGCTTTTAAAGGTCATCAAAATCCACAC
LL0115-5F	Forward oligo to amplify *yjgN* wildtype	GGCTGCGAATTCATGGCTCAAGTTATTAATGAAATG

LL0115-5R	Reverse oligo to amplify *yjgN* wildtype	CGGGCAAGCTTTTACAGAAATGGTAAAGAAGGC
LL0115-6F	Forward oligo to amplify *yjgN* with disrupted promoters	GGCTGCGAATTCATGGCTCAAGTCCTTAATGAAATG
LL0115-6R	Reverse oligo to amplify *yjgN* with disrupted promoters	CGGGCAAGCTTTTACAGAAATGGCCAAGAAGGC
LL0115-7F	Forward oligo to amplify *ykgH* wildtype and disrupted promoters	GGCTGCGAATTCATGCGTGAACAAATCAAACAGG
LL0115-7R	Reverse oligo to amplify *ykgH* wildtype and disrupted promoters	CGGGCAAGCTTTTACTGCCCAGAGTTTTGTTGG
LL0115-8F	Forward oligo to amplify *yjgL* wildtype	GGCTGCGAATTCATGAGCAAAATATCAGATTTAAATTATTC
LL0115-8R	Reverse oligo to amplify *yjgL* wildtype	CGGGCAAGCTTTTATGCAGTGATAGGTCTATTC
LL0115-9F	Forward oligo to amplify *yjgL* with disrupted promoters	GGCTGCGAATTCATGAGCAAAACCTCAGATTGGAATGGTTC
LL0115-9R	Reverse oligo to amplify *yjgL* with disrupted promoters	CGGGCAAGCTTTTATGCAGTGATAGGTCGGTTC

Oligos used to do ChIP-PCR on yccE WT and disrupted		
LL0715-5R	*yccE* WT and disrupted chip	GAATATGTTTGATTTTTTTCGTAC
Oligos used to amplify other promoters + yccE WT/disrupted		
LL0715-4F	*yccE* disrupted HindIII site	GGCTGCAAGCTTATGGGTAGGGATATACATGGAATTAG

buffer. DNA samples were first incubated at 90°C for 2 minutes, then samples were loaded onto a 6% sequencing gel (National Diagnostics) in 1X TBE buffer with a 60 mA conducting current. Footprints were visualised by exposing the dried gel to a phosphor screen (Bio-rad), scanned with Molecular FX phosphorimager and analysed using the Bio-Rad Quantity One software.

2.25. Calibration of sequencing gels (G+A ladder)

Maxam-Gilbert G+A ladders were used to calibrate DNA sequencing gels. Each G+A ladder is derived from the DNA fragment that is used in the DNase I footprinting experiments. Fifty (50) µl formic acid was added to 12 µl of radiolabelled DNA fragment, and the mixture was allowed to stand for 2-3 minutes at room temperature. Then, DNA was precipitated using 20 µl sodium acetate and 700 µl 100% ethanol. The precipitated DNA was subsequently vacuum-dried. Next, the DNA pellet was resuspended in piperidine, after which the resuspended DNA was incubated for 30 minutes at 90°C. After the reaction, the DNA was precipitated once more, washed twice with 70% ethanol, and subsequently dried. The resulting DNA pellet was resuspended in 20 µl loading buffer.

2.26. M13 sequencing reactions (for primer extension)

The M13 sequencing reactions were done using the T7 Sequencing Kit (USB). First, 2 µg of the DNA control template (single stranded M13mp18 DNA) was diluted to a volume of 32 µl with DEPC-treated ddH_2O. Next, 8 µl of 2 M sodium hydroxide was added and the sample was incubated for 10 minutes at room temperature. An additional 7 µl 3 M sodium acetate (pH 4.8), 4

µl DEPC-treated ddH_2O, and 120 µl 100% ethanol was then added to this mixture before incubation at -80°C for 30 minutes. After incubation, DNA was precipitated using the ethanol precipitation method, and the resulting dried DNA pellet was re-dissolved in 10 µl DEPC-treated ddH_2O. Universal primer (10 pM), and 2 µl annealing buffer, were then added and the sample was subjected to vigorous mixing. The reaction was then subjected to the following steps: incubation at 65°C for 5 minutes, incubation at 37°C for 10 minutes, and finally incubation at room temperature for 5 minutes, before undergoing brief centrifugation. This process results in an annealed primer-template mix.

At this point, 2.5 µl of the "short" read of "A", "C", "G", and "T"-mix were pipetted individually into four microfuge tubes. The enzyme T7 polymerase was also prepared by diluting 1 µl (8 units/µl) with 4 µl dilution buffer. Next, 3 µl labelling mix, 1 µl α-P^{32} dATP (10 µCi/µl), and 2 µl of the prepared T7 polymerase dilution were added to 14-15 µl of the annealed primer-template mix. This mixture was then incubated at room temperature for 5 minutes. At this point, the tubes carrying the "short" read of "A", "C", "G", and "T"-mix were pre-warmed at 37°C for 5 minutes. The reaction mixture was terminated by transferring 4.5 µl of the mixture into the pre-warmed "A", "C", "G", and "T"-mix short tubes, after which the terminated reactions were gently mixed by pipetting. Termination reactions were incubated at 37°C for 5 minutes, then were stopped by adding 5 µl of the "STOP" solution. The tubes were subjected to brief mixing and centrifugation, and at this point they can now be stored at -20°C. Prior to use (i.e. before loading on a 6% denaturing PAGE), a 3 µl aliquot from each tube was heated at 80°C for 2 minutes.

2.27. Primer extension

RNA was first extracted and purified from cells carrying the relevant pRW50 plasmid construct. Each experiment is typically carried out over two days:

Day 1:

The D49724 primer, which anneals downstream of the *Hin*dIII restriction site in pRW50, was radiolabelled with T4 polynucleotide kinase and γ-P^{32} labelled ATP as described above. Then, to 40 µg of purified RNA, 1 µl of radiolabelled primer (100-400 nM) was added. This mixture was subsequently precipitated with 1/10 volume of 3 M sodium acetate and 2.5 volumes of cold 100% (w/v) ethanol. Next, the sample was subjected to ethanol precipitation, and the resulting pellet was re-suspended in 30 µl hybridization buffer. This pellet was vigorously mixed in the buffer and was incubated at 50°C for 5 minutes. The primer was then annealed by incubating the RNA-primer mixture at 75°C for 15 minutes, then at 50°C for 3 hours. After annealing, 75 µl of cold 100% ethanol was added, and, after vigorous mixing, was left at -80°C overnight.

Day 2:

After completion of the ethanol precipitation, the pellet was in 31 µl DEPC-treated ddH_2O. The primer extension reaction was completed in a 50 µl solution containing the annealed primer-RNA reaction, 5X reverse transcriptase buffer, 1 mM DTT, and 0.2 mM (final concentration) dNTPs and 0.6 µl RNasin (an RNase inhibitor obtained from Promega) and 2.5 µl AMV reverse transcriptase (obtained from Promega). The reaction was then incubated at 37°C for 1 hour, and subsequently inactivated at 72°C for 10 minutes and centrifuged briefly. Residual RNA was degraded by adding 1 µl of 10 mg/ml RNase A and incubating the reaction at 37°C for 30 minutes. Then, to the precipitated DNA, 6.7 µl 3 M sodium acetate pH 4.8 and 125 µl 100% ethanol was

added, after which the reaction was subjected to the final ethanol precipitation step. The resulting pellet was re-suspended in 4 µl STOP solution and 2 µl of the primer extension reaction was run alongside M13 sequencing reactions in a 6% (w/v) denaturing gel. This gel was run at 60 W for 2 hours.

2.28. β-galactosidase assays

A day prior to the assay, an overnight culture, in the appropriate liquid medium with antibiotic supplements, was prepared from a fresh colony of cells carrying the required pRW50 plasmid construct. From this overnight culture, 50 µl was used to inoculate 5 ml of the same liquid medium with antibiotic supplements. This culture was then incubated at 37°C, with shaking, until mid-log phase (i.e. OD_{650} 0.3-0.5). After reaching the desired OD, two drops each of toluene and 1% (w/v) sodium deoxycholate was added to the culture. The culture was then vortexed vigorously in order to lyse the cells. The lysed cultures were put in the incubator (37°C) for 20 minutes in order for the toluene to evaporate. Once done, a 100 µl aliquot of the lysates was assayed for β-galactosidase activity. To do this, 2.5 ml (13 mM) of 2-Nitrophenyl β-D-galactopyranoside (ONPG) in Z-buffer was added to the lysate aliquot. The reaction was allowed to turn yellow before it was terminated with the addition of 1 ml 1 M sodium carbonate. Subsequently, the OD_{420} of the reaction was measured.

The β-galactosidase activity (in Miller Units) was calculated according to the following formula:

$$Activity\ (Miller\ Units) = \frac{1000\ x\ 2.5\ x\ 3.6\ x\ OD_{420}}{OD_{650}\ x\ 4.5\ x\ T\ x\ V}\ \frac{\frac{nmol}{min}}{mg\ bacterial\ mass}$$

Where:

2.5 – a conversion factor that converts OD_{650} into dry protein mass (mg). This assumes that OD_{650} of 1 is equivalent to 0.4 mg/ml bacteria (dry weight)

3.6 – final assay volume (ml)

1000/4.5 – a conversion factor that converts OD_{650} into nmol O-nitrophenol. This assumes that 1 nmol/ml O-nitrophenol has an OD_{420} of 0.0045.

T – time (mins)

V – volume of lysate used (0.1 ml)

Assays were done in triplicate. Activities were plotted in bar graphs and standard deviations were calculated accordingly.

2.29. *In vitro* multi-round transcription assays

The *in vitro* multi-round transcription assays were performed as described (Kolb *et al.*, 1995). Prior to the assay, supercoiled pSR plasmid DNA carrying the desired DNA fragment was purified in large scale using the Maxiprep Kit (Qiagen), according to the manufacturer's instructions. The resulting purified plasmid was used as a template for *in vitro* transcription. Each reaction is composed of 16 μg/ml plasmid DNA template in transcription buffer (20 mM pH 7.9, 5 mM $MgCl_2$,

500 µM DTT, 50 mM KCl, 100 µg/ml BSA, 200 µM ATP/GTP/CTP, 10 mM UTP, and 5 µCi α-P^{32}-UTP). To this mixture, 400 nM RNA polymerase holoenzyme (Cambio) with σ^{70} was added to start the reaction. When required RNA polymerase holoenzyme (Cambio) with σ^{32} was also added. The reaction was incubated at 37°C for 10 minutes, and was subsequently terminated with 20 µl of "STOP" solution. The resulting labelled RNA products were analysed on a denaturing polyacrylamide gel (run at 60 W for 1 hour). The run gel was subsequently dried and exposed to the Bio-Rad phosphorscreen. The exposed gel was imaged using the Bio-Rad FX phosphoimager and the Bio-Rad Quantity One software.

2.30. Chromatin immunoprecipitation followed by PCR (ChIP-PCR)

The chromatin immunoprecipitation (ChIP) assay was done according to the protocol described in our lab (Singh *et al.*, 2014). A day prior to the assay, an overnight culture, in the appropriate liquid medium with antibiotic supplements, was prepared from a fresh colony of cells carrying the required pRW50 plasmid construct. From the overnight culture, 400 µl was used to inoculate 40 ml of fresh liquid media. This culture was then allowed to grow to mid-log phase (i.e. OD_{650} 0.3-0.6). Once the desired optical density was acquired, cells were cross-linked with the addition of formaldehyde to a final concentration of 1% (w/v). This solution was mixed gently and incubated at room temperature for 20 minutes. Then, the crosslinking was quenched with 10 ml of 2.5 M glycine, to achieve a final concentration of 0.5 M glycine. Cells were collected by centrifugation for 5 minutes at 1,600 x g, and re-suspended in approximately the original culture volume of 1X TBS. This washing step (i.e. centrifugation and re-suspension) was repeated, but this

time the cells were re-suspended in 1 ml FA lysis buffer (150 mM NaCl) containing 4 mg/ml lysozyme. This mixture was incubated at 37°C for 30 minutes in order for cells to lyse. Next, the lysate was left to chill on ice for at least 5 minutes before sonication. The sonication step was done at 4°C using a Bioruptor (Diagenode), which is set to two 15-minute cycles (30 seconds on, 30 seconds off). The resulting cell debris was cleared by centrifugation at 17,000 x g for 5 minutes. The supernatant carrying the lysate was diluted in FA lysis buffer (150 mM NaCl), to a volume that would make sure that there was 1 ml of lysate for every 20 ml of the original culture.

Immunoprecipitation then proceeds by preparing cocktails containing 500 µl diluted lysate, 300 µl FA lysis buffer, and 25 µl Protein A beads (obtained from GE Healthcare, GE17-0780-01) made up in a 50% (v/v) slurry with TBS. In the following steps, it is important to use blunted tips when pipetting Protein A beads in order to avoid shearing of the beads. To this cocktail, 2 µl of anti-H-NS (kindly donated by Jay Hinton) was added. No antibody was used for "mock" immunoprecipitations. Cocktails were mixed on a rotating wheel at room temperature for 90 minutes. After incubation, cells were collected by centrifugation at 1,600 x g for 1 minute and the supernatants were removed carefully in order to make sure the Protein A beads were collected. Once more, 700 µl FA lysis buffer (150 mM NaCl) was added and the beads were re-suspended gently using blunt tips. The suspension was transferred to Spin-X columns, and the columns were rotated at room temperature for 3 minutes. Again, the beads were collected by centrifugation at 1,600 x g for 1 minute and flow-throughs were discarded. At this point, suspensions were subjected to several wash steps, with each step involving the addition of 750 µl of the appropriate buffer, sealing column lids with parafilm, rotating the columns at room temperature for 30 minutes, discarding the supernatant (unless otherwise stated), and briefly re-spinning the columns for 1

minute to remove excess liquid. The following wash buffers were used in this specific order: FA lysis buffer (150 mM NaCl), FA lysis buffer (500 mM NaCl), ChIP wash buffer, and TE. Finally, 300 µl TE was flushed through the Spin-X columns in order to get rid of detergent bubbles. The Spin-X columns were then transferred to fresh dolphin-nosed tubes, after which 100 µl elution buffer was added and the tubes were incubated at 65°C for 10 minutes. DNA was subsequently eluted by centrifugation at 4,000 rpm for 1 minute, and the flow-through was transferred to a new microcentrifuge tube.

DNA samples were then subjected to amplification by PCR. Each reaction contains 25 µl MyTaq Red Mix, 1 µl template (DNA sample), 1 µl forward oligo, 1 µl reverse oligo, and 22 µl ddH_2O. The cycling parameters used for the reactions are listed in Table 2.6. Note that after each segment, 5 µl of the PCR product was set aside to be loaded simultaneously on a 1% (w/v) agarose gel.

2.31. Bioinformatic analyses

Stringent search criteria that selected for putative σ^{70}-dependent promoters was previously described in our lab (Singh *et al.*, 2014). Hence, sequences that were selected matched the following motifs: 5'-TAnAAT-3', 5'-TATnAT-3' or 5'-TATAnT-3'. The relaxed search selected the sequence 5'-TAnnnT-3'. To negate the activity of potential promoters, a guanine substitution was introduced in the first two positions of each motif. A FASTA file containing the *E. coli* MG1655 open reading frames was provided by David Grainger, and copied manually into the Microsoft

Office package (Word or Excel). Motifs were then manually searched using the "Ctrl + F" (Control Key + Find) option from the Microsoft Office package.

2.32. Data analyses

Where relevant, quantitative data were analysed using ANOVA (Single Factor) test or the t-tests and $P < 0.05$ was considered statistically significant. Each statistical analysis was done using Microsoft Excel.

Table 2.6. Cycling parameters used for ChIP-PCR

SEGMENT	DESCRIPTION	CYCLES	TEMPERATURE	TIME
1	Initial melting of DNA	25	94°C	1 min
	Melting of DNA		94°C	30 sec
	Primer annealing		55°C	30 sec
	Elongation		72°C	35 sec

*****At this point, 5 µl samples are set aside to be loaded onto a 1% (w/v) agarose gel.

SEGMENT	DESCRIPTION	CYCLES	TEMPERATURE	TIME
1	Melting of DNA	3	94°C	30 sec
	Primer annealing		55°C	30 sec
	Elongation		72°C	35 sec

*At this point, 5 µl samples are set aside after each segment to be loaded onto a 1% (w/v) agarose gel.

Characterisation of the Fis binding element at the *cbpA* regulatory region

Chapter 3

3.1. Introduction

CbpA is a nucleoid associated protein and co-chaperone that is expressed during stationary phase. The *cbpAM* operon is divergent to the adjacent gene *yccE* and can be transcribed from one of four promoters (Figure 3.1). The P1 (σ^{70}- dependent) and P2 (σ^{38}-dependent) promoters overlap and are located close to the *cbpA* start codon. Of these promoters P2 is dominant and responsible for the expression of *cbpA* in stationary phase. The P4 (σ^{70}-dependent) promoter is located 157 bp upstream of *cbpA* and is poorly active. Hence, P4 has little effect on *cbpA* transcription. However, P6 (σ^{70}-dependent), which is located within the neighbouring *yccE* gene, is highly active. In order to prevent P6 from driving high levels of *cbpA* transcription during growth phase, Fis binds between *cbpA* and P6 to protect *cbpA* from P6 activity (Chintakayala *et al.*, 2013). The Fis binding element extends over 52 bp in DNase I footprinting experiments. Given that Fis binds a 15 bp DNA site, it seems likely that the Fis binding element is occupied by three Fis dimers. However, only one high affinity Fis binding site (Fis I) can be deduced from the DNA sequence (Figure 3.1). The aim of this chapter is to better understand interactions between Fis and its binding element at the *cbpA* regulatory region.

3.2. Location of Fis binding sites at the *cbpA* regulatory region

Fis binds a 15 bp, AT-rich DNA tract. Fis binding sites typically have a G at position 1 and a C at position 15 (Hengen *et al.*, 1997, Hirvonen *et al.*, 2001, Shao *et al.*, 2008). However, this

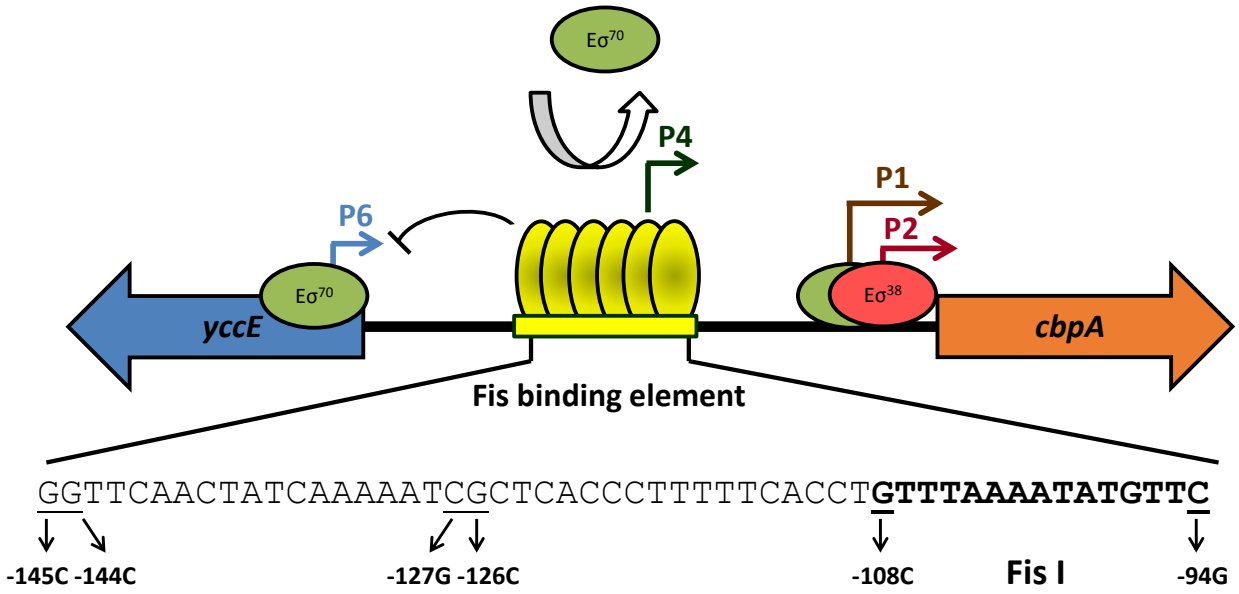


Figure 3.1. The regulation of *cbpA* by Fis.

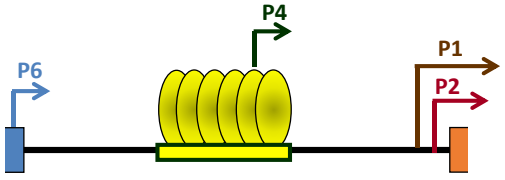
The starting model of *cbpA* regulation by Fis. Coloured block arrows indicate genes, and the black line indicates the intergenic region. Thin bent arrows signify promoters. Green ovals indicate σ^{70}-dependent RNA polymerase. The pink oval indicates σ^{38}-dependent RNA polymerase. Finally, the yellow ovals indicate the Fis protein (Chintakayala *et al.*, 2013). The Fis I site is indicated in bold text. Mutations made in this study are indicated by black arrows.

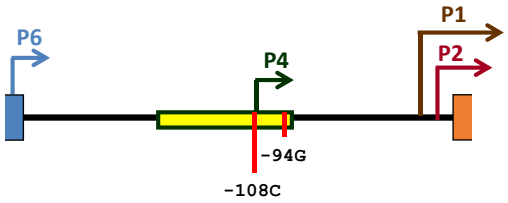
sequence is highly degenerate. In particular, the initial G or terminal C can be missing. As such, only one Fis binding site identified within the Fis binding element upstream of *cbpA* is examined on the basis of DNA sequence (labelled Fis I in Figure 3.1) (Chintakayala *et al.*, 2013).

To identify additional binding sites for Fis, point mutations were made in the Fis binding element. These point mutations were designed on the assumption that i) Fis sites begin with a G or end with a C ii) degenerate sites still have a span of 15 bp and iii) Fis sites cannot overlap. The mutations made are highlighted in Figure 3.1. Note that the -94G-108C mutation removes the high affinity Fis I site (Chintakayala *et al.*, 2013). Mutations in additional sites were made after first mutating Fis I. DNA fragments containing the full *cbpA* regulatory region, with or without the various mutations, are illustrated in Figure 3.2. These DNA fragments were radiolabelled and incubated with increasing concentrations of Fis. The protein-DNA complexes formed were analysed by electrophoretic mobility shift assays (EMSAs) or DNAse I footprinting.

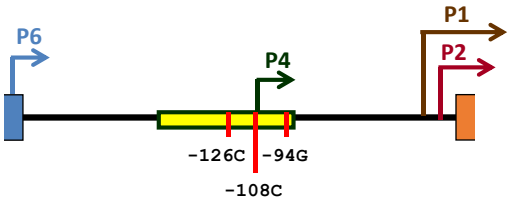
In the EMSA experiments, the wildtype *cbpA* regulatory region shows three distinct shifts in mobility upon the addition of Fis (Figure 3.3A). These complexes are referred to as Fis I, Fis II, and Fis III. Note that no free DNA is observed, even at the lowest Fis concentration with the wildtype DNA fragment (Figure 3.3A). Conversely, The DNA fragment with the -94G-108C mutation, which affects the high affinity Fis I target, never becomes saturated by Fis (Figure 3.3B). DNA fragments with mutations -94G-108C-126C (Figure 3.3C) or -94G-108C-127G (Figure 3.3D) had subtly different Fis binding patterns when compared to DNA with only the -94G-108C mutations. Thus, the Fis II complex is formed less readily (compare relative intensity of bands for free DNA and the Fis II complex in panels B, C and D). This was not the case for DNA fragments

***cbpA* (wildtype)** ***cbpA*-94G-108C**

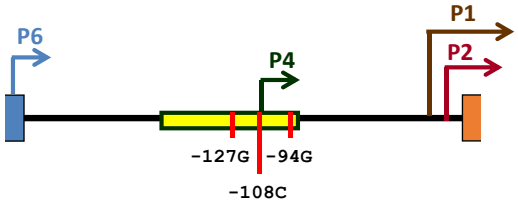




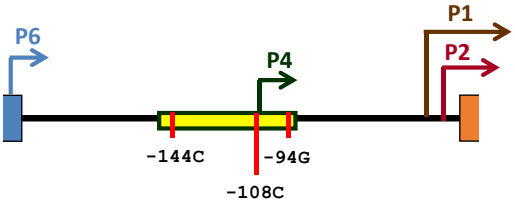
cbpA-94G-108C-126C



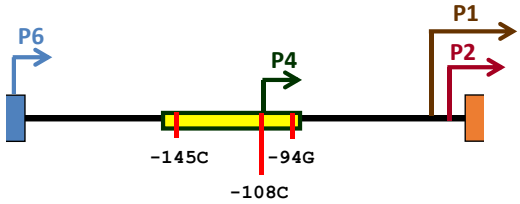
cbpA-94G-108C-126C



cbpA-94G-108C-144C



cbpA-94G-108C-145C



cbpA-126C-127G

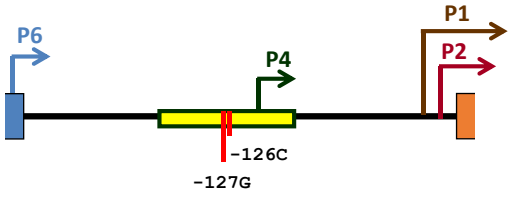


Figure 3.2. Schematic diagrams of the *cbpA* regulatory region and derivatives.

The figure shows a series of 302 bp DNA fragments carrying the *cbpA* regulatory region. The 5' end of the *yccE* gene is shown in blue blocks, and the *cbpA* start codon is shown in peach blocks. The *cbpA* P1 promoter is shown as a bent brown arrow. The Fis binding element, consisting of a 52 bp DNA sequence and likely bound by 3 Fis dimers, is shown as a yellow box. Mutations within the Fis binding element (94, 108, 126, 127, 144, or 145 bp upstream of the P1 transcription start site) are labelled and highlighted by red lines. The full sequences of these DNA fragments can be found in the Appendices.

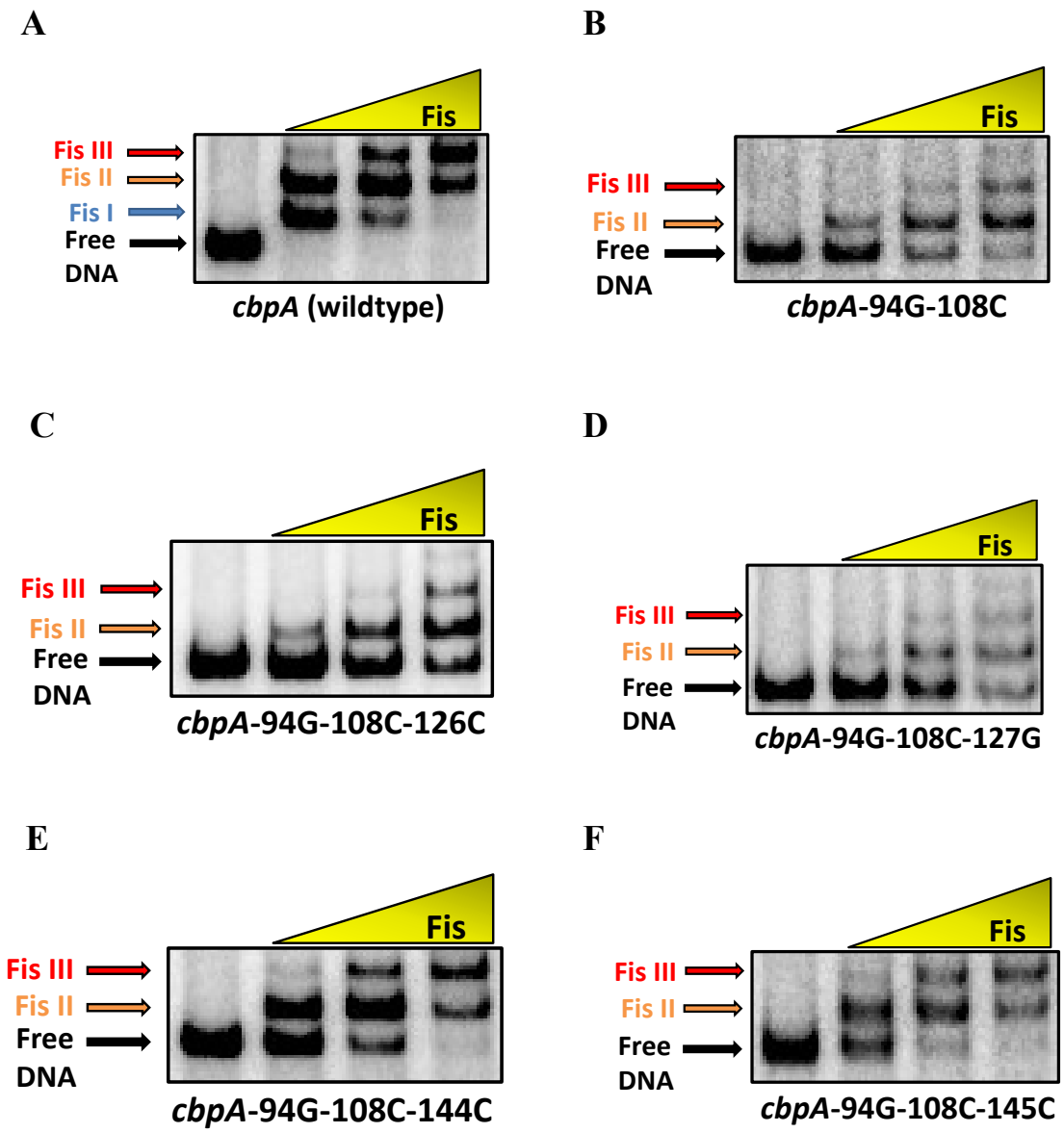


Figure 3.3. Binding of Fis to the *cbpA* regulatory region analysed by electrophoretic mobility shift assays.

The figure shows images of polyacrylamide gels on which Fis:DNA complexes were separated. The different panels represent equivalent data for the (A) wildtype *cbpA* regulatory region (B) *cbpA* regulatory region with mutations at positions -94 and -108 (C) *cbpA* regulatory region with

mutation at positions -94, -108, and -126 (D) *cbpA* regulatory region with mutations at positions -94, -108, and -127 (E) *cbpA* regulatory region with mutations at positions -94, -108, and -144 (F) *cbpA* regulatory region with mutations at positions -94, -108, and -145. All positions mentioned are numbered with respect to the P1 transcription start site. Fis was added in the following concentrations: 0 μM, 1.04 μM, 3.11 μM, 9.33 μM. Fis I, II, and III refers to the first, second, and third sites of the protein-DNA complex, respectively.

combining the -94G-108C mutation with base changes -144C or -145C (compare panels B, E, and F).

To aid interpretation of the EMSA data, DNase I footprinting assays were used to examine effects of mutations -126C-127G and -94G-108C-126C-127G. This was necessary because the EMSA experiments do not distinguish between Fis binding at different sites. Purified Fis was added at increasing concentrations to each fragment and DNase I was added to digest the DNA. The result is shown in Figure 3.4. Note that Fis binding typically results in the disappearance of observable bands (protection) and the appearance of a DNase I hypersensitive band. Thus, the wildtype *cbpA* regulatory region has a large footprint extending across the whole 52 bp Fis binding element. This large footprint can be divided into three distinct sites. The first site, Fis I (highlighted in blue) extends from -94 to -108 displays 3 regions of protection and 1 site of hypersensitivity. The second site, Fis II (highlighted in orange), extends from -112 to -126 and has 1 site of protection and 1 site of hypersensitivity. Finally, the third site, Fis III (highlighted in red), extends from -127 to -141, and has 2 sites of protection and 1 site of hypersensitivity. The -126C-127G mutation negated the formation of the Fis II and Fis III footprints but the Fis I footprint was not altered. However, the footprint was completely lost for the -94G-108C-126C-127G DNA fragment. Taken together, these results suggest that positions -126 and -127 are part of two low affinity binding sites for Fis.

The data show that the -94G-108C-126C-127G version of the *cbpA* regulatory region cannot bind Fis. To examine effects on *cbpA* expression, the various DNA fragments were cloned upstream of *lacZ* into pRW50. Thus, measurements of β-galactosidase activity can be used to measure *cbpA* expression. The data are shown in Figure 3.5. As expected, transcription driven by

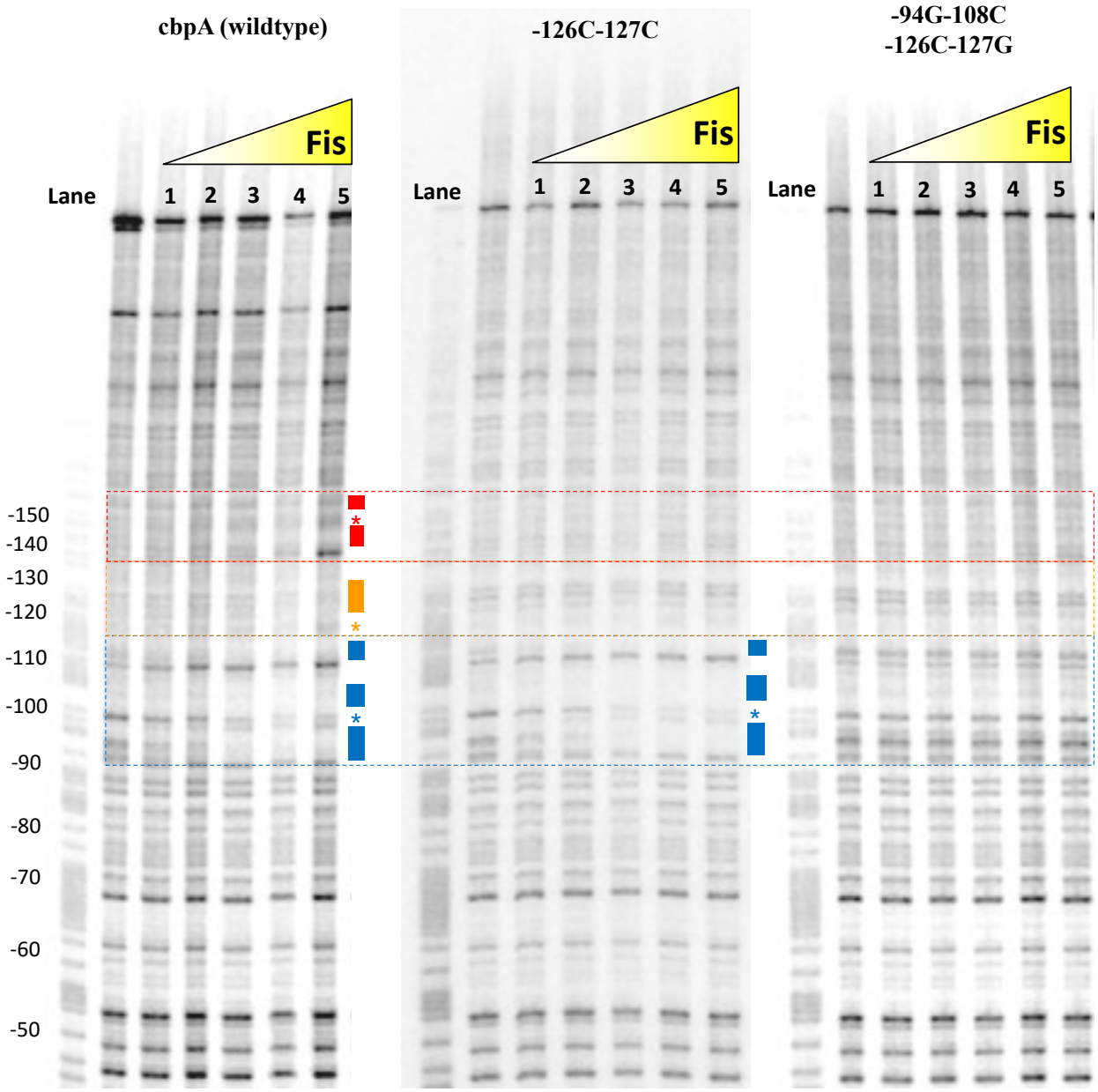


Figure 3.4. DNase I footprinting of Fis binding to the *cbpA* regulatory region.

The figure shows images of polyacrylamide sequencing gels on which DNase I digestion of 302 bp (A) wildtype *cbpA* regulatory region (B) *cbpA* regulatory region with mutations at positions -126 and -127 (C) *cbpA* regulatory region with mutations at positions -94, -108, -126, and -127 fragments were compared in the presence and absence of 14 nM, 28 nM, 56 nM, 112 nM, and 280

nM Fis. Fis I (blue dotted box), Fis II (orange dotted box), and Fis III (red dotted box) are binding sites within the Fis binding element. Stars indicate regions of hypersensitivity, bars indicate regions which are protected from DNAse I.

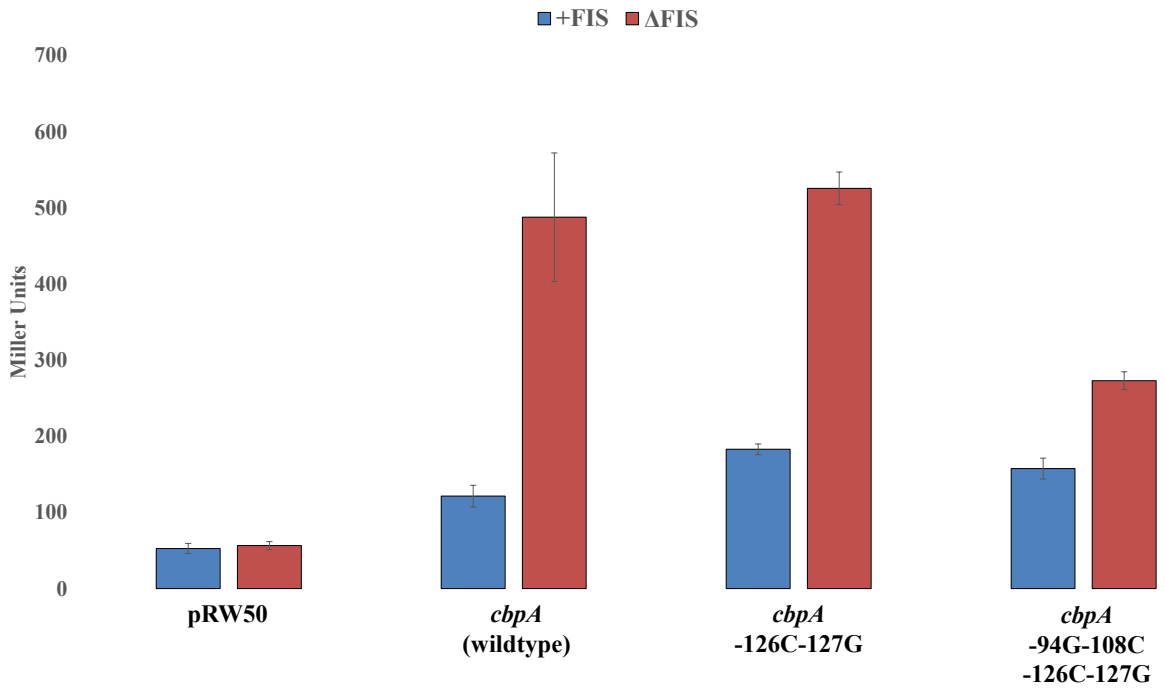


Figure 3.5. Expression of the *cbpA* regulatory region containing mutations in the Fis binding element.

Results of LacZ assays using JCB387, or the Δ*fis* derivative, transformed with pRW50 containing the wildtype *cbpA* regulatory region, the *cbpA* regulatory region with mutations at positions -126 and -127, or the *cbpA* regulatory region with mutations at positions -94, -108, -126, and -127. Empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of *E. coli* strain JCB387 are shown in blue bars. Activities of the Fis mutant derivative strain are shown in red bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

the wildtype *cbpA* regulatory region was repressed 3-fold by Fis. The same regulatory region containing mutations at positions -126 and -127, and hence, lacking the Fis II and III binding sites, behaved in a similar way to the wildtype *cbpA* fragment. However, removing all Fis binding sites had greatly reduced the effect of deleting Fis.

3.3. Orientation effects of the Fis binding element

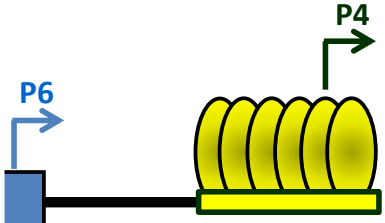
The next objective was to determine whether orientation of the Fis binding element was important. To do this, truncated DNA fragments were synthesised so that the *cbpA* regulatory region lacked the P1 and P2 promoters. Two versions of this fragment were made. In the first (*cbpA*Δ203.1), the Fis binding element was correctly orientated. In the second (*cbpA*Δ203.1.Reversed), the Fis binding was in the reverse orientation. In the context of the reversed Fis binding element the -10 region of the P6 promoter was also mutated (*cbpA*Δ203.1.Reversed-216G-217G). A schematic diagram of these fragments is shown in Figure 3.6. All fragments were cloned into pRW50 in order to measure *cbpA* expression *in vivo.* Figure 3.7 shows the results of the resulting LacZ assays. A large increase in *lacZ* expression can be observed when the orientation of Fis binding element is reversed. This occurs in both the presence and absence of Fis. Importantly, inactivation of P6 (by the mutation of -10 hexamer) abolishes transcription in all cases. Hence, the P6 promoter is responsible for the observed transcriptional effects. To further confirm this observation, derivatives of the *cbpA*Δ203.1.Reversed fragment, containing sequential 10 bp deletions from the 5' end were also made and cloned upstream of *lacZ* in pRW50 (Figure 3.8). My logic was that any residual promoter activity remaining after removal of P6 would be apparent in this assay. Thus, consistent with P6

being the sole promoter present, deletions of between 10 and 130 bp all abolished LacZ expression (Figure 3.9). Taken together, these results suggest that the orientation of the Fis binding element is important for repression of P6.

3.4. Discussion

The aim of this chapter was to fully characterise the location of Fis binding sites at the *cbpA* gene regulatory region. Previously, it was shown that the Fis binding element contained three Fis binding sites that could prevent transcription from the strong σ^{70}-dependent P6 promoter located within *yccE* (Chintakayala et al., 2013). The data presented here are consistent with one high affinity site (Fis I) located between 94 and 108 bp upstream of the *cbpA* P1 transcription start site. Two further lower affinity sites (Fis II and Fis III) are likely located between 112 and 126 bp, and between 127 and 141 bp, upstream of the *cbpA* P1 transcription start site (Figure 3.10). The position of Fis II and Fis III are proposed on the basis that mutations -126C and -127G prevent binding of two Fis molecules upstream upstream of Fis I (Figures 3.3-3.4). When suggesting the location of Fis II and III, it has been taken into account that Fis binding sites are 15 bp long, must begin or end with a G or C and cannot overlap (Hengen *et al.*, 1997, Hirvonen *et al.*, 2001, Shao *et al.*, 2008). The Fis II binding site has 54% AT-content and starts with a G, while the Fis III binding site has 74% AT-content and ends with a C. Consistent with this, Fis III fills before Fis II in DNase I footprinting assays (Figure 3.4). It is possible that binding to Fis II is facilitated by cooperative interactions between bound Fis dimers. Unexpectedly, mutating the

***cbpA*Δ203.1** ***cbpA*Δ203.1.Reversed**



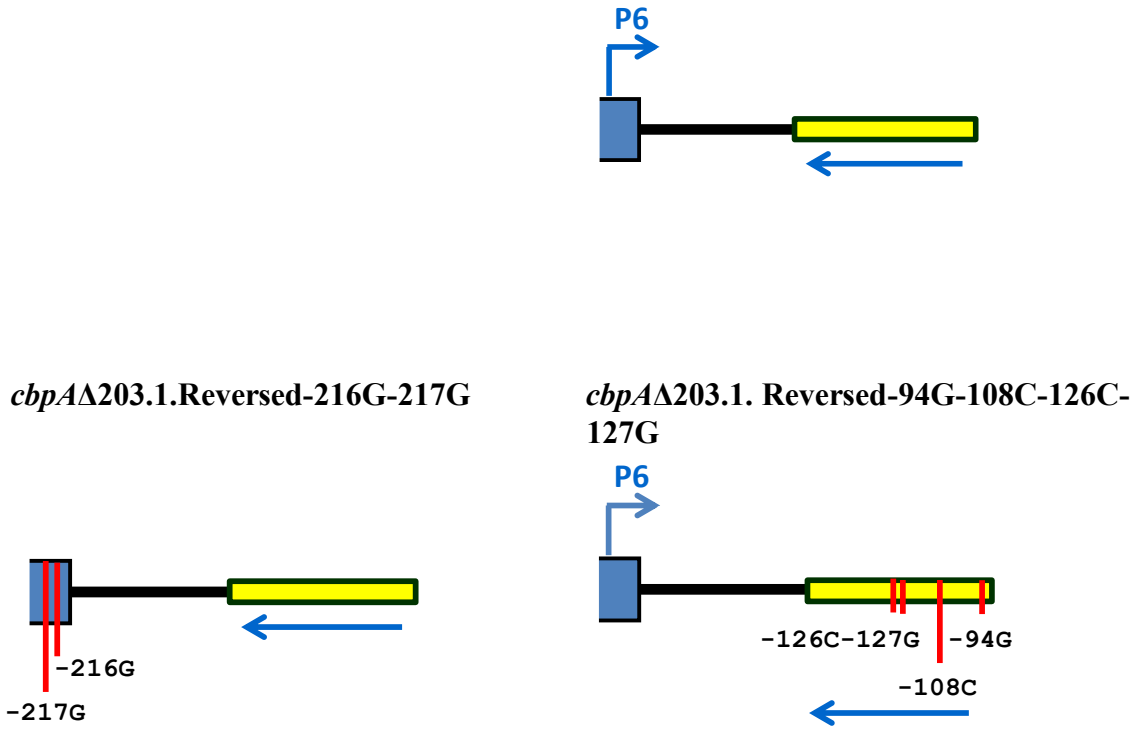


Figure 3.6. Schematic diagrams of *cbpA*Δ203.1 and its derivatives.

The figure shows a series of 209 bp DNA fragments carrying *cbpA*Δ203.1. The 5' end of the *yccE* gene is shown in blue blocks. The *cbpA* P4 promoter is a bent dark green arrow. The *cbpA* P6 promoter is a bent blue arrow. The Fis binding element, consisting of a 52 bp DNA sequence and likely bound by 3 Fis dimers, is shown as a yellow box. DNA sequence orientation is indicated by a straight blue arrow. Mutations within the Fis binding element (94, 108, 126, and 127 upstream of the P1 transcription start site) and the P6 promoter (126 and 126 upstream of the P1 transcription start site) are labelled and highlighted by red lines. The full sequences of these fragments can be found in the Appendices.

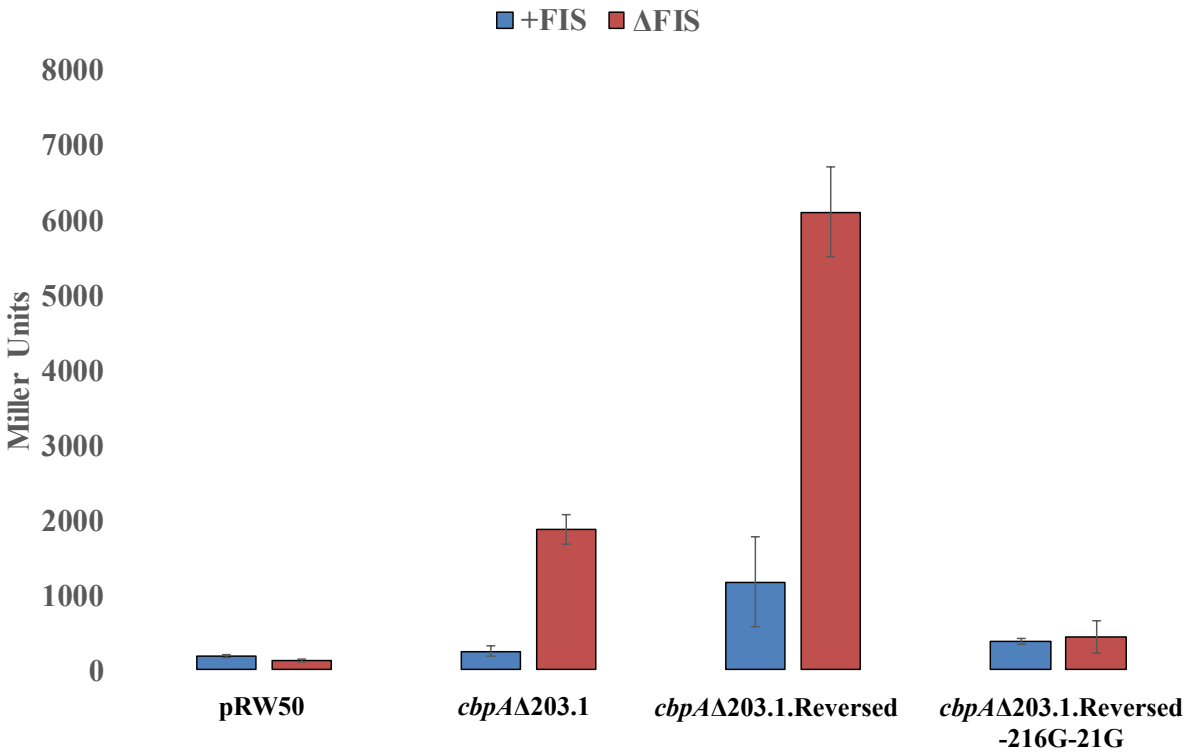


Figure 3.7. Expression comes from the *cbpA* P6 promoter when the Fis binding site is reversed.

Results of LacZ assays using JCB387, or the Δ*fis* derivative, transformed with pRW50 containing the wildtype *cbpA*Δ203.1, c*bpA*Δ203.1 with the Fis binding site in the reverse orientation, and *cbpA*Δ203.1 with the Fis binding site in the reverse orientation and mutations at positions -216 and -217 (-10 element of P6). The empty pRW50 was used as a control. Assays were done using cultures in mid-log phase. Activities of *E. coli* strain JCB387 are shown in blue bars. Activities of the Fis mutant derivative strain are shown in red bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

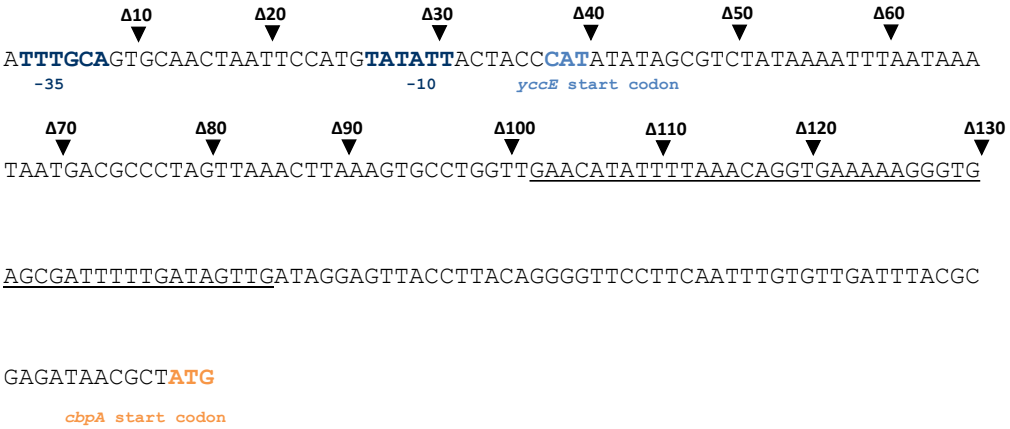


Figure 3.8. Sequence of *cbpA*Δ203.1 with the Fis binding site reversed and its deletion derivatives.

The figure shows the sequence of the "*cbpA*Δ203.1.Reversed" DNA fragment, which is a version of the *cbpA* regulatory region which lacks the P1 and P2 promoters as well as a reversed Fis binding site. The 5' end of *cbpA*Δ203.1.Reversed fragment derivatives, with sequential 10 base pair deletions, are indicated by inverted triangles and labelled according to the size of the truncation. Hence, Δ10 has a ten base pair deletion and so on.

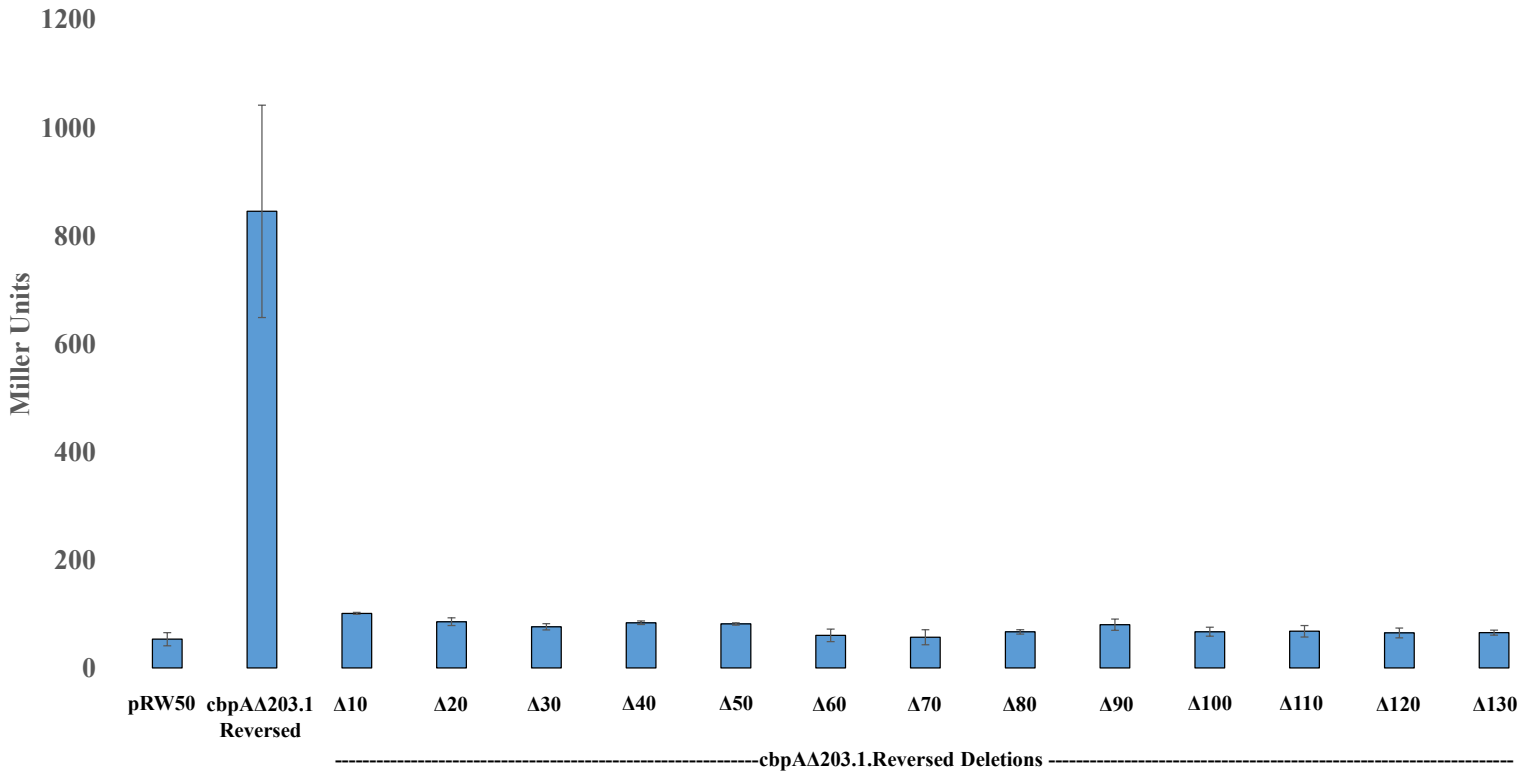


Figure 3.9. The promoter for *yccE* is not located within the Fis binding element in the reversed orientation.

Results of LacZ assays using JCB387 transformed with pRW50 containing *cbpA*Δ203.1 with the Fis binding site in reversed orientation and its sequential 10 bp deletions. The empty pRW50 was used as a control. Assays were done using cultures in mid-log phase. Activities of *E. coli* strain JCB387 are shown in blue bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

three Fis binding sites mitigates, rather than mimics, the effect of deleting the Fis gene (Figure 3.5). This suggests that the role of Fis is redundant and that, in the absence of Fis, another factor binds the Fis element. Intriguingly, the entire *yccE* gene is bound by H-NS in ChIP-seq experiments and H-NS spreading into the intergenic region between *yccE* and *cbpA* abuts the Fis binding element (Chintakayala *et al.*, 2013).

Finally, data in this chapter also shows that orientation of the Fis binding element is important for function (Figures 3.7 and 3.9). Orientation specificity is most likely linked to relative changes in the positioning of Fis and P6. Interestingly, the occurrence of three similarly positioned Fis binding sites is needed to "trap" RNA polymerase at the *tyrT* promoter (Muskhelishvili *et al.*, 1995). The formation of this complex is facilitated by collaborative interactions between the Fis protein and RNA polymerase, as well as between Fis molecules.

In summary, this chapter shows that the Fis binding element at the *cbpA* intergenic region contains one high affinity site and two low affinity sites. The orientation of these sites is essential for correct repression of P6 (Figure 3.10).

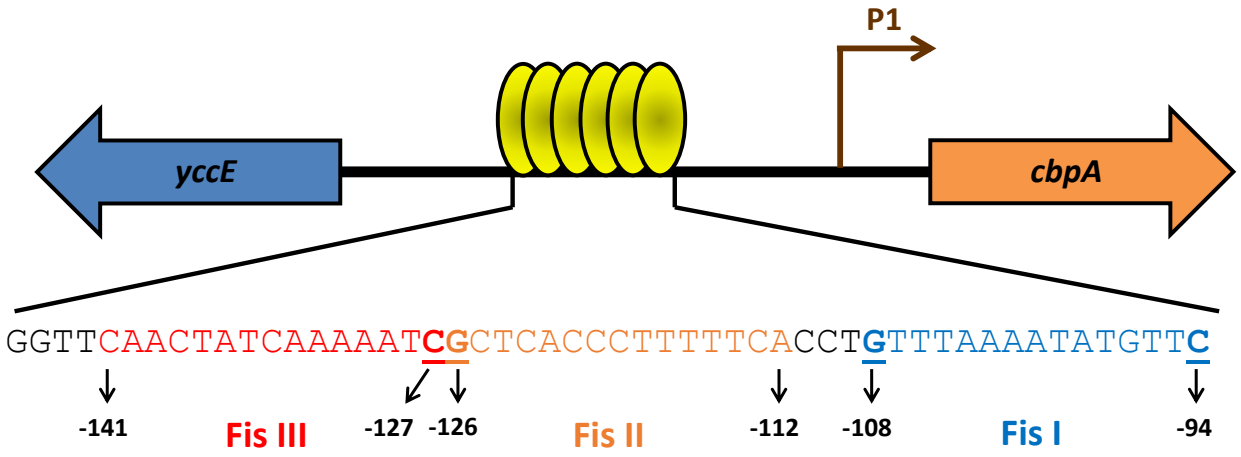


Figure 3.10. The location of the Fis binding element at the *cbpA* intergenic region.

This shows the 52-bp Fis binding element at the *cbpA* intergenic region containing the sequences of the three identified Fis binding sites. In a previous study, Fis I (blue) was located between positions 94 and 108 bp upstream of the P1 transcription start site (Chintakayala *et al.*, 2013). In this study, Fis II (orange) and Fis III (red) were located between positions 112 and 126, and between positions 127 and 141, respectively, upstream of the P1 transcription start site.

Identification of the *yccE* promoter

Chapter 4

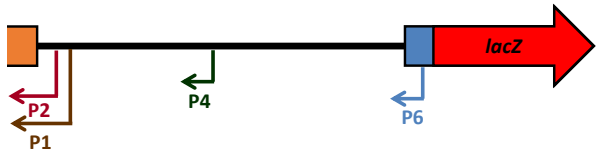
4.1. Introduction

The previous chapter focused on understanding how Fis binding to the intergenic region between *cbpA* and *yccE* influences *cbpA* expression. Whilst the P6 promoter is well defined, little is known about the *yccE* gene, its regulation, or the impact of P6 (that overlaps the *yccE* start codon) on this process. Transposon mutagenesis previously showed that mutations in *yccE* cause an increased resistance to chloramphenicol (Duo *et al.*, 2008). Furthermore, Wade and colleagues identified the 5' end of *yccE* as a target for σ^{32} (Wade *et al.*, 2006). This chapter identifies a σ^{32}-dependent promoter upstream of *yccE* and investigates the interplay between this and the nearby P6 promoter.

4.2. Identification of the *yccE* promoter and transcription start site

To crudely map the location of the *yccE* promoter, a library of DNA fragments, carrying the full *yccE* regulatory region, or truncated derivatives, was generated. The fragment carrying the full regulatory region is illustrated in Figure 4.1A and the sequence of this DNA fragment is shown in Figure 4.1B. The truncated DNA fragments have sequential 10-bp deletions at the 5' end (marked in Figure 4.1B). Oligonucleotides used to construct these fragments are listed in Table 2.5 (Chapter 2). The DNA fragments were cloned into the pRW50 plasmid, upstream of *lacZ*, to create *lacZ* fusions. The resulting plasmids were used to transform *E. coli* strain JCB387. Transformants were then used to inoculate LB media and lysates of these cultures were assayed for LacZ activity. Data are shown in Figure 4.2. The full *yccE* regulatory region, and derivatives with between 10 and 200

A



B

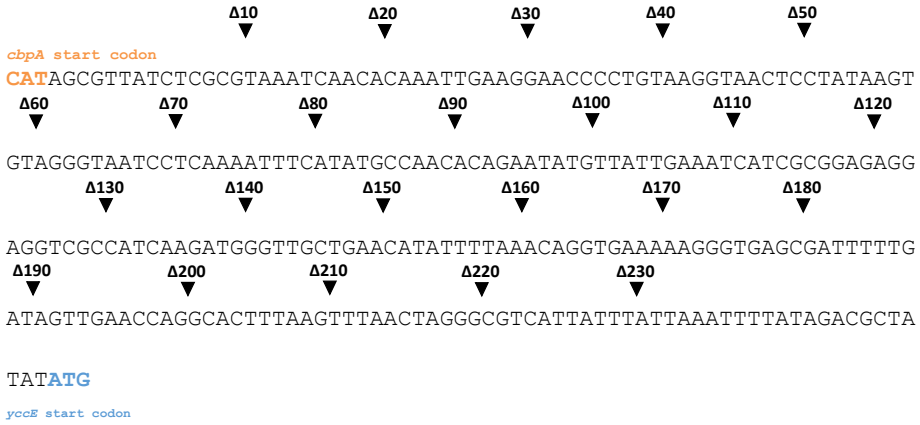


Figure 4.1. Schematic diagram and sequence of the *yccE* regulatory region and its deletion derivatives.

The figure shows the *yccE* regulatory region DNA fragment. Panel (A) shows a schematic diagram of a DNA fragment containing the intragenic region between *cbpA* and *yccE*. The 5' end of the *yccE* gene is shown as a blue block, *lacZ* is shown in red, and the *cbpA* start codon is shown in peach. Promoters are shown as bent arrows. Panel (B) shows the sequence of the DNA fragment. The 5’ end of the *yccE* mRNA, and the *yccE* start codon, are highlighted in blue. Sequential 10 base pair deletions are indicated by inverted triangles and labelled according to the size of the truncation. Hence, Δ10 has a ten base pair deletion and so on.

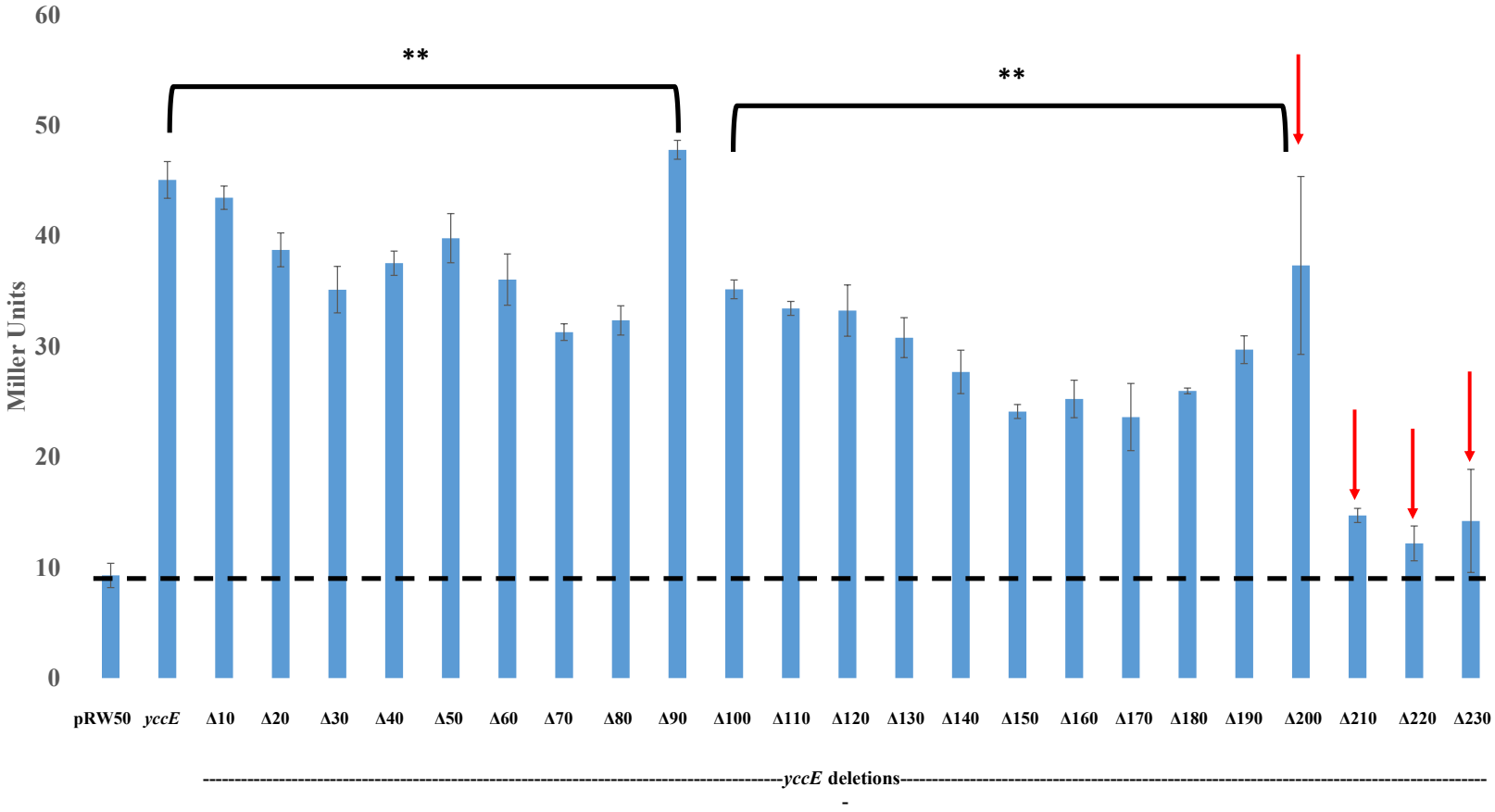


Figure 4.2. Deletion analysis of the *yccE* regulatory region.

Results of LacZ assays using JCB387 transformed with pRW50 containing the full *yccE* regulatory region or derivatives with different 10-bp deletions at the 5' end. Activities of the *E. coli* strain JCB387 are shown in blue bars. The empty pRW50 was used as a control. Background activity generated by pRW50 is indicated by black dashed lines. Assays were done using overnight

cultures. Activity data are reported as means ± standard deviations of the mean of three replicates. Constructs used in subsequent *in vitro* transcription assays (Δ200, Δ210, Δ220, Δ230) are highlighted by red arrows. Data were analysed statistically using the ANOVA (Single Factor) test (** $P < 0.01$).

bp removed from the 5' end, had similar activity levels varying less than 0.5-fold. The construct carrying the *yccE* full intergenic region and the constructs containing deletions of 10 bp to 90 bp were found to be significantly different from each other. Similarly, the constructs carrying deletions of 100 bp to 200 bp were also found to be significantly different from each other. In contrast, deletion of 210 bp or more reduced LacZ activity by >6-fold. As expected, the differences between these particular constructs were found to be not significant. Hence, the minimal *yccE* promoter is likely located within the Δ200 construct.

To understand if the potential promoter for *yccE* requires σ^{32}, the Δ200, Δ210, Δ220, and Δ230 DNA fragments were cloned into the plasmid, pSR. Thus, the potential *yccE* promoter is placed upstream of the factor independent *λoop* transcription terminator. This DNA can be used as a template for *in vitro* transcription. Thus, RNA polymerase should produce transcripts from the *yccE* promoter that are terminated at *λoop*. Note that the RNAI transcript, which is produced from the pSR plasmid replication origin, is apparent as a 109/110 nt doublet in these experiments. As expected, σ^{70} containing RNA polymerase is more efficient at transcribing RNAI than σ^{32} holoenzyme (compare Lanes 1 and 2 in Figure 4.3). Crucially, in Lanes 1 and 2, the pSR template did not contain a cloned *yccE* promoter fragment. Thus, the 106 nt mRNA transcript generated by the σ^{32} holoenzyme when the full *yccE* intergenic region was cloned in pSR likely represents the *yccE* mRNA (Figure 4.3, Lane 3). This same 106 nt transcript was observed using the Δ200 DNA fragment (compare Figure 4.3, Lanes 3 and 4). However, the 106 nt mRNA was absent in reactions using the Δ210, Δ220, or Δ230 DNA fragments (Figure 4.3, Lanes 5-7). These data are consistent with β-galactosidase expression patterns *in vivo*.

To precisely map the *yccE* transcription start site, mRNA primer extension assays were used. Here, RNA was extracted from *E. coli* JCB387 cells carrying pRW50 construct with the full *yccE* regulatory region (Figure 4.1). The radiolabelled nucleotide D49724, which binds to a small (37 bp) region within *lacZ*, is incubated with extracted RNA and primes reverse transcription. Hence, defining the size of the resulting cDNA allows the precise 5' end of the *yccE* mRNA to be defined. An image of the gel is shown in Figure 4.4A. The extension product is 108 nt in length and was present at lower levels in cells lacking Fis (Figure 4.4A, Lanes 1 and 2). The deduced *yccE* transcription start site is labelled as "+1" in Figure 4.4B. A potential σ^{32} dependent promoter is located just upstream. Consistent with my deletion analysis, this promoter, referred to from this point as P*yccE*, falls just inside the boundary of the Δ200 DNA fragment.

In order to check correct identification of P*yccE*, site-directed mutagenesis was used to mutate the proposed-10 and -35 elements for σ^{32}. These mutations, numbered with respect to the P*yccE* transcription start site, were made in the context of the Δ200 DNA fragment. Three derivatives of the Δ200 DNA fragment were created. The mutations -29G, -28G, -26G, and -25G alter the -35 promoter element. Mutations -5G and -4G alters the -10 promoter element. When all mutations are combined, both the -35 and -10 elements should be removed. The three Δ200 fragment derivatives, along with the original Δ200 fragment, were cloned upstream of *lacZ* in pRW50. Cells transformed with these constructs were then used in β-galactosidase assays to determine promoter activity. Data are shown in Figure 4.5. As expected, the mutated Δ200 fragments all had greatly reduced activity compared to the wildtype Δ200 DNA fragment.

Lane	1	2	3	4	5	6	7
DNA template	empty pSR	empty pSR	*yccE* pSR	Δ200 pSR	Δ210 pSR	Δ220 pSR	Δ230 pSR
σ factor	70	32	32	32	32	32	32

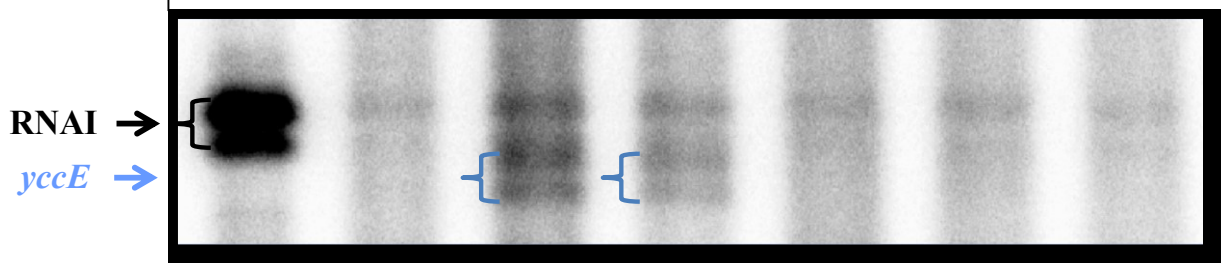


Figure 4.3. *In vitro* transcription assays of the *yccE* full regulatory region and its deletion constructs.

The figure shows mRNA transcripts generated from plasmid pSR or derivatives carrying the full *yccE* regulatory region ("*yccE*") or truncated derivatives. The *yccE* mRNA transcript, shown in blue brackets, is identified as a 106 nt product and is generated by σ^{32}-associated RNA polymerase (800 nM). Control transcripts encoded in pSR, RNAI (109-110 nts), shown in black brackets, were generated by σ^{70} associated RNA polymerase (200 nM). Lane 1 – pSR on its own + σ^{70}-associated RNAP, 2 – pSR on its own + σ^{32}-associated RNAP, 3 – *yccE* + σ^{70}-associated RNAP, 4 – Δ200 + σ^{70}-associated RNAP, 5 – Δ210 + σ^{70}-associated RNAP, 6 – Δ220 + σ^{70}-associated RNAP, 7 – Δ230 + σ^{70}-associated RNAP.

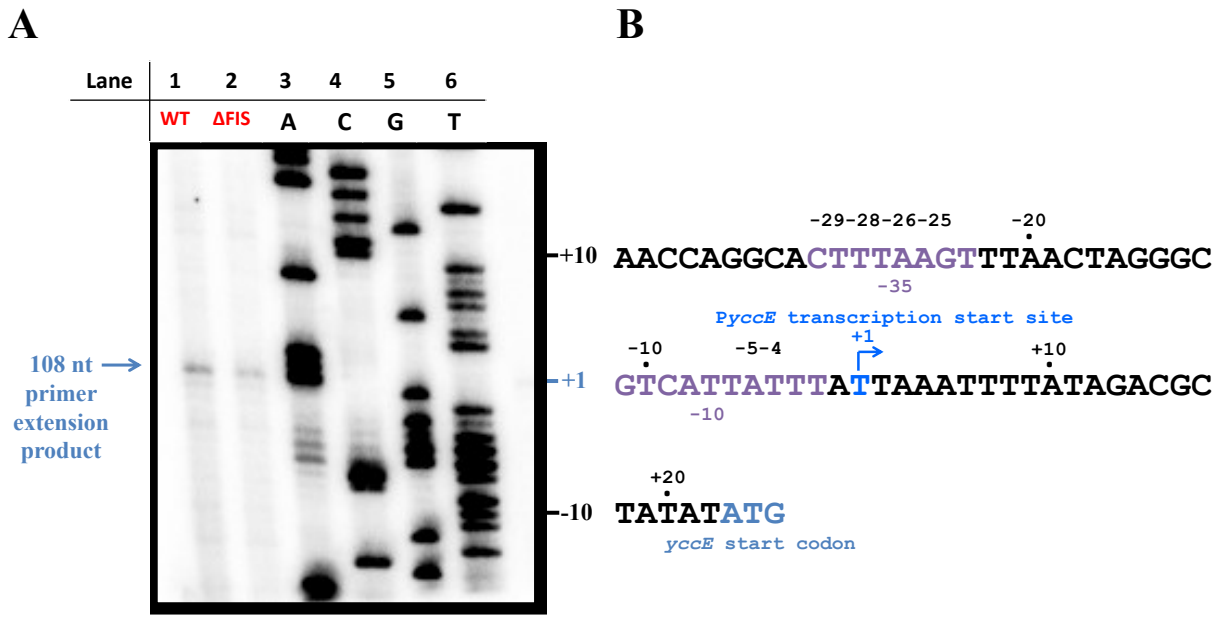


Figure 4.4. Location of the *yccE* transcription start site *in vivo*.

This result maps the *yccE* transcription start site *in vivo*.

(A) The gel shows results of the *yccE* mRNA primer extension analysis. The transcription start site (+1, shown in blue text) is indicated by a 108 nt primer extension product. The gel was calibrated using arbitrary size standards (A, C, G, T). Primer extension was done in both wildtype JCB387 and Fis mutant derivative strains.

(B) The non-template strand sequence of the *yccE* regulatory region. The *yccE* transcription start site is labelled as +1. Nucleotides where site specific mutations were subsequently introduced are in black numbered texts. Both the *yccE* transcription start site and the *yccE* start codon are shown in blue text.

4.3. Occurrence of transcriptional interference between the *yccE* promoter and the *cbpA* P6 promoter

P*yccE* and the *cbpA* P6 promoter are separated by 35 bp and orientated towards each other. This arrangement implies that transcription driven by P*yccE* and *cbpA* P6 might be antagonistic.

To understand how P*yccE* and *cbpA* P6 interact, a further set of DNA fragments was generated. In all of these fragments P6 is orientated to drive transcription of a downstream gene (Figure 4.6). I made a derivative of this fragment containing a 20-bp deletion in P*yccE* and a derivative where the P6 -10 element had been mutated as described in the previous chapter (Figure 4.6). Full sequences of the DNA fragments are provided in the Appendices. The DNA fragments were cloned into pRW50 plasmid, upstream of *lacZ*, in order to produce *lacZ* fusions. Cells transformed with these plasmids were then used in the β-galactosidase assays to measure *lacZ* expression. Data obtained from the assays are shown in Figure 4.7. Deletion of P*yccE* leads to a moderate increase (not statistically significant) in transcription that is lost when the *cbpA* P6 promoter is mutated.

In confirmatory experiments, *in vitro* transcription assays were used to better visualise interference between P*yccE* and *cbpA* P6. Thus, the same DNA fragments containing P6 with or without P*yccE* were cloned upstream of the *λoop* terminator in pSR. Therefore, RNA polymerase generates a transcript from *cbpA* P6, which is terminated at the *λoop* signal. This transcript can be visualised and quantified by denaturing PAGE. Note that any transcript from P*yccE* is not evident since there is no terminator in the opposite orientation. It was expected that σ^{70}-dependent transcription from the P6 promoter would be reduced by σ^{32}-dependent transcription from P*yccE*.

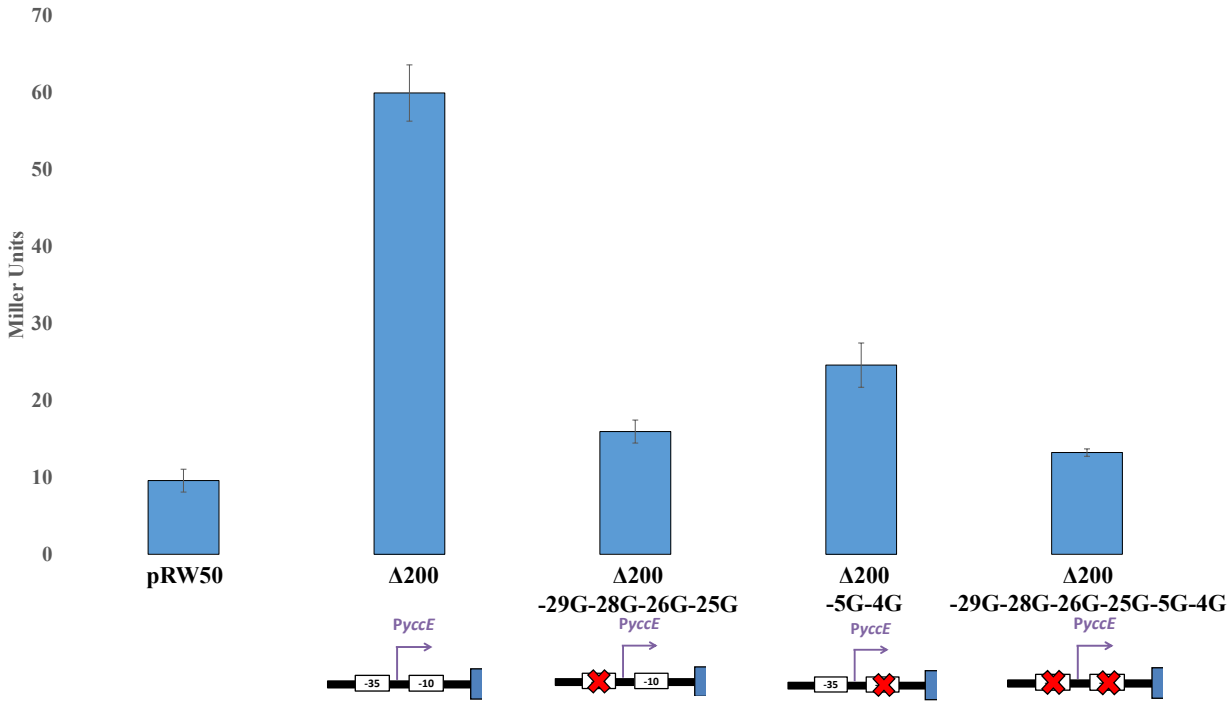


Figure 4.5. Expression levels of constructs containing the P*yccE* -35 and -10 promoter elements.

Results of LacZ assays using JCB387 transformed with pRW50 containing the *yccE* promoter region, derivatives with mutations in the -35 or -10 promoter elements, and a derivative with mutations in both the -35 and -10 promoter elements. Schematic diagrams are shown under the graph. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of the *E. coli* strain JCB387 are shown in blue bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

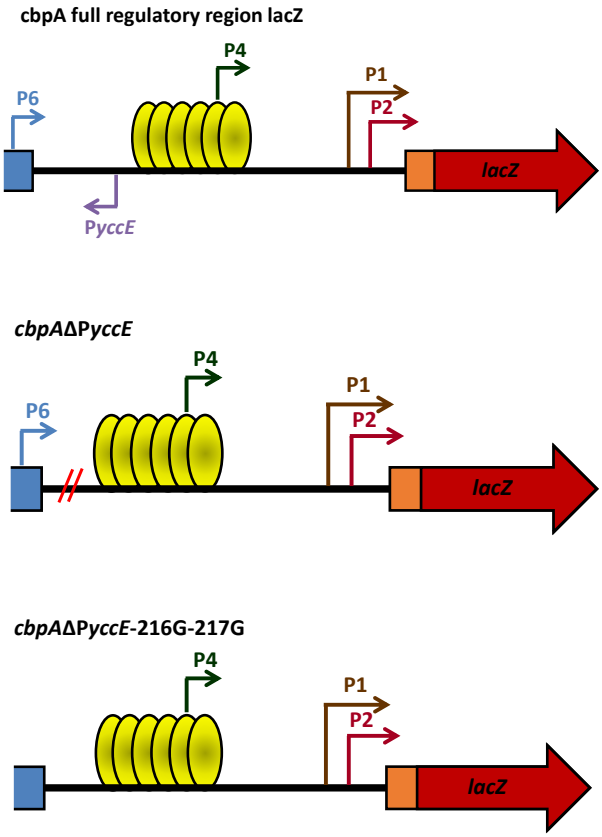


Figure 4.6. The *cbpA* regulatory region and its *lacZ* derivatives.

The figure shows schematic diagrams of the *cbpA* regulatory region, its *lacZ* derivative which has the P*yccE* deletion, and another derivative which has the P*yccE* deletion (indicated by red lines) and the P6 mutation. The 5'end of the *yccE* gene is shown as blue blocks, the *cbpA* start codon is shown as peach blocks, and *lacZ* is shown as red block arrows. The P*yccE*, *cbpA* P1, P2, P4, and promoters are shown as bent arrows. The yellow circles indicate Fis protein monomers. The direction of each arrowhead indicates the direction of expression. Full sequences of the DNA fragments can be found in the Appendices.

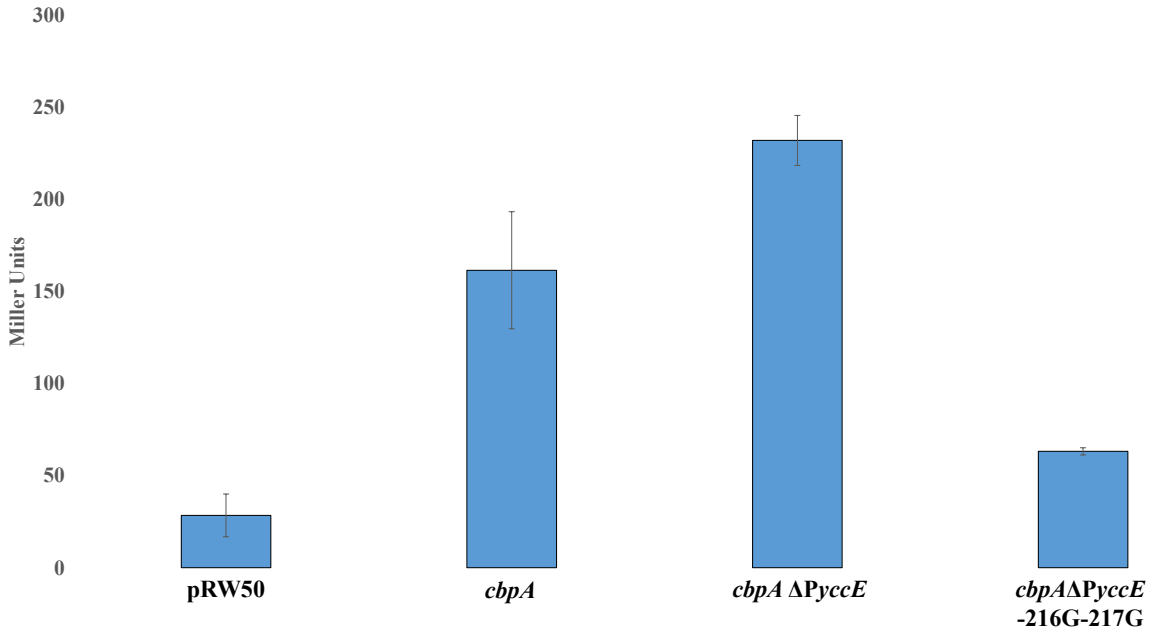


Figure 4.7. Expression levels of constructs containing the *cbpA* full regulatory region, the *cbpA* regulatory region with the *yccE* promoter region deleted and its derivative.

Results of LacZ assays using JCB387 transformed with pRW50 containing the *cbpA* regulatory region, the *cbpA* regulatory region with the *yccE* promoter region deleted, and derivative containing mutations in the *cbpA* P6 promoter. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of the *E. coli* strain JCB387 are shown in blue bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

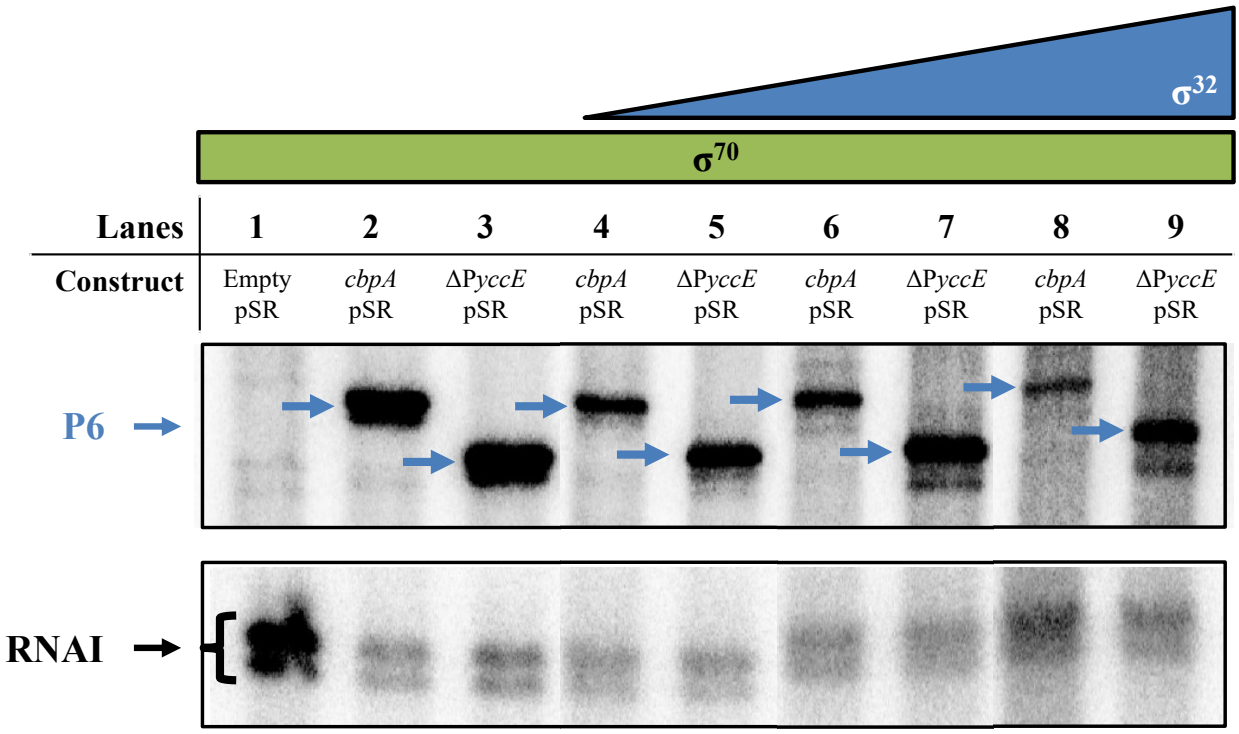


Figure 4.8. *In vitro* transcription assays of the *cbpA* regulatory region and its derivative with the *yccE* promoter deleted.

The figure shows mRNA transcripts generated from plasmid pSR or derivatives carrying the *cbpA* regulatory region (*cbpA*) and the *cbpA* regulatory region with the *yccE* promoter deleted (Δ*yccE*). Transcripts were generated using σ^{70}-associated RNA polymerase (σ^{70} in a green block), σ^{32}-associated RNA polymerase (σ^{32} in a blue triangle), or a mixture of both. The concentration of the σ^{70}-associated RNA polymerase was maintained at 200 nM throughout the assay. On the other hand, σ^{32}-associated RNA polymerase was added in increasing concentrations of 200 nM, 400 nM, and 700 nM. The *cbpA* P6 mRNA transcripts, shown in blue arrows, are identified as 334 nt and 354 nt products. Control transcripts encoded in pSR, RNAI (109-110 nts), shown in black brackets, were generated by σ^{70} associated RNA polymerase (200 nM).

Furthermore, any such effect would be abated by the 20-bp deletion that removes P*yccE*. Data are shown in Figure 4.8. In Figure 4.8, the *cbpA* P6 transcript is 354 nt in length (Lane 2) or 334 nt in length if derived from the DNA template with the 20-bp deletion in P*yccE* (Lane 3). Production of the 354 nt transcript appeared to be sensitive to σ^{32} (compare Lanes 2, 4, 6, and 8) but the 334 nt transcript did not appear to be sensitive to σ^{32} (compare Lanes 3, 5, 7, and 9).

4.4. Discussion

This chapter identifies P*yccE* and determines how P*yccE* influences transcription from the nearby P6 promoter. The data show that P*yccE* is located 50 bp upstream of the *yccE* start codon and is recognised by σ^{32} containing RNAP (Figures 4.2 and 4.3). Surprisingly, although P*yccE* binds σ^{32} *in vitro* (Figure 4.3) and *in vivo* (Wade *et al.*, 2006), and has a sequence consistent with recognition by σ^{32} (Figure 4.4B), I found no apparent induction of P*yccE* by heat shock (data not shown). This may suggest that additional factors are needed for full induction of P*yccE*. Even so, P*yccE* is located close to the convergent P6 promoter. Data presented here suggest that P*yccE* interferes with transcription from the P6 promoter. Thus, both biochemical and genetic assays show that deletion of the P*yccE* promoter results in increased P6 activity (Figures 4.7 and 4.8). The occurrence of interference between convergent promoters is not uncommon, as shown in a number of studies both in prokaryotes and eukaryotes (Moinier *et al.*, 2014, Russo *et al.*, 2015).

In summary, this chapter identifies further complexities in the intergenic region between *yccE* and *cbpAM*. Thus, the region contains multiple promoters recognised by 3 different σ factors. The experimental data show that transcription from the σ^{70}-dependent P6 promoter is reduced by both

σ^{32}-mediated transcriptional interference and direct repression by Fis. Hence, this work extends our understanding of how *cbpA* expression can be controlled (Figure 4.9).

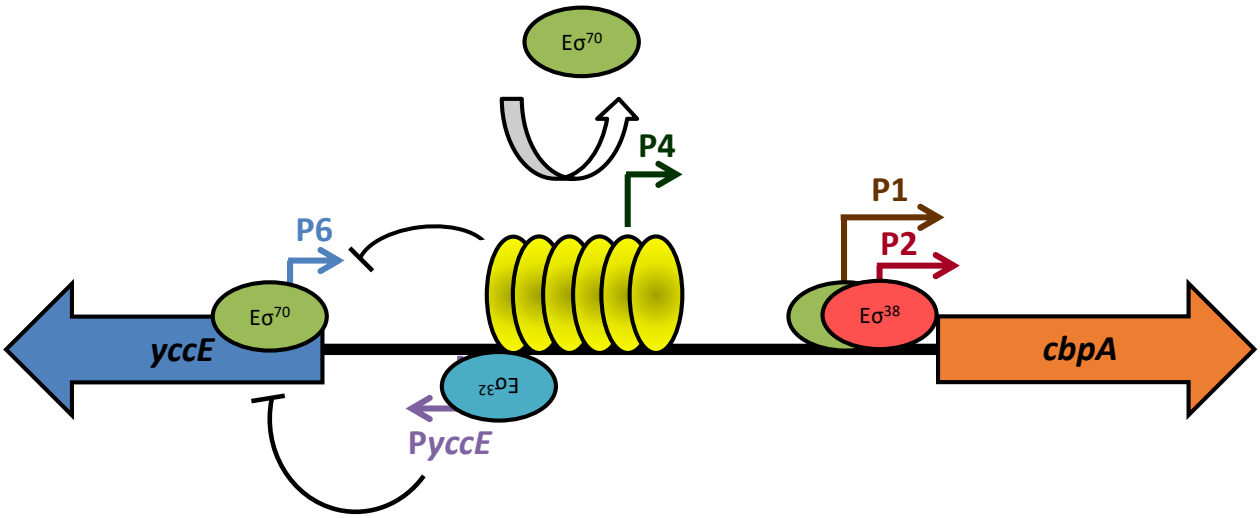


Figure 4.9. The proposed new model of the regulation of *cbpA*.

This figure shows the proposed new model of the regulation of *cbpA*. Coloured block arrows indicate gene regions, and the black line indicates the intergenic region. Thin bent arrows signify promoters, and green ($E\sigma^{70}$), pink ($E\sigma^{38}$) and blue ($E\sigma^{32}$) ovals signify RNA polymerase holoenzymes. Finally, the yellow ovals indicate the Fis protein.

"Pseudo-regulation" of *yccE* and other solitary AT-rich genes by H-NS

Chapter 5

5.1. Introduction

A number of studies have noted that AT-rich genes contain a large number of DNA sequences that resemble promoters (Huang *et al.*, 2012, Singh and Grainger, 2013, Singh *et al.*, 2014, Lam and Charles, 2015). Such genes tend to be bound by H-NS and, when *hns* is deleted, intragenic promoters can stimulate transcription (Singh *et al.*, 2014). These observations are relevant here because *yccE* has an AT-content of 65.87% and is bound by H-NS. AT-rich genes are likely to be acquired through horizontal gene transfer. As such, *yccE* can only be found within the *Escherichia coli* and *Shigella flexneri* species (Appendix A.16), showing the distinct lack of conservation within the enterobacteriaceae family. Furthermore, our work to date has already defined at least one intragenic promoter (P6) within *yccE*. This chapter will test the hypothesis that many "regulatory" effects of H-NS on RNA synthesis may be an illusion resulting from widespread intragenic transcription. I will refer to this predicted phenomenon as "pseudo-regulation" (Figure 5.1A). This differs from canonical gene regulation (Figure 5.1B) because most of the transcripts regulated are inside genes and are non-coding. In this chapter, in order to distinguish the different effects of H-NS, I have utilised the properties of *yccE*.

5.2. The number of promoter elements observed per base has a strong, positive correlation to the AT-content within the *E. coli* MG1655 genome

Promoters are AT-rich DNA sequences containing the hexamers 5'-TTGACA-3' (the -35 element) and 5'-TATAAT-3' (the -10 element). Thus, AT-rich genes have a tendency to contain promoter-like sequences. In order to understand the relationship between gene AT-content, and promoter occurrence, I analysed the complete sequence of the *E. coli* genome. For each gene the number of potential -10 elements was predicted. To identify possible promoters I utilised a stringent search criteria. Hence, the sequences 5'-TAnAAT-3', 5'-TATnAT-3', or 5'-TATAnT-3' were selected. Figure 5.2 shows the data obtained from the analysis. Each data point represents a gene in the *E. coli* MG1655 genome. The position of each data point reflects the gene AT-content as a function of the number of predicted promoter -10 elements per base. The plot shows a clear positive correlation between the AT-content of a gene and the number of promoter -10 motifs. In this plot, *yccE* is shown in blue, and data points for four other similarly sized AT-rich genes are also highlighted. Importantly, unlike many AT-rich genes, *yccE* and the other highlighted ORFs are not part of larger sections of AT-rich DNA (e.g. prophages). Interestingly, clusters of promoter motifs have also been found below the normal distribution (i.e. low AT-rich genes with a high number of observed promoter motifs).

5.3. The canonical promoter for *yccE* is not required for the increase in transcription of *yccE* in cells lacking H-NS

The analysis shown in Figure 5.2 is consistent with *yccE* being enriched for intragenic promoters. The next goal was to understand if these intragenic promoters, rather than P*yccE*, were responsible for increased transcription at the *yccE* locus in cells lacking H-NS. Hence, I fused the

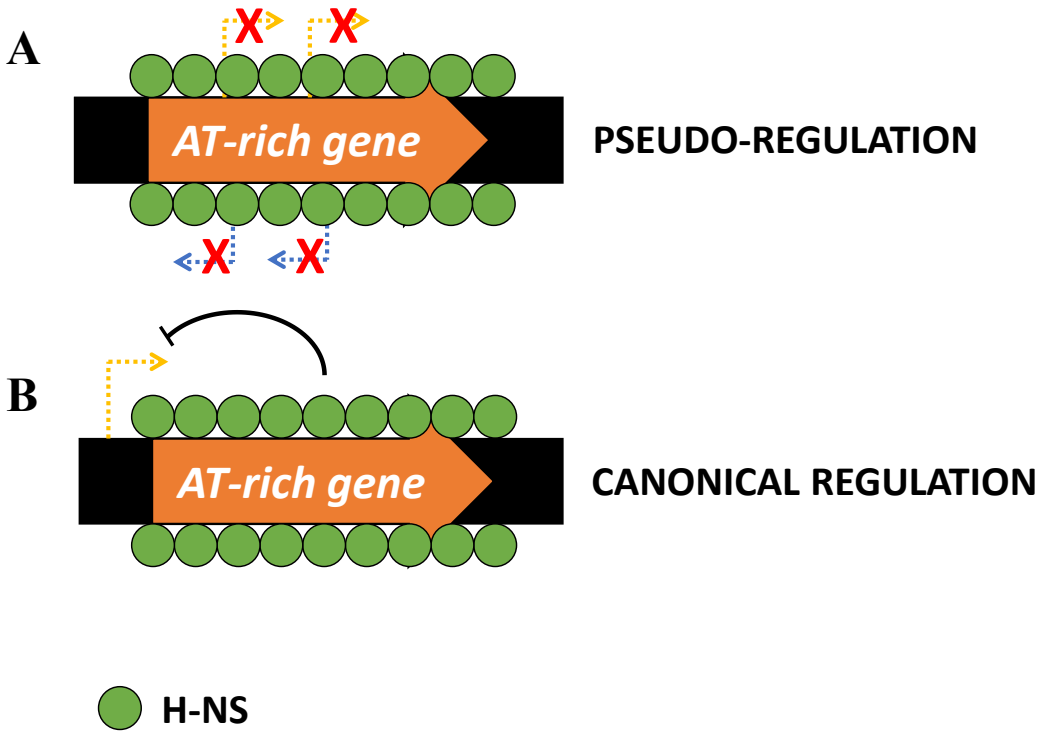


Figure 5.1. Canonical regulation vs. pseudo-regulation.

This figure shows a schematic diagram for the models of canonical regulation vs. pseudo-regulation of AT-rich genes by H-NS. The black block lines indicate double-stranded DNA, coloured block arrows indicate genes, green circles indicate H-NS monomers, and bent dotted arrows indicate promoters.

entire *yccE* gene to *lacZ*, with or without P*yccE*, in plasmid pRW50. Thus, *yccE* and *lacZ* form an operon where *lacZ* expression is under the control of P*yccE*. Consequently, when P*yccE* is removed, *lacZ* can only be expressed if *yccE* contains internal promoters. The pRW50 derivatives were used to transform the *E. coli* strain M182 or the Δ*hns* variant. Remarkably, in cells lacking H-NS, LacZ activity is dramatically increased for both constructs (Figure 5.3B).

5.4. The *yccE* coding sequence contains a number of active internal promoters

The genome-wide search for potential -10 elements within genes (Figure 5.2) identified 21 potential promoters inside *yccE* (listed in Table 5.1). I next sought to determine how many of these predicted promoters were able to stimulate transcription. Thus, the identified promoters were individually isolated on short (56 bp) DNA fragments. Note that, in order to bind DNA, H-NS must interact with itself by oligomerisation. Hence, the short 56-bp DNA fragments were not expected to bind H-NS. After being cloned upstream of *lacZ* in pRW50, the promoters were tested for activity. Note that one DNA fragment had an internal *Eco*RI restriction site and another resisted cloning for unknown reasons. Thus, of the 21 potential promoters, 19 were tested for activity (Figure 5.4). Of the 19 promoters, 11 showed activity twofold or more above background levels (Figure 5.4, labelled "A" – "K"). Also, most of the active promoters were in the sense orientation relative to *yccE*. DNA fragment "A" contains the strong antisense promoter (*cbpA* P6) described in previous chapters. To confirm functionality, predicted promoter -10 elements of the 11 transcriptionally active promoters were mutated. Hence, the first two bases of the -10 element

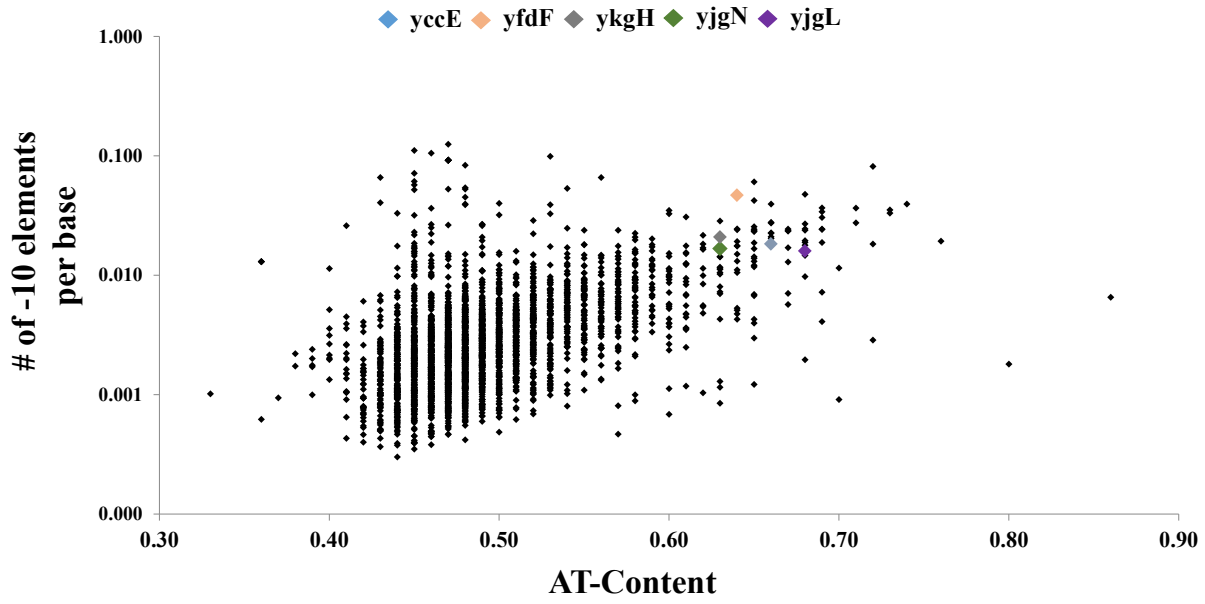


Figure 5.2. Relationship between AT-content and the number of promoter motifs (-10 element) per base.

This scatter plot shows the relationship between the AT-content and the number of -10 promoter elements observed per base. Each data point (◆) represents the number of -10 promoter motifs observed given a gene with specified % AT-content. The criteria used to define a -10 promoter motif is a 3 out of 6 match to the perfect -10 promoter motif, which is 5'-TAnnnT. The genes used in this study are shown in coloured data points: *yccE* (◆), *yfdF* (◆), *ykgH* (◆), *yjgN* (◆), *yjgL* (◆).

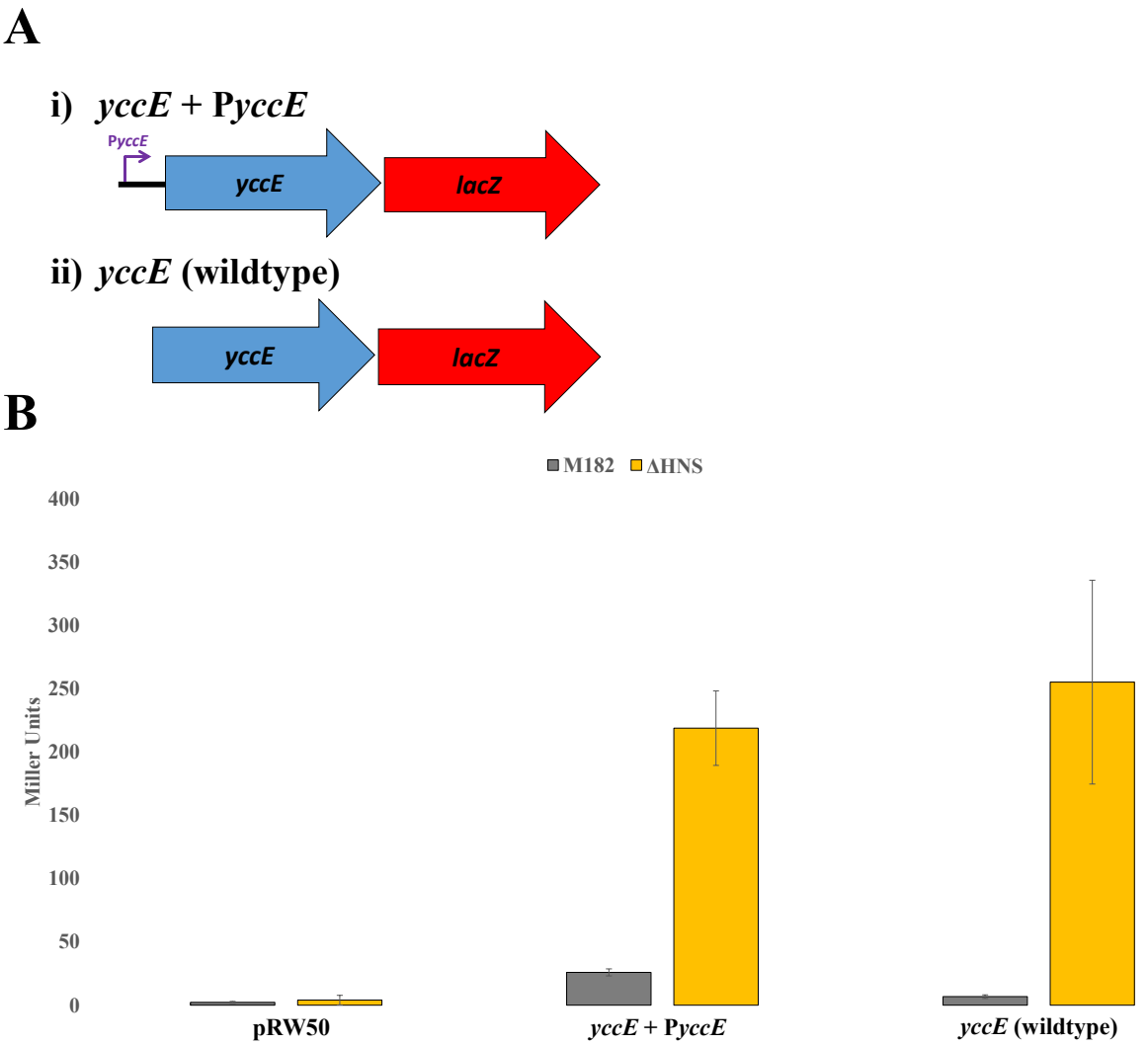


Figure 5.3. Transcription of *yccE* does not require the canonical promoter *in vivo*.

The figure shows schematic diagrams and LacZ assay results of the *yccE* coding region and derivative with the canonical promoter cloned upstream of the gene. Panel (A) shows a schematic diagram. The black lines indicate double stranded DNA, coloured block arrows indicate genes and bent solid arrows indicate promoters. The red block arrows indicate *lacZ* encoded in the pRW50 plasmid. The direction of each arrowhead indicates the direction of expression. Full sequences are

found in the Appendices. Panel (B) shows results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of the *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

Table 5.1. DNA fragments made from the identified promoters within the yccE coding region. Promoter elements (-35 and -10) are shown in blue text. Mutated -10 promoter elements are shown in underlined text. The Shine-Dalgarno sequence is shown in bold text.

Fragment	Type	Sequence	Orientation
M10 *yccE* 1.2	Wildtype	TAATTAT**TTGCAG**TGCAACTAATTCCATG**TATA** **TT**ACTACCCATATATAG**AGGAGG**	Antisense
	Disrupted	TAATTAT**TTGCAG**TGCAACTAATTCCATG**GGTA** **TT**ACTACCCATATATAGAGGAGG	
M10 *yccE* 2	Wildtype	AAAACACAC**TGGTGT**GTTTTATGCGGTTA**TACA** **AT**GAACTCAGACGTAAA**AGGAGG**	Sense
	Disrupted	AAAACACAC**TGGTGT**GTTTTATGCGGTTA**GGCA** **AT**GAACTCAGACGTAAA**AGGAGG**	
M10 *yccE* 3.2	Wildtype	CAATA**TTCTCA**TAAATAATGATAGATCAT**TATC** **AT**TCTGCATTTCCTCAA**AGGAGG**	Antisense
	Disrupted	CAATA**TTCTCA**TAAATAATGATAGATCAT**GGTC** **AT**TCTGCATTTCCTCAA**AGGAGG**	
M10 *yccE* 4	Wildtype	AT**TTGAGG**AAATGCAGAATGATAATGATC**TATC** **AT**TATTTATGAGAATAT**AGGAGG**	Sense
	Disrupted	AT**TTGAGG**AAATGCAGAATGATAATGATC**GGTC** **AT**TATTTATGAGAATAT**AGGAGG**	
M10 *yccE* 5.2	Wildtype	ACCTC**CTGAAA**CCCCTGAATAAATTTGTG**TATC** **AT**TAGTACGCAATATTC**AGGAGG**	Antisense
	Disrupted	ACCTC**CTGAAA**CCCCTGAATAAATTTGTG**GGTC** **AT**TAGTACGCAATATTC**AGGAGG**	
M10 *yccE* 6	Wildtype	TTTA**TTCAGG**GGTTTCAGGAGGTATTACA**TATA** **CT**ATACAATATGTTCGA**AGGAGG**	Sense
	Disrupted	TTTA**TTCAGG**GGTTTCAGGAGGTATTACA**GGTA** **CT**ATACAATATGTTCGA**AGGAGG**	
M10 *yccE* 7.2	Wildtype	CTAACAA**TATCAA**TATCTCGAACATATTG**TATA** **GT**ATATGTAATACCTCC**AGGAGG**	Antisense
	Disrupted	CTAACAA**TATCAA**TATCTCGAACATATTG**GGTA** **GT**ATATGTAATACCTCC**AGGAGG**	
M10 *yccE* 8	Wildtype	AGGGGT**TTCAGG**AGGTATTACATATACTA**TACA** **AT**ATGTTCGAGATATTG**AGGAGG**	Sense
	Disrupted	AGGGGT**TTCAGG**AGGTATTACATATACTA**GGCA** **AT**ATGTTCGAGATATTG**AGGAGG**	
M10 *yccE* 9.2	Wildtype	CTCCATAAAG**TTATAC**CAATAATAACCTT**TAAA** **AT**CTGTGATAGACTCTG**AGGAGG**	Antisense
	Disrupted	CTCCATAAAG**TTATAC**CAATAATAACCTT**GGAA** **AT**CTGTGATAGACTCTG**AGGAGG**	

Table 5.1. DNA fragments made from the identified promoters within the *yccE* coding region (continued)

Fragment	Type	Sequence	Orientation
M10 *yccE* 10	Wildtype	CTTCAGAGT**CTATCA**CAGATTTTAAAGGT**TATT** **AT**TGGTATAACTTTATG**AGGAGG**	Sense
	Disrupted	CTTCAGAGT**CTATCA**CAGATTTTAAAGGT**GGTT** **AT**TGGTATAACTTTATG**AGGAGG**	
M10 *yccE* 11.2	Wildtype	AACATC**ATCACA**CGCATTAATGTTTTCAA**TATA** **CT**CCATAAAGTTATACC**AGGAGG**	Antisense
	Disrupted	AACATC**ATCACA**CGCATTAATGTTTTCAA**GGTA** **CT**CCATAAAGTTATACC**AGGAGG**	
M10 *yccE* 12	Wildtype	AAGGTTATTA**TTGGTA**TAACTTTATGGAG**TATA** **TT**GAAAACATTAATGCG**AGGAGG**	Sense
	Disrupted	AAGGTTATTA**TTGGTA**TAACTTTATGGAG**GGTA** **TT**GAAAACATTAATGCG**AGGAGG**	
M10 *yccE* 13.2	Wildtype	GGCGTA**TTTATC**CGCTCTGGCTGGACACT**TATA** **TT**TTCATCATCAAAACA**AGGAGG**	Antisense
	Disrupted	GGCGTA**TTTATC**CGCTCTGGCTGGACACT**GGTA** **TT**TTCATCATCAAAACA**AGGAGG**	
M10 *yccE* 14	Wildtype	TCAGC**GTGAAA**CTAACCAGGCATTAGGAT**TAAA** **AT**ATGCTCCTGTAGATG**AGGAGG**	Sense
	Disrupted	TCAGC**GTGAAA**CTAACCAGGCATTAGGAT**GGAA** **AT**ATGCTCCTGTAGATG**AGGAGG**	
M10 *yccE* 15	Wildtype	GATGGCGA**TGGATA**TTGTCTGTTAAGAGC**TATA** **CT**GGTTTTAAAACAACA**AGGAGG**	Sense
	Disrupted	GATGGCGA**TGGATA**TTGTCTGTTAAGAGC**GGTA** **CT**GGTTTTAAAACAACA**AGGAGG**	
M10 *yccE* 16	Wildtype	ATGCAAAAGGAAG**TTTACA**ACGAATTCAT**TAAA** **AT**GGTTGATAAAAAAAC**AGGAGG**	Sense
	Disrupted	ATGCAAAAGGAAG**TTTACA**ACGAATTCAT**GGAA** **AT**GGTTGATAAAAAAAC**AGGAGG**	
M10 *yccE* 17.2	Wildtype	ATAACGTC**TTTACA**TCTTCCCTGAGATTA**TAGA** **AT**GCCGTATCAACAAGA**AGGAGG**	Antisense
	Disrupted	ATAACGTC**TTTACA**TCTTCCCTGAGATTA**GGGA** **AT**GCCGTATCAACAAGA**AGGAGG**	
M10 *yccE* 18	Wildtype	CGATCGAGGCTCTTG**TTGATA**CGGCATTC**TATA** **AT**CTCAGGGAAGATGTA**AGGAGG**	Sense
	Disrupted	CGATCGAGGCTCTTG**TTGATA**CGGCATTC**GGTA** **AT**CTCAGGGAAGATGTA**AGGAGG**	

Table 5.1. DNA fragments made from the identified promoters within the *yccE* coding region (continued)

Fragment	Type	Sequence	Orientation
M10 *yccE* 19	Wildtype	AGA**TGTAAA**GACGTTATTTGGCGTTGATC**TACA AT**CTGACAACCAAATTC**AGGAGG**	Sense
	Disrupted	AGA**TGTAAA**GACGTTATTTGGCGTTGATC**GGCA AT**CTGACAACCAAATTC**AGGAGG**	
M10 *yccE* 20	Wildtype	GAAAAA**TGTATC**CTGCATTTACCCGAGTT**TATA TT**TAATGATAACAAGAA**AGGAGG**	Sense
	Disrupted	GAAAAA**TGTATC**CTGCATTTACCCGAGTT**GGTA TT**TAATGATAACAAGAA**AGGAGG**	
M10 *yccE* 21.2	Wildtype	GTCGGTA**TCTAAA**GCAAGCAAGTTCTTGT**TATC AT**TAAATATAAACTCGG**AGGAGG**	Antisense
	Disrupted	GTCGGTA**TCTAAA**GCAAGCAAGTTCTTGT**GGTC AT**TAAATATAAACTCGG**AGGAGG**	

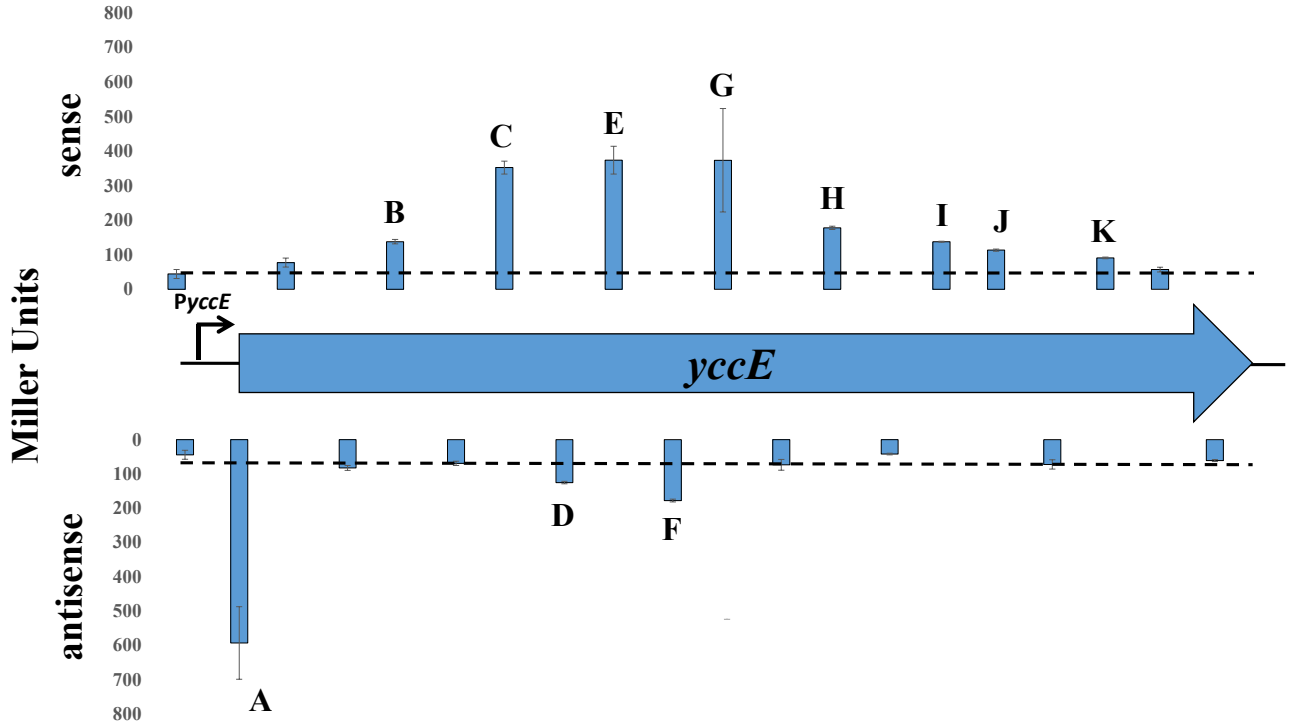


Figure 5.4. Identification of promoters embedded within the coding sequence of *yccE*.

Results of LacZ assays using JCB387 transformed with pRW50 containing the identified promoters within the *yccE* gene sequence. The upper graph labelled "sense" shows the activities of promoters in the sense direction, while the lower graph labelled "antisense" shows the activities of promoters in the antisense direction. Data points labelled "a" – "k" are promoters that were shown to have at least 2-fold activity above background levels. In addition, mutant versions of these promoters were generated in the next experiments that will be described later on. The canonical promoter, P*yccE*, is shown upstream of *yccE*. The *yccE* gene is shown

in blue block arrows. Background activity generated by the vector only construct (pRW50) is shown by the dashed lines, first upper bar. Activity data are reported as means ± standard deviations of the mean of three replicates.

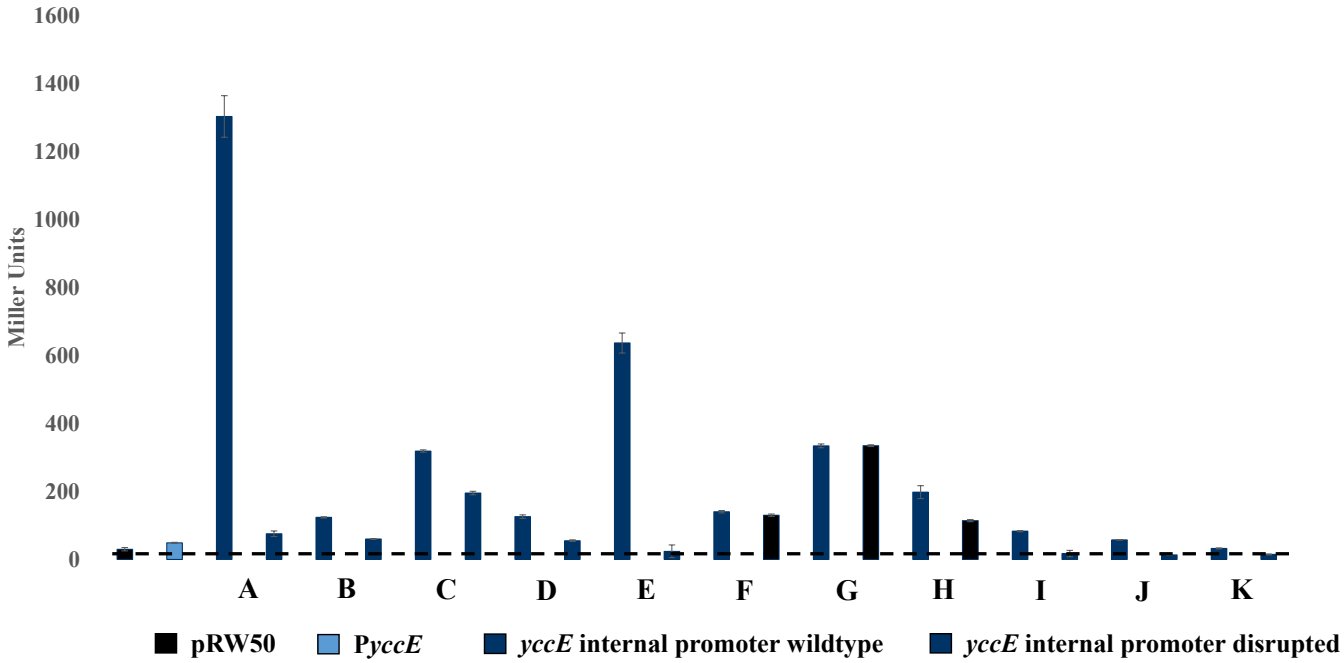


Figure 5.5. Inactivation of individual internal promoters of *yccE* with by point mutation.

Results of LacZ assays using JCB387 transformed with pRW50 containing the identified intragenic promoters of *yccE* along with their corresponding disrupted counterparts. Data points labelled "a" – "k" are the promoters identified in the previous experiment to have at least 2-fold activity above background levels (Figure 5.2). Activity of the canonical promoter, P*yccE*, is shown in blue bars. Wildtype intragenic promoters are shown in dark blue bars and disrupted intragenic promoters are shown in grey bars. Background activity generated by the vector only construct (pRW50) is shown by the black bar and dashed lines. Activity data are reported as means ± standard deviations of the mean of three replicates.

were mutated to G. The DNA fragments were again cloned upstream of *lacZ* in pRW50, and transformants were tested for LacZ activity. Nine out of the 11 mutated DNA fragments stimulated less transcription in comparison to the wildtype version of the DNA fragment (Figure 5.5).

5.5. The observed increase of *yccE* transcription in cells lacking H-NS requires internal promoters

To further confirm that transcription initiates within the coding region of *yccE*, a series of *yccE* derivatives, carrying point mutations in internal promoters, was generated. Hence, the 11 most active promoters identified by the stringent search criteria were mutated. I also relaxed the search criteria to identify, and subsequently mutate, a further 22 potential promoters within *yccE*. This relaxed search criteria identified all sequences matching the motif 5'-TAnnnT-3'. I also cloned *yccE* in the reverse orientation and mutated the only "antisense" promoter with substantial activity (i.e. P6). The different constructs are illustrated under the bar graph in Figure 5.6. Mutating the 11 promoters identified by the stringent search reduced but did not abolish intragenic transcription (Figure 5.6), suggesting that the stringent bioinformatics search had failed to identify all internal *yccE* promoters. However, mutating all internal promoters identified using the relaxed search criteria greatly reduced *lacZ* expression induced in the absence of H-NS (Figure 5.6). When *yccE* was in the reverse orientation, mutation of P6 promoter (i.e. the promoter in DNA fragment "A" in Figure 5.4) led to a large decrease of transcription in cells lacking H-NS (Figure 5.6). Taken together, examination of the *yccE* locus using biochemical and genetic techniques reveals an

important role of H-NS in silencing intragenic transcription initiation. This phenomenon will from here onwards be referred to as "pseudo-regulation". Interestingly, the P6 promoter is only partially silenced by H-NS. This may explain the requirement for Fis binding to the adjacent *cbpA* regulatory DNA.

5.6. Transcription initiation within *yccE* is terminated by Rho

The ATP-dependent translocase Rho specifically targets untranslated RNAs for premature transcription termination (Cardinale *et al.*, 2008, Peters *et al.*, 2009). Previous studies have shown that Rho often targets transcripts that are intragenic in origin since they contain no translation initiation signal. I expected that transcription initiating within *yccE* should be sensitive to Rho, but that this effect should be abolished when internal promoters were mutated. In order to investigate the effects of Rho, I used the same DNA fragments containing the *yccE* gene, and derivatives without internal promoters, cloned upstream of *lacZ* in pRW50. The plasmids were again used to transform M182, and the derivative lacking H-NS. Transformants were then treated with sublethal concentrations of bicyclomycin (BCM), a specific Rho inhibitor, during growth to an OD_{650} of 0.4-0.6 in liquid culture. Measurements of LacZ activity were then taken. The results of the experiment are shown in Figure 5.7. Upon the addition of BCM, a very small increase in *lacZ* expression levels is observed with the wildtype *yccE* coding region in sense orientation, both with and without H-NS. As expected, there was no effect of BCM using the *yccE* coding region lacking intragenic -10 elements. Similar observations were made for the construct containing

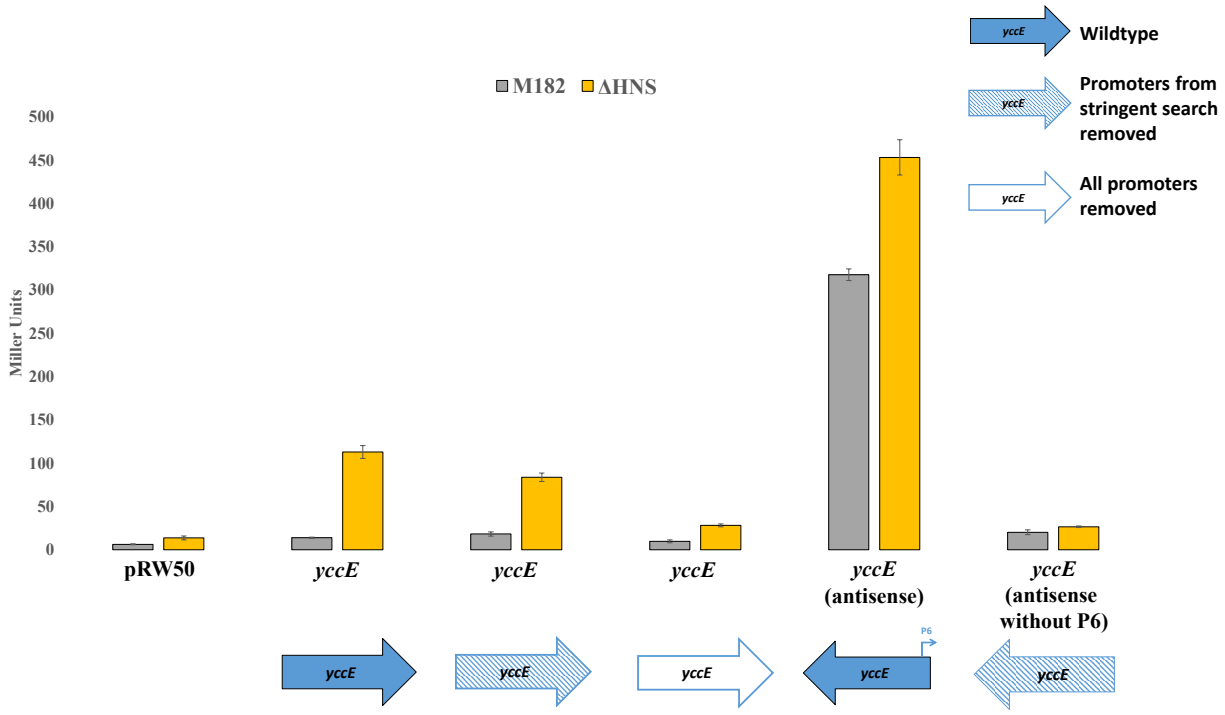


Figure 5.6. Inactivation of *yccE* internal promoters by point mutation.

Results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50 containing the *yccE* coding region and derivatives containing mutated intragenic promoters, and the *yccE* coding region in reverse orientation and derivative containing mutated *cbpA* P6 promoter. The mutated intragenic promoters were identified using a stringent search criteria (5'-TAnAAT-3', 5'-TATnAT-3', or 5'-TATAnT-3') or a relaxed search criteria (5'-TAnnnT-3'). Schematic diagrams of different constructs are shown under the bar graph. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of the *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

yccE in the antisense orientation but addition of BCM led to a more substantial increase in *lacZ* expression in both wildtype cells and cells lacking H-NS. Again, little effect was seen when intragenic promoters were mutated. Taken together, these results suggest that Rho helps to terminate spurious transcription which initiates from within the *yccE* coding region.

5.7. Many AT-rich genes are subject to pseudo-regulation by H-NS *in vivo*

The bioinformatic analysis in Figure 5.2 suggests that pseudo-regulation by H-NS may be common; many AT-rich genes contain internal promoters. Hence, I next focused on *yfdF*, *ykgH*, *yjgN*, and *yjgL*. Recall that these genes are all AT-rich, bound by H-NS, and enriched for predicted promoters (Figure 5.2). These genes, with an AT-content between 63% and 68%, were cloned upstream of *lacZ* in pRW50. Derivatives of each gene, with point mutations in all predicted intragenic -10 elements, were also cloned. Additionally, the *fepE* locus was cloned as a control. The *fepE* locus has an AT-content of 55.03%, and is not expected to bind H-NS or contain internal promoters. For all constructs, expression of *lacZ* was measured in M182 or M182*Δhns*. Assays were also done in the presence of bicyclomycin (0 or 10 μg/ml BCM) to check for termination of intragenic transcription by Rho. Results of these assays are shown in Figure 5.8. As expected, the *fepE* locus was unable to promote *lacZ* expression in the presence or absence of H-NS or BCM. Conversely, all of the AT-rich genes (*yfdF*, *ykgH*, *yjgN*, and *yjgL*) stimulated expression of *lacZ* when H-NS was absent. Furthermore, addition of bicyclomycin caused increased expression of *lacZ* in these cells.

5.8. Many AT-rich genes are subject to pseudo-regulation by H-NS *in vitro*

Finally, to confirm observation from genetic experiments, *in vitro* transcription assays were done using the different genes as templates. The results of the assays are shown in Figure 5.9. Consistent with previous observations, RNA polymerase failed to initiate transcription within the coding sequence of *fepE* (Lanes 1 and 2). However, for all of the other alleles, intragenic transcriptions were observed (*yfdF*, *ykgH*, *yjgN*, *yjgL*, and *yccE*) (Lanes 3, 5, 7, 9, and 11). Furthermore, mutation of the intragenic promoters resulted in a significant reduction of transcription within these genes (Lanes 4, 6, 8, 10, and 12).

5.9. Discussion

This chapter demonstrates the phenomenon of "pseudo-regulation", whereby H-NS-targeted genes retain the illusion of H-NS control in the absence of their canonical promoters. This occurs because intragenic promoters occur frequently within H-NS targeted AT-rich genes. Thus, increased transcription of *yccE* in Δ*hns* cells does not require the presence of the genuine *yccE* promoter. Furthermore, *yfdF*, *ykgH*, *yjgN*, and *yjgL* behave the same way. Reassuringly, my observations of these effects are the same using genetic and biochemical approaches (Figures 5.8 and 5.9). The observation that Rho terminates transcripts originating within H-NS targeted genes supports the previous proposition that Rho and H-NS act cooperatively in order to silence transcription at the genomic loci (Peters *et al.*, 2012, Kotlajich *et al.*, 2015). In particular, Kotlajich and colleagues have investigated the mechanism for this cooperative interaction (Kotlajich *et al.*,

2015). These authors suggest that H-NS directly inhibits transcription elongation, and as such, increase the time available for the effective action of Rho. That a Rho effect can be observed is also consistent with the notion that intragenic transcripts are non-coding.

In summary, it is likely that "pseudo-regulation" of genes by H-NS is a widespread phenomenon in bacteria; the AT-rich nature of promoters is highly conserved. Recently, it was shown that intragenic promoters are the source of the H-NS regulated transcription of the *pilE* locus in *Neisseria gonorrhoeae* (Masters *et al.*, 2015). Therefore, it is important to caution against the use of genome-scale data alone in order to identify true gene regulation by H-NS.

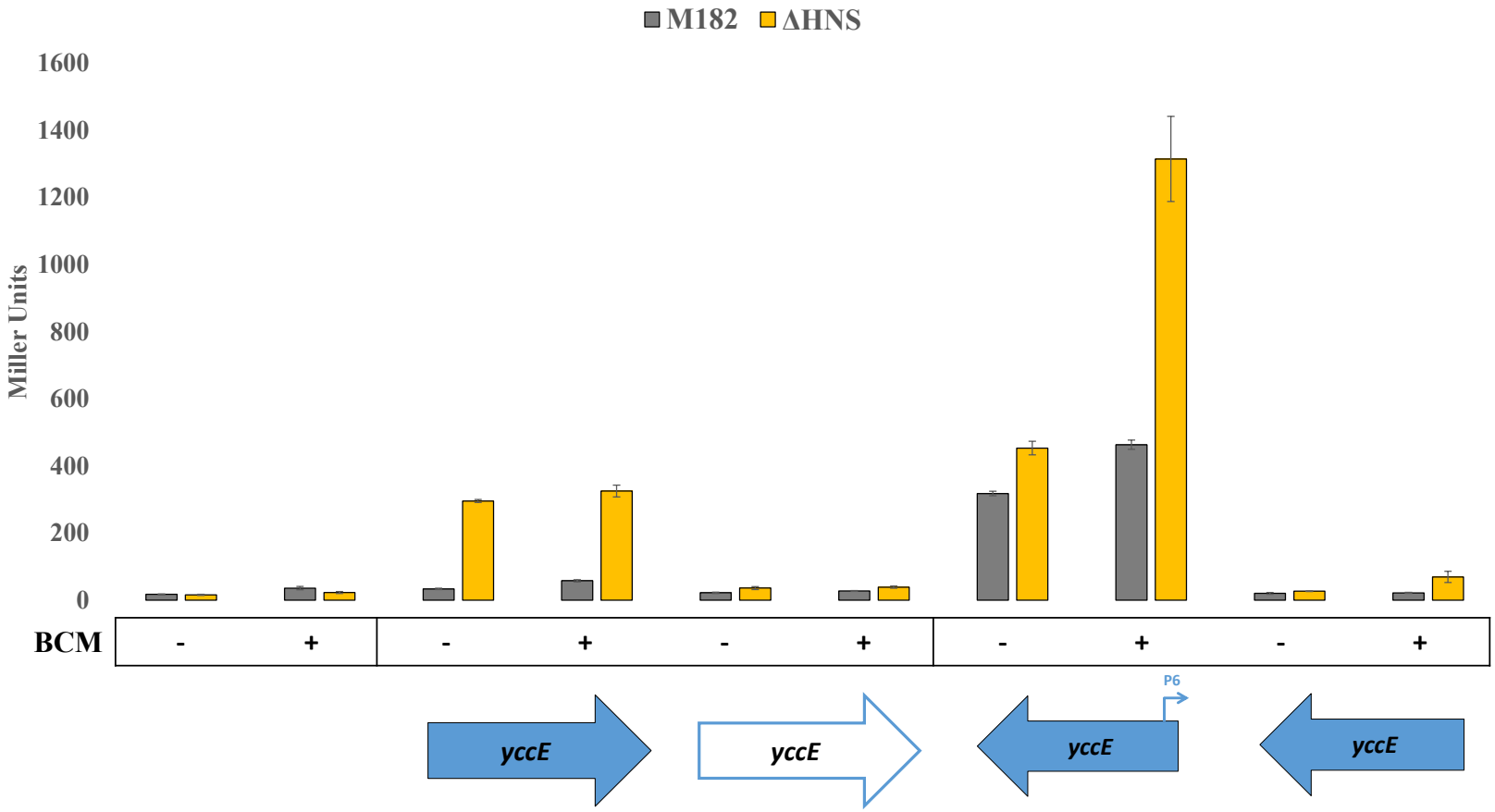


Figure 5.7. Transcription initiation within *yccE* is repressed by Rho.

Results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50 containing *yccE* gene variations. Cells carrying these constructs were treated with 0 (-) or 10 (+) μg/ml bicyclomycin (BCM) before these cells were allowed to grow to an OD_{650} of 0.4-0.6, after which the usual β-galactosidase assay protocol follows. The empty pRW50 was used as a control. Assays

were done using overnight cultures. Activities of the *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

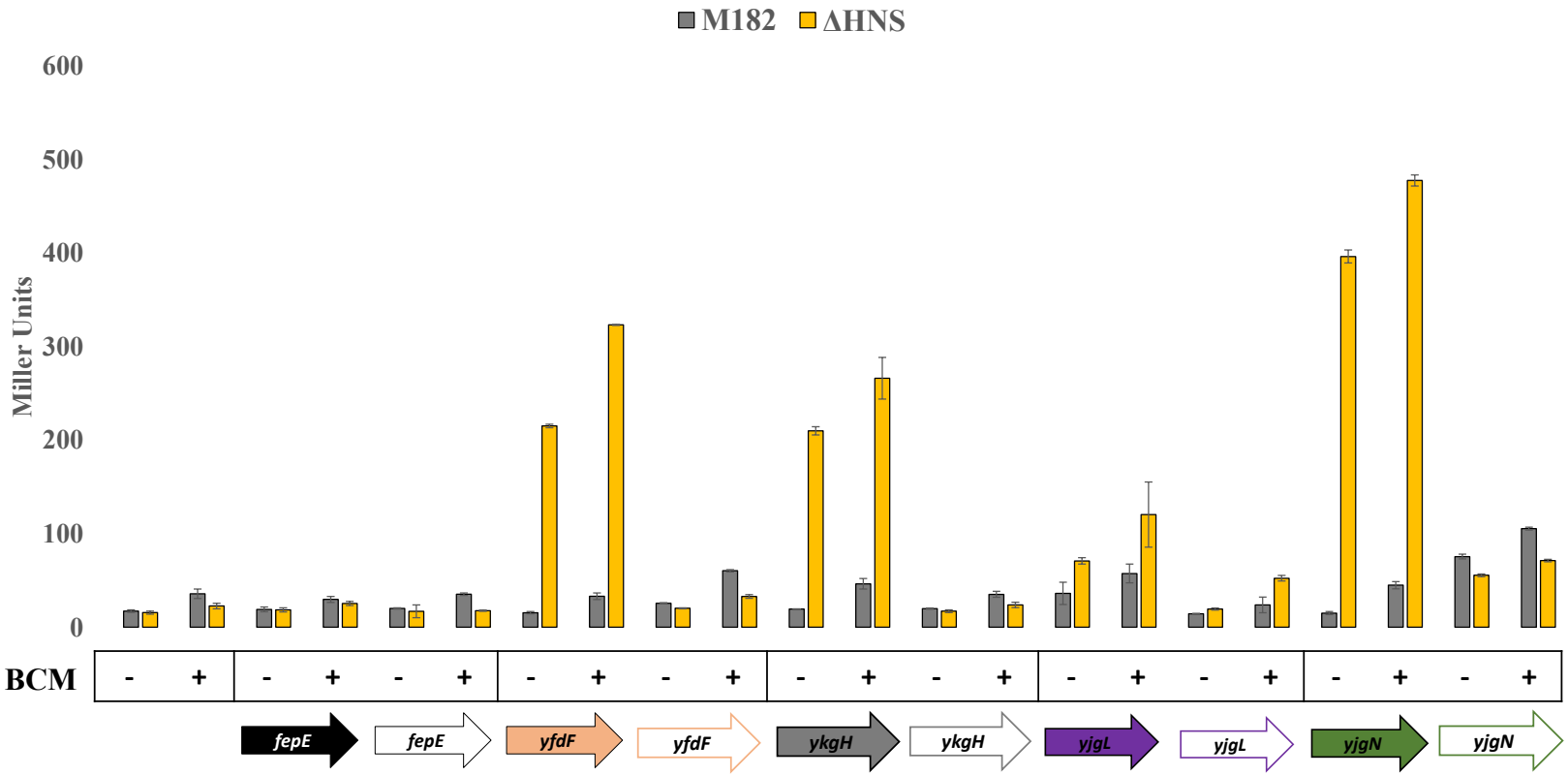


Figure 5.8. Intragenic transcription occurs in other genes targeted by H-NS *in vivo*.

Results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50 containing the AT-rich gene variations. Cells carrying these constructs were treated with 0 (-) or 10 (+) μg/ml bicyclomycin (BCM) before these cells were allowed to grow to an OD_{650} of 0.4-0.6, after which the usual β-galactosidase assay protocol follows. The empty pRW50 was used as a control. Assays

were done using overnight cultures. Activities of the *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

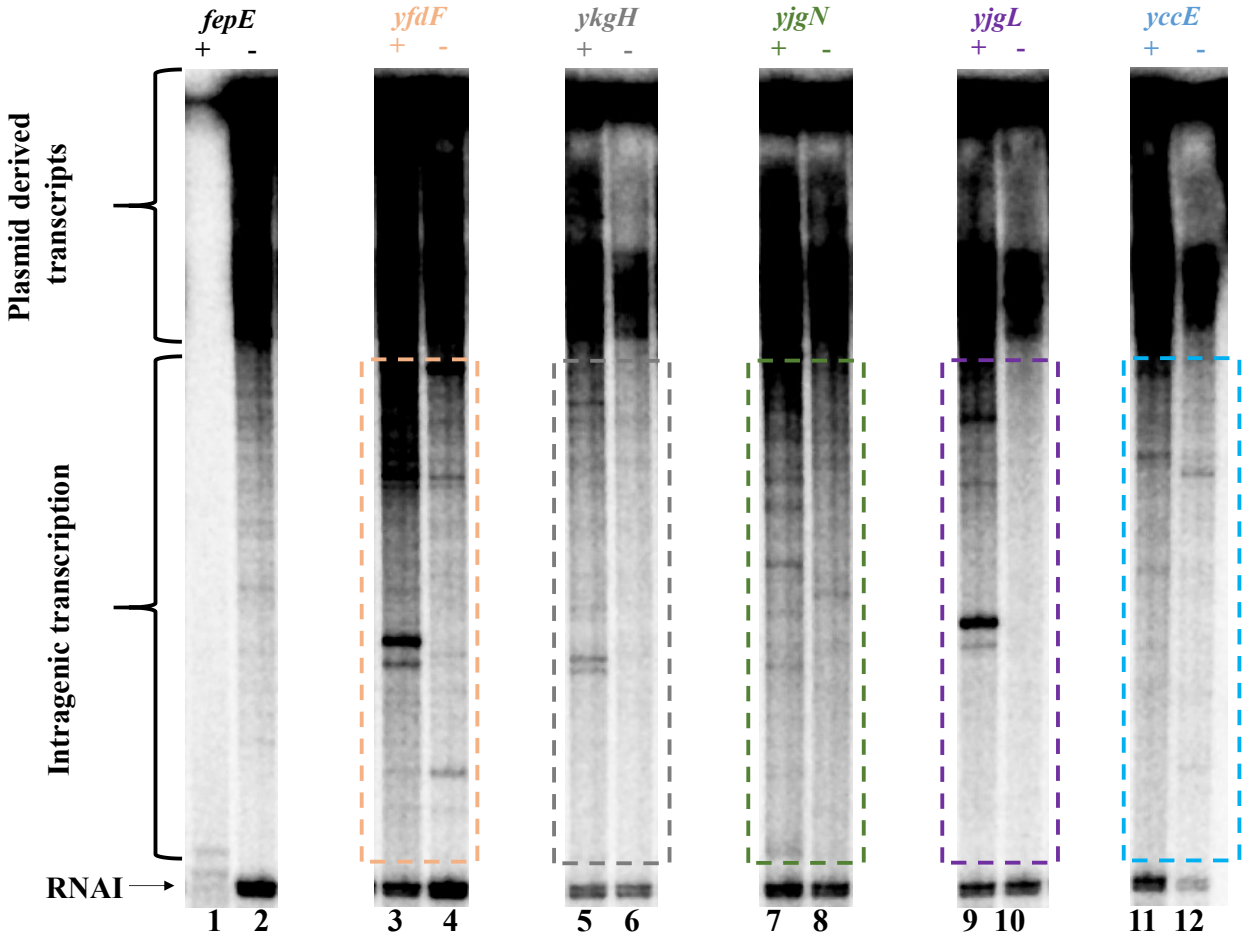


Figure 5.9. Intragenic transcription occurs in other genes targeted by H-NS *in vitro*.

The figure shows mRNA transcripts generated from plasmid pSR or derivatives carrying the intragenic mRNA transcripts from AT-rich genes targeted by H-NS. These genes have an AT-content of 64.5% (*yfdF*, shown in light orange), 63.08% (*ykgH,* shown in grey), 62.90% (*yjgN*, shown in dark green), 67.82% (*yjgL,* shown in purple), and 65.87% (*yccE*, shown in blue). Each of these genes have two cloned derivatives—the wildtype (+) and the variation with all internal promoters mutated (-). Note that *fepE* (shown in black), which has an AT-content of 55%, is used as a control. Transcripts were generated by σ^{70}-associated RNA polymerase (400 nM). RNAI, indicated by a black arrow, is a plasmid encoded transcript which serves as an assay control.

Monitoring transit of RNA polymerase across the H-NS-bound *yccE*

Chapter 6

6.1. Introduction

The previous chapter showed that multiple promoters are embedded within the coding sequence of AT-rich genes such as *yccE*. Therefore, H-NS plays a major role in bacteria by preventing intragenic transcription initiation. However, these observations have implications for our understanding of H-NS as a regulator of mRNA transcription. For example, is increased transcription of a gene in the absence of H-NS a consequence of intragenic promoters alone or is mRNA production also affected? Previous DNA microarray based experiments lacked the resolution to differentiate between spurious intragenic RNAs and mRNAs. This problem is also likely to affect recent RNA-seq data unless individual promoters are mapped. Hence, many studies conclude increased transcription in cells lacking H-NS is due to mRNA production. My data argues that this may frequently be an incorrect, or minimally an incomplete, explanation. Thus, a pertinent question is to understand if a promoter upstream of an H-NS bound gene can drive transcription of that gene, or if H-NS binding blocks transit of RNAP. Intriguingly, a recent biochemical study re-examined the role of H-NS in blocking transcription elongation during mRNA synthesis (Kotlajich *et al.*, 2015). This work concludes that the effect of H-NS on elongation can be negligible when H-NS forms filaments with DNA. Consistently, a large effect is apparent in conditions when H-NS forms DNA bridges.

Importantly, the AT-rich genes lacking intragenic promoters described in the previous chapter provide a useful tool for investigating the effects of H-NS on mRNA synthesis without the confounding influence of intragenic promoters. Hence, this chapter aims to determine whether RNA polymerase can elongate across the promoterless, but H-NS bound, derivative of *yccE*.

6.2. H-NS binds to *yccE* lacking intragenic promoters

In the previous chapter I described a derivative of the *yccE* gene where internal promoters had been inactivated by point mutations. The goal here was to understand if this gene derivative could bind H-NS and block transit of RNA polymerase across the transcription unit. As a first step, I aimed to determine if H-NS is still able to bind the mutated *yccE* allele. Hence, DNA fragments containing the wildtype *yccE* coding region, or the derivative with mutated intragenic promoter -10 elements, were generated. These DNA fragments were then tested for H-NS binding both *in vitro* and *in vivo*. To determine the binding affinity of H-NS *in vitro*, an electrophoretic mobility shift assay (EMSA) was carried out using increasing concentrations of H-NS. The result of this assay is shown in Figure 6.1; *yccE* and its mutated derivative showed almost identical patterns of H-NS binding. Thus, mutation of *yccE* to remove intragenic promoters has no effect on H-NS binding *in vitro*. To confirm this observation *in vivo*, a chromatin immunoprecipitation (ChIP) assay, coupled with PCR, was applied to the same two *yccE* sequences. In this assay, each *yccE* DNA fragment was cloned upstream of *lacZ* in the plasmid, pRW50. These plasmids were used to separately transform *E. coli* strain JCB387. Cells containing either plasmid were then grown to mid-log phase, fixed with formaldehyde, and harvested. After fragmenting the DNA by sonication, immunoprecipitations were done using an anti-H-NS antibody. In parallel, mock immunoprecipitations (i.e. with no antibody) were done as a control. Following immunoprecipitation, oligonucleotides were used to specifically amplify either the pRW50 encoded *yccE* alleles (i.e. the chromosomal *yccE* was not detected) or the control *yabN* locus that is not bound by H-NS. Samples were taken at different stages of PCR amplification (i.e. after 25,

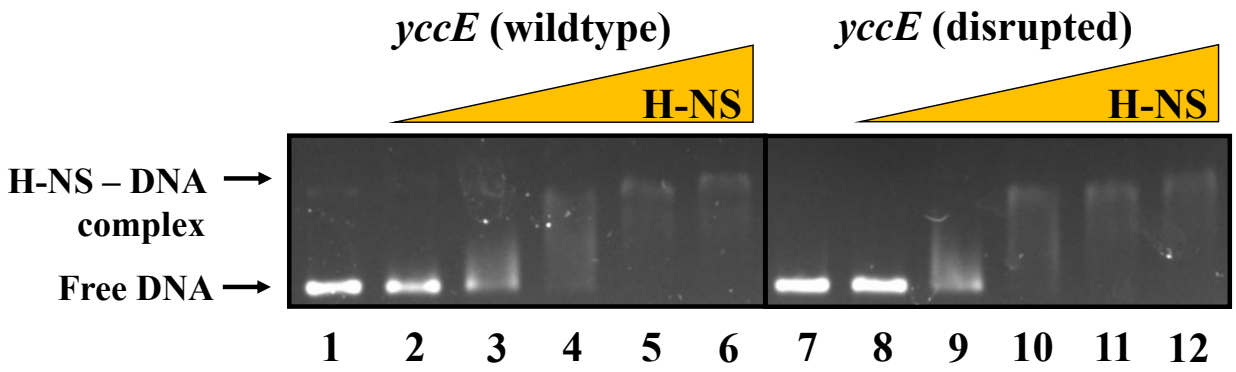


Figure 6.1. *In vitro* analysis of H-NS binding to the *yccE* coding region and its derivative containing mutations in the identified intragenic -10 promoter elements.

Results of *in vitro* binding of purified H-NS to DNA fragments carrying the *yccE* coding region in wildtype (Lanes 1-6) and mutated intragenic -10 promoter elements (Lanes 7-12). Reactions were incubated for 20 minutes at 37°C, after which they were run on an agarose gel (1% w/v) stained with ethidium bromide (1% w/v). The gel was run at 50 V for 1 hour before visualisation. Purified H-NS was added in the following concentrations: 0 µM (Lanes 1 and 7), 2.89 µM (Lanes 2 and 8), 3.21 µM (Lanes 3 and 9), 8.67 µM (Lanes 4 and 10), 11.57 µM (Lanes 5 and 11), and 14.25 µM (Lanes 6 and 12). Free DNA and H-NS-DNA complex are indicated by solid black arrows.

28, 31, or 34 cycles of PCR) and analysed by agarose gel electrophoresis to check amplification efficiency. Results of the experiment are shown in Figure 6.2. As expected, the wildtype *yccE* gene was amplified more efficiently than *yabN* when DNA was immunoprecipitated using anti-H-NS (Figure 6.2A, Lanes 1-4) but not when DNA was obtained from mock immunoprecipitations (Figure 6.2A, Lanes 5-8). Similar results were obtained for the *yccE* coding region with mutations in intragenic promoters (Figure 6.2B). Consistent with the *in vitro* experiments, this further confirmed H-NS binding to the mutated *yccE* gene.

6.3. RNA polymerase can elongate across the H-NS-bound *yccE* allele

Binding of H-NS to *yccE* is unaffected by mutations in intragenic promoters (Figures 6.1 and 6.2). Hence, this *yccE* allele is a good tool to investigate the ability of RNA polymerase to transcribe through H-NS bound DNA (i.e. H-NS binds but there is no possibility of intragenic promoters having confounding effects). Recall that, in Chapter 4, it was shown that the canonical *yccE* promoter is poorly active, and as such, not ideal for these experiments. Instead, two unrelated promoters were cloned upstream of the mutated *yccE*. Hence, the factor independent *cbpA* P6 promoter, or a derivative of the *melAB* promoter (called JK22) (Grainger *et al.*, 2004) were used. These promoters were cloned, upstream of *yccE* and *lacZ*, as illustrated in Figure 6.3. Note that *yccE* was cloned in either the forward or reverse orientation with respect to *lacZ* and the promoter.

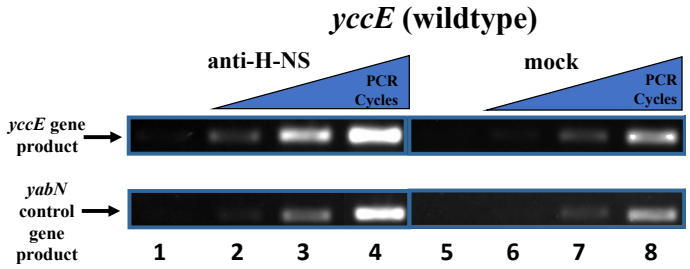
Results of β-galactosidase assays with *yccE* cloned in the forward orientation are shown in Figure 6.4. High levels of *lacZ* expression were observed in wildtype M182 cells (i.e. in the presence of H-NS). This expression was reduced dramatically in cells lacking H-NS. Whilst the reduced

transcription in cells lacking H-NS was surprising these data suggest that H-NS does not prevent transcription of the *yccE* mRNA from an ectopic promoter. Remarkably, the opposite result was observed when *yccE* was cloned in the reverse orientation (Figure 6.5). Here, *lacZ* expression was increased in the cells lacking H-NS compared to the wildtype M182 cells. Hence, repression of transcription by H-NS is only apparent when *yccE* is cloned in the reverse orientation.

6.4. Discussion

This chapter investigates the ability of RNA polymerase to transcribe across an H-NS bound tract of DNA. When *yccE* was cloned in the forward orientation, RNA polymerase was able to transcribe the *yccE* DNA in the presence of H-NS (Figure 6.4). Hence, binding of H-NS to *yccE* does not pose a barrier to mRNA synthesis under the conditions of the experiment. Remarkably, the opposite is true when *yccE* is in the reverse orientation; transcription increased in the absence of H-NS (Figure 6.5). Hence, the orientation of *yccE* is critical. This may be a consequence of translation being coupled to transcription only when *yccE* is in the forward orientation. In this case, transcribing RNA polymerase is followed closely by a ribosome which stimulates forward motion of RNA polymerase and prevents termination by Rho (Proshkin *et al.*, 2010). Conversely, when cloned in the reverse orientation, *yccE* is not translated and a long tract of naked RNA is produced. The presence of untranslated RNA enables binding of the termination factor Rho. Thus, in recent studies, it was shown that Rho and H-NS function together to terminate transcription because H-NS reduces the rate of elongation (Peters *et al.*, 2012, Kotlajich *et al.*, 2015).

A



B

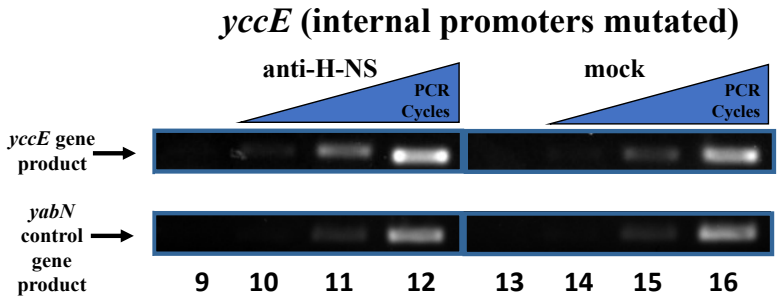


Figure 6.2. ChIP-PCR analysis of the association between H-NS and the *yccE* coding region derivatives.

Results of the PCR amplification of the wildtype *yccE* coding region (Panel A, Lanes 1 – 8) and its disrupted derivative containing mutations in the intragenic -10 promoter elements (Panel B, Lanes 9 – 16) from the anti-H-NS (Panel A, Lanes 1 – 4 and Panel B, 9 – 12) and mock (Panel A, Lanes 5 – 8 and Panel B, 13 – 16) immunoprecipitates. Upper panel: amplified products using the oligonucleotide set specific to the pRW50 plasmid and the *yccE* coding region derivatives. Lower panel: amplified products using the control oligonucleotide set specific to *yabN*. The blue coloured

block triangles represent the increasing number of PCR cycles used to generate the amplification products.

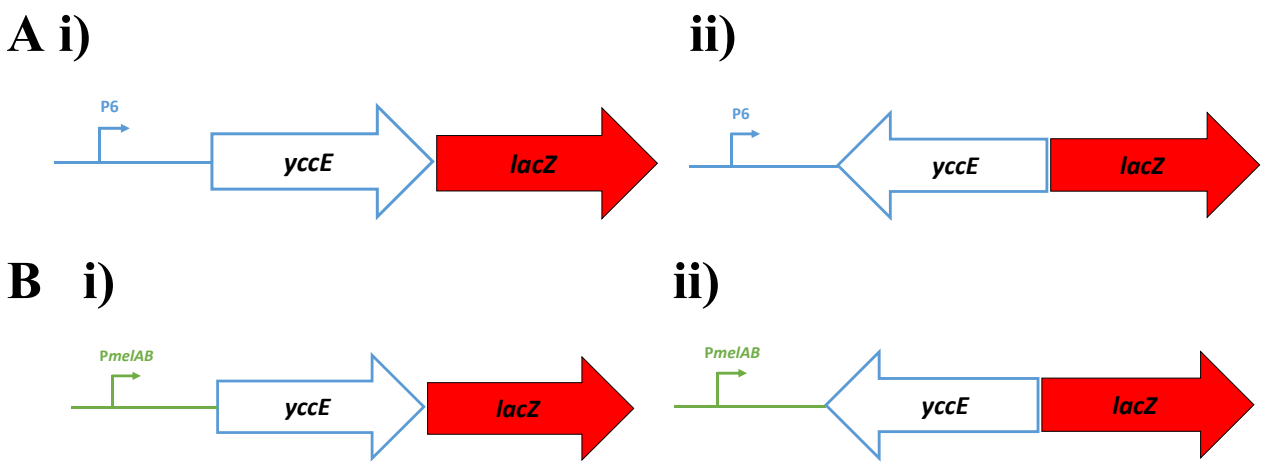


Figure 6.3. Schematic diagrams of different strong promoters cloned upstream of the *yccE* coding region containing the mutated intragenic -10 promoter elements.

The figure shows schematic diagrams and corresponding sequences of the promoters (A) P6 (shown in blue) and (B) P*melAB* (shown in green) cloned upstream of the *yccE* coding region containing the mutated -10 intragenic -10 promoter elements. The red block arrows indicate the *lacZ* gene. Coloured bent arrows indicate promoters. The -35 and -10 elements are shown in coloured text. The Shine-Dalgarno sequence is shown in black bold text. The direction of the arrows indicate the direction of gene expression.

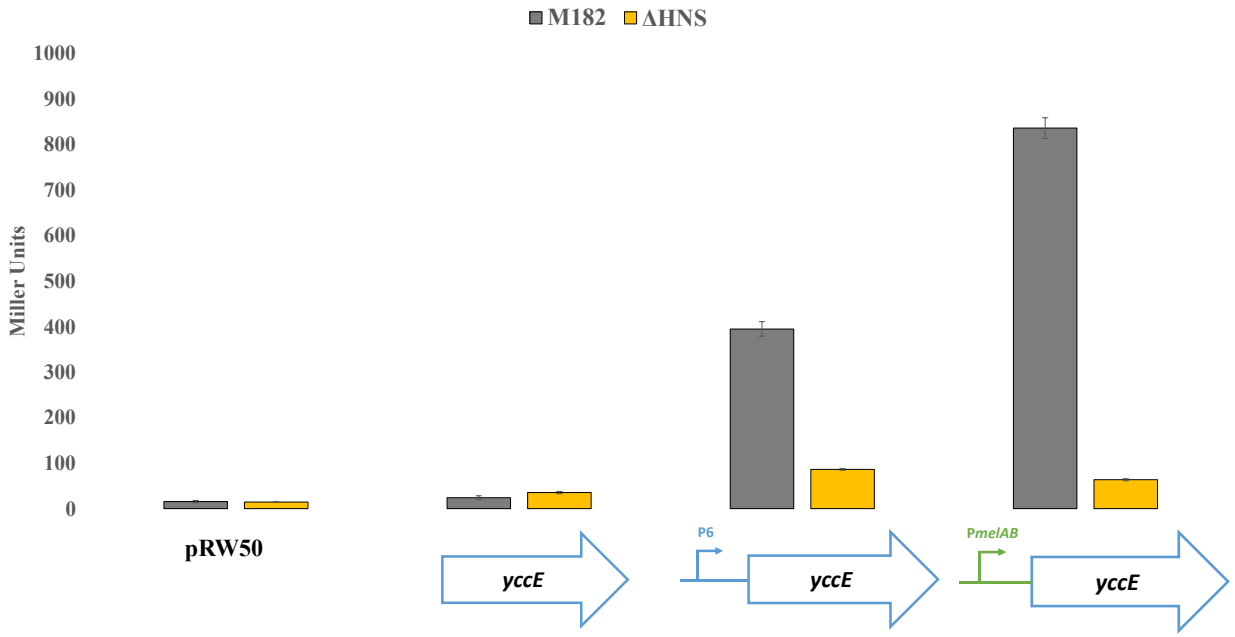


Figure 6.4. RNA polymerase reads through the H-NS filament bound at the *yccE* coding region containing mutations at the intragenic -10 promoter elements.

Results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50 containing different strong promoters cloned upstream of the disrupted derivative of the *yccE* coding region in sense orientation containing mutations at the intragenic -10 promoter elements. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative strain are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

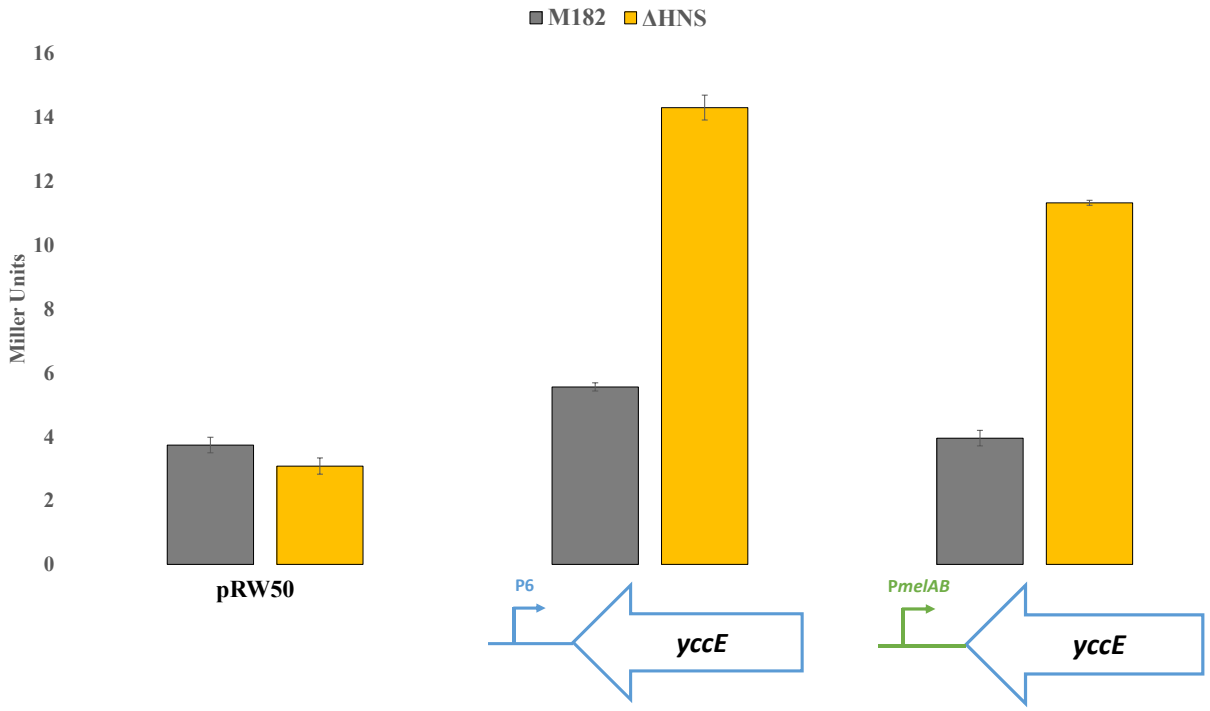


Figure 6.5. Repression by H-NS at the *yccE* coding region in the antisense orientation containing mutations at the intragenic -10 promoter elements.

Results of LacZ assays using M182, or the Δ*hns* derivative, transformed with pRW50 containing different strong promoters cloned upstream of the disrupted derivative of the *yccE* coding region in antisense orientation containing mutations at the intragenic -10 promoter elements. The empty pRW50 was used as a control. Assays were done using overnight cultures. Activities of *E. coli* strain M182 are shown in grey bars. Activities of the H-NS mutant derivative strain are shown in gold bars. Activity data are reported as means ± standard deviations of the mean of three replicates.

Final Conclusions

Chapter 7

The regulation of gene expression is mostly controlled at the level of transcription. This process allows the cell to respond to various environmental cues, ranging from nutrient availability to changes in temperature. The transcriptional process is divided into three steps: initiation, elongation, and termination. This study focuses on the first step, which is transcription initiation. Transcription initiation is dependent on two key factors: the strength of a promoter, and the presence and absence of transcription factors. The bacterial RNA polymerase binds specifically to the template DNA when one of its subunits, the σ factor, recognises sequence motifs called promoters. In addition to promoter recognition by RNA polymerase, transcription can be either activated or repressed by the presence (i.e. binding) and absence (i.e. not binding) of transcription factors to sequences upstream of the promoter.

A central presumption is that transcription initiation sites are located mainly at the 5' end of genes. However, as illustrated in this thesis, there are instances in which this particular paradigm is incorrect. Thus, this work provides alternative explanations for longstanding observations. Firstly, the nucleoid-associated protein CbpA, which is normally expressed when the cell is placed under conditions of stress, was previously found to be under the control of the σ^{38}-dependent P2 promoter. However, further dissection of this locus has shown that *cbpA* can also be transcribed from a strong σ^{70}-dependent promoter (P6) found further upstream within the coding region of the divergent gene, *yccE*. The *cbpA* gene is protected from P6 by binding of Fis (factor for inversion stimulation) to a specific sequence upstream of the *cbpA* gene. This study provides an additional level of control; the weak but convergent σ^{32}-dependent promoter for *yccE* (P*yccE*) can also interfere with P6 activity. Secondly, horizontally acquired genes, such as *yccE*, are transcriptionally silenced by another nucleoid-associated protein, H-NS (histone-like nucleoid structuring protein).

Canonical regulation of an AT-rich gene dictates that H-NS binds and inhibits activity of a genuine promoter of mRNA production. However, this study argues otherwise. Instead, H-NS frequently silences intragenic promoters. This type of regulation, termed in this study as “pseudo-regulation”, points towards intragenic promoters as the source of transcription at many loci in cells lacking H-NS. Spurious intragenic transcription likely depletes the cells resources, and is therefore implicated in the toxicity of horizontally-acquired genes.

Appendices

Figure A.1. Sequence of the *cbpA* regulatory region.

This figure shows the sequence of the *cbpA* regulatory region used in this study. Promoter elements (-35 and -10), transcription start sites, and start codons are written in coloured texts. Nucleotides where site specific mutations are introduced are written in bold black texts. The Fis binding element is underlined.

Figure A.2. Alignment of the *cbpA* regulatory region and its Fis binding site derivatives.

Alignment of the *cbpA* regulatory region and its Fis binding site derivatives. The cbpA wildtype contains the identified Fis binding sites that have been mutated in the other derivatives. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space.

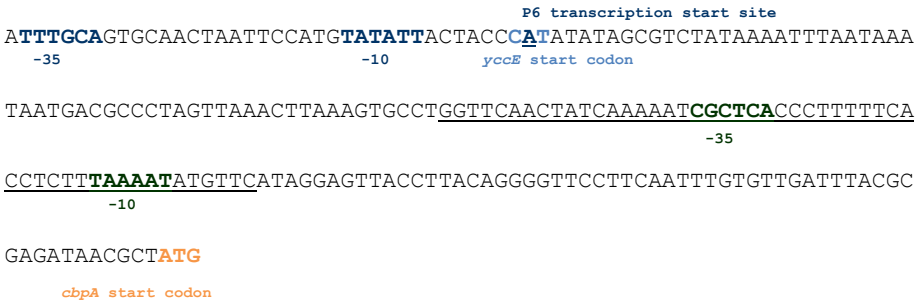


Figure A.3. Sequence of the cbpAΔ203.1 DNA fragment

This figure shows the sequence of the *cbpA* regulatory region used in this study. Promoter elements (-35 and -10), transcription start sites, and start codons are written in coloured texts. The Fis binding element is underlined.

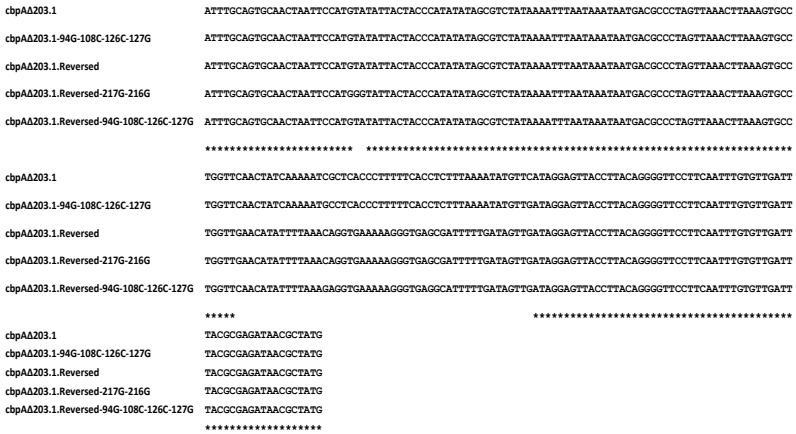


Figure A.4. Alignment of the cbpAΔ203.1 DNA fragment

Alignment of the cbpAΔ203.1 DNA fragment and its derivatives. The cbpAΔ203.1 contains the identified Fis binding sites that have been mutated as well as the Fis binding element that has been shown in reverse orientation in the other derivatives. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space.

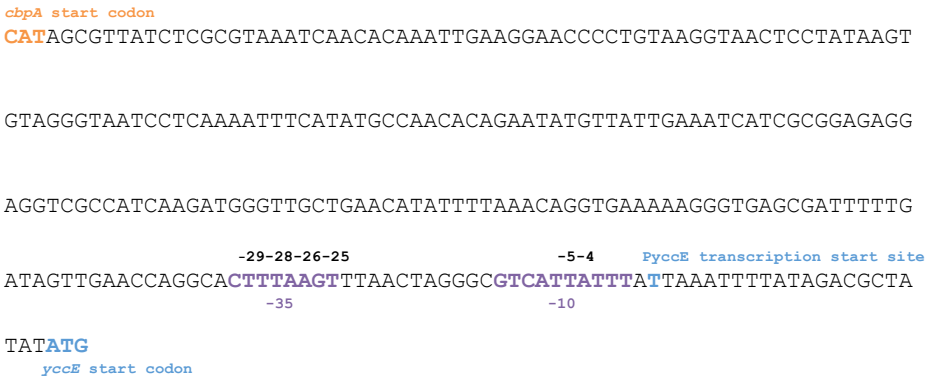


Figure A.5. Sequence of the *yccE* regulatory region.

This figure shows the sequence of the *yccE* regulatory region used in this study. Promoter elements (-35 and -10), transcription start sites, and start codons are written in coloured texts. Nucleotides where site specific mutations are introduced are written in bold black texts.

```
Δ200                        GCACTTTAAGTTTAACTAGGGCGTCATTATTTATTAAATTTTATAGACGCTATATATG
Δ200-29G-28G-26G-25G-5G-4G  GCACGGTGGGTTTAACTAGGGCGTCATGGTTTATTAAATTTTATAGACGCTATATATG
                            ****  *  ******************  ****************************
```

Figure A.6. Alignment of the Δ200 DNA fragment

Alignment of the Δ200 DNA fragment and its derivative. The Δ200 contains the identified -35 and -10 promoter elements that have been mutated. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space.

```
cbpA (wildtype)          ATTTGCAGTGCAACTAATTCCATGTATATTACTACCCATATATAGCGTCTATAAAATTTAATAAATAATGACGCCCTAGTTAAACTTAAAGTGCC
cbpAΔPyccE               ATTTGCAGTGCAACTAATTCCATGTATATTACTACCCATATATAGCGTCTATAAAATTTAA----------TAAACTTAAAGTGCC
cbpAΔPyccE-217G-216G     ATTTGCAGTGCAACTAATTCCATGGGTATTACTACCCATATATAGCGTCTATAAAATTTAA----------TAAACTTAAAGTGCC
                         ************************  ************************************          ***************
cbpA (wildtype)          TGGTTCAACTATCAAAAATCGCTCACCCTTTTTCACCTGTTTAAAATATGTTCAGCAACCCATCTTGATGGCGACCTCCTCTCCGCGATGATTTC
cbpAΔPyccE               TGGTTCAACTATCAAAAATCGCTCACCCTTTTTCACCTGTTTAAAATATGTTCAGCAACCCATCTTGATGGCGACCTCCTCTCCGCGATGATTTC
cbpAΔPyccE-217G-216G     TGGTTCAACTATCAAAAATCGCTCACCCTTTTTCACCTGTTTAAAATATGTTCAGCAACCCATCTTGATGGCGACCTCCTCTCCGCGATGATTTC
                         ************************************************************************************************
cbpA (wildtype)          AATAACATATTCTGTGTTGGCATATGAAATTTTGAGGATTACCCTACACTTATAGGAGTTACCTTACAGGGGTTCCTTCAATTTGTGTTGATTTA
cbpAΔPyccE               AATAACATATTCTGTGTTGGCATATGAAATTTTGAGGATTACCCTACACTTATAGGAGTTACCTTACAGGGGTTCCTTCAATTTGTGTTGATTTA
cbpAΔPyccE-217G-216G     AATAACATATTCTGTGTTGGCATATGAAATTTTGAGGATTACCCTACACTTATAGGAGTTACCTTACAGGGGTTCCTTCAATTTGTGTTGATTTA
                         ************************************************************************************************
cbpA (wildtype)          CGCGAGATAACGCTATG
cbpAΔPyccE               CGCGAGATAACGCTATG
cbpAΔPyccE-217G-216G     CGCGAGATAACGCTATG
                         *****************
```

Figure A.7. Alignment of the *cbpA* regulatory region, the cbpAΔPyccE and its derivative.

Alignment of the *cbpA* regulatory region, the cbpAΔPyccE and its derivative. The cbpAΔPyccE DNA fragment has the *yccE* promoter deleted and contains the identified -35 and -10 *ccbpA* P6 promoter elements that have been mutated. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space.

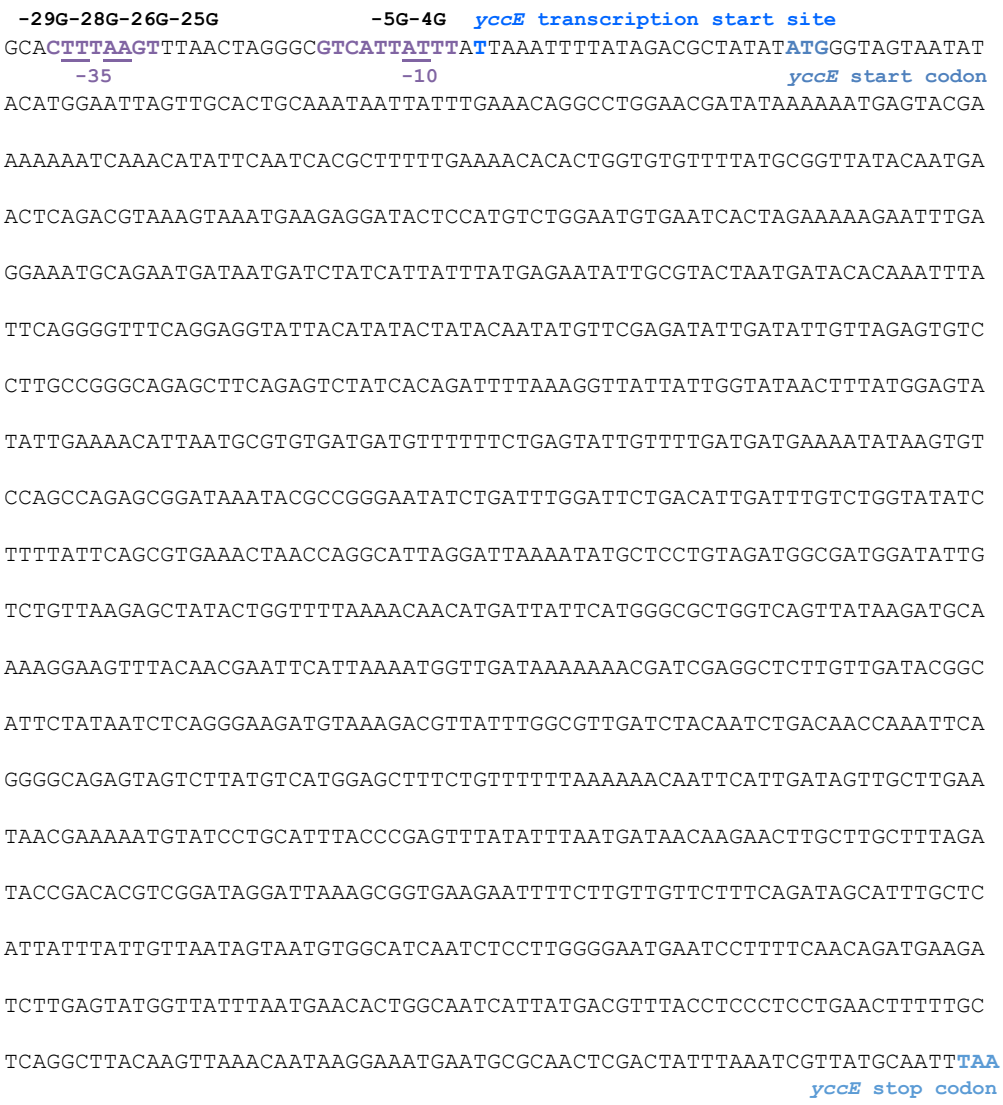


Figure A.8. Sequence of the *yccE* coding region with the upstream PyccE region.

This figure shows the sequence of the *yccE* coding region with the upstream PyccE region used in this study. Promoter elements (-35 and -10), transcription start sites, start and stop codons are written in coloured texts. Nucleotides where site specific mutations are introduced are written in bold black texts.

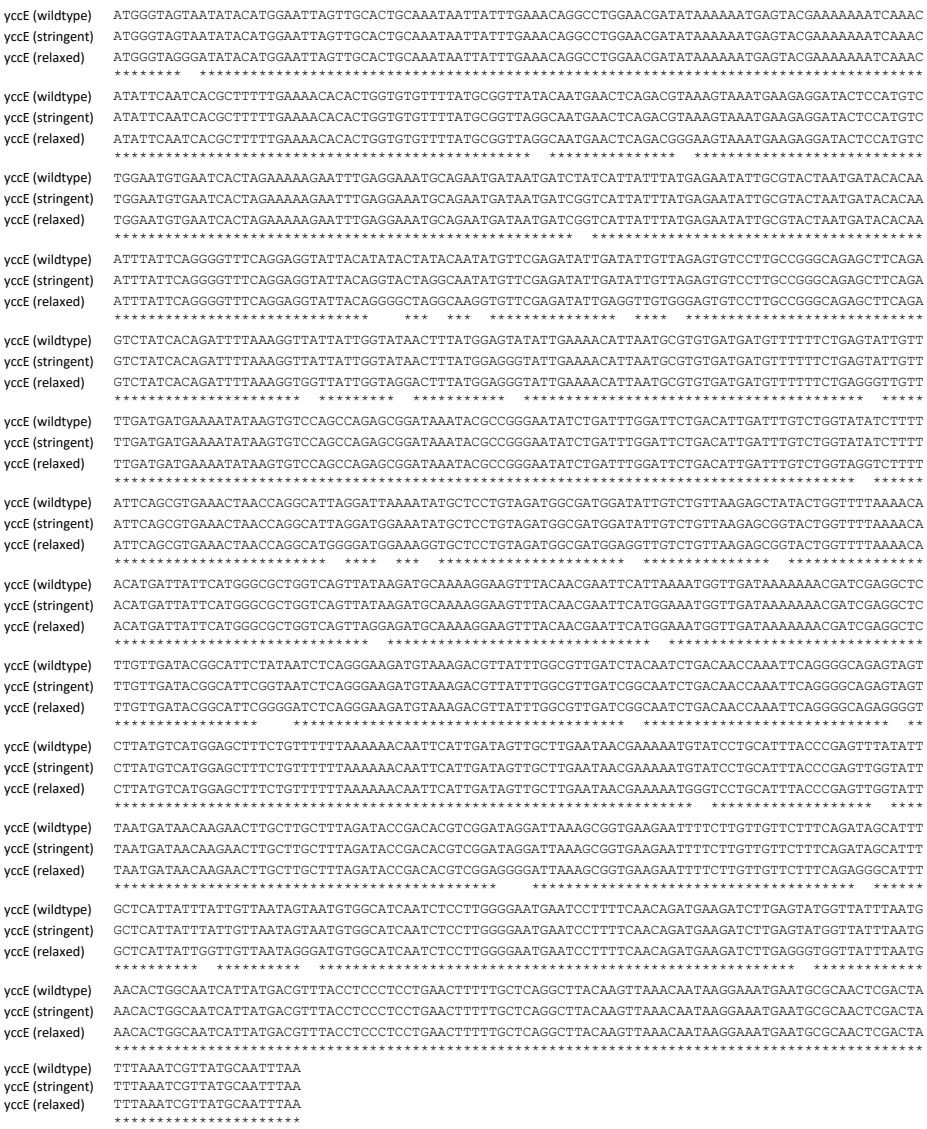


Figure A.9. Alignment of the *yccE* coding region and its derivatives.

Alignment of the *yccE* coding region and its different derivatives. The *yccE* "wildtype" contains the identified internal promoters that have been mutated in the other two derivatives. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The stringent search criteria for the -10 promoter elements include the sequences 5'-

TAnAAT-3', 5'-TATnAT-3' or 5'-TATAnT-3'. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

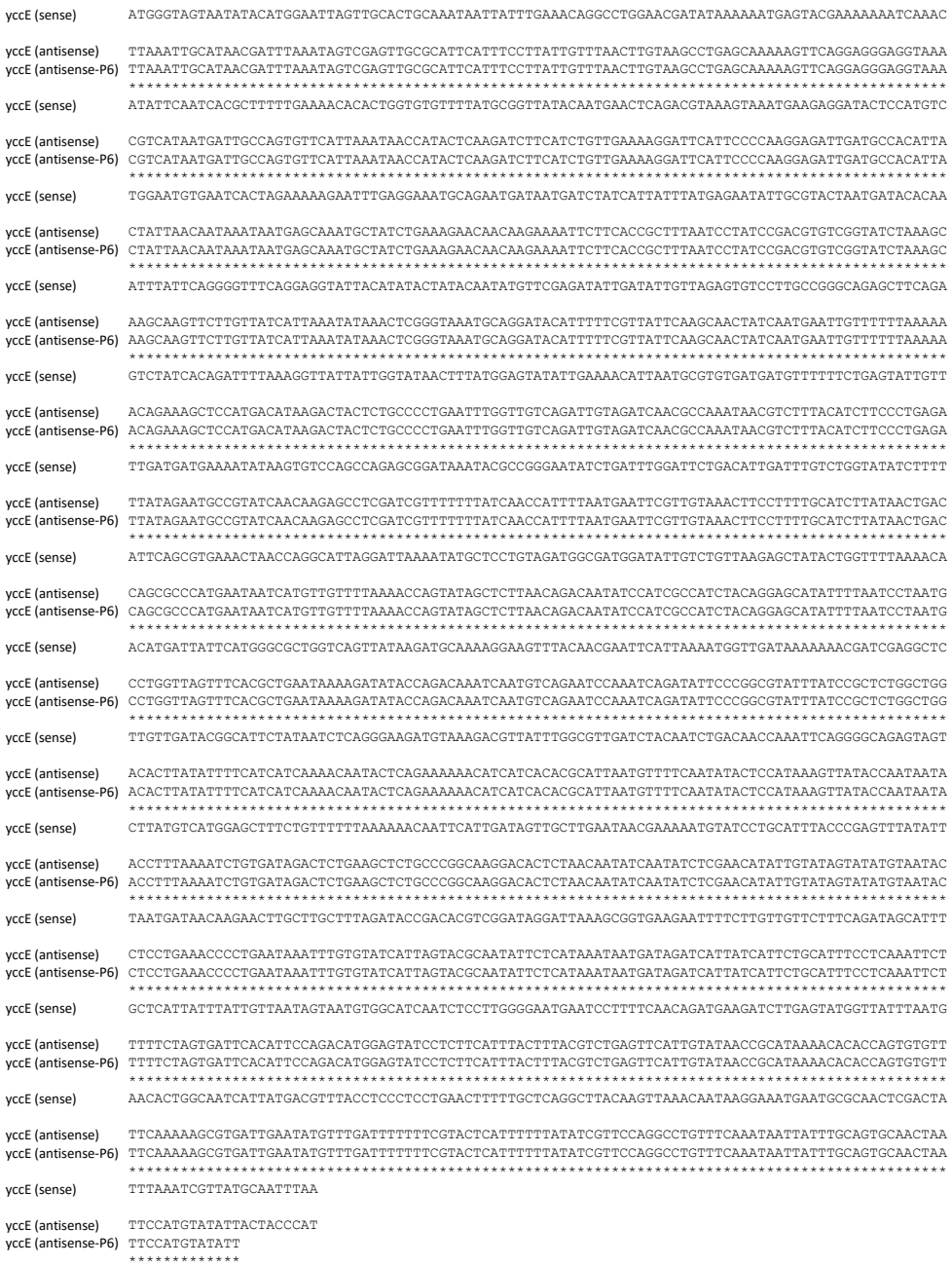


Figure A.10. Alignment of the *yccE* coding region and antisense derivatives.

Alignment of the *yccE* coding region and its antisense derivatives. The sequence labelled "sense" pertains to the *yccE* coding region in sense orientation. Consequently the sequences labelled

"antisense" and "antisense-P6" pertain to the derivatives of the *yccE* coding region in antisense orientations. Bases that remain unchanged in the two antisense derivatives are marked by stars and mismatches are marked by a space.

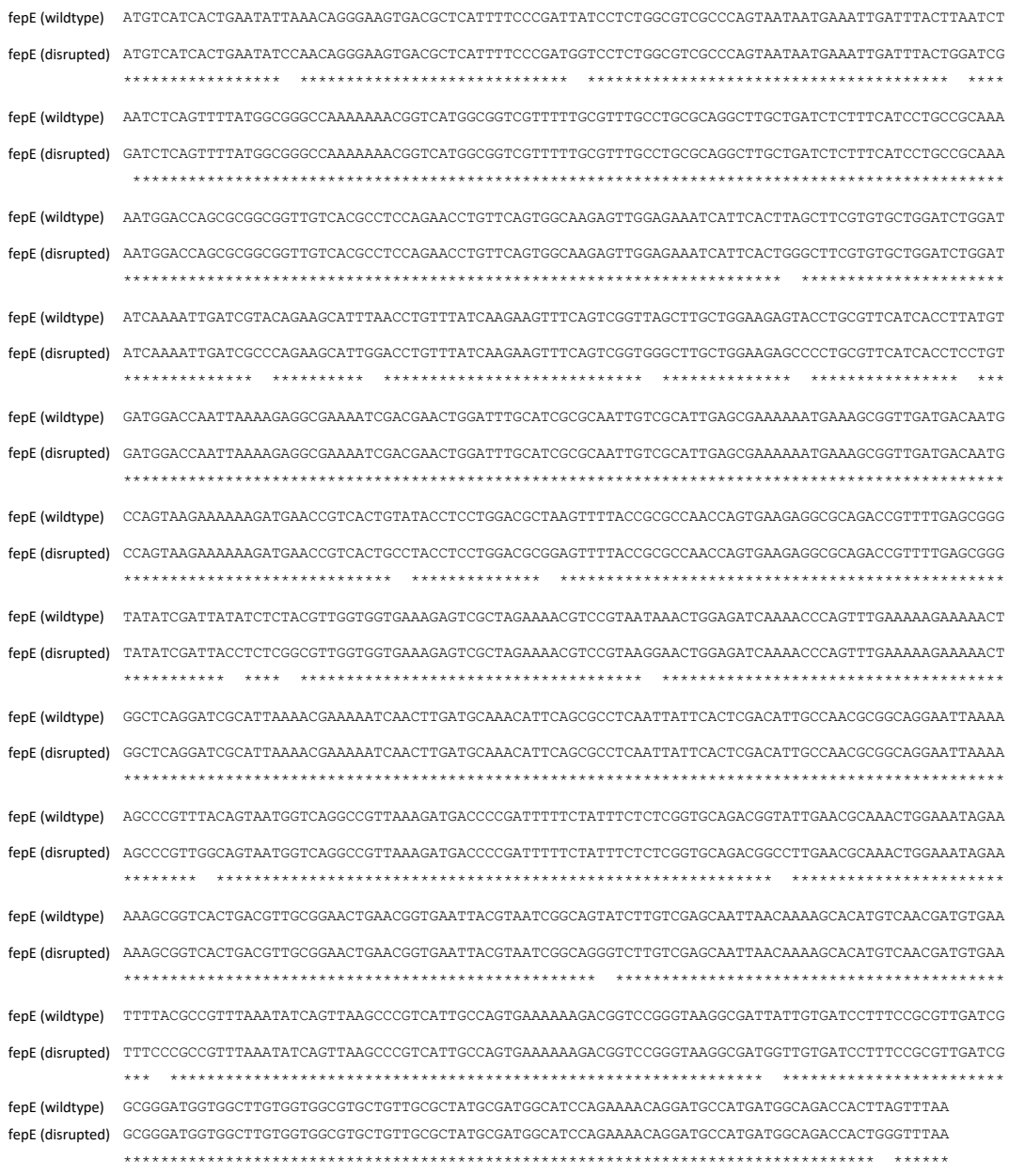


Figure A.11. Alignment of the *fepE* coding region and its derivatives.

An alignment of the *fepE* coding region and its different derivatives. The *fepE* "wildtype" contains the identified internal promoters that have been mutated in the other derivative. Bases that remain

unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

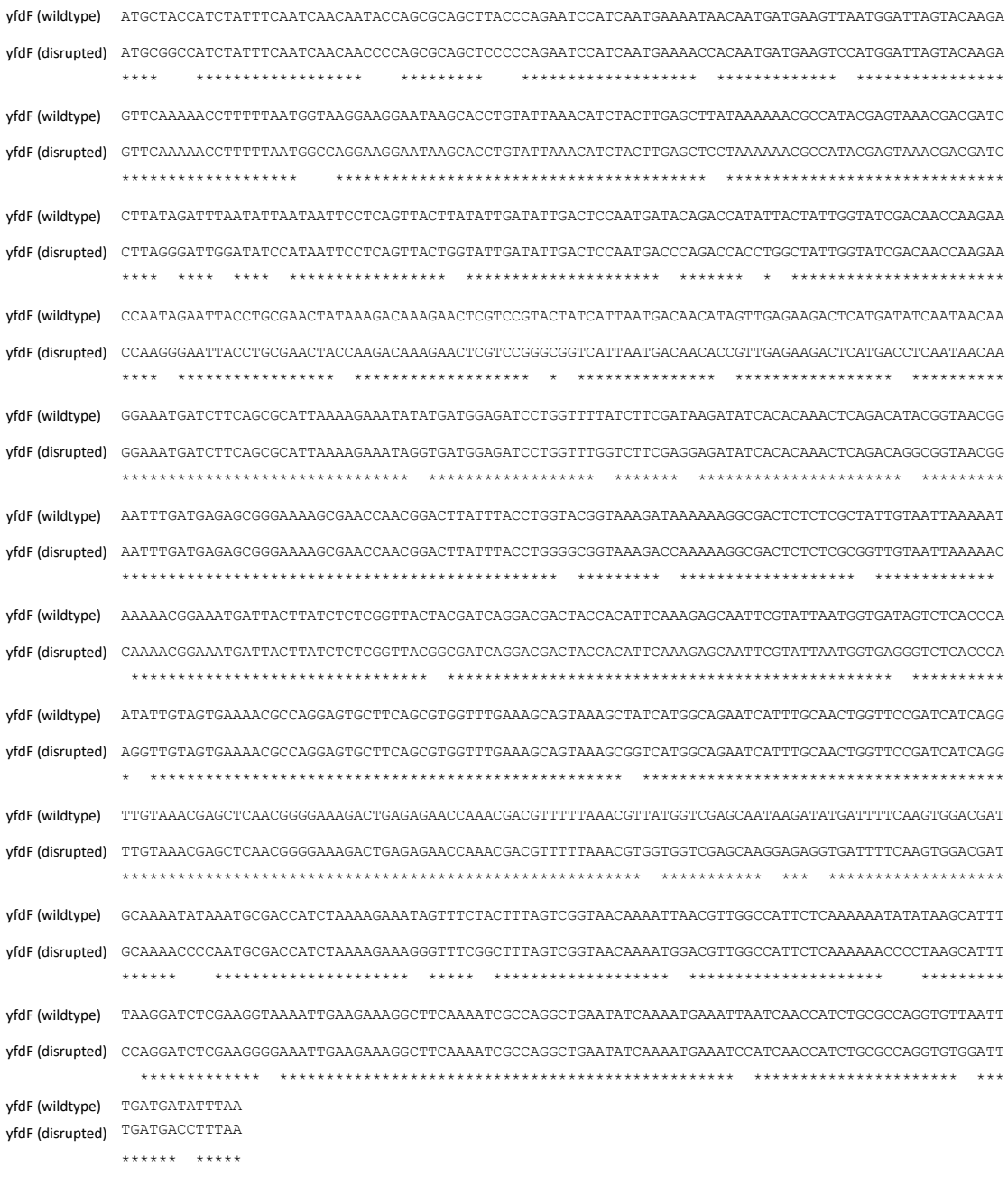


Figure A.12. Alignment of the *yfdF* coding region and its derivatives.

An alignment of the *yfdF* coding region and its different derivatives. The *yfdF* "wildtype" contains the identified internal promoters that have been mutated in the other derivative. Bases that remain

unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

```
ykgH (wildtype)    ATGCGTGAACAAATCAAACAGGATATCGATCTGATTGAGATTTTATTTTATCTGAAGAAAAAGATTCGTGTTATCCTTTTTATTATGGCTATATG
ykgH (disrupted)   ATGCGTGAACAAATCAAACAGGACCTCGATCTGATTGAGATTTGGTTTCCTCTGAAGAAAAAGATTCGTGTGGTCCTTTTGGTTATGGCTATATG
                   ************************  ******************  ***  **********************  *******  *************

ykgH (wildtype)    TATGGCTATGGTGCTGTTGTTTCTGTATATCAATAAAGACAATATAAAAGTGATTTACAGCCTAAAAATAAACCAGACAACGCCAGGTATACTTG
ykgH (disrupted)   TATGGCGGTGGTGCTGTTGTTTCTGTATATCAATAAAGACAACCCCAAAGTGATTTACAGCCCCAAAACCAACCAGACAACGCCAGGGGTACTTG
                   ******  ***********************************    ****************  ****  *****************  ******

ykgH (wildtype)    TTAGCTGTGATAGCAATAATAATTTTGCCTGTCAGACTACAATGACTGAAGATGTTATTCAGCGAATTACTACATTTTTTCACACCAGCCCAGAT
ykgH (disrupted)   TTAGCTGTGATAGCAATAAGGATTTTGCCTGTCAGACGGCAATGACTGAAGATGTCCTTCAGCGAATTACGGCATTTTTTCACACCAGCCCAGAT
                   ******************  ***************  ***************  *************  ***********************

ykgH (wildtype)    GTCAAGAACAGAGAAATAAGGCTGGAATGGTCAGGAGATAAGAGAGCTTTACCAACTGCTGAAGAGGAAATATCTCGCGTGCAGGCCTCTATTAT
ykgH (disrupted)   GTCAAGAACAGAGAAATAAGGCTGGAATGGTCAGGAGATAAGAGAGCTTTACCAACTGCTGAAGAGGAAATATCTCGCGTGCAGGCCTCGGTTAT
                   ********************************************************************************  ****

ykgH (wildtype)    CAAATGGTATGCGTCAGAATATCATAATGGCAGGCAAGTTCTCGATGAGATACAAACGCCTTCAGCAATTAACAGTGAGCTTTATACAAAAATGA
ykgH (disrupted)   CAAATGGCCTGCGTCAGAAGGTCACCATGGCAGGCAAGTTCTCGATGAGATACAAACGCCTTCAGCAATTAACAGTGAGCTTTATACAAAAATGA
                   *******  **********  ***  ********************************************************************

ykgH (wildtype)    TATACCTGACCAGGAACTGGTCATTGTATCCGAACGGTGATGGCTGTGTAACTATAAGCTCACCAGAAATAAAAAATAAATACCCTGCTGCCATT
ykgH (disrupted)   CCTACCTGACCAGGAACTGGTCATTGCCTCCGAACGGTGATGGCTGTGTAACTAGGAGCTCACCAGAAATAAAAAACCAAGGCCCTGCTGCCATT
                     ************************  *************************  ******************  **  *************

ykgH (wildtype)    TGCCTGGCTCTGGGATTTTTTCTAAGCATTGTAATTTCTGTAATGTTTTGCCTTGTCAAAAAAATGGTAGATGAATACCAACAAAACTCTGGGCA
ykgH (disrupted)   TGCCTGGCTCTGGGATTTTTTCTAAGCATTGGGATTTCTGGGATGTTTTGCCTTGTCAAAAAAATGGCCGATGAATACCAACAAAACTCTGGGCA
                   ********************************  *******  *************************  *************************

ykgH (wildtype)    GTAA
ykgH (disrupted)   GTAA
                   ****
```

Figure A.13. Alignment of the *ykgH* coding region and its derivatives.

An alignment of the *ykgH* coding region and its different derivatives. The *ykgH* "wildtype" contains the identified internal promoters that have been mutated in the other derivative. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

Figure A.14. Alignment of the *yjgL* coding region and its derivative.

An alignment of the *yjgL* coding region and its different derivatives. The *yjgL* "wildtype" contains the identified internal promoters that have been mutated in the other derivative. Bases that remain

unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

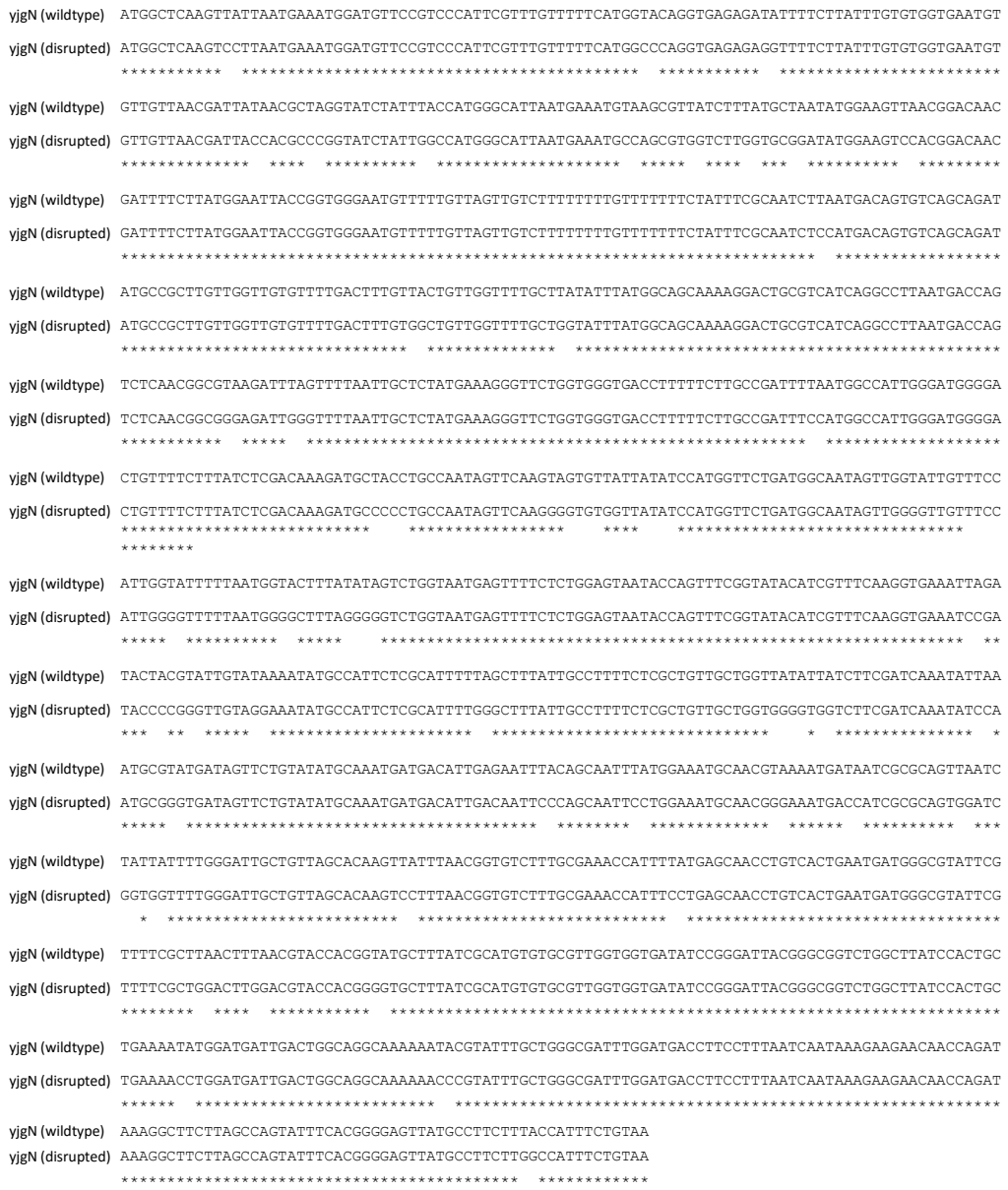


Figure A.15. Alignment of the *yjgN* coding region and its derivative.

An alignment of the *yjgN* coding region and its different derivatives. The *yjgN* "wildtype" contains the identified internal promoters that have been mutated in the other derivative. Bases that remain unchanged in the three derivatives are marked by stars and mismatches are marked by a space. The relaxed search criteria for -10 promoter elements selected the sequence 5'-TAnnnT-3'.

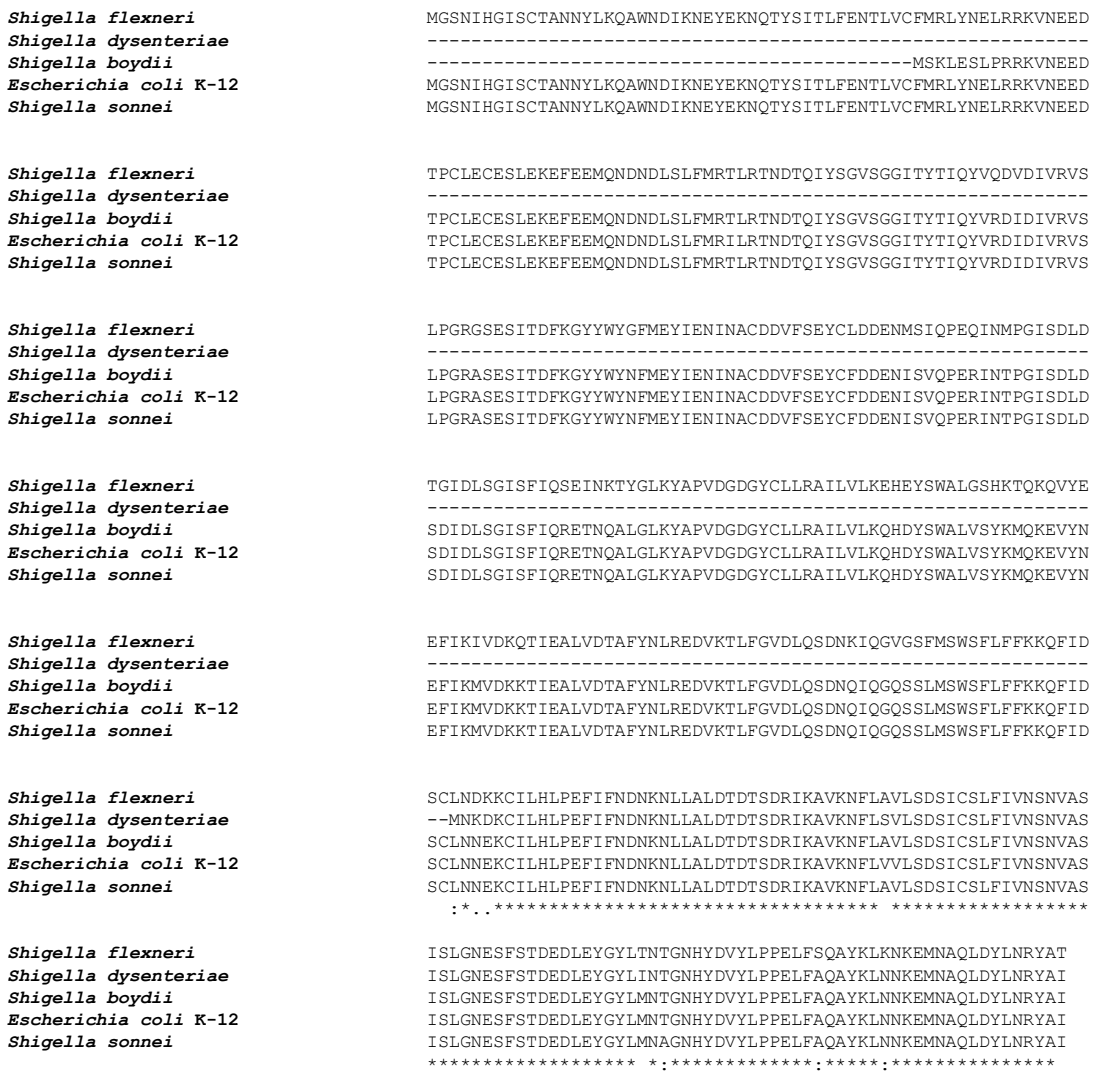


Figure A.16. Alignment of yccE homologues.

This figure shows a selection of 5 yccE sequences from a larger alignment of 500 yccE homologues. The homologues shown in this figure were among the top five hundred hits in a protein blast search using *Escherichia coli* K-12 yccE as the query sequence.

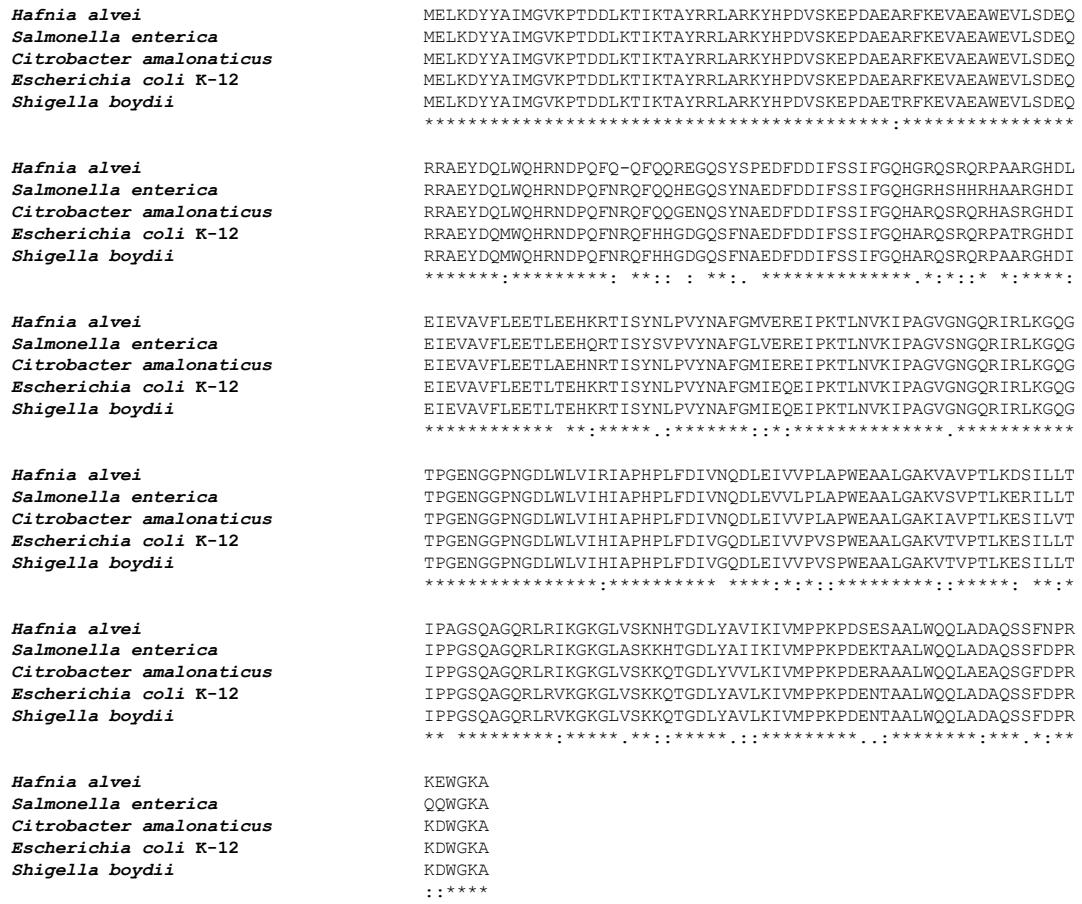


Figure A.17. Alignment of CbpA homologues.

This figure shows a selection of 5 CbpA sequences from a larger alignment of 500 CbpA homologues. The homologues shown in this figure were among the top five hundred hits in a protein blast search using *Escherichia coli* K-12 CbpA as the query sequence.

List of References

AIYAR, S. E., MCLEOD, S. M., ROSS, W., HIRVONEN, C. A., THOMAS, M. S., JOHNSON, R. C. & GOURSE, R. L. 2002. Architecture of Fis-activated transcription complexes at the Escherichia coli rrnB P1 and rrnE P1 promoters. *J Mol Biol,* 316**,** 501-16.

ALBERTS, B., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K. & WALTER, P. 2002. *Molecular Biology of the Cell, Fifth Edition,* New York, USA, Garland Science.

AROLD, S. T., LEONARD, P. G., PARKINSON, G. N. & LADBURY, J. E. 2010. H-NS forms a superhelical protein scaffold for DNA condensation. *Proc Natl Acad Sci U S A,* 107**,** 15728-32.

AZAM, T. A. & ISHIHAMA, A. 1999. Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J Biol Chem,* 274**,** 33105-13.

BAE, B., DAVIS, E., BROWN, D., CAMPBELL, E. A., WIGNESHWERARAJ, S. & DARST, S. A. 2013. Phage T7 Gp2 inhibition of Escherichia coli RNA polymerase involves misappropriation of sigma70 domain 1.1. *Proc Natl Acad Sci U S A,* 110**,** 19772-7.

BALL, C. A., OSUNA, R., FERGUSON, K. C. & JOHNSON, R. C. 1992. Dramatic changes in Fis levels upon nutrient upshift in Escherichia coli. *J Bacteriol,* 174**,** 8043-56.

BARKER, M. M., GAAL, T. & GOURSE, R. L. 2001a. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *J Mol Biol,* 305**,** 689-702.

BARKER, M. M., GAAL, T., JOSAITIS, C. A. & GOURSE, R. L. 2001b. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. *J Mol Biol,* 305**,** 673-88.

BEATTY, C. M., BROWNING, D. F., BUSBY, S. J. & WOLFE, A. J. 2003. Cyclic AMP receptor protein-dependent activation of the Escherichia coli acsP2 promoter by a synergistic class III mechanism. *J Bacteriol,* 185**,** 5148-57.

BECKER, G. & HENGGE-ARONIS, R. 2001. What makes an Escherichia coli promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). *Mol Microbiol,* 39**,** 1153-65.

BELL, C. E. & LEWIS, M. 2001. The Lac repressor: a second generation of structural and functional studies. *Curr Opin Struct Biol,* 11**,** 19-25.

BELYAEVA, T. A., BOWN, J. A., FUJITA, N., ISHIHAMA, A. & BUSBY, S. J. 1996. Location of the C-terminal domain of the RNA polymerase alpha subunit in different open complexes at the Escherichia coli galactose operon regulatory region. *Nucleic Acids Res,* 24**,** 2242-51.

BELYAEVA, T. A., RHODIUS, V. A., WEBSTER, C. L. & BUSBY, S. J. 1998. Transcription activation at promoters carrying tandem DNA sites for the Escherichia coli cyclic AMP receptor protein: organisation of the RNA polymerase alpha subunits. *J Mol Biol,* 277**,** 789-804.

BIRD, J. G., SHARMA, S., ROSHWALB, S. C., HOSKINS, J. R. & WICKNER, S. 2006. Functional analysis of CbpA, a DnaJ homolog and nucleoid-associated DNA-binding protein. *J Biol Chem,* 281**,** 34349-56.

BOWN, J., BARNE, K., MINCHIN, S. & BUSBY, S. 1997. Extended -10 Promoters. *Mechanisms of Transcription.* Springer Berlin Heidelberg.

BROWN, N. L., STOYANOV, J. V., KIDD, S. P. & HOBMAN, J. L. 2003. The MerR family of transcriptional regulators. *FEMS Microbiol Rev,* 27**,** 145-63.

BROWNING, D. F., BEATTY, C. M., WOLFE, A. J., COLE, J. A. & BUSBY, S. J. 2002. Independent regulation of the divergent Escherichia coli nrfA and acsP1 promoters by a nucleoprotein assembly at a shared regulatory region. *Mol Microbiol,* 43**,** 687-701.

BROWNING, D. F. & BUSBY, S. J. 2004. The regulation of bacterial transcription initiation. *Nat Rev Microbiol,* 2**,** 57-65.

BROWNING, D. F., COLE, J. A. & BUSBY, S. J. 2000. Suppression of FNR-dependent transcription activation at the Escherichia coli nir promoter by Fis, IHF and H-NS: modulation of transcription initiation by a complex nucleo-protein assembly. *Mol Microbiol,* 37**,** 1258-69.

BROWNING, D. F., GRAINGER, D. C. & BUSBY, S. J. 2010. Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr Opin Microbiol,* 13**,** 773-80.

BUCKLE, M., PEMBERTON, I. K., JACQUET, M. A. & BUC, H. 1999. The kinetics of sigma subunit directed promoter recognition by E. coli RNA polymerase. *J Mol Biol,* 285**,** 955-64.

BUSBY, S. & EBRIGHT, R. H. 1994. Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell,* 79**,** 743-6.

BUSBY, S. & EBRIGHT, R. H. 1997. Transcription activation at class II CAP-dependent promoters. *Mol Microbiol,* 23**,** 853-9.

BUSBY, S. & EBRIGHT, R. H. 1999. Transcription activation by catabolite activator protein (CAP). *J Mol Biol,* 293**,** 199-213.

BUSBY, S., KOTLARZ, D. & BUC, H. 1983. Deletion mutagenesis of the Escherichia coli galactose operon promoter region. *J Mol Biol,* 167**,** 259-74.

CAMERON, A. D., KROGER, C., QUINN, H. J., SCALLY, I. K., DALY, A. J., KARY, S. C. & DORMAN, C. J. 2013. Transmission of an oxygen availability signal at the Salmonella enterica serovar Typhimurium fis promoter. *PLoS One,* 8**,** e84382.

CARDINALE, C. J., WASHBURN, R. S., TADIGOTLA, V. R., BROWN, L. M., GOTTESMAN, M. E. & NUDLER, E. 2008. Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science,* 320**,** 935-8.

CASHEL, M., GENTRY, D. R., HERNANDEZ, V. H. & VINELLA, D. 1996. The stringent response. *In:* NEIDHARDT, F. C., CURTIS III, R., INGRAHAM, J. L., LIN, E. C. C., LOW, K. B., MAGASANIK, B., REZNIKOFF, W. S., RILEY, M., SCHAECHTER, M. & UMBARGER, H. E. (eds.) *Escherichia coli and Salmonella: cellular and molecular biology* 2nd ed.: ASM Press.

CHAE, C., SHARMA, S., HOSKINS, J. R. & WICKNER, S. 2004. CbpA, a DnaJ homolog, is a DnaK co-chaperone, and its activity is modulated by CbpM. *J Biol Chem,* 279**,** 33147-53.

CHECROUN, C., BORDES, P., LEROY, O., KOLB, A. & GUTIERREZ, C. 2004. Interactions between the 2.4 and 4.2 regions of sigmaS, the stress-specific sigma factor of Escherichia coli, and the -10 and -35 promoter elements. *Nucleic Acids Res,* 32**,** 45-53.

CHINTAKAYALA, K. & GRAINGER, D. C. 2011. A conserved acidic amino acid mediates the interaction between modulators and co-chaperones in enterobacteria. *Journal of molecular biology,* 411**,** 313-20.

CHINTAKAYALA, K., SINGH, S. S., ROSSITER, A. E., SHAHAPURE, R., DAME, R. T. & GRAINGER, D. C. 2013. E. coli Fis protein insulates the cbpA gene from uncontrolled transcription. *PLoS Genet,* 9**,** e1003152.

CHO, B. K., ZENGLER, K., QIU, Y., PARK, Y. S., KNIGHT, E. M., BARRETT, C. L., GAO, Y. & PALSSON, B. O. 2009. The transcription unit architecture of the Escherichia coli genome. *Nat Biotechnol,* 27**,** 1043-9.

CHODAVARAPU, S., GOMEZ, R., VICENTE, M. & KAGUNI, J. M. 2008. Escherichia coli Dps interacts with DnaA protein to impede initiation: a model of adaptive mutation. *Mol Microbiol,* 67**,** 1331-46.

CHOY, H. & ADHYA, S. 1996. *Escherichia coli and Salmonella: Cellular and Molecular Biology* Washington, ASM Press.

COSGRIFF, S., CHINTAKAYALA, K., CHIM, Y. T., CHEN, X., ALLEN, S., LOVERING, A. L. & GRAINGER, D. C. 2010. Dimerization and DNA-dependent aggregation of the Escherichia coli nucleoid protein and chaperone CbpA. *Molecular microbiology,* 77**,** 1289-300.

DAME, R. T., NOOM, M. C. & WUITE, G. J. 2006. Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature,* 444**,** 387-90.

DAME, R. T., WYMAN, C. & GOOSEN, N. 2000. H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Res,* 28**,** 3504-10.

DARWIN, A. J. & STEWART, V. 1996. *Regulation of Gene Expression in E. coli,* New York, R.G. Landes.

DAVIS, C. A., BINGMAN, C. A., LANDICK, R., RECORD, M. T., JR. & SAECKER, R. M. 2007. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A,* 104**,** 7833-8.

DEHASETH, P. L., ZUPANCIC, M. L. & RECORD, M. T., JR. 1998. RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J Bacteriol,* 180**,** 3019-25.

DEMPLE, B. 1996. Redox signaling and gene control in the Escherichia coli soxRS oxidative stress regulon--a review. *Gene,* 179**,** 53-7.

DILLON, S. C., CAMERON, A. D., HOKAMP, K., LUCCHINI, S., HINTON, J. C. & DORMAN, C. J. 2010. Genome-wide analysis of the H-NS and Sfh regulatory networks in Salmonella Typhimurium identifies a plasmid-encoded transcription silencing mechanism. *Mol Microbiol,* 76**,** 1250-65.

DILLON, S. C. & DORMAN, C. J. 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Microbiol,* 8**,** 185-95.

DORMAN, C. J. 2014. Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *J Mol Microbiol Biotechnol,* 24**,** 316-31.

DORNENBURG, J. E., DEVITA, A. M., PALUMBO, M. J. & WADE, J. T. 2010. Widespread antisense transcription in Escherichia coli. *MBio,* 1.

DOVE, S. L., DARST, S. A. & HOCHSCHILD, A. 2003. Region 4 of sigma as a target for transcription regulation. *Mol Microbiol,* 48**,** 863-74.

DUO, M., HOU, S. & REN, D. 2008. Identifying Escherichia coli genes involved in intrinsic multidrug resistance. *Appl Microbiol Biotechnol,* 81**,** 731-41.

ESPINOSA-URGEL, M., CHAMIZO, C. & TORMO, A. 1996. A consensus structure for sigma S-dependent promoters. *Mol Microbiol,* 21**,** 657-9.

FEKLISTOV, A. & DARST, S. A. 2011. Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell,* 147**,** 1257-69.

FREDERIX, M., EDWARDS, A., MCANULLA, C. & DOWNIE, J. A. 2011. Co-ordination of quorum-sensing regulation in Rhizobium leguminosarum by induction of an anti-repressor. *Mol Microbiol,* 81**,** 994-1007.

FREUNDLICH, M., RAMANI, N., MATHEW, E., SIRKO, A. & TSUI, P. 1992. The role of integration host factor in gene expression in Escherichia coli. *Mol Microbiol,* 6**,** 2557-63.

GAAL, T., ROSS, W., ESTREM, S. T., NGUYEN, L. H., BURGESS, R. R. & GOURSE, R. L. 2001. Promoter recognition and discrimination by EsigmaS RNA polymerase. *Mol Microbiol,* 42**,** 939-54.

GIANNOUKOS, G., CIULLA, D. M., HUANG, K., HAAS, B. J., IZARD, J., LEVIN, J. Z., LIVNY, J., EARL, A. M., GEVERS, D., WARD, D. V., NUSBAUM, C., BIRREN, B. W. & GNIRKE, A. 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol,* 13**,** R23.

GOURSE, R. L., ROSS, W. & 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**,** 687-95.

GOWRISHANKAR, J. & HARINARAYANAN, R. 2004. Why is transcription coupled to translation in bacteria? *Mol Microbiol,* 54**,** 598-603.

GRAINGER, D. C., GOLDBERG, M. D., LEE, D. J. & BUSBY, S. J. 2008. Selective repression by Fis and H-NS at the Escherichia coli dps promoter. *Mol Microbiol,* 68**,** 1366-77.

GRAINGER, D. C., HURD, D., GOLDBERG, M. D. & BUSBY, S. J. 2006. Association of nucleoid proteins with coding and non-coding segments of the Escherichia coli genome. *Nucleic Acids Res,* 34**,** 4642-52.

GRAINGER, D. C., WEBSTER, C. L., BELYAEVA, T. A., HYDE, E. I. & BUSBY, S. J. 2004. Transcription activation at the Escherichia coli melAB promoter: interactions of MelR with its DNA target site and with domain 4 of the RNA polymerase sigma subunit. *Mol Microbiol,* 51**,** 1297-309.

GRALLA, J. D. 1996. Activation and repression of E. coli promoters. *Curr Opin Genet Dev,* 6**,** 526-30.

GROSS, C. E. A. 1998. *Mechanisms of Transcription* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

GROSSMAN, A. D., STRAUS, D. B., WALTER, W. A. & GROSS, C. A. 1987. Sigma 32 synthesis can regulate the synthesis of heat shock proteins in Escherichia coli. *Genes Dev,* 1**,** 179-84.

GRUBER, T. M. & GROSS, C. A. 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol,* 57**,** 441-66.

GUR, E., BIRAN, D., SHECHTER, N., GENEVAUX, P., GEORGOPOULOS, C. & RON, E. Z. 2004. The Escherichia coli DjlA and CbpA proteins can substitute for DnaJ in DnaK-mediated protein disaggregation. *J Bacteriol,* 186**,** 7236-42.

HANCOCK, S. P., STELLA, S., CASCIO, D. & JOHNSON, R. C. 2016. DNA Sequence Determinants Controlling Affinity, Stability and Shape of DNA Complexes Bound by the Nucleoid Protein Fis. *PLoS One,* 11**,** e0150189.

HARDY, C. D. & COZZARELLI, N. R. 2005. A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure. *Mol Microbiol,* 57**,** 1636-52.

HAYKINSON, M. J. & JOHNSON, R. C. 1993. DNA looping and the helical repeat in vitro and in vivo: effect of HU protein and enhancer location on Hin invertasome assembly. *EMBO J,* 12**,** 2503-12.

HELDWEIN, E. E. & BRENNAN, R. G. 2001. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature,* 409**,** 378-82.

HENGEN, P. N., BARTRAM, S. L., STEWART, L. E. & SCHNEIDER, T. D. 1997. Information analysis of Fis binding sites. *Nucleic Acids Res,* 25**,** 4994-5002.

HENGGE-ARONIS, R. 1996. Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in Escherichia coli. *Mol Microbiol,* 21**,** 887-93.

HERRING, C. D., RAFFAELLE, M., ALLEN, T. E., KANIN, E. I., LANDICK, R., ANSARI, A. Z. & PALSSON, B. O. 2005. Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J Bacteriol,* 187**,** 6166-74.

HIRVONEN, C. A., ROSS, W., WOZNIAK, C. E., MARASCO, E., ANTHONY, J. R., AIYAR, S. E., NEWBURN, V. H. & GOURSE, R. L. 2001. Contributions of UP elements and the transcription factor FIS to expression from the seven rrn P1 promoters in Escherichia coli. *J Bacteriol,* 183**,** 6305-14.

HOCHSCHILD, A. & DOVE, S. L. 1998. Protein-protein contacts that activate and repress prokaryotic transcription. *Cell,* 92**,** 597-600.

HOOK-BARNARD, I. G. & HINTON, D. M. 2009. The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A,* 106**,** 737-42.

HUANG, Q., CHENG, X., CHEUNG, M. K., KISELEV, S. S., OZOLINE, O. N. & KWAN, H. S. 2012. High-density transcriptional initiation signals underline genomic islands in bacteria. *PLoS One,* 7**,** e33759.

ISHIHAMA, A. 2000. Functional modulation of Escherichia coli RNA polymerase. *Annu Rev Microbiol,* 54**,** 499-518.

JAIN, D., NICKELS, B. E., SUN, L., HOCHSCHILD, A. & DARST, S. A. 2004. Structure of a ternary transcription activation complex. *Mol Cell,* 13**,** 45-53.

JORDI, B. J. & HIGGINS, C. F. 2000. The downstream regulatory element of the proU operon of Salmonella typhimurium inhibits open complex formation by RNA polymerase at a distance. *J Biol Chem,* 275**,** 12123-8.

JOUNG, J. K., LE, L. U. & HOCHSCHILD, A. 1993. Synergistic activation of transcription by Escherichia coli cAMP receptor protein. *Proc Natl Acad Sci U S A,* 90**,** 3083-7.

KAHRAMANOGLOU, C., SESHASAYEE, A. S., PRIETO, A. I., IBBERSON, D., SCHMIDT, S., ZIMMERMANN, J., BENES, V., FRASER, G. M. & LUSCOMBE, N. M. 2011. Direct and indirect effects of H-NS and Fis on global gene expression control in Escherichia coli. *Nucleic Acids Res,* 39**,** 2073-91.

KEANE, O. M. & DORMAN, C. J. 2003. The gyr genes of Salmonella enterica serovar Typhimurium are repressed by the factor for inversion stimulation, Fis. *Mol Genet Genomics,* 270**,** 56-65.

KESELER, I. M., COLLADO-VIDES, J., GAMA-CASTRO, S., INGRAHAM, J., PALEY, S., PAULSEN, I. T., PERALTA-GIL, M. & KARP, P. D. 2005. EcoCyc: a comprehensive database resource for Escherichia coli. *Nucleic Acids Res,* 33**,** D334-7.

KOLB, A., KOTLARZ, D., KUSANO, S. & ISHIHAMA, A. 1995. Selectivity of the Escherichia coli RNA polymerase E sigma 38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res,* 23**,** 819-26.

KOSTREWA, D., GRANZIN, J., KOCH, C., CHOE, H. W., RAGHUNATHAN, S., WOLF, W., LABAHN, J., KAHMANN, R. & SAENGER, W. 1991. Three-dimensional structure of the E. coli DNA-binding protein FIS. *Nature,* 349**,** 178-80.

KOTLAJICH, M. V., HRON, D. R., BOUDREAU, B. A., SUN, Z., LYUBCHENKO, Y. L. & LANDICK, R. 2015. Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria. *Elife,* 4.

LACOUR, S., KOLB, A. & LANDINI, P. 2003. Nucleotides from -16 to -12 determine specific promoter recognition by bacterial sigmaS-RNA polymerase. *J Biol Chem,* 278**,** 37160-8.

LACOUR, S., LEROY, O., KOLB, A. & LANDINI, P. 2004. Substitutions in region 2.4 of sigma70 allow recognition of the sigmaS-dependent aidB promoter. *J Biol Chem,* 279**,** 55255-61.

LAM, K. N. & CHARLES, T. C. 2015. Strong spurious transcription likely contributes to DNA insert bias in typical metagenomic clone libraries. *Microbiome,* 3**,** 22.

LANDINI, P., EGLI, T., WOLF, J. & LACOUR, S. 2014. sigmaS, a major player in the response to environmental stresses in Escherichia coli: role, regulation and mechanisms of promoter recognition. *Environ Microbiol Rep,* 6**,** 1-13.

LAWSON, C. L., SWIGON, D., MURAKAMI, K. S., DARST, S. A., BERMAN, H. M. & EBRIGHT, R. H. 2004. Catabolite activator protein: DNA binding and transcription activation. *Curr Opin Struct Biol,* 14**,** 10-20.

LEE, D. J., MINCHIN, S. D. & BUSBY, S. J. 2012. Activating transcription in bacteria. *Annu Rev Microbiol,* 66**,** 125-52.

LEIBMAN, M. & HOCHSCHILD, A. 2007. A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J,* 26**,** 1579-90.

LEMAUX, P. G., HERENDEEN, S. L., BLOCH, P. L. & NEIDHARDT, F. C. 1978. Transient rates of synthesis of individual polypeptides in E. coli following temperature shifts. *Cell,* 13**,** 427-34.

LIU, D., HANIFORD, D. B. & CHALMERS, R. M. 2011. H-NS mediates the dissociation of a refractory protein-DNA complex during Tn10/IS10 transposition. *Nucleic Acids Res,* 39**,** 6660-8.

LIU, Y., CHEN, H., KENNEY, L. J. & YAN, J. 2010. A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes Dev,* 24**,** 339-44.

LODGE, J., FEAR, J., BUSBY, S., GUNASEKARAN, P. & KAMINI, N. R. 1992. Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. *FEMS Microbiol Lett,* 74**,** 271-6.

LONETTO, M., GRIBSKOV, M. & GROSS, C. A. 1992. The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol,* 174**,** 3843-9.

LUCCHINI, S., ROWLEY, G., GOLDBERG, M. D., HURD, D., HARRISON, M. & HINTON, J. C. 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog,* 2**,** e81.

MADAN BABU, M. & TEICHMANN, S. A. 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic acids research,* 31**,** 1234-44.

MAEDA, H., FUJITA, N. & ISHIHAMA, A. 2000. Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. *Nucleic acids research,* 28**,** 3497-503.

MAKRIS, J. C., NORDMANN, P. L. & REZNIKOFF, W. S. 1990. Integration host factor plays a role in IS50 and Tn5 transposition. *J Bacteriol,* 172**,** 1368-73.

MARTIN, R. G. & ROSNER, J. L. 2001. The AraC transcriptional activators. *Curr Opin Microbiol,* 4**,** 132-7.

MARTINEZ-ANTONIO, A. & COLLADO-VIDES, J. 2003. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol,* 6**,** 482-9.

MASTERS, T. L., WACHTER, S., WACHTER, J. & HILL, S. A. 2015. H-NS suppresses pilE intragenic transcription and antigenic variation in Neisseria gonorrhoeae. *Microbiology*.

MATHEW, R. & CHATTERJI, D. 2006. The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol,* 14**,** 450-5.

MAURER, S., FRITZ, J. & MUSKHELISHVILI, G. 2009. A systematic in vitro study of nucleoprotein complexes formed by bacterial nucleoid-associated proteins revealing novel types of DNA organization. *J Mol Biol,* 387**,** 1261-76.

MCLEOD, S. M., AIYAR, S. E., GOURSE, R. L. & JOHNSON, R. C. 2002. The C-terminal domains of the RNA polymerase alpha subunits: contact site with Fis and localization during co-activation with CRP at the Escherichia coli proP P2 promoter. *J Mol Biol,* 316**,** 517-29.

MEKLER, V., KORTKHONJIA, E., MUKHOPADHYAY, J., KNIGHT, J., REVYAKIN, A., KAPANIDIS, A. N., NIU, W., EBRIGHT, Y. W., LEVY, R. & EBRIGHT, R. H. 2002. Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell,* 108**,** 599-614.

MITCHELL, J. E., ZHENG, D., BUSBY, S. J. & MINCHIN, S. D. 2003. Identification and analysis of 'extended -10' promoters in Escherichia coli. *Nucleic Acids Res,* 31**,** 4689-95.

MOINIER, D., SLYEMI, D., BYRNE, D., LIGNON, S., LEBRUN, R., TALLA, E. & BONNEFOY, V. 2014. An ArsR/SmtB family member is involved in the regulation by arsenic of the arsenite oxidase operon in Thiomonas arsenitoxydans. *Appl Environ Microbiol,* 80**,** 6413-26.

MOONEY, R. A., DAVIS, S. E., PETERS, J. M., ROWLAND, J. L., ANSARI, A. Z. & LANDICK, R. 2009. Regulator trafficking on bacterial transcription units in vivo. *Mol Cell,* 33**,** 97-108.

MOORE, J. M., MAGNAN, D., MOJICA, A. K., NUNEZ, M. A., BATES, D., ROSENBERG, S. M. & HASTINGS, P. J. 2015. Roles of Nucleoid-Associated Proteins in Stress-Induced Mutagenic Break Repair in Starving Escherichia coli. *Genetics,* 201**,** 1349-62.

MULLER-HILL, B. 1996. *The Lac Operon: A Short History of a Genetic Paradigm,* New York, Walter de Gruyter.

MULLER-HILL, B. 1998. Some repressors of bacterial transcription. *Curr Opin Microbiol,* 1**,** 145-51.

MURAKAMI, K., OWENS, J. T., BELYAEVA, T. A., MEARES, C. F., BUSBY, S. J. & ISHIHAMA, A. 1997. Positioning of two alpha subunit carboxy-terminal domains of RNA polymerase at promoters by two transcription factors. *Proc Natl Acad Sci U S A,* 94**,** 11274-8.

MURAKAMI, K. S. 2013. X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem,* 288**,** 9126-34.

MURAKAMI, K. S. & DARST, S. A. 2003. Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol,* 13**,** 31-9.

MURAKAMI, K. S., MASUDA, S., CAMPBELL, E. A., MUZZIN, O. & DARST, S. A. 2002a. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science,* 296**,** 1285-90.

MURAKAMI, K. S., MASUDA, S. & DARST, S. A. 2002b. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science,* 296**,** 1280-4.

MUSKHELISHVILI, G., TRAVERS, A. A., HEUMANN, H. & KAHMANN, R. 1995. FIS and RNA polymerase holoenzyme form a specific nucleoprotein complex at a stable RNA promoter. *EMBO J,* 14**,** 1446-52.

NAVARRE, W. W., PORWOLLIK, S., WANG, Y., MCCLELLAND, M., ROSEN, H., LIBBY, S. J. & FANG, F. C. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science,* 313**,** 236-8.

NICKELS, B. E., DOVE, S. L., MURAKAMI, K. S., DARST, S. A. & HOCHSCHILD, A. 2002. Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambdacI. *J Mol Biol,* 324**,** 17-34.

NIU, W., KIM, Y., TAU, G., HEYDUK, T. & EBRIGHT, R. H. 1996. Transcription activation at class II CAP-dependent promoters: two interactions between CAP and RNA polymerase. *Cell,* 87**,** 1123-34.

O CROININ, T. & DORMAN, C. J. 2007. Expression of the Fis protein is sustained in late-exponential- and stationary-phase cultures of Salmonella enterica serovar Typhimurium grown in the absence of aeration. *Mol Microbiol,* 66**,** 237-51.

OSHIMA, T., ISHIKAWA, S., KUROKAWA, K., AIBA, H. & OGASAWARA, N. 2006. Escherichia coli histone-like protein H-NS preferentially binds to horizontally acquired DNA in association with RNA polymerase. *DNA Res,* 13**,** 141-53.

PAGET, M. S. 2015. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules,* 5**,** 1245-65.

PAGET, M. S. & HELMANN, J. D. 2003. The sigma70 family of sigma factors. *Genome Biol,* 4**,** 203.

PEREZ-RUEDA, E. & COLLADO-VIDES, J. 2000. The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic acids research,* 28**,** 1838-47.

PEREZ-RUEDA, E., GRALLA, J. D. & COLLADO-VIDES, J. 1998. Genomic position analyses and the transcription machinery. *J Mol Biol,* 275**,** 165-70.

PETERS, J. M., MOONEY, R. A., GRASS, J. A., JESSEN, E. D., TRAN, F. & LANDICK, R. 2012. Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes Dev,* 26**,** 2621-33.

PETERS, J. M., MOONEY, R. A., KUAN, P. F., ROWLAND, J. L., KELES, S. & LANDICK, R. 2009. Rho directs widespread termination of intragenic and stable RNA transcription. *Proc Natl Acad Sci U S A,* 106**,** 15406-11.

PETERSEN, C., MOLLER, L. B. & VALENTIN-HANSEN, P. 2002. The cryptic adenine deaminase gene of Escherichia coli. Silencing by the nucleoid-associated DNA-binding protein, H-NS, and activation by insertion elements. *J Biol Chem,* 277**,** 31373-80.

PLUMBRIDGE, J. 2002. Regulation of gene expression in the PTS in Escherichia coli: the role and interactions of Mlc. *Curr Opin Microbiol,* 5**,** 187-93.

POTRYKUS, K. & CASHEL, M. 2008. (p)ppGpp: still magical? *Annu Rev Microbiol,* 62**,** 35-51.

PRIBNOW, D. 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *J Mol Biol,* 99**,** 419-43.

PROSHKIN, S., RAHMOUNI, A. R., MIRONOV, A. & NUDLER, E. 2010. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science,* 328**,** 504-8.

RAGHAVAN, R., GROISMAN, E. A. & OCHMAN, H. 2011. Genome-wide detection of novel regulatory RNAs in E. coli. *Genome Res,* 21**,** 1487-97.

RENAULT, M., GARCIA, J., CORDEIRO, T. N., BALDUS, M. & PONS, M. 2013. Protein oligomers studied by solid-state NMR--the case of the full-length nucleoid-associated protein histone-like nucleoid structuring protein. *FEBS J,* 280**,** 2916-28.

REYES-CABALLERO, H., CAMPANELLO, G. C. & GIEDROC, D. P. 2011. Metalloregulatory proteins: metal selectivity and allosteric switching. *Biophys Chem,* 156**,** 103-14.

RHODIUS, V. A. & BUSBY, S. J. 2000a. Interactions between activating region 3 of the Escherichia coli cyclic AMP receptor protein and region 4 of the RNA polymerase sigma(70) subunit: application of suppression genetics. *J Mol Biol,* 299**,** 311-24.

RHODIUS, V. A. & BUSBY, S. J. 2000b. Transcription activation by the Escherichia coli cyclic AMP receptor protein: determinants within activating region 3. *J Mol Biol,* 299**,** 295-310.

ROSS, W., SCHNEIDER, D. A., PAUL, B. J., MERTENS, A. & GOURSE, R. L. 2003. An intersubunit contact stimulating transcription initiation by E coli RNA polymerase: interaction of the alpha C-terminal domain and sigma region 4. *Genes Dev,* 17**,** 1293-307.

ROSS, W., VRENTAS, C. E., SANCHEZ-VAZQUEZ, P., GAAL, T. & GOURSE, R. L. 2013. The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Mol Cell,* 50**,** 420-9.

RUSSO, J., HARRINGTON, A. W. & STEINIGER, M. 2015. Antisense Transcription of Retrotransposons in Drosophila: The Origin of Endogenous Small Interfering RNA Precursors. *Genetics,* 202**,** 107-121.

SAECKER, R. M., RECORD, M. T., JR. & DEHASETH, P. L. 2011. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol,* 412**,** 754-71.

SANDERSON, A., MITCHELL, J. E., MINCHIN, S. D. & BUSBY, S. J. 2003. Substitutions in the Escherichia coli RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. *FEBS Lett,* 544**,** 199-205.

SARRAF, N. S., BAARDSNES, J., CHENG, J., O'CONNOR-MCCOURT, M., CYGLER, M. & EKIEL, I. 2010. Structural basis of the regulation of the CbpA co-chaperone by its specific modulator CbpM. *J Mol Biol,* 398**,** 111-21.

SAVERY, N. J., LLOYD, G. S., KAINZ, M., GAAL, T., ROSS, W., EBRIGHT, R. H., GOURSE, R. L. & BUSBY, S. J. 1998. Transcription activation at Class II CRP-dependent promoters: identification of determinants in the C-terminal domain of the RNA polymerase alpha subunit. *EMBO J,* 17**,** 3439-47.

SCHLEIF, R. 2010. AraC protein, regulation of the l-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action. *FEMS Microbiol Rev,* 34**,** 779-96.

SCHNETZ, K. 1995. Silencing of Escherichia coli bgl promoter by flanking sequence elements. *EMBO J,* 14**,** 2545-50.

SCLAVI, B., ZAYCHIKOV, E., ROGOZINA, A., WALTHER, F., BUCKLE, M. & HEUMANN, H. 2005. Real-time characterization of intermediates in the pathway to open complex formation by Escherichia coli RNA polymerase at the T7A1 promoter. *Proc Natl Acad Sci U S A,* 102**,** 4706-11.

SEGAL, G. & RON, E. Z. 1998. Regulation of heat-shock response in bacteria. *Ann N Y Acad Sci,* 851**,** 147-51.

SHAO, Y., FELDMAN-COHEN, L. S. & OSUNA, R. 2008. Functional characterization of the Escherichia coli Fis-DNA binding sequence. *J Mol Biol,* 376**,** 771-85.

SHIMADA, T., YAMAZAKI, Y., TANAKA, K. & ISHIHAMA, A. 2014. The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli. *PLoS One,* 9**,** e90447.

SHIN, M., KANG, S., HYUN, S. J., FUJITA, N., ISHIHAMA, A., VALENTIN-HANSEN, P. & CHOY, H. E. 2001. Repression of deoP2 in Escherichia coli by CytR: conversion of a transcription activator into a repressor. *EMBO J,* 20**,** 5392-9.

SIEBENLIST, U., SIMPSON, R. B. & GILBERT, W. 1980. E. coli RNA polymerase interacts homologously with two different promoters. *Cell,* 20**,** 269-81.

SINGH, S. S. & GRAINGER, D. C. 2013. H-NS can facilitate specific DNA-binding by RNA polymerase in AT-rich gene regulatory regions. *PLoS Genet,* 9**,** e1003589.

SINGH, S. S., SINGH, N., BONOCORA, R. P., FITZGERALD, D. M., WADE, J. T. & GRAINGER, D. C. 2014. Widespread suppression of intragenic transcription initiation by H-NS. *Genes Dev,* 28**,** 214-9.

SRIVATSAN, A. & WANG, J. D. 2008. Control of bacterial transcription, translation and replication by (p)ppGpp. *Curr Opin Microbiol,* 11**,** 100-5.

STELLA, S., CASCIO, D. & JOHNSON, R. C. 2010. The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes Dev,* 24**,** 814-26.

STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000. Two-component signal transduction. *Annu Rev Biochem,* 69**,** 183-215.

STOEBEL, D. M., FREE, A. & DORMAN, C. J. 2008. Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology,* 154**,** 2533-45.

SWINGLE, B., O'CARROLL, M., HANIFORD, D. & DERBYSHIRE, K. M. 2004. The effect of host-encoded nucleoid proteins on transposition: H-NS influences targeting of both IS903 and Tn10. *Mol Microbiol,* 52**,** 1055-67.

TAKAHASHI, K., KASAI, K. & OCHI, K. 2004. Identification of the bacterial alarmone guanosine 5'-diphosphate 3'-diphosphate (ppGpp) in plants. *Proc Natl Acad Sci U S A,* 101**,** 4320-4.

TANAKA, K., TAKAYANAGI, Y., FUJITA, N., ISHIHAMA, A. & TAKAHASHI, H. 1993. Heterogeneity of the principal sigma factor in Escherichia coli: the rpoS gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase Escherichia coli. *Proc Natl Acad Sci U S A,* 90**,** 8303.

TEBBUTT, J., RHODIUS, V. A., WEBSTER, C. L. & BUSBY, S. J. 2002. Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiol Lett,* 210**,** 55-60.

TOMSIC, M., TSUJIKAWA, L., PANAGHIE, G., WANG, Y., AZOK, J. & DEHASETH, P. L. 2001. Different roles for basic and aromatic amino acids in conserved region 2 of Escherichia coli sigma(70) in the nucleation and maintenance of the single-stranded DNA bubble in open RNA polymerase-promoter complexes. *J Biol Chem,* 276**,** 31891-6.

TYPAS, A. & HENGGE, R. 2006. Role of the spacer between the -35 and -10 regions in sigmas promoter selectivity in Escherichia coli. *Mol Microbiol,* 59**,** 1037-51.

UEGUCHI, C., KAKEDA, M., YAMADA, H. & MIZUNO, T. 1994. An analogue of the DnaJ molecular chaperone in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America,* 91**,** 1054-8.

VALENTIN-HANSEN, P., SOGAARD-ANDERSEN, L. & PEDERSEN, H. 1996. A flexible partnership: the CytR anti-activator and the cAMP-CRP activator protein, comrades in transcription control. *Mol Microbiol,* 20**,** 461-6.

VASSYLYEV, D. G., SEKINE, S., LAPTENKO, O., LEE, J., VASSYLYEVA, M. N., BORUKHOV, S. & YOKOYAMA, S. 2002. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. *Nature,* 417**,** 712-9.

VUTHOORI, S., BOWERS, C. W., MCCRACKEN, A., DOMBROSKI, A. J. & HINTON, D. M. 2001. Domain 1.1 of the sigma(70) subunit of Escherichia coli RNA polymerase modulates the formation of stable polymerase/promoter complexes. *J Mol Biol,* 309**,** 561-72.

WACHTER, J., MASTERS, T. L., WACHTER, S., MASON, J. & HILL, S. A. 2015. pilS loci in Neisseria gonorrhoeae are transcriptionally active. *Microbiology,* 161**,** 1124-35.

WADE, J. T., CASTRO ROA, D., GRAINGER, D. C., HURD, D., BUSBY, S. J., STRUHL, K. & NUDLER, E. 2006. Extensive functional overlap between sigma factors in Escherichia coli. *Nat Struct Mol Biol,* 13**,** 806-14.

WANG, W., LI, G. W., CHEN, C., XIE, X. S. & ZHUANG, X. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science,* 333**,** 1445-9.

WANG, Y. & DEHASETH, P. L. 2003. Sigma 32-dependent promoter activity in vivo: sequence determinants of the groE promoter. *J Bacteriol,* 185**,** 5800-6.

WINARDHI, R. S., YAN, J. & KENNEY, L. J. 2015. H-NS Regulates Gene Expression and Compacts the Nucleoid: Insights from Single-Molecule Experiments. *Biophys J,* 109**,** 1321-9.

WOSTEN, M. M. 1998. Eubacterial sigma-factors. *FEMS Microbiol Rev,* 22**,** 127-50.

WU, H., TYSON, K. L., COLE, J. A. & BUSBY, S. J. 1998. Regulation of transcription initiation at the Escherichia coli nir operon promoter: a new mechanism to account for co-dependence on two transcription factors. *Mol Microbiol,* 27**,** 493-505.

WYBORN, N. R., STAPLETON, M. R., NORTE, V. A., ROBERTS, R. E., GRAFTON, J. & GREEN, J. 2004. Regulation of Escherichia coli hemolysin E expression by H-NS and Salmonella SlyA. *J Bacteriol,* 186**,** 1620-8.

YUAN, H. S., FINKEL, S. E., FENG, J. A., KACZOR-GRZESKOWIAK, M., JOHNSON, R. C. & DICKERSON, R. E. 1991. The molecular structure of wild-type and a mutant Fis protein: relationship between mutational changes and recombinational enhancer function or DNA binding. *Proc Natl Acad Sci U S A,* 88**,** 9558-62.

ZHANG, X., REEDER, T. & SCHLEIF, R. 1996. Transcription activation parameters at ara pBAD. *J Mol Biol,* 258**,** 14-24.

ZHANG, Y., FENG, Y., CHATTERJEE, S., TUSKE, S., HO, M. X., ARNOLD, E. & EBRIGHT, R. H. 2012. Structural basis of transcription initiation. *Science,* 338**,** 1076-80.

ZHAO, K., LIU, M. & BURGESS, R. R. 2005. The global transcriptional response of Escherichia coli to induced sigma 32 protein involves sigma 32 regulon activation followed by inactivation and degradation of sigma 32 in vivo. *J Biol Chem,* 280**,** 17758-68.

ZHOU, Y., PENDERGRAST, P. S., BELL, A., WILLIAMS, R., BUSBY, S. & EBRIGHT, R. H. 1994. The functional subunit of a dimeric transcription activator protein depends on promoter architecture. *EMBO J,* 13**,** 4549-57.

ZUO, Y., WANG, Y. & STEITZ, T. A. 2013. The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex. *Mol Cell,* 50**,** 430-6.