RNA polymerase-DNA interactions at complex gene regulatory regions

by

Shivani Shatrughana Singh

Thesis submitted to University of Birmingham for the degree of

Doctor of Philosophy

March, 2014

School of Biosciences
University of Birmingham
Edgbaston
Birmingham
B15 2TT
United Kingdom

Acknowledgements

I would like to thank my supervisor, David Grainger, for his guidance and for being a great supervisor. I would like to thank all the members of ‘Power Graingers’, past and present, for making work as fun as it can be. In particular, I would like to thank Kiran for his sarcastic humour and James for his ‘hard to define’ humour. I would also like to thank Yousra and Yui for being super understanding and listening to me moan and complain especially in the last few months. Huge thanks to all of you for being awesome friends.

I am what I am thanks to my family’s continuous love and support. Thank you ma and pa for your blessings and always believing in me. Special thanks to Di and Amit for tolerating me over weekends and trying to show me the bigger picture always. Thank you Bhai for the never ending laugh riots and Suvir for just being there.

Finally, I would like to thank Sunindra, my husband. Thank you for being a perfect partner and standing by me as a rock!

Table of Contents

Synopsis

RNA polymerase must recognise and bind to specific DNA elements in order to initiate transcription. This binding specificity is mediated by the RNA polymerase σ factor. At regulatory regions, with more than one σ factor binding site, RNA polymerase has to distinguish between the different targets to initiate transcription correctly. Moreover, at regulatory regions that are very AT rich, RNA polymerase faces further difficulties because the σ factor binding elements are also AT-rich. Two regulatory regions are studied in this work: i) the *cbpA* regulatory DNA that has overlapping binding sites for σ^{70} and σ^{38} associated RNA polymerase ii) the regulatory region for *ehxCABD* operon that has an AT content of 71 % and contains many "decoy" RNA polymerase binding sites. At both of these regulatory regions I have examined how correct RNA polymerase binding is ensured.

For the *cbpA* regulatory region, containing overlapping promoters, it was found that the shared promoter spacer region played a key role. Briefly, the promoter spacer region is the section of DNA between the elements recognised by the σ factor. The sequence of the spacer is generally considered unimportant. Using random mutagenesis, I identified a location in the shared spacer region that differently affected the overlapping *cbpA* promoters. Further investigation confirmed a base preference of T>A>C>G at this position of the spacer. I determined that σ^{70} side chain R451 senses the base sequence changes. I propose that alterations in spacer sequence modulate conformation, making it either easier, or more difficult, for R451 to interact with the DNA. Interestingly, it was found that R451 only plays a role at weak promoters. Thus, σ side chain R451 is not required at promoters with a high affinity σ binding element. Using a single molecule analysis, I also measured the DNA compaction properties of the *cbpA* gene product; curved DNA binding protein A (CbpA).

CbpA is known to form aggregates with DNA and I was able to monitor this aggregation using tethered particle motion analysis.

At the *ehxCABD* regulatory region, which has an AT content of 71 %, many sequences that resemble σ factor binding elements were found. I discovered that RNA polymerase is capable of binding to the correct promoter elements in this region only in the presence of a chromosome folding protein called H-NS. This protein binds AT-rich DNA and was found to "coat" the *ehxCABD* regulatory region. Thus, in presence of H-NS, RNA polymerase binds to the *ehxCABD* regulatory DNA with greater specificity. Finally, many intragenic promoters within the *ehxCABD* operon were identified. We thus propose that H-NS plays a role in silencing this pervasive intragenic transcription.

Index of Tables

Index of Figures

Introduction

Chapter 1

1.1 Transcription initiation in *Escherichia coli* (*E. coli*)

The first step in gene expression is transcription. This involves synthesis of an mRNA that is complementary to the template strand of DNA. Information in this mRNA is then used to make a polypeptide (Figure 1.1). The resulting polypeptides may participate in any aspect of cell function. Thus, by regulating gene expression, bacteria can control all aspects of cell biology.

1.1.1 Bacterial RNA polymerase

Transcription is catalyzed by RNA polymerase (RNAP). In order to transcribe a gene, RNA polymerase must first recognise a "promoter". Hence, promoters are located upstream of the gene to be transcribed. Once bound to a promoter, RNA polymerase initiates separation of the DNA strands thus forming an 'open complex'. RNA polymerase is therefore responsible for two key steps in transcription initiation; promoter binding and open complex formation (Griffiths *et. al.,* 2005).

RNA polymerase core enzyme (E) is made up of five subunits; two identical alpha (α) subunits, one beta (β) subunit, one beta prime (β') subunit, and one omega (ω) subunit (Figure 1.2). The α amino-terminal domain (αNTD) and the α carboxyl-terminal domain (αCTD) are connected to each other by a flexible linker. The αNTD drives α subunit dimerisation. Additionally, αNTD interacts with other RNA polymerase subunits, to stimulate assembly of the core enzyme (Igarashi *et. al*., 1991; Kimura *et. al*., 1994). The αCTD can play a major role in DNA binding during transcription initiation (Figure 1.2) (Browning and Busby, 2004). The β and β' subunits form the RNA polymerase active site that catalyse the synthesis of RNA. Thus, β and β' drive both initiation and elongation of the mRNA chain. As part of the core enzyme β and β' are capable of binding non-specifically to DNA (Young *et. al.*, 2001).

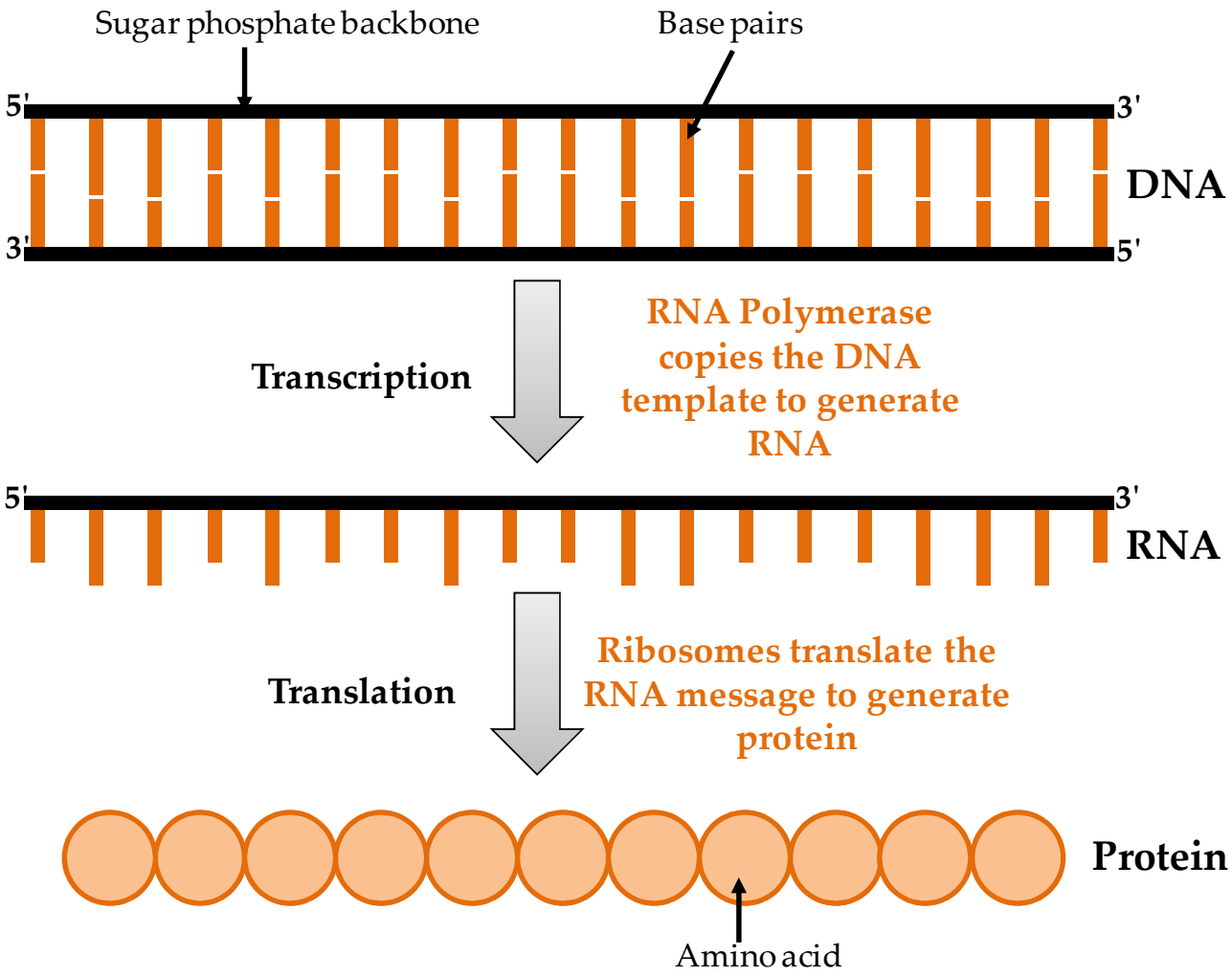


Figure 1.1 Overview of gene expression

The DNA is shown as a ladder where the individual steps (orange) are base pairs and the rails (black) represent the sugar phosphate backbone. Proteins consist of amino acids (orange circles) joined by peptide bonds.

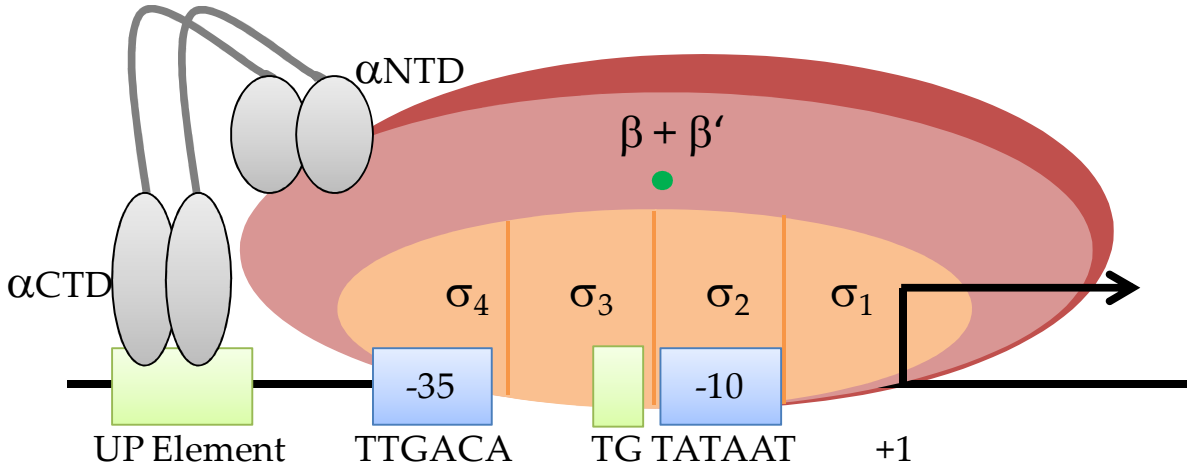


Figure 1.2 RNA polymerase holoenzyme interacting with a promoter

A cartoon based on crystallographic and genetic studies of RNA polymerase-promoter interactions. The black line indicates DNA. The -10 and -35 elements are highlighted in blue. The transcription start site "+1" is shown by an arrow. The TG motif and UP elements are highlighted in green. The β and β' subunits of RNA polymerase are coloured in light and dark red, respectively. The αNTDs and αCTDs are shown in grey and different domains of σ are highlighted in orange. The Mg^{2+} ion, (green dot) represents the RNA polymerase active site (Based on Browning and Busby, 2004).

The ω subunit facilitates assembly and maturation of the β' subunit and renders RNA polymerase sensitive to guanosine tetraphosphate (Matthew and Chatterji, 2006; Hampsey, 2001). Guanosine tetraphosphate (ppGpp) is an intracellular signalling molecule that enables bacteria to regulate gene expression based on environmental cues (Ross *et. al.*, 2013). The ppGpp binding site locates to the interface between β' and ω subunit (Mechold *et. al.*, 2013). Thus, by binding at the β'/ω interface, ppGpp weakens RNA polymerase-DNA interactions. This inhibits the transition from a closed to an open initiation complex (Ross *et. al.*, 2013, Zuo *et. al.*, 2013).

1.1.2 Sigma factors

The RNA polymerase core enzyme can catalyze RNA synthesis but cannot bind to promoter elements specifically. DNA binding specificity is conferred by the RNA polymerase sigma (σ) factor. Thus, the core enzyme must interact with the σ factor before binding promoter DNA. RNA polymerase associated with a σ factor is known as the holoenzyme (Eσ) and initiates transcription specifically (Wosten, 1998). Multiple σ factors are present in bacteria with each recognising different subsets of promoters. Thus, a simple mechanism via which bacteria can switch transcription between different sets of genes is to use alternative transcription factors (Browning and Busby, 2004).

There are 7 σ factors in *E. coli*:

1. σ^{70} (RpoD) - the "housekeeping" sigma factor (Burgess *et. al.*, 1969)
2. σ^{54} (RpoN) - the nitrogen-limitation sigma factor (Hunt and Magasanik, 1985)
3. σ^{38} (RpoS) - the starvation/stationary phase sigma factor (Mulvey *et. al.*, 1990)
4. σ^{32} (RpoH) - the heat shock sigma factor (Grossman *et. al.*, 1984)

5. σ^{28} (RpoF) - the flagellar sigma factor (Arnosti and Chamberlin, 1989)

6. σ^{24} (RpoE) - the extracytoplasmic/extreme heat stress sigma factor (Erikson *et. al.*, 1987)

7. σ^{19} (FecI) - the ferric citrate sigma factor, which regulates the *fec* gene for iron transport (Pressler *et. al.*, 1988)

1.1.3 RNA polymerase-DNA interactions

The first step in the process of transcription initiation by RNA polymerase is sequence specific recognition of the promoter by the σ factor. The two major recognition elements for σ^{70} are the -10 element (or the Pribnow box, 5'-TATAAT-3') and the -35 element (5'-TTGACA-3'). The -10 and -35 sequences are located ~ 10 bp and ~ 35 bp upstream of the transcription start site (+1) respectively. Note that both the consensus sequence and position of promoter elements, for the different σ factors varies. Once bound to DNA, the σ factor further contributes by playing a major role in the formation of the open complex (Browning and Busby, 2004).

Promoter recognition by RNA polymerase can be improved by additional DNA sequence elements. Thus, the AT-rich UP element, ~ 20 bp in length, is found upstream of the -35 hexamer at some promoters. The RNA polymerase αCTD recognises and binds to UP elements. The αCTD residues that make contact with UP element DNA are grouped and collectively referred to as the "265 determinant" (that includes residues R265, N268, N294, G296, K298, S299 and E302) (Busby and Ebright, 1999). The 265 determinant of αCTD interacts with the DNA backbone, along the minor groove, leading to tighter RNA polymerase-promoter interactions (Rao *et al.,* 1994; Ross *et. al.*, 2001).

Some -10 elements are "extended" and have the sequence 5'-TGnTATAAT-3'. Thus, the extended -10 element is signified by the presence of a "TG motif" at promoter positions -14 and -15. This "TG motif" is present in ~ 20 % of *E. coli* promoters and significantly increases promoter activity. The "TG motif" has been proposed to compensate for a poor -35 sequence and is recognised by residues I439 and R441 in σ^{70} region 2.4. Residues H455 and E458, in σ^{70} region 3.0, also contribute (Figure 1.3) (Burr *et. al.,* 2000; Sanderson *et. al.*, 2003). There are no promoters with a full complement of perfect binding elements for RNA polymerase in *E. coli*. This is likely because such a promoter would be difficult to regulate and may bind RNA polymerase too tightly to allow efficient promoter escape.

1.1.4 σ factor domain organisation

A σ factor has three structured domains; σ_2, σ_3, and σ_4. These domains consist of conserved regions 1.2-2.4, 3.0-3.1 and 4.1-4.2 respectively. Region 1.2-2.4 in σ_2 is the most highly conserved region in the σ^{70} family of σ factors (Figure 1.3). Consequently, residues L384, V387, L402, D403, Q406, E407, N409 and M413 (in σ region 2.1-2.2) along with M487, P504 and S506 (in region 3.1) play a critical role in binding of σ to the core RNA polymerase (Figure 1.3). Moreover, region 2.4 residues Q437, T440 and R441 recognise and bind to the promoter -10 element. Feklistov and Darst (2011) showed that σ_2 interacts with bases on the non-template strand. Recognition of the -35 promoter element is mediated by region 4.2 of σ_4. However, as mentioned above, interactions with the TG motif can substitute for this contact. The TG motif is recognised by σ_3 region 3.0 (Figure 1.3). Residues Y430, W433 and W434, in region 2.3, contribute to unwinding of the promoter DNA (Figure 1.3). Thus, Campbell *et. al.* (2002) found that σ regions 1.2-3.1 are sufficient to mediate DNA binding and melting of

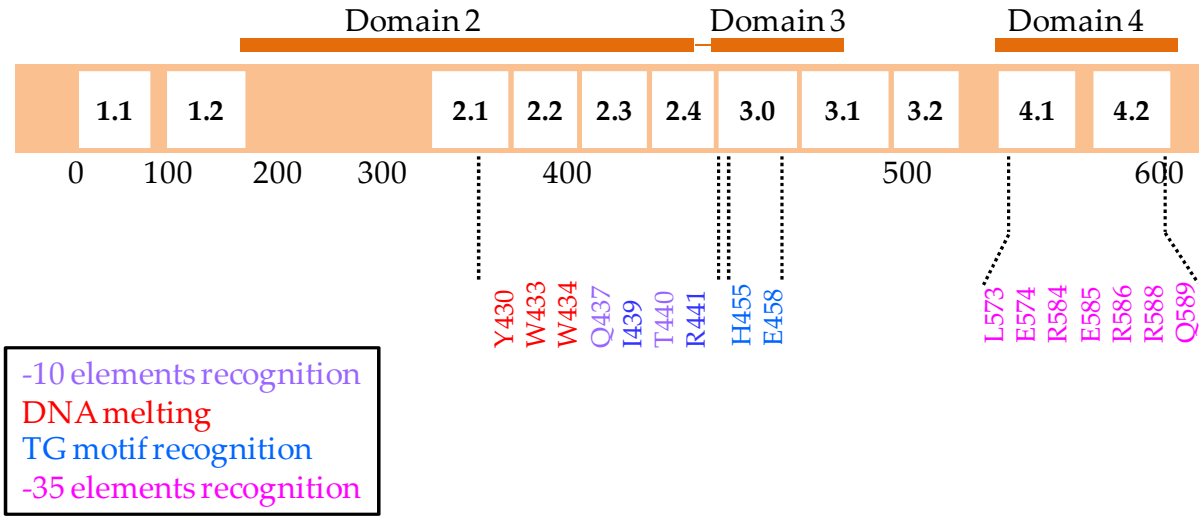


Figure 1.3 σ^{70} residues that interact with different promoter elements

The σ^{70} residues that interact with various promoter elements and drive DNA melting, are highlighted. The orange bar represents σ^{70} with white boxes representing the different regions of the four domains. The dark orange bar at the top highlights the different domains. The numbers below the σ^{70} schematic represent amino acid residues. The colour of the residue corresponds to the promoter element it interacts with and/or the role it plays in DNA melting (Based on Campbell *et. al.*, 2002).

the DNA around the transcription start site. Please note that the residue numbering given here refers to σ^{70} although the structural work has been carried out on Thermus aquaticus (Taq) σ^{A}.

1.1.5 Overlapping promoters for different σ factors are common

As mentioned above, σ factors regulate transcription of specific sets of genes under specific conditions. For many years it was thought that the set of genes, regulated by each σ factor, was distinct. However, for *E. coli*, it is becoming increasingly apparent that this is not the case. Thus, the sets of genes regulated by different σ factors in *E. coli* overlap (Wade *et. al.*, 2006). Consequently genes, which are transcribed by multiple holoenzymes, often have overlapping sets of promoter elements specific to each σ factor.

In vitro, σ^{70} and σ^{38} are known to recognise identical promoter elements. However, in vivo, σ^{38} and σ^{70} are capable of differentiating and binding specific promoters based on difference in the consensus sequence. Typas and Hengge (2006) showed that at naturally occurring σ^{38} dependent promoters, σ^{38} outcompetes σ^{70} by sensing the spacer length. These promoters contain -35 elements that are non-optimally spaced which enhances σ^{38} binding via regions 2.5 and 4.

Promoter elements recognised by σ^{32} (the heat-shock specific σ factor) comprise 5'-CCCCATNT-3' around position -10 and 5'-CTTGAAA-3' around position -35 (Wade *et. al.*, 2006). Strikingly, over 99 % of σ^{32} targets are located within 300 bp of σ^{70} targets. Hence, it is likely that most σ^{32} regulated genes can also be targeted by σ^{70}. As the upstream half of the σ^{32} -35 consensus (5'-CTTGAA-3') is quite similar to the σ^{70} -35 consensus hexamer (5'-TTGACA-3'), many promoters for σ^{32} and σ^{70} could overlap. Wade *et. al.* (2006) selected three σ^{32} promoters and measured binding with σ^{70} before and after heat shock. It was found

that σ^{70} binding decreased about two-fold after heat shock while σ^{32} binding increased. Also, σ^{70} was capable of initiating transcription from the same nucleotide, and as efficiently as σ^{32}, at five heat shock promoters. Zhou *et. al.* (1988) found that cells lacking *rpoH* (σ^{32}) were capable of transcribing heat-shock genes. The widespread overlap between σ^{32} and σ^{70} regulated genes suggests that the heat-shock genes in *rpoH* deficient cells are being transcribed by σ^{70} (Wade *et. al.*, 2006).

Another σ factor, σ^{24}, also has targets that extensively overlap with σ^{70} promoters. Hence, ~ 40 % of σ^{24} promoters are also bound by σ^{70} (Wade *et. al.*, 2006). Five σ^{24} dependent promoters; (*rpoH*, *mdoG*, *ycbK*, *sixA* and *yieE*) bound to σ^{70} *in vivo*. The binding was measured before and after the cells were heat shocked. Moreover, σ^{70} binding at two of these five promoters (*sixA* and *mdoG*) increased substantially after the cells were heat shocked. Another two σ^{24} dependent promoters, *rpoE* and *yfiO*, produced transcription products with σ^{70} *in vitro* (Wade *et. al.*, 2006). Thus, targets for σ^{70} overlap with σ^{32} targets and targets for other σ factors like σ^{24}. This means expression of a gene, in different growth conditions, can be induced by different σ factors. This phenomenon is not limited to *E. coli*. In *B. subtilis*, three promoters are already known to be regulated by both the housekeeping σ factor and an alternative σ factor (Popham and Setlow, 1993; Ramirez *et. al.*, 2004; Paul *et. al.*, 2004).

1.2 Regulation by Transcription Factors

Although gene regulatory networks controlled by σ factors contain more layers of complexity than previously thought, they cannot explain all gene regulatory responses. Hence, transcription factors also play a role. Transcription factors up or down regulate gene expression in response to environmental signals. Thus, transcription factors bind to specific

consensus sequences at target promoters and this binding is often controlled by a specific ligand. In *E. coli*, DNA binding proteins, encoded by more than 300 genes, represent a set of transcription factors that may be activators, repressors or dual regulators. Some regulators, termed "global" regulators, each control expression of hundreds of genes. Conversely, specific regulators control only a handful of genes.

1.2.1 Transcription activation factors

No promoters have a perfect match to all of the elements that facilitate RNA polymerase binding. Thus, many promoters require additional factors, known as transcription activators, to stimulate transcription initiation. Transcription activators are sequence specific DNA binding proteins whose activity is regulated in response to environmental factors. The cAMP receptor protein (CRP) and fumarate and nitrate reductase (FNR) are two of the best characterized global transcriptional regulators in *E. coli* (Savery *et. al.,* 1996b; Browning *et. al.,* 2002). Transcription activation factors typically bind upstream of the core promoter elements and make contacts with RNA polymerase. Promoters regulated by activators can be divided into three groups based on these contacts:

(i) Class I activator dependent promoters

At Class I activator dependent promoters, the activator protein binds upstream of the core promoter elements and functions by enhancing the binding of the RNA polymerase αCTD to the promoter (Figure 1.4a). The position of Class I activator binding sites is not fixed as αCTD is connected to core RNA polymerase by a flexible linker (Blatter *et. al*., 1994; Zhou *et. al*., 1994). The αCTD residues not required for αCTD-DNA interactions, but necessary for transcription at Class I promoters, are grouped and known as the "261 determinant" (that includes residues V257, D258, D259 and E261) or the "287 determinant" (that includes

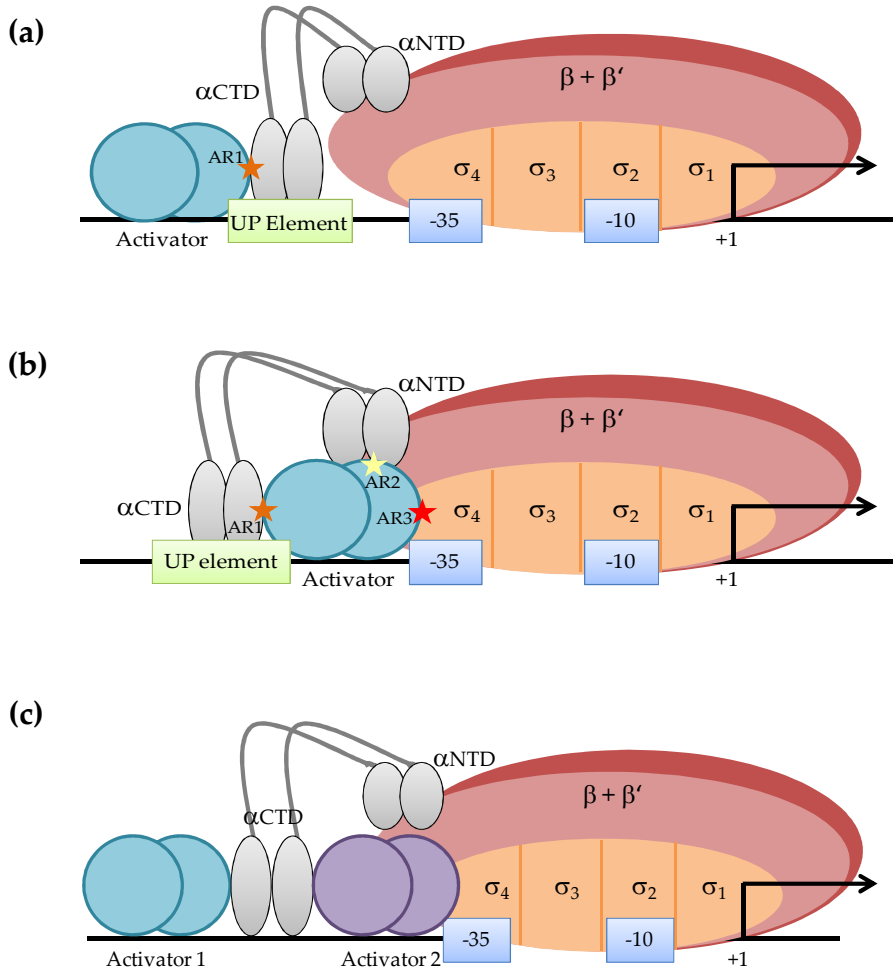


Figure 1.4 Activation at simple promoters

The Figure illustrates activation of simple promoters by transcriptional activators. Activators are shown as dimers (blue or purple ovals) and subunits of RNA polymerase are labelled.

a. Class I activation: activator binds upstream of the core promoter and enables RNA polymerase binding by making a contact with αCTD (AR1)
b. Class II activation: activator binds at or around promoter position -41.5, adjacent to the -35 element, and interacts at AR1 as well as with αNTD (AR2) and domain 4 of σ^{70} (AR3).
c. Class III activation: one or more activator proteins are bound and function through a combination of Class I and Class II mechanisms (Based on Browning and Busby, 2004).

residues T285, E286, V287, E288, L289, G315, R317 and L318). These determinants facilitate protein-protein interactions between αCTD and the transcription activator (Busby and Ebright, 1999). Activation of the *lac* promoter by CRP is an example of Class I regulation. In this case, the CRP binding site is centered at promoter position -61.5. When bound at this site, activating region 1 (AR1) of CRP interacts with one of the two copies of αCTD via the 287 determinant (Figure 1.4a). This protein-protein interaction facilitates the 265 determinant, of same αCTD subunit, to bind to promoter UP elements. Thus, CRP aids recruitment of RNA polymerase at the *lac* promoter (Tang *et. al*., 1994). Crucially, the Class I activator binding site must reside on the same face of the DNA helix as RNA polymerase; else the activation will be lost (Newlands *et. al*., 1992). Another example of Class I activation is found at the *malT* promoter. Here, the binding site for CRP is situated further away from the -35 promoter element at position -70.5. Moreover, an additional inhibitory binding site for αCTD is situated at position -47. AR1-αCTD, and αCTD-DNA interactions, similar to those at the *lac* promoter inhibits αCTD binding to the inhibitory site and thus avoids formation of a non-productive RNA polymerase-promoter complex (Tagami and Aiba, 1999).

(i) Class II activator dependent promoters

In a Class II situation the activator protein binds proximal to the core promoter elements (Figure 1.4b). Thus, the activator has more opportunities to interact with the transcriptional machinery. However, the positioning of the activator binding site is less flexible than at Class I promoters. This is because Class II activators make contact with domain 4 of the σ^{70} subunit of RNA polymerase, the positioning of which is fixed by the promoter -35 element. Thus, the binding sites of Class II activators generally overlap the -35 promoter element and are centered at promoter position -41.5 (Figure 1.4b) (Dove *et. al*., 2003).

An example of Class II activation by CRP is found at the *galP1* promoter. Here, CRP interacts with RNA polymerase at two positions. First, AR1 of CRP interacts with 287 determinant of one of the two αCTDs. This enables the 265 determinant of αCTD to bind UP elements situated further upstream. Second, activating region 2 (AR2) of CRP interacts with residues 162-165 of αNTD (Figure 1.4b) (Busby and Ebright, 1999). Another example of Class II activation is activation by FNR at the *narG* promoter. At this promoter, FNR contacts αCTD, and σ^{70} domain 4, via AR1 and AR3 respectively (Figure 1.4b) (Lamberg and Kiley, 2000; Blake *et. al.,* 2002).

(ii) Class III activator dependent promoters

A third mechanism of activation requires RNA polymerase to interact with two or more activators. This can either be multiple copies of the same activator or a combination of different activators. Thus, Class III promoters function through a mixture of Class I and Class II mechanisms (Figure 1.4c) (Barnard *et. al.*, 2004). Unsurprisingly, Class III activated promoters have diverse architectures with varying positions for activator binding sites. An example of a Class III activated promoter is the *melAB* promoter, which is activated by combination of MelR and CRP. MelR, a specific activator, binds as a dimer to pairs of sites (1, 1′ and 2, 2′) separated by a site for CRP, a global activator (Figure 1.5) (Grainger *et. al.*, 2003). In the presence of CRP, MelR bound to site 2' acts via a Class II mechanism to make contacts with σ^{70} domain 4 (Figure 1.5) (Grainger *et*. *al*., 2004). However, CRP is capable of binding only when MelR is prebound to sites 1 and 2. Thus, at this promoter, two signals (i.e. melibiose and cAMP) are integrated to regulate transcription (Wade *et. al.*, 2001).

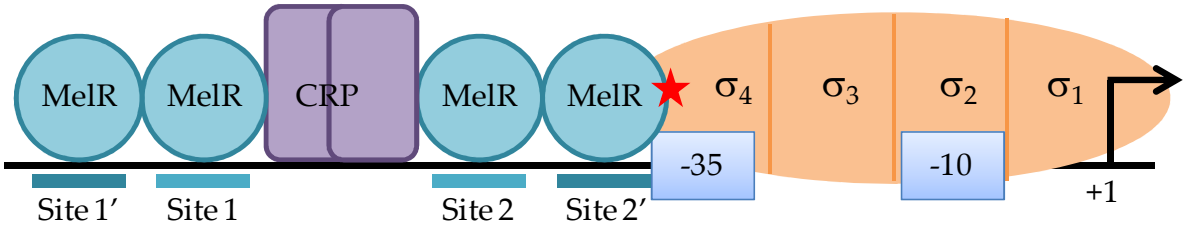


Figure 1.5 Regulation at *melAB* promoter

The figure shows activation at the *melAB* promoter with the help of MelR (blue circles) and CRP (purple rectangles). Transcription is activated when CRP and MelR are bound co-operatively. The relative locations for MelR and CRP binding are shown. In the presence of CRP, MelR acts as a class II activator and makes contact with σ^{70} domain 4 (AR3), shown as red star (Based on Barnard *et. al*., 2004).

1.2.2 Regulation by transcriptional repressors

An alternative mechanism for the precise control of gene expression is transcriptional repression. There are three simple mechanisms used by repressors to block transcription initiation at target promoters; steric hindrance of RNA polymerase, steric hindrance of transcriptional activators, and DNA looping (Browning and Busby, 2004). Steric hindrance requires the repressor to bind at a position overlapping either the core promoter elements or an activator binding site. By doing this, the repressor physically blocks the binding of other key proteins (Figure 1.6a) (Müller-Hill, 1996). DNA loops are formed by repressor proteins interacting when bound at distal sites. This leads to repression by excluding RNA polymerase from, or trapping RNA polymerase at, a promoter in the intervening DNA (Figure 1.6b). Both steric hindrance and DNA looping are observed at the *lac* promoter. The *lac* repressor is a product of the *lacI* gene and binds to operator sequences overlapping the RNA polymerase binding site at the *lac* promoter. Because LacI-DNA interactions are controlled by lactose, the *lac* operon is repressed and de-repressed in the presence and absence of lactose respectively (Figure 1.7) (Bell and Lewis, 2001). Additionally, the *lac* repressor binds to O1, and either O2 or O3, to form repression loops (Figure 1.7b) (Wilson *et al.,* 2007).

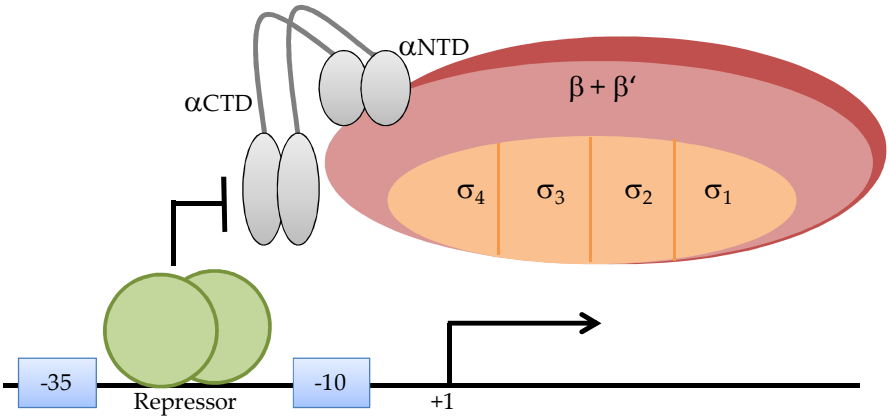
1.3 Global Studies of transcription initiation

Many biochemical and genetic approaches have revealed the location of individual promoters. More recently, the availability of genome sequences, and methods to rapidly analyse them, have allowed transcription to be studied on a global scale for the first time.

1.3.1 Chromatin Immunoprecipitation

A powerful method to identify promoters on a chromosome-wide scale is chromatin

(a)



(b)

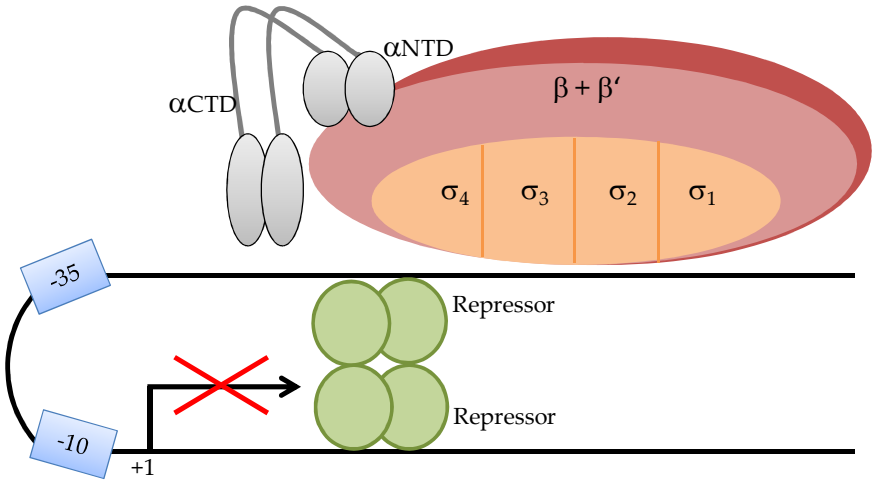


Figure 1.6 Repression at promoters

a. Repression by steric hindrance: repressor protein physically blocks the binding of RNA polymerase

b. Repression by forming loops: repressor proteins bind at distal sites and interact to form a DNA loop thus blocking RNA polymerase binding or trapping RNA polymerase at a promoter (Based on Browning and Busby, 2004).

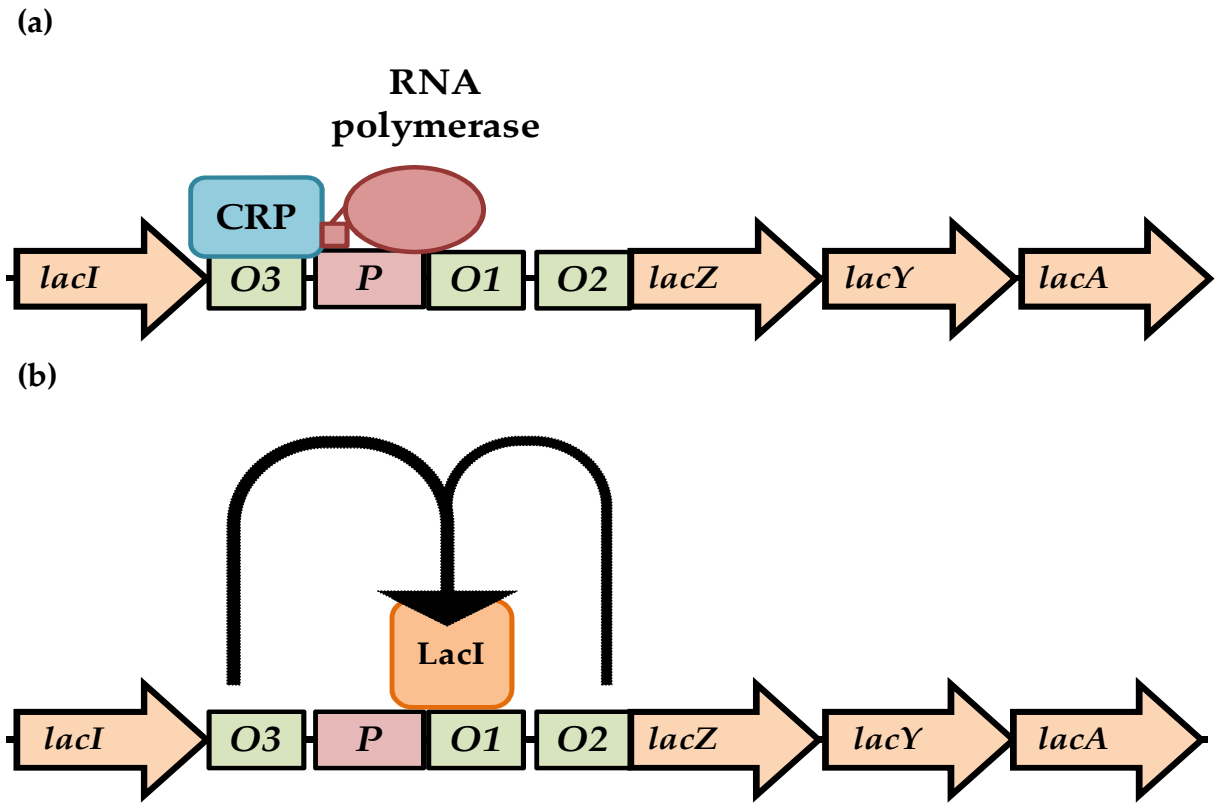


Figure 1.7 Regulation of the *lac* operon

a. A high *lacZYA* expression level is achieved when glucose is absent and lactose is available. In this situation, CRP stimulates binding of RNA polymerase to the lac promoter (P)

b. When glucose levels are high and lactose absent, the LacI repressor protein binds to O1 and forms repression loops by also interacting with O2 and O3.

immunoprecipitation (ChIP) coupled with DNA microarray analysis (ChIP-chip) or whole genome sequencing (ChIP-seq). ChIP is a technique that measures protein-DNA interactions *in vivo*, irrespective of their effects on transcription. In a ChIP experiment, proteins are covalently attached to their target DNA sequence by fixing cells with formaldehyde. Following cross-linking, the nucleoprotein is extracted from cells and sonicated to obtain DNA fragments. An antibody is then used to immunoprecipitate the protein of interest and the cross-linked DNA. The protein-DNA cross-links are then reversed. Finally, the DNA fragments are labelled and hybridised to a suitable DNA microarray or sequenced. A control sample, from a mock immunoprecipitation, is usually prepared in parallel for comparison. Once the immunoprecipitated DNA has been characterised, protein binding targets, across the genome, can be mapped. Figure 1.8 explains the step by step process of a ChIP-chip experiment. Recent studies, using ChIP-chip, have mapped genome-wide location of RNA polymerase (Herring *et. al.*, 2005) and transcription factors (Grainger *et. al.*, 2005). They have also provided an insight into the transition between transcription initiation and elongation (Reppas *et. al.*, 2006).

(i) ChIP reveals binding sites for RNA polymerase in unusual locations

In their analysis, Herring *et. al.* (2005) mapped RNA polymerase binding in *E. coli* cells treated with rifampicin. Rifampicin is capable of diffusing through the bacterial cell membrane where it binds to RNA polymerase (Herring *et. al.*, 2005). Rifampicin associated RNA polymerase is able to bind DNA but unable to synthesise RNA beyond 2 to 3 nucleotides in length (McClure *et. al.*, 1978). Thus, Herring *et. al.* identified 1139 binding sites for RNA polymerase out of which 721 were previously unidentified targets. Of these 1139 RNA polymerase binding sites, 501 were located within intergenic regions and 638 in open reading frames. This suggests that transcription regulation is more complicated than

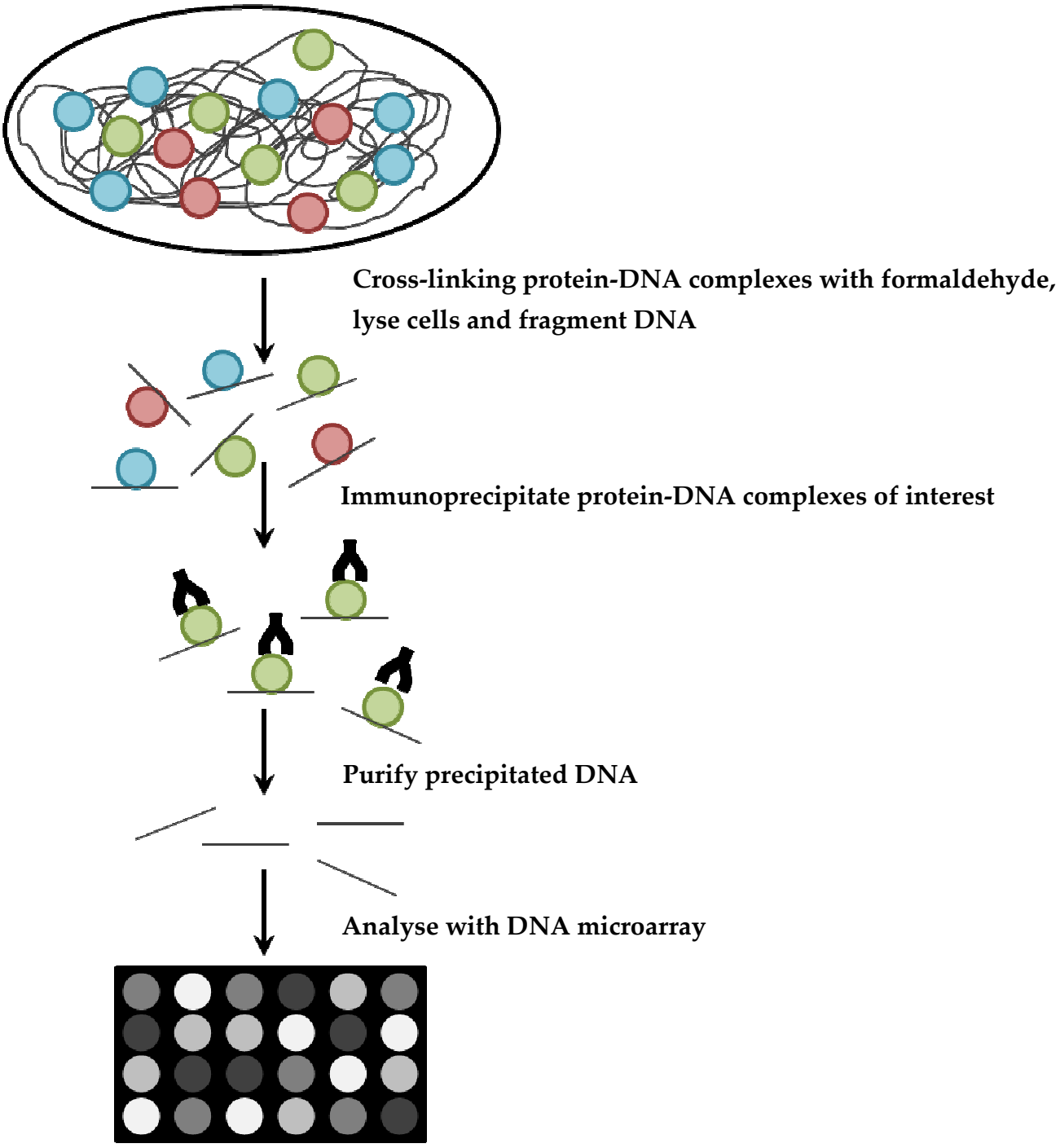


Figure 1.8: Transcription factor-binding site analysis using ChIP-chip technique

The figure shows a step-by-step account of the ChIP-chip technique. Blue, red and green circles indicate protein bound to DNA (black line) (Based on Grainger and Busby, 2008 b)

previously thought and that many genes may contain promoters. Out of 368 σ^{70}-specific promoters, 74 % (273) were detected within 1000 bp upstream of an expressed gene. Although, they had a 26 % negative-detection rate (i.e. many known promoters were not detected), this study provided a strong base for further unexpected aspects of transcription to be uncovered in *E. coli*.

(ii) ChIP analyses of RNA polymerase occupancy

Extending the work done by Herring *et. al.* (2005) Grainger *et. al.* (2005) used the ChIP-chip technique to detect RNA polymerase binding targets under specific conditions, for example, in the presence of IPTG or salicylic acid. When cells were treated with IPTG, the data showed an induction in RNA polymerase binding to the *lacZYA* operon. Interestingly, more RNA polymerase was located towards the 5' end of the operon. Similarly, in the presence of salicylic acid, RNA polymerase was redistributed from genes encoding stable RNA, towards genes encoding proteins for stress response. In this study, Grainger *et. al.* (2005), also identified 68 specific binding targets for the transcriptional regulator, CRP as well as many thousands of weak DNA sites. This background binding could be due to CRP acting as a chromosome shaping protein. This study also mapped RNA polymerase distribution in rapidly growing cells. Approximately 90 transcription units, including many genes required for protein synthesis, are the major RNA polymerase binding targets.

Reppas *et. al.* (2006) used the ChIP-chip technique to investigate the transition between transcription initiation and elongation. Hence, they determined the genomic binding profiles of β and σ^{70}. The study finds that, after the transition from initiation to elongation, the majority of σ^{70} is released very rapidly from RNA polymerase. Thus, only a small fraction of elongating RNA polymerase is associated with σ^{70}. This study also identified 1286 σ^{70}

binding targets and, at 300 of these sites, RNA polymerase bound promoters were not associated with any transcriptional activity. The melting points (T_m), of these promoters are higher than expected and this could act as an intrinsic energetic barrier to promoter escape. It could also be that RNA polymerase is only able to initiate transcription under appropriate environmental conditions. This study demonstrates that the transition from transcription initiation to elongation is rate-limiting and highly variable.

1.3.2 Pervasive transcription

As described above, ChIP has identified unusual binding locations for RNA polymerase and transcription factors. More recently, next generation sequencing has shown that transcription can be initiated from unusual locations, such as within genes and in the anti-sense direction (Selinger *et. al*. 2000; Dornenburg *et. al*., 2010; Thomason and Storz , 2010; Raghavan *et. al*., 2012). This is sometimes referred to as pervasive transcription. Pervasive transcription describes a model whereby genomes are subjected to low level non-canonical transcription both in the sense and antisense orientations. Inefficient termination, promoters within genes, and overactive intergenic regions, may all contribute to the phenomenon of pervasive transcription.

(i) Pervasive transcription due to promoters within genes

As discussed above, many transcription factor and RNA polymerase binding sites have been identified within genes (Herring *et. al.*, 2005; Grainger *et. al.*, 2005; Reppas *et. al.*, 2006). Hence, one source of pervasive transcription is transcription that initiates within genes. RNA resulting from antisense intragenic transcription (aRNAs) is capable of transcriptional interference by base pairing with the corresponding mRNA (Figure 1.9). As aRNAs are present in wide range of bacterial species, pervasive antisense transcription could play an

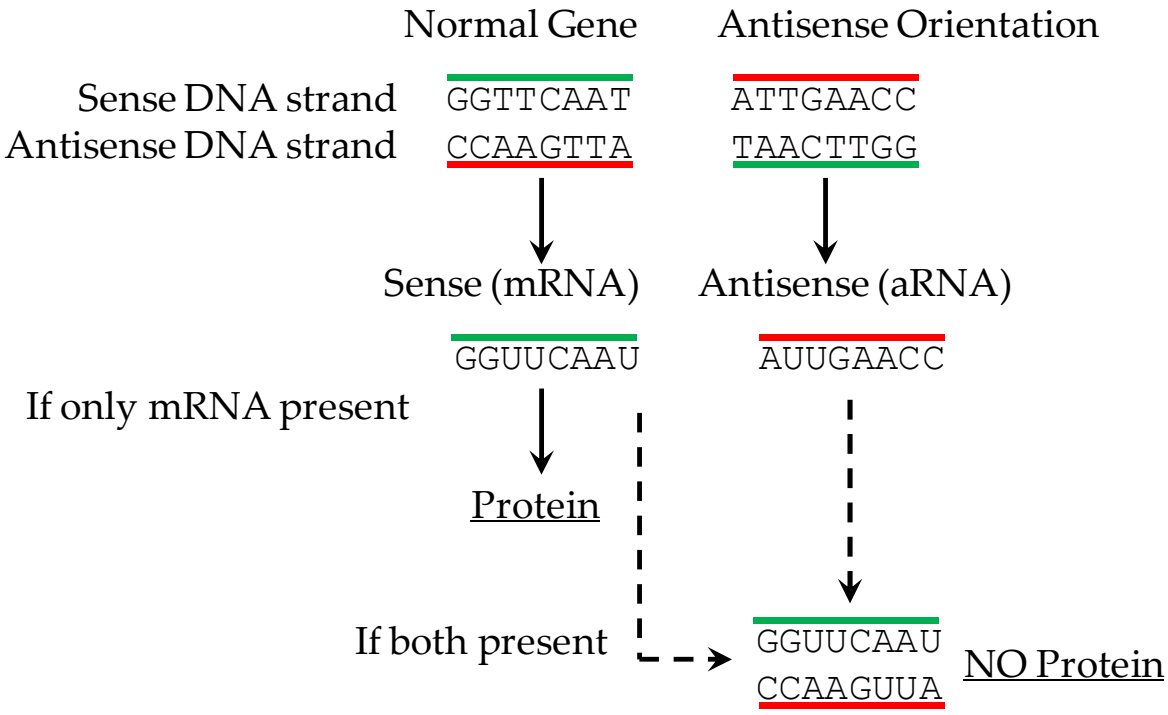


Figure 1.9: Model outline for downregulation of a gene by antisense RNA

(Based on Tucker, 1996)

important role in regulating gene expression in all bacteria (Dornenburg *et. al.*, 2010). Conversely, pervasive anti-sense transcription may simply represent "noise" that needs to be silenced.

Widespread antisense transcription in bacteria was first detected in *Helicobacter pylori* (*H. pylori*). Hence, 27 % of the transcription start sites (TSS) in this bacterium are antisense. Furthermore, 46 % of all ORFs have at least one associated antisense TSS. Consistent with this, antisense regulation of genes for both surface structure, and host cell interactions, has been observed in *H. pylori*. This hints that antisense transcription throughout the *H. pylori* genome could be an important, and poorly understood, aspect of transcriptional regulation in this organism (Sharma *et. al.*, 2010). Subsequently, Dornenburg *et. al.* (2010) identified about 1005 antisense TSS in *E. coli*. Strikingly, the DNA sequence properties of canonical promoters, and those promoters associated with antisense TSS, are identical. Thus, the -10 element score of 471 published canonical promoters is 3.28 out of 6. Similarly, for the 1005 antisense TSS, a score of, 3.27 out of 6 was observed. Of published TSS in *E. coli* 46 % have an 'A' at position +1. This compares well with 48 % of the 1005 antisense TSS that also have 'A' at position +1. As expected, mutating the -10 hexamer in 9 out of 10 antisense TSS in *E. coli* resulted in reduced transcription. Interestingly, Dornenburg *et. al*. (2010) also found that many aRNAs regulate expression of the overlapping gene. For example, *rplJ* overlaps a putative aRNA and mutating the -10 hexamer/+1 start site of the aRNA resulted in increased *rplJ* gene expression.

(ii) Inefficient transcription termination

Transcription termination is either factor dependent or factor independent. Factor dependent termination requires proteins like Rho, NusG and NusA. This ensures transcripts that are not being translated, like the ends of various genes and operons, are terminated (Peters *et. al.*,

2011). When termination is not efficient it can result in R-loops. Here, the untranslated transcript reanneals with the 5' upstream DNA. R-loops are known to affect many essential functions, like blocking the DNA replication fork. Hence, R-loops can be lethal (Leela *et. al.*, 2013). Peters *et. al.* (2012) found that Rho and NusG suppress pervasive antisense transcription in *E. coli*. Rho binds to C-rich mRNA and initiates transcription termination. At AT-rich DNA, NusG aids both sense and antisense Rho-dependent termination (Peters *et*. *al*., 2012). Thus, there is a direct relationship between inefficient transcription termination and pervasive transcription.

(iii) AT-rich DNA

Promoter elements, particularly the -10 hexamer, recognised by RNA polymerase are AT-rich. This is because the weak bonds associated with A:T base pairing facilitate unwinding of the DNA. Thus, DNA regions that are AT-rich are challenging for RNA polymerase to recognise correctly. This can result in RNA polymerase binding to cryptic promoters or transcription being initiated promiscuously. Hence, AT-rich DNA sequences are a frequent source of pervasive transcription (Sharma *et*. *al*., 2010).

1.4 The Prokaryotic Nucleoid

All the processes described in the sections above take place in the context of the folded chromosome. The single circular *E. coli* chromosome is embedded in the cytoplasm and is highly condensed. As this nucleoprotein-dense area is similar in function to the eukaryotic nucleus, it is known as the nucleoid. The *E. coli* K-12 genome contains about 4.6 Mbp of DNA and has a physical length of 1.6 mm. Several mechanisms are required to compact this DNA into a cell that is about 1-3 μm in length. The compaction results in DNA occupying only about 3 % of nucleoid mass (Dame, 2005; Thanbichler *et al.,* 2005). Until recently, very

little was known about the bacterial nucleoid. Thus, processes that maintain nucleoid organisation during DNA replication, transcription and chromosome segregation are currently of great interest (Reyes-Lamothe *et al.,* 2008). This interest has been stimulated by advances in high resolution imaging techniques and cell biology methods. These have helped to show that bacteria maintain a highly organised interior to support various cellular processes (Reyes-Lamothe *et al.,* 2008). Factors such as DNA supercoiling, macromolecular crowding, and nucleoid associated proteins, all play important roles in compacting bacterial DNA. Importantly, these factors also regulate the expression of genes via diverse mechanisms (Reyes-Lamothe *et al.,* 2008; Thanbichler *et al.,* 2005).

1.5 Nucleoid Associated Proteins (NAPs)

1.5.1 An overview of nucleoid associated proteins

Plant and animal cells use histone proteins to facilitate DNA packaging in the nucleus. The detailed understanding of histones, their ability to alter chromatin structure, and influence transcription, resulted in a search for similar proteins in bacteria. Hence, bacterial nucleoid proteins are referred to as 'histone-like' proteins. However, as the different nature of these bacterial proteins has been revealed, this term is becoming less appropriate. Collectively known as nucleoid associated proteins, based on their location inside the cell (Dillon and Dorman, 2010), these factors are involved in DNA folding, repair, replication and protection in prokaryotes (Dame, 2005). Nucleoid associated proteins differ from histones in that they do not form nucleosomes, where the DNA is wrapped around a protein core. Instead, nucleoid proteins use other mechanisms to promote compaction of DNA. Thus, DNA bending and DNA-DNA bridging are common. However, the major difference between nucleoid

associated proteins and histones is that nucleoid associated proteins can also act as specific transcriptional regulators to regulate expression of genes (Mcleod and Johnson, 2001).

There are about 12 different Nucleoid associated proteins encoded by the *E. coli* genome (Dillon and Dorman, 2010). These are expressed at different time periods during growth (Figure 1.10). Hence, the nucleoid associated proteins expressed during rapid growth are different to those expressed during stationary phase. This leads to changes in chromosome structure as cells grow (Ali Azam *et al.,* 1999). Thus, the different nucleoid associated proteins differently influence DNA supercoiling, transcription and DNA compaction (Rimsky and Travers, 2011).

The most widely studied nucleoid associated proteins, traditionally characterised on the basis of their abundance and transcriptional regulation properties, are:

1. Histone-like Nucleoid Structuring protein (H-NS) (Cukier-Kahn *et. al*., 1972)
2. Factor for Inversion Stimulation (Fis) (Koch and Kahmann, 1986)
3. Histone-like protein from *E. coli* strain U93 (HU) (Rouvière-Yaniv and Gross, 1975)
4. Suppression of td^- phenotype (StpA) (Zhang and Belfort, 1992)
5. DNA- binding protein from starved cells (Dps) (Almirón *et. al*., 1992) and
6. Integration host factor (IHF) (Miller and Friedman, 1980)

Other less extensively studied proteins, classified as nucleoid associated, are:

7. Leucine responsive protein (Lrp) (Willins *et. al*., 1991)
8. Curved DNA binding protein (CbpA) (Ueguchi *et. al*., 1994)

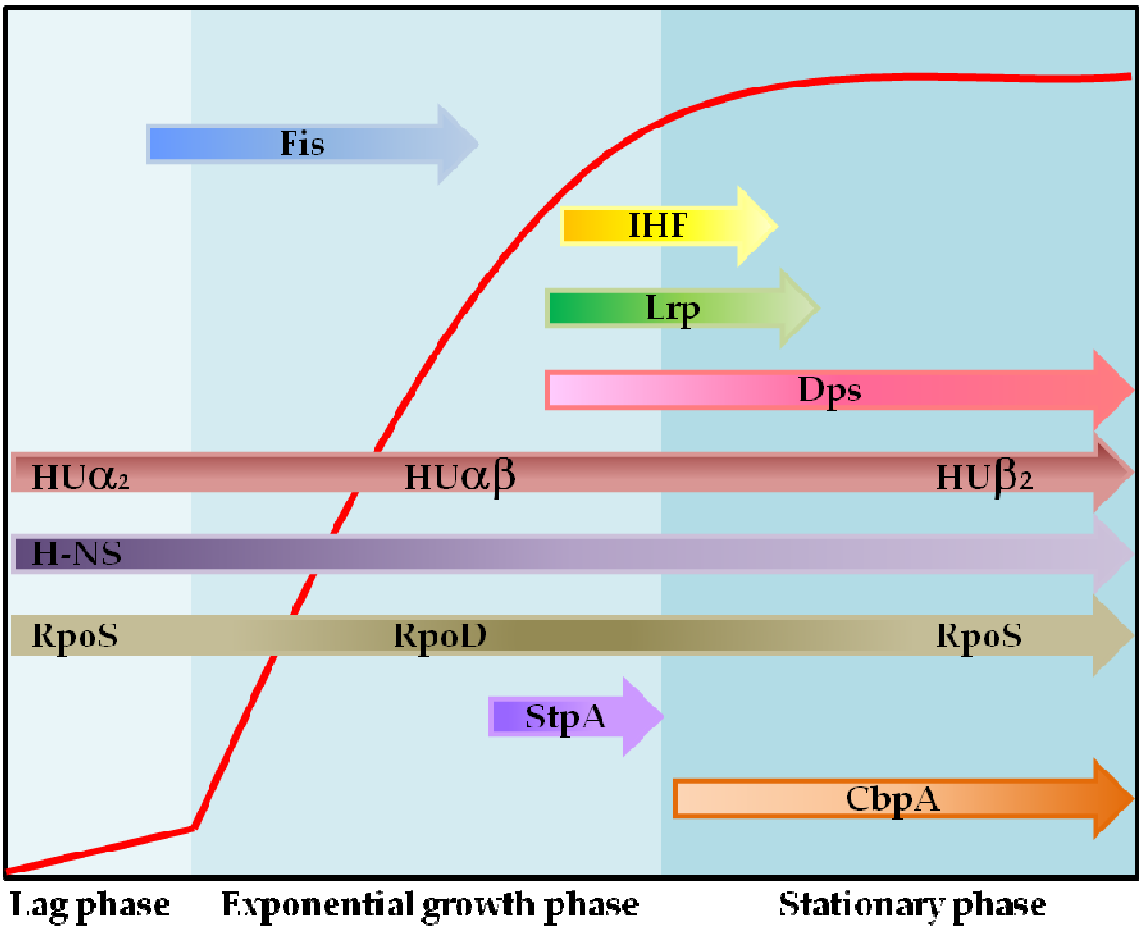


Figure 1.10 Temporal order of NAP expression during the *E. coli* growth cycle.

The red line indicates a typical growth curve for *E. coli*. The cells go through a lag phase, exponential growth phase, and then stationary phase. Nucleoid associated proteins expressed throughout the growth phase are shown as different coloured arrows. The light green arrow indicates expression of different RNA polymerase σ-factors. In exponential growth phase Fis (blue), homo (α_2) and heterodimeric (αβ) HU (brown) as well as H-NS (purple) are present. IHF (yellow), Lrp (dark green) and StpA (purple) are expressed in late exponential phase and Dps (pink) and CbpA (orange) are expressed in stationary phase. HU (αβ) and HU (β_2) are also present throughout (Based on Dorman, 2013 and Ali Azam, 1999).

1.5.2 Nucleoid associated proteins and compaction of bacterial nucleoprotein

Nucleoid associated proteins have been shown to play a role in the organisation and compaction of bacterial nucleoprotein (Dame, 2005). This organisation varies with growth and the nucleoid is less compact in growing cells compared to starved cells (Meyer and Grainger, 2013). The nucleoid associated proteins most prominent during rapid growth are; HU, IHF, Fis, Lrp and H-NS (Drlica and Bendich, 2009).

(i) HU

HU is the most abundant, and the best conserved, nucleoid associated protein. It is present at between 60000 and 100000 dimers per cell (Luijsterburg *et. al.*, 2006). HU can be present as homo or heterodimers (α_2, $\alpha\beta$, β_2) depending on the growth phase (Figure 1.10). Although, HU binding to DNA is non-specific, HU shows a preference for supercoiled DNA structures. HU bends the DNA rather than wrapping it like histones (Drlica and Bendich, 2009). Remarkably, the bend introduced by HU can be up to 180^o. However, on average, it introduces bends of about 100^o (Figure 1.11 a, f). At low concentrations of HU *in vitro*, DNA bending is observed (Figure 1.11 f, top). However, at higher concentrations of HU, rigid nucleoprotein filaments have been observed *in vitro* (Figure 1.11 f, bottom) (Luijsterburg *et. al.*, 2006). However, the biological significance of these filaments is unknown.

(ii) IHF

Integration host factor, or IHF, is known to bind specifically to the consensus sequence 5'-YAACTTNTTGATTTW-3' and bends DNA by 160^o (Figure 1.11 b, g) (Ussery *et. al.*, 2001). There are about 1000 specific IHF binding sites present in the *E. coli* genome (Ussery *et. al.*, 2001). However, IHF is highly abundant in the cell (30000-60000 copies) compared to the

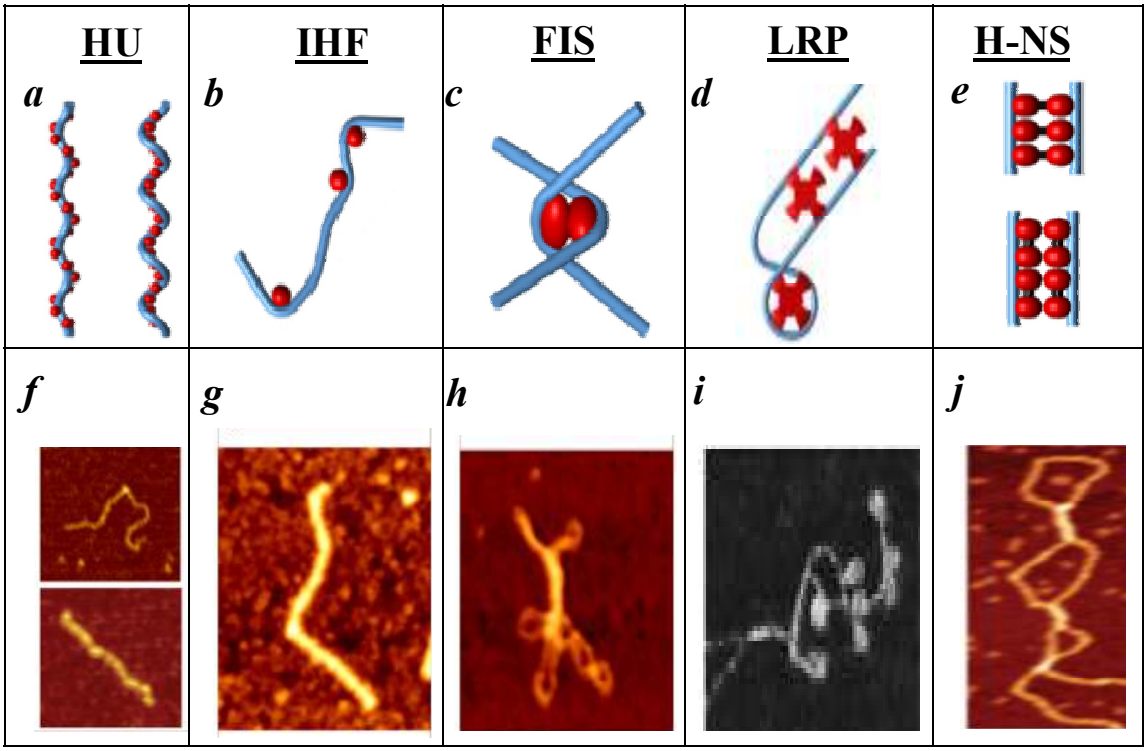


Figure 1.11 Architectural properties of DNA bending nucleoid associated proteins.

Cartoon (a-e) and AFM images (f-j) of DNA binding/bending/bridging by nucleoid associated proteins.

a,f- HU–DNA complexes. (f) HU bends (top) DNA at low concentrations and induces the formation of rigid filaments (bottom) at high HU concentrations

b,g- IHF–DNA complex

c,h- Fis–DNA complexes with Fis bound at the nodes of supercoiled plasmids. (h) Node formation due to Fis–Fis interactions

d,i- Lrp–DNA duplex bridging and wrapping mediated by Lrp

e,j- DNA loop formation as a consequence of DNA duplex bridging mediated by H-NS (Luijsterburg *et al.,* 2006)

specific IHF binding site. Hence, IHF may have a sequence independent mechanism of DNA binding which aids DNA compaction. *In vitro* IHF can compact DNA by 30 %, probably through such non-specific DNA binding (Ali *et. al*., 2001).

(iii) Fis

Factor for inversion stimulation (Fis) binds as a dimer to adjacent major grooves of the DNA. This creates a bending angle of 50^{o} to 90^{o} (Figure 1.11 c, h) (Dillon and Dorman, 2010). The Fis binding site is 15 bp in length and consists of a 5'-GNNNNNNNNNNNNC-3' sequence. In this sequence, the central 5-7 bp are A:T rich (Hancock *et. al.*, 2013). Fis expression is significantly increased as cells enter log phase, reaching a maximum copy number of 60000 copies per genome. Fis levels drop to almost zero when cells are in stationary phase. Fis is known to regulate transcription both as an activator and as a repressor (Drlica and Bendich, 2009; Dame 2005).

(iv) Lrp

Leucine-responsive regulatory protein, or Lrp, responds to the nutrient status of the cell (mainly amino acid levels). Interestingly, Lrp wraps DNA at its own promoter by oligomerising into a nucleosome like structure (Jafri *et. al*., 1999). Lrp also has a DNA-bridging activity. Thus, Lrp molecules, bound at two distal DNA sites, can interact with each other (Figure 1.11 d, i). Lrp acts as both repressor and activator of gene expression and targets genes involved in metabolism (Drlica and Bendich, 2009; Dame 2005).

(v) H-NS

Histone-like nucleoid structuring (H-NS) protein is probably the best studied bacterial chromosome folding protein. H-NS binds to AT-rich, intrinsically bent, DNA and helps stabilize the DNA bend. Thus H-NS consists of two domains connected by a flexible linker.

The C-terminal domain binds to DNA. The N-terminal domain drives dimerisation of H-NS (Dame *et. al.*, 2006). Each H-NS dimer has two DNA-binding domains that facilitates the formation of bridges between regions of double stranded DNA (Figure 1.11 e, j) (Drlica and Bendich, 2009; Dame 2005).

1.5.3 Nucleoid associated proteins and gene regulation

Gene regulation is highly sensitive to the DNA folding processes described above. Hence, it is no surprise that nucleoid associated proteins can act to influence transcription on a global scale (Browning *et al.,* 2010). Note that, although nucleoid associated proteins bind to DNA with less specificity than transcription factors (like FNR and CRP), they can recognize discrete targets in promoter DNA.

(i) IHF

IHF is known to regulate transcription primarily by acting as an architectural factor rather than by interacting with RNA polymerase directly. Hence, at σ^{54} dependent promoters, IHF positions activator proteins so that they can easily contact RNA polymerase (Goosen and Van de Putte, 1995). However, the action of IHF is not limited to σ^{54} dependent promoters. Thus, IHF may also remodel promoters to facilitate σ^{70} containing RNA polymerase binding. For instance, at the *Pseudomonas putida Pu* promoter, IHF correctly positions an upstream αCTD binding site by introducing a bend in the DNA (Bertoni *et. al*., 1998). Similarly, at λ P_L promoter, IHF facilitates RNA polymerase interactions with an UP element situated 90 bp upstream of the start site (Giladi *et. al*., 1998).

(ii) Fis

Fis can affect gene expression at individual regulatory regions by at least six different mechanisms. Hence, Fis can act as both an activator and repressor of transcription. *E. coli* has

seven ribosomal RNA operons and transcription of all seven is stimulated by Fis. In each case, Fis binds at a Class I position and aids RNA polymerase-DNA binding by making specific contacts with αCTD (Aiyar *et al.,* 2002). Similarly, Fis acts as a typical Class II activator at the *proP* P2 promoter by binding close to the core promoter elements (Mcleod *et al.,* 2002). During rapid growth, Fis also plays a major role as a repressor. Hence, Fis shuts off the expression of many unimportant gene products. Fis does this via various mechanisms. Thus, Fis may displace an activator or prevent an activator from functioning correctly. At some promoters, Fis represses transcription by RNA polymerase jamming (Browning *et al.,* 2010). For example, at the *E. coli dps* promoter, Fis prevents DNA opening by binding to the promoter in conjugation with RNA polymerase (Grainger *et al.,* 2008).

(iii) H-NS/StpA

H-NS is mainly thought to function as a transcriptional silencing protein. H-NS is therefore referred to as a 'universal repressor' (Owen Hughes *et al.,* 1992; Navarre *et al.,* 2006). Most promoters regulated by H-NS have two H-NS binding elements separated by several hundred base pairs. A key feature in gene regulation by H-NS is interaction between H-NS molecules binding at these loci. Hence, H-NS stabilises a loop in the promoter that can trap RNA polymerase and block transcription. Schröder and Wagner (2000) found that, at the *rrnB* P1 promoter, H-NS does not interfere with RNA polymerase binding. Rather, H-NS inhibits promoter escape. In further studies, Dame *et. al*. (2002) concluded that H-NS traps RNA polymerase at *rrnB* P1 by stabilising a loop in the DNA. RNA polymerase may itself contribute towards looping as it wraps DNA around itself when forming an open complex. Hence, RNA polymerase can bring H-NS binding sites into closer proximity and facilitate their interaction (Dame *et. al*., 2002). Shin *et al.* (2005) found that, at the *hdeAB* promoter, σ factor specificity is driven by H-NS. Thus, when $E\sigma^{70}$ is bound to the DNA, H-NS forms

bridges and traps RNA polymerase. This terminates transcription. However, this effect is not observed when $E\sigma^{38}$ is bound to the promoter (Shin *et al.,* 2005). At other promoters, such as *proU,* H-NS blocks the binding of RNA polymerase to the DNA (Nagarajavel *et al.,* 2007). In such cases, H-NS is not thought to form looped structures. Instead, H-NS oligomerises across the DNA to block access of RNA polymerase.

StpA is a paralogue of H-NS that has similar oligomerisation domains. Thus, StpA and H-NS can form heterodimers. Although StpA has no DNA binding activity (StpA is best defined as an RNA-binding protein and RNA chaperone) StpA occupies the same DNA targets as H-NS (Ali Azam *et*. *al*. 1999). Thus, StpA likely recognises DNA indirectly by first binding to H-NS to form a heterodimer. Interestingly, *stpA* knockouts have very little effect in *E. coli*. It is thus suggested that StpA primarily acts as a "reserve" H-NS like protein (Browning *et al.,* 2010; Dillon and Dorman, 2010).

1.6 Single molecule Techniques

Single molecule techniques have opened up new areas in molecular and cellular biology. Thus, single protein-DNA complexes can be manipulated and imaged in real time. Molecular mechanisms driving biological processes are often best understood at the single molecule level as this provides access to the dynamics of individual biomolecules. Thus, many laboratories have developed single-molecule approaches to investigate a wide range of DNA transactions (Pouget *et al*., 2004). Atomic force microscopy (AFM), Magnetic tweezers (MT) and Tethered particle motion (TPM) have all been effectively combined to study transcriptional regulation and chromosome folding (Finzi and Dunlap, 2010).

AFM is a technique that permits DNA-protein complexes to be visualised in real time. The protein-DNA complex is first deposited on a mica sheet which is atomically flat. A nanoscale tip is then used to scan the surface to reveal its topography. The tip is mounted on a piezoelectric scanner, which can move precisely in three dimensions. A laser beam, emitted onto the back of the tip by a laser diode, monitors tip movement and, with the help of a photodiode, a map of surface topography is built (Dorobantu and Gray, 2010; Liu and Wang, 2010).

MT allows the control of DNA supercoiling and tension to be controlled. Thus, it is possible to monitor the DNA conformations available for protein binding. A pair of magnets is mounted above the microscope that can be both translated along, and rotated around, the optical axis of the microscope. This enables both, stretching and twisting of the DNA tether. Manipulation of the thermodynamic equilibria using force is brought about by applying general tension to the DNA tether. Application of tension in the range of subpiconewtons has been made possible with the help of new configurations of optical tweezers (Finzi and Dunlap, 2010).

In TPM, a DNA tether is anchored to a glass surface at one end and an optically visible bead at the other. Tracking microscopy is then used to monitor the Brownian motion of the bead. Conformational changes in the DNA, including loop formation and bending, will alter the length of the DNA tether. As a result, the Brownian excursions of the bead will change accordingly. A telegraph like TPM signal is generated over time indicating stochastic, protein-mediated, changes in DNA conformation (Finzi and Dunlap, 2010; Towles *et al*., 2009).

1.7 Objectives

Recognition of DNA by the RNA polymerase sigma factor is crucial for specific transcription initiation. Whilst this process appears simple, RNA polymerase faces challenges. For example, at regulatory regions with i) multiple sigma factor binding sites or ii) a very high A:T content. The main aim of this work is to study transcription initiation, in particular promoter recognition, by RNA polymerase at complex gene regulatory regions.

First, I have analysed a gene regulatory region with overlapping binding sites for more than one sigma factor. This gene regulatory region (for *cbpA*) has overlapping binding sites for σ^{38} and σ^{70} but is σ^{38} dependent. My aim was to understand how this specificity is achieved. An additional aim was to better understand the function of the *cbpA* gene product. Thus, CbpA is known to bind to curved DNA and form aggregates in starved cells. I used a single molecule technique, TPM analysis, to monitor CbpA-DNA aggregates and DNA compaction.

My second aim was to understand AT-rich gene regulatory regions, which contain many "decoy" promoter elements, are recognised by RNA polymerase. Hence, I have also examined binding of the *ehxCABD* regulatory region by RNA polymerase. I have determined how RNA polymerase binding specificity is achieved at this regulatory region and what factors, if any, of bacterial chromatin play a role in maintaining this specificity.

Materials and Methods

Chapter 2

2.1 General reagents and solutions

2.1.1 Reagents, buffers and solutions

All the chemicals were obtained from Sigma-Aldrich Co. Ltd., Fisher Scientific, VWR and Bioline unless otherwise stated. Radioactive nucleotides were obtained from MP biochemicals and Perkin Elmer. Phosphor imager screens were purchased from Biorad. When needed, solutions were sterilised by autoclaving for 15-20 minutes at 120 °C.

2.1.2 Antibiotics (Stock Solutions)

Ampicillin: 100 mg/ml in distilled H_2O (dH_2O) stored at -20 °C

Tetracycline: 35 mg/ml in Methanol stored at -20 °C

Kanamycin: 50 mg/ml in dH_2O stored at -20 °C

The above antibiotics were added (from stock solutions) to liquid or solid media, after autoclaving, to a final concentration of 100 µg/ml (Ampicillin), 35 µg/ml (Tetracycline) or 50 µg/ml (Kanamycin).

2.1.3 Enzymes

All enzymes (Restriction enzymes, T4 polynucleotide kinase, T4 DNA ligase, and Calf Intestinal Alkaline Phosphatase) were purchased from New England Biolabs and stored at -20 °C. The reactions were carried out according to the manufacturer's instructions.

2.2 Extraction, Purification and radiolabelling of nucleic acids

2.2.1 Phenol/Chloroform extraction of DNA

Used to purify plasmids and PCR products. Also used to purify DNA fragments resulting from footprinting experiments.

Reagents and Chemicals used:

Phenol/Chloroform/Isoamyl alcohol: pH 6.7/8.0

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

Method:

An equal amount of phenol/chloroform was added to aqueous DNA solutions to remove contaminating proteins. After addition of phenol/chloroform the sample was vortexed for 10 seconds and centrifuged at maximum speed (20000 g) in a benchtop centrifuge for 2 minutes. The upper aqueous layer, containing the DNA, was transferred into a sterile tube taking care not to disturb proteins at the interface between the aqueous and organic layer. The DNA in the aqueous layer was then precipitated with ethanol, recovered and resuspended in the required volume of water (Section 2.2.2)

2.2.2 Ethanol Precipitation

Used to precipitate and concentrate DNA from aqueous solution after phenol/chloroform extraction.

Reagents and Chemicals used:

Ethanol: 100 % *v/v* stored at -20 °C

Ethanol: 70 % *v/v* diluted in dH_2O stored at -20 °C

Sodium acetate: 3 M pH 5.2

Method:

Two volumes of ice cold ethanol, and 0.1 volumes of 3 M sodium acetate (pH 5.2), were added to the aqueous DNA solution and vortexed. The sample was then incubated overnight at -20 °C or for 20 minutes at -80 °C. The precipitated DNA was then collected by centrifugation at 4 °C at maximum speed (20000 g) in a benchtop centrifuge for 15 minutes. The supernatant was removed and the DNA pellet was washed with 1 ml of ice cold 70 % ethanol by inverting the tube several times. The pellet was then recovered by centrifugation at 20000 g at 4 °C as described above and the supernatant was again discarded. The DNA pellet was dried for 10 minutes, at low or medium heat, in a vacuum manifold. Finally, the DNA was resuspended in the required volume of dH_2O.

2.2.3 Labelling DNA fragments

DNA fragments were radio-labelled for use in electrophoretic mobility shift assays (EMSA) and DNA footprinting analysis

Reagents and chemicals used:

T4 polynucleotide kinase

T4 polynucleotide kinase buffer: 10 x

Tris-EDTA (TE) buffer: 10 mM Tris-HCl, 1 mM EDTA (pH 8.0)

Sephadex G-50: 12 % *v/v* Sephadex G-50 in TE buffer. Solutions were autoclaved and cooled. The supernatant was discarded and equal volume of fresh TE buffer was added.

Method:

16 µl of DNA fragment (100-400 nM), 1 µl T4 polynucleotide kinase, 2 µl 10 x kinase buffer and 1 µl [γ-^{32}P]-ATP (10 µCi/µl) was mixed together and the reactions were incubated at 37 °C for 30 minutes. The reactions were then passed through a 200 µl volume Sephadex G-50 spin column to remove the unincorporated [γ-^{32}P]-ATP. The flow through was collected and the labelled fragment was stored at -20 °C.

2.3 Bacterial Growth Media

2.3.1 Solid media

Used to grow bacteria and for short-term storage purposes

Reagents and chemicals used:

MacConkey-Lactose Agar: 52 g/l (Oxoid)

Nutrient Agar (LB agar): 20 g/l (Sigma)

Method:

For bacterial growth on a solid medium we used either MacConkey or Nutrient agar. Typically, MacConkey-Lactose agar was used to monitor cell phenotypes (i.e. *lac*$^+$ or *lac*$^-$) during construction and analysis of promoter::*lacZ* fusions. Nutrient agar was used at all other times. Solid media was prepared by dissolving in water according to the manufacturer's instructions and then autoclaving. Where appropriate, once the molten agar had cooled, antibiotics were added. Whilst still molten, the solution was then poured into a petri dish next to a lit Bunsen burner. Once set, the agar was dried by leaving at room temperature overnight.

2.3.2 Liquid media

Reagents and chemicals used:

Lennox-Broth Media (LB): 20 g/l tryptone, 10 g/l yeast, 10 g/l NaCl

Method:

LB broth was used to grow liquid bacterial cultures. When selecting for antibiotic resistance, appropriate antibiotics were used to supplement growth media.

2.4 Bacterial Strains and Plasmids

2.4.1 Bacterial strains and growth conditions

Table 2.1 lists all the bacterial strains used in this study. The strains were stored and maintained in glycerol stocks at -80 °C. When in regular use, the strains were streaked and stored on MacConkey-lactose agar or LB agar plates. When liquid cultures were required a single fresh colony was used to inoculate the required medium. These cultures were then incubated overnight at 37 °C in a shaking incubator. Typically overnight cultures were diluted 1:100 into fresh media on the day of an experiment. The Optical Density (OD at 650 nm) was used to monitor growth of liquid cultures using a Jenway 6300 Spectrophotometer (Thermo Fisher Scientific Inc.). Bacteria grown on solid media were also incubated at 37 °C.

2.4.2 Plasmids

Table 2.2 lists all plasmids used in this study. Plasmids were prepared as described below and stored at -20 °C in dH_2O. Plasmids shown in Figure 2.1, 2.2 and 2.3 were used for analysis of gene expression either *in vivo* (Figure 2.1 and 2.2) or *in vitro* (Figure 2.3).

Table 2.1 list of bacterial strains used in this study

Strain	Genotype	Source
JCB387	Δ*nir* Δ*lac*	Typas and Hengge (2006)
MC4100	F$^-$ *araD*139 Δ(*argF-lac*)U169 *rpsL*150 *relA*1 *deoC*1 *ptsF*25 *rbsR flbB*5301	Fenton *et. al.,* (2000)
MC4100*rpoS::*kan	MC4100*rpoS::*kan	Busby *et. al.,* (1984)
M182	Δ*lac galK galU strA*	Busby *et. al.* (1983)
M182*Δhns* (JRG4864)	M182*Δhns* Cm^R	Wyborn *et. al.* (2004)
T7 express	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*–Tet^S*)2 [dcm] R(zgb-210::Tn10*–Tet^S*) endA1* Δ*(mcrC-mrr)114::IS10*	(Invitrogen)
MG1655	F$^-$ λ$^-$ rph-1	Blattner *et. al.* (1997)
MG1655*Δhns*	F$^-$ λ$^-$ rph-1 *Δhns::*kan	

Table 2.2 list of plasmids used in this study

Plasmid	Description	Source
pSR	pBR322-derived plasmid containing an *Eco*RI – *Hin*dIII fragment upstream of the λoop transcription terminator	Kolb *et. al.*, 1995
pRW50	low copy number broad-host-range *lac* fusion vector for cloning promoters on *Eco*RI–*Hin*dIII fragments: contains the RK2 origin of replication and encodes Tc^R	Thouvenot *et. al.,*(2004)
pLux	pCS26 derivative; STOP codons, ribosome binding site, and NcoI ction site switch; parent plasmid for pLUX series	Burton *et. al*. (2010)
pVRσ	pBR322 derivative encoding *rpoD* and mutant derivatives	Gaal *et. al.,* (1989)
pET21b	T7 Expression vector containing 6xHis tag, used to overexpress σ^{70} and derivatives	(Novagen)

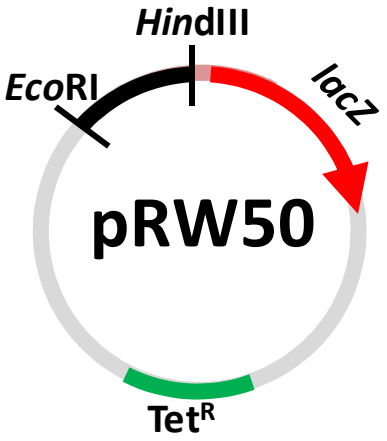


Figure 2.1 The pRW50 plasmid used as *in vivo* reporter vector for β-galactosidase analysis

This plasmid was used as a reporter for promoter activity in *E. coli* K-12. The black bar indicates the promoter DNA fragment cloned using the restriction sites *Eco*RI and *Hin*dIII. The *lacZ* gene is shown as a red arrow. The plasmid encodes resistance to tetracycline (green).

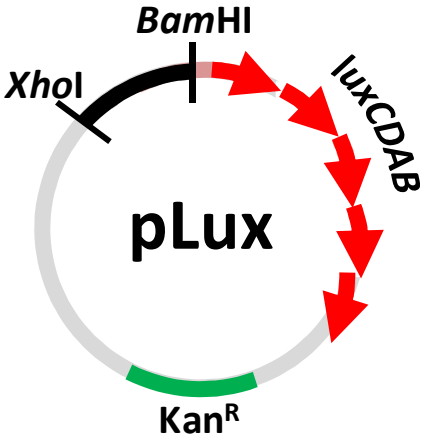


Figure 2.2 The pLux plasmid used as *in vivo* reporter vector for luciferase assays

This plasmid was used as a reporter for promoter activity in *E. coli* O157:H7. The black bar indicates the promoter DNA fragment cloned using the restriction sites *Xho*I and *Bam*HI. The *luxCDAB* operon is shown in red arrows. The plasmid encodes resistance to kanamycin (green).

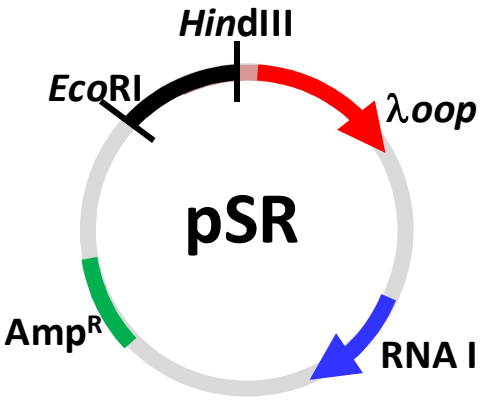


Figure 2.3 The pSR plasmid

This plasmid was used as a reporter for transcription *in vitro*. The black bar indicates the promoter DNA fragment cloned using the restriction sites *Eco*RI and *Hin*dIII upstream of the *λoop* transcription terminator. The RNAI acts as an internal control. The plasmid encodes resistance to ampicillin (green).

2.5 DNA purification using QIAquick PCR Purification kit (QIAgen)

The QIAquick PCR purification kit (QIAgen) was used to purify DNA after PCR synthesis or restriction digestion. This was done according to the manufacturer's instructions. Purified DNA was eluted in 30 µl sterile dH_2O from QIAquick columns. Purification of up to 20 µg of high-copy plasmid DNA was achieved using this kit.

2.6 Plasmid DNA preparation

2.6.1 QIAprep Spin Miniprep kit (QIAgen)

LB (5 ml) supplemented with the appropriate antibiotic was used to grow an overnight culture of the strain carrying the plasmid DNA to be extracted. The plasmid DNA was extracted using the QIAprep Spin Miniprep kit (QIAgen), according to the manufacturer's instructions. Plasmid DNA was eluted from the QIAprep column in 30 µl sterile dH_2O.

2.6.2 QIAprep Spin Maxiprep kit (QIAgen)

LB (100 ml to extract high copy number plasmid and 400 ml to extract low copy number plasmid) supplemented with the appropriate antibiotic was used to grow an overnight culture of the strain carrying the plasmid DNA to be extracted. The plasmid DNA was extracted using the QIAprep Spin Maxiprep kit (QIAgen), according to the manufacturer's instructions. Plasmid DNA was eluted in 100 µl sterile dH_2O and quantified using Nanodrop.

2.7 Gel Electrophoresis

Reagents and chemicals used:

Gel loading dye: 0.025 % (*w/v*) Bromophenol Blue; 20 % (*v/v*) glycerol; 0.025 % (*w/v*) Xylene Cyanol F; 1 mM EDTA; 10 mM Tris, pH 7.5

Ehidium Bromide: 10 mg/ml

DNA Ladder: 100 bp and 1 kb DNA ladder purchased from NEB and diluted 6-fold in DNA loading dye.

2.7.1 Agarose Gel Electrophoresis

Reagents and chemicals used:

5 x TBE Buffer 1.1 M Tris; 900 mM Borate; pH 8.3, 10 mM Na_2EDTA (Fisher Scientific)

1 x TBE Buffer: Diluted from 5 x TBE Buffer in dH_2O

Agarose Solutions: 1 % (*v/v*) in 1 x TBE Buffer, heated for 2 minutes on medium power in a microwave. Cooled until hand hot. Ethidium bromide added to a final concentration of 0.2 - 0.5 µg/ml. Poured into gel cast.

Method:

For electrophoresis, DNA samples were mixed in a 1:1 ratio with gel loading dye and loaded onto a 1 % (*v/v*) agarose gel. Electrophoresis was done at 120 V for the required time in 1 x TBE buffer. The gel, once run, was visualized using a UV-transilluminator.

Extraction of DNA fragments from agarose gels

DNA samples were run on 1 % agarose gels as described above. The DNA bands to be extracted were excised with a clean razor blade. The QIAgen gel extraction kit was used to recover DNA from the band according to manufacturer's instructions. The DNA fragments recovered were eluted in 50 µl dH_2O from the QIAquick column.

2.7.2 Polyacrylamide gel electrophoresis

Reagents and chemicals used:

Protogel: 30 % (*w/v*) (Geneflow), 0.8 % (*w/v*) Bisacrylamide, Stock Solution 37.5:1

N,N,N′,N′-Tetramethylethylenediamine (TEMED)

Ammonium persulphate (APS): 10 % (*w/v*) solution (diluted in dH_2O) prepared fresh for every use.

5 x TBE Buffer: 0.445 M Tris borate pH 8.3, 10 mM Na_2EDTA (Fisher Scientific)

Method:

First, 18.8 ml of protogel and 20 ml of 5 x TBE was diluted to 100 ml with dH_2O. Polymerisation of the gel was initiated by addition of 100 µl 10 % APS and 40 µl TEMED. DNA samples for electrophoresis were loaded in 1:1 ratio with gel loading dye. The gels were run at about 30 mV for about 25 minutes in 1 x TBE. Once run, the gel was visualized using UV-transilluminator after staining it in ethidium bromide solution.

2.8 Transformation of *E. coli* cells with plasmid

2.8.1 Preparation of $CaCl_2$ competent cells

Reagents and chemicals used:

Calcium Chloride ($CaCl_2$): 0.1M

Glycerol: 50 % *v/v*

Method:

One ml of an overnight culture of the strain to be transformed was used to inoculate 50 ml of fresh LB. This sub-culture was incubated at 37 °C in a shaking incubator until the culture reached an OD_{650} of 0.3-0.5 (mid-logarithmic phase). The culture was transferred into a sterile centrifuge tube and kept on ice for 10 minutes. The cells were then harvested at 4 °C for 20 minutes at 4000 rpm in a Thermo Scientific 16R Megafuge centrifuge. After resuspending the pellet into 25 ml of ice-cold 0.1 M $CaCl_2$, the cells were kept on ice for 30 minutes. The cells were again harvested by centrifugation as described above. The cells were resuspended in three ml ice-cold 0.1 M $CaCl_2$ and kept on ice for 60 minutes before use. 0.3 volumes of ice-cold 50 % (*v/v*) glycerol was added before storing cells at -70 °C as aliquots for later use.

2.8.2 Preparation of Rubidium Chloride ($RbCl_2$) competent cells

Reagents and chemicals used:

Buffer 1: 2-(*N*-morpholino)ethanesulfonic acid (MES) 0.39 g, $CaCl_2$ (1 M) 2 ml, $MnCl_2.4H_2O$ 1.98 g, $RbCl_2$ 2.42 g. MES was dissolved in 100 ml of dH_2O and the pH was adjusted to 6.2. The other 3 ingredients were added to 50 ml dH_2O and the pH was adjusted to 5.8 using acetic acid. The above two solutions were combined and made up to 200 ml using dH_2O.

Buffer 2: 3-(N-morpholino)propanesulfonic acid (MOPS) 0.41 g, 1 M $CaCl_2$ 15 ml, 15 % Glycerol 30 ml, $RbCl_2$ 0.242 g. The above ingredients were dissolved in 100 ml dH_2O and the pH was adjusted to 6.5 using KOH. The solution was made up to 200 ml using dH_2O.

Both the buffers were filter sterilized.

Method:

One ml of overnight culture of the strain to be transformed was added to 50 ml LB. This sub culture was incubated at 37 °C in a shaking incubator until the culture reached an OD_{650} of 0.3-0.5 (mid-logarithmic phase). The culture was then transferred into a sterile centrifuge tube and kept on ice for 10 minutes. The cells were then harvested at 4 °C for 20 minutes at 4000 rpm as for $CaCl_2$ competent cells. The pellet was resuspended in 8 ml ice cold Buffer 1 and incubated for 20 minutes on ice. The cells were then harvested by centrifugation for 10 minutes at 4000 rpm. The pellet was then resuspended in 1.6 ml ice-cold Buffer 2 and the cells kept on ice for at least 15 minutes before use. The competent cells were stored as 200 µl aliquots at -70 °C.

2.8.3 Transformation of plasmid DNA into chemically competent cells

One µl of plasmid DNA (approximately 0.2 µg) was added to 100 µl of competent cells and incubated on ice for 30 minutes. The cells were then heat shocked for 2 minutes at 42 °C and kept on ice for a minute. One ml LB was then added to the heat shocked cells. The cells were then kept at 37 °C in a shaking incubator for 60 minutes. The cells were then harvested by centrifugation at 20000 g and resuspended in 100 µl LB. This resuspension was spread on the nutrient agar or MacConkey agar supplemented with appropriate antibiotics. The agar plates were incubated at 37 °C overnight.

2.9 DNA Manipulations

Table 2.3 lists all the Oligonucleotides used in this study

2.9.1 Polymerase Chain Reaction (PCR)

Phusion DNA polymerase (New England Biolabs) and the buffer provided were used for PCR. One µl of template DNA, which is typically plasmid DNA, was used for PCR reactions.

Table 2.3 Oligonucleotides

Oligonucleotides used in this thesis are mentioned below. The *Eco*RI, *Hin*dIII, *Xho*I and *Bam*HI restriction sites are highlighted in bold. All the mutations are highlighted in bold and underlined. Shine dalgarno sequence is underlined.

Name	Sequence (5'-3')	Source
Primers to introduce random single base substitutions into the *cbpA* regulatory region spacer DNA element		
-22N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAA**N**TTTGAGGATTACCC	This work
-21N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAAT**N**TTGAGGATTACCCT	This work
-20N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATT**N**TGAGGATTACCCTA	This work
-19N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTT**N**GAGGATTACCCTAC	This work
-18N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTT**N**AGGATTACCCTACA	This work
-17N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTG**N**GGATTACCCTACAC	This work
-16N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTGA**N**GATTACCCTACACT	This work
-15N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTGAG**N**ATTACCCTACACTT	This work
-14N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTGAGG**N**TTACCCTACACTTA	This work
-13N cbpAΔ45	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTGAGGA**N**TACCCTACACTTAT	This work
Primers used for site-directed mutagenesis of the *cbpA* regulatory region		
cbpAΔ45 -18C	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTCAGGATTACCC	This work
cbpAΔ45-9A -10T	GGCTGC**GAATTC**ATATTCTGTGTTGGCATATGAAATTTTGAGGATTATACTACACTTATAGG	This work
Primers used to amplify *cbpA* promoter inserts cloned in plasmid pSR		
pSR up	GCATTTATCAGGGTTATTGTCTC	This work
pSR down	CATCACCGAAACGCGCGAGG	This work
Primers used to introduce alanine codons into RpoD		
*Hin*DIII oligo	GGGG**AAGCTT**TTAATCGTCCAGGAAGCTACGCAGCACTTCAGAACGGCTCGGGTGACGCAGTTTGCGCAGCGCCTTCGCTTC	This work
R451A	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATC**GCT**ATTCCGGTGCATATGATTGAGACC	This work
I452A	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATCCGT**GCT**CCGGTGCATATGATTGAGACC	This work
P453A	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATCCGTATT**GCG**GTGCATATGATTGAGACC	This work
V454A	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATCCGTATTCCG**GCG**CATATGATTGAGACC	This work
H455A	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATCCGTATTCCGGTG**GCT**ATGATTGAGACC	This work
Primer used to amplify the RpoD RA451 allele for cloning into pET21b		

RpoD pET21b Up	AGCTCA**GCTAGC**GAGCAAAACCCGCAGTCACAGCTGAAAC	This work
Primers used to amplify the *lee*1 promoter and derivatives		
LEE up	**GAATTC**TTGACATTTAATGATAATGTATTTTACACATTAGAAAAAAG	This work
LEE up-18A	**GAATTC**TTGACATTTAATGATAAT**A**TATTTTACACATTAGAAAAAAG	This work
LEE up -18C	**GAATTC**TTGACATTTAATGATAAT**C**TATTTTACACATTAGAAAAAAG	This work
LEE up -18T	**GAATTC**TTGACATTTAATGATAAT**T**TATTTTACACATTAGAAAAAAG	This work
LEE down	**AAGCTT**ATTCTCTTTTTTCTAATGTGTAAAA	This work
R451K	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATC**AAG**ATTCCGGTGCATATGATTGAGACC	This work
R451E	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATC**GAA**ATTCCGGTGCATATGATTGAGACC	This work
R451W	CCTGGTGGATCCGTCAGGCGATCACCCGCTCTATCGCGGATCAGGCGCGCACCATC**T**G**G**ATTCCGGTGCATATGATTGAGACC	This work
Radiolabelled primer used for primer extension analysis		
D49724	GGTTGGACGCCCGGCATAGTTTTTCAGCAGGTCGTTG	
ehxCABD F3 WT and mutations in the -10 element and overlapping decoy promoter elements		
ehxCABD F3 up	GGCTGC**GAATTC**TATCTTACAAATCAATCATCTGAGTGTTATAATATAACTTAGCTGTGATATGTGTAAGAATGTTTAGGCAAT	This work
ehxCABD F3 down	CGCCCG**AAGCTT**CATCTCTCCCAACCAAAACAACATTAGCGATAATAATATATTGCCTAAACATTCTTACACATATCA	This work
-13G up	GGCTGC**GAATTC**TATCTTACAAATCAATCATCTGAGTGT**G**ATAATATAACTTAGCTGTGATATGTGTAAGAATGTTTAGGCAAT	This work
-40G-7T-5T-4T up	GGCTGC**GAATTC**GATCTTACAAATCAATCATCTGAGTGTTATAAT**T**T**TT**CTTAGCTG	This work
-40G up	GGCTGC**GAATTC**GATCTTACAAATCAATCATCTGAGTGTTATAATATAACTTAGCTG	This work
ehxCABD F2		
ehxCABD F2 up	GGCTGC**GAATTC**TGTTTTTAGATGCTTCTTGCTTAAAAGAATATAATTCCTGTTCTTTTATATAGAGTTCTTTACATTTAC	This work
ehxCABD F2 down	**AAGCTT**CATAATGTTTAAACAAATAAGAAAATTCAGTAAATGTAAAGAACTCTATATAAAAGAACAGGAA	This work
ehxCABD F3 and F2 for pLUX analysis		
ehxCABD F3 up	GGCTGC**CTCGAG**TATCTTACAAATCAATCATCTGAGTGTTATAATATAACTTAGCTGTGATATG	This work
ehxCABD F3 down	CGCCCG**GGATCC**CATCTCTCCCAACCAAAACAACATTAGCG	This work
ehxCABD F2 up	GGCTGC**CTCGAG**TGTTTTTAGATGCTTCTTGCTTAAAAGG	This work
ehxCABD F2 down	CGCCCG**GGATCC**CATAATGTTTAAATAAATAAGAAAACTCAG	This work
ehxCABD from different *E. coli* isolates		
TW15838 Forward	GGCTGC**GAATTC**CATTTGTTATCTTACAAAACAATCATCTGAGTATTATAATATAACTCATCTGTAATATGCATAAGAGTTGTAGCCAATAT	This work
TW15838 Reverse	CGCCCG**AAGCTT**TACCGCAAGGAGCTCCCCTCCTAACACAAACAACTATAATATATTGGCTACAACTCTTATGCATATTACA	This work
1.2264 Forward	GGCTGC**GAATTC**TTTGTCGCAGGGTGATTCATCTGAATGTTATAATATAACCTATCTAGGTGTGCATAAGAATACATAGCTAATGT	This work
1.2264 Reverse	CGCCCG**AAGCTT**TACCGCAAGGAGAATCCCTCCTACAAAAAAACAACATTAGAAACAGTAATACATTAGCTATGTATTCTTATGCACACCTA	This work
TW15838 Forward -12G	GGCTGC**GAATTC**CATTTGTTATCTTACAAAACAATCATCTGAGTAT**G**ATAATATAACTCATCTGTAATATGCATAAGAGTTGTAGCCAATAT	This work

1.2264 Forward -12G	GGCTGC**GAATTC**TTTGTCGCAGGGTGATTCATCTGAATGT**G**ATAATATAACCTATCTAGG TGTGCATAAGAATACATAGCTAATGT	This work

Primers used to generate fragments containing -10 like hexamers within *ehxCABD* operon (forward, 5'-3' orientation)

1F up	GGCTGC**GAATTC**CTGGGCCAGTTTAGATCTGGAATGTGAGGTAAAAT	This work
2F up	GGCTGC**GAATTC**TTGAGAAAGAAAACTAATATGACAGTAAATAAAAT	This work
3F up	GGCTGC**GAATTC**AAGCGTACGTTCCGCTGGAAAAAAACTCATATTAT	This work
4F up	GGCTGC**GAATTC**CTTCTGCAGAAGTATCAGAAAGTTGGGAGTAAAAT	This work
5F up	GGCTGC**GAATTC**TTGGAAATGTTGGTAAAGCTGTTTCGCAATATATT	This work
6F up	GGCTGC**GAATTC**ATGCTGATCGCAGTCAGAGTGGTAAGGCATATATT	This work
7F up	GGCTGC**GAATTC**TAAAAGGCATAAAAAATCATAAAGGTGTATATGAT	This work
8F up	GGCTGC**GAATTC**CTCATCAGAAGTATTTGCTGGTGAAGGTTATGATACCGTATCTTATAA T	This work
9F up	GGCTGC**GAATTC**GCAACAGGAGCATCAAAACCTGGTGAGTATATAGT	This work
10F up	GGCTGC**GAATTC**GGTGTCAGTAGGGAAGCGAACAGAGAAAATACAAT	This work
11F up	GGCTGC**GAATTC**ATTTTGAATTCAGAACCGGTGGAATTCCTTATGAT	This work
12F up	GGCTGC**GAATTC**GATGAATTCAAAGGCGGTAAGTTTAATGATATATT	This work
13F up	GGCTGC**GAATTC**GTAATGATGAGCTTCAGGCACACGGAGCTTATAAT	This work
14F up	GGCTGC**GAATTC**TATGTTTATGGGCAAAACTATGGTCATCATACAAT	This work
15F up	GGCTGC**GAATTC**ATCGCATTTAAGAGAGTTGGAAATGATCTTATCAT	This work
16F up	GGCTGC**GAATTC**ATTCAGGTGATGAACGATTAGCATCTTTATATAAT	This work
17F up	GGCTGC**GAATTC**GGCCATTCCAGGAGGAATGGTCATCGTTATATGAT	This work
18F up	GGCTGC**GAATTC**GTCTGTATGCACTGATATTGCTTGCACAATATCAT	This work
19F up	GGCTGC**GAATTC**CTGTCAATGCTGAAACTATAAGGCATCAGTATAAT	This work
20F up	GGCTGC**GAATTC**TTGATATGGCGGGATGACGGTAAGCATTATATATT	This work
21F up	GGCTGC**GAATTC**TTACTAAAGATTCATCACGCTATCTTGTTTATGAT	This work
22F up	GGCTGC**GAATTC**TTAGTTCACCGGGGGTTTTCAACGTTAAATATTAT	This work
23F up	GGCTGC**GAATTC**ACGTTAAATATTATCACAATAGCATTTATTATAGT	This work
24F up	GGCTGC**GAATTC**TTGAAGTGATATTAACCGGAGCCAGAACTTATATT	This work
25F up	GGCTGC**GAATTC**TGGTATTACAGCCCTCAATTAACACTGGTTATATTAT	This work
26F up	GGCTGC**GAATTC**TTTCTTGTCGAAACGGTAACAGCAATAAATACAAT	This work
27F up	GGCTGC**GAATTC**ACACCGGTAGAGAAAAAGTCAGGCAGAAATATACT	This work
28F up	GGCTGC**GAATTC**TTATACATATCAAAGGGGGATGTTATCGGTATAGT	This work
29F up	GGCTGC**GAATTC**CGCCAGATTGGTGTTGTATTGCAGGAAAATATACT	This work
30F up	GGCTGC**GAATTC**TTGCAGGAAAATATACTATTAAATCGTAGTATTAT	This work
31F up	GGCTGC**GAATTC**ATGATTTTATTAGAGAACTAAAAGAAGGGTACAAT	This work
32F up	GGCTGC**GAATTC**ACCAGTGCTCTTGATTATGAGTCGGAAAATATAAT	This work
33F up	GGCTGC**GAATTC**TCAAGAATATGTAAGAACAGAACCGTAATTATTAT	This work
34F up	GGCTGC**GAATTC**ACGTTCTCATTTTGTTGTATGGAGTATTTTATTAT	This work
35F up	GGCTGC**GAATTC**GACGGTGCCTGGTGTTGAATCTGATATTTTAAAAT	This work
36F up	GGCTGC**GAATTC**AACATTCAGTTATGGAGCAGGAAAATAGCTATATT	This work
1F down	CGCCCG**AAGCTT**CCTCCTTTACATCATTTATATATTTTACCTCACATTCCAGA	This work
2F down	CGCCCG**AAGCTT**CCTCCTTTGAAAATGTTCTTTATTTTATTTACTGTCATATT	This work
3F down	CGCCCG**AAGCTT**CCTCCTAATTATCAGGTATTAATAATATGAGTTTTTTTCCA	This work
4F down	CGCCCG**AAGCTT**CCTCCTTCAGCGGTTCCTCCTATTTTACTCCCAACTTTCTG	This work
5F down	CGCCCG**AAGCTT**CCTCCTCATTCTCTGAGCCAGAATATATTGCGAAACAGCTT	This work
6F down	CGCCCG**AAGCTT**CCTCCTATTTTCCAGATAATTAATATATGCCTTACCACTCT	This work
7F down	CGCCCG**AAGCTT**CCTCCTAATCAATTTTGAATAATCATATACACCTTTATGAT	This work
8F down	CGCCCG**AAGCTT**CCTCCTACCAACATCCGTCTTATTATAAGATACGGTATCATAACCTTC ACC	This work
9F down	CGCCCG**AAGCTT**CCTCCTTACATATTTTTTGAAACTATATACTCACCAGGTTT	This work
10F down	CGCCCG**AAGCTT**CCTCCTATTCAAAATCACGATATTGTATTTTCTCTGTTCGC	This work
11F down	CGCCCG**AAGCTT**CCTCCTAAGATTATCTATTACATCATAAGGAATTCCACCGG	This work

12F down	CGCCCG**AAGCTT**CCTCCTCCATCTGCGCCATGGAATATATCATTAAACTTACC	This work
13F down	CGCCCG**AAGCTT**CCTCCTACCACCTGACAGAATATTATAAGCTCCGTGTGCCT	This work
14F down	CGCCCG**AAGCTT**CCTCCTCCTCCTTCATCTGCAATTGTATGATGACCATAGTT	This work
15F down	CGCCCG**AAGCTT**CCTCCTTTAATGGCTTTATTCATGATAAGATCATTTCCAAC	This work
16F down	CGCCCG**AAGCTT**CCTCCTTTGATGTAAGGATAAATTATATAAAGATGCTAATC	This work
17F down	CGCCCG**AAGCTT**CCTCCTGAACTACATTTACTCATCATATAACGATGACCATT	This work
18F down	CGCCCG**AAGCTT**CCTCCTATTGACAGTTATATTATGATATTGTGCAAGCAATA	This work
19F down	CGCCCG**AAGCTT**CCTCCTATCTTGTGTGTGGGTATTATACTGATGCCTTATAG	This work
20F down	CGCCCG**AAGCTT**CCTCCTTTAGTAATACGAGACAATATATAATGCTTACCGTC	This work
21F down	CGCCCG**AAGCTT**CCTCCTCTGATGTTGTTCTGGATCATAAACAAGATAGCGTG	This work
22F down	CGCCCG**AAGCTT**CCTCCTATAAATGCTATTGTGATAATATTTAACGTTGAAAA	This work
23F down	CGCCCG**AAGCTT**CCTCCTACTTCAAAAAGTATCACTATAATAAATGCTATTGT	This work
24F down	CGCCCG**AAGCTT**CCTCCTTGTAGTATGAGAGAAAATATAAGTTCTGGCTCCGG	This work
25F down	CGCCCG**AAGCTT**CCTCCTAACAAGGTAGTGACAATAATATAACCAGTGTTAAT	This work
26F down	CGCCCG**AAGCTT**CCTCCTACTGCCATGGATTTGATTGTATTTATTGCTGTTAC	This work
27F down	CGCCCG**AAGCTT**CCTCCTCCCTGAATTTCCGGCAGTATATTTCTGCCTGACTT	This work
28F down	CGCCCG**AAGCTT**CCTCCTGAACCAGAACGGCCAACTATACCGATAACATCCCC	This work
29F down	CGCCCG**AAGCTT**CCTCCTATACTACGATTTAATAGTATATTTTCCTGCAATAC	This work
30F down	CGCCCG**AAGCTT**CCTCCTAATGTAATATTATCGATAATACTACGATTTAATAG	This work
31F down	CGCCCG**AAGCTT**CCTCCTTTCTCCAACAATAGTATTGTACCCTTCTTTTAGTT	This work
32F down	CGCCCG**AAGCTT**CCTCCTGACATATTTTTCATTATTATATTTTCCGACTCATA	This work
33F down	CGCCCG**AAGCTT**CCTCCTGACAACCTGTGTGCAATAATAATTACGGTTCTGTT	This work
34F down	CGCCCG**AAGCTT**CCTCCTGAGATATAATTACAAATAATAAAATACTCCATACA	This work
35F down	CGCCCG**AAGCTT**CCTCCTACAAAGATGCTTCTGATTTTAAAATATCAGATTCA	This work
36F down	CGCCCG**AAGCTT**CCTCCTTTCATTTTTTGCCTGAATATAGCTATTTTCCTGCT	This work

Primers used to generate fragments containing -10 like hexamers within *ehxCABD* operon (reverse, 3'-5' orientation)

1R up	GGCTGC**GAATTC**ATGTTACATCATTTATATATTTTACCTCACATTCC	This work
2R up	GGCTGC**GAATTC**TTTCTTTCTCAATATTAAAATTAAAATCTTGTCTGTTTTTT	This work
3R up	GGCTGC**GAATTC**GGTTTGCAATCGCTGTATCATCTCGTTCAGTACGG	This work
4R up	GGCTGC**GAATTC**ATAAAGCAATCCCCGTAAAATTCTGTAGTGCTGAG	This work
5R up	GGCTGC**GAATTC**TACTACTTTTGGCAATATCATTCTGACTTATATCC	This work
6R up	GGCTGC**GAATTC**ATCAACGGTACTGTTTATACTTGATACTGTATCTA	This work
7R up	GGCTGC**GAATTC**TTCTCTGAGCCAGAATATATTGCGAAACAGCTTTA	This work
8R up	GGCTGC**GAATTC**TTAAATCGTTCAGAATATGATTCAAGCTGCTTAGC	This work
9R up	GGCTGC**GAATTC**CAATAAGAGATGTCATATATTCATATTTACCACTC	This work
10R up	GGCTGC**GAATTC**TGTATCCTTACCATTTACAATAAGAGATGTCATAT	This work
11R up	GGCTGC**GAATTC**AATCATATACACCTTTATGATTTTTTATGCCTTTT	This work
12R up	GGCTGC**GAATTC**TACCAACATCCGTCTTATTATAAGATACGGTATCATAACCTTCACCAG CA	This work
13R up	GGCTGC**GAATTC**TATTTTTTGAAACTATATACTCACCAGGTTTTGAT	This work
14R up	GGCTGC**GAATTC**CTGCGCCATGGAATATATCATTAAACTTACCGCCT	This work
15R up	GGCTGC**GAATTC**CATAATTACCTTCGATATAATCGTTCCCATCTGCG	This work
16R up	GGCTGC**GAATTC**TCCCCATCATCGCCGTATAGTCGATCATTACCATA	This work
17R up	GGCTGC**GAATTC**CCTGTCCTCCGGATATATAATCATCCCCATCATCG	This work

18R up	GGCTGC**GAATTC**TACCACCTGACAGAATATTATAAGCTCCGTGTGCC	This work
19R up	GGCTGC**GAATTC**CACCACCATAAAGTTTATCATCACCAGTACCACCT	This work
20R up	GGCTGC**GAATTC**GCCCATAAACATAAATATCATTACCAAAACCACCA	This work
21R up	GGCTGC**GAATTC**CTTCATCTGCAATTGTATGATGACCATAGTTTTGC	This work
22R up	GGCTGC**GAATTC**CTCTCTTAAATGCGATATCATCAAAGCTAATATCA	This work
23R up	GGCTGC**GAATTC**AACCTGACCGTTCATTATTATTATGAGGTATCTTATCA	This work
24R up	GGCTGC**GAATTC**GTTTTTTTCAGATGCTATATTACTGGCTTTTATAT	This work
25R up	GGCTGC**GAATTC**AAATAATCTTATTAATATCATTGGCAACACTGGTG	This work
26R up	GGCTGC**GAATTC**TCAGTGCATACAGACTATTATGAGAACTACATTTA	This work
27R up	GGCTGC**GAATTC**TTCAGCATTGACAGTTATATTATGATATTGTGCAAGCAAT	This work
28R up	GGCTGC**GAATTC**AATCTTGTGTGTGGGTATTATACTGATGCCTTATAGTTTCAGCATTGACAG	This work
29R up	GGCTGC**GAATTC**TAATACGAGACAATATATAATGCTTACCGTCATCC	This work
30R up	GGCTGC**GAATTC**AGTTAACACCTCAAGTAAAATCCTCCTGTATTTCA	This work
31R up	GGCTGC**GAATTC**TTCAAAAAGTATCACTATAATAAATGCTATTGTGA	This work
32R up	GGCTGC**GAATTC**CATTCCTGAGAAACTTATCATCAAGACGTCGACGT	This work
33R up	GGCTGC**GAATTC**CTGAATTTCCGGCAGTATATTTCTGCCTGACTTTT	This work
34R up	GGCTGC**GAATTC**ATATGTATAAATTAATATTATTCAAAATAACATTA	This work
35R up	GGCTGC**GAATTC**ATCATGCCCATCAATTAAAATCTGTCCGGTCTCTG	This work
36R up	GGCTGC**GAATTC**ACTACGATTTAATAGTATATTTTCCTGCAATACAA	This work
37R up	GGCTGC**GAATTC**GAGAAGCTAATGTAATATTATCGATAATACTACGA	This work
38R up	GGCTGC**GAATTC**CTTTTAGTTCTCTAATAAAATCATGGGCACCTGCA	This work
39R up	GGCTGC**GAATTC**TGACATATTTTTCATTATTATATTTTCCGACTCATAAT	This work
40R up	GGCTGC**GAATTC**AGATAAGAGAAGAGATATAATTACAAATAATAAAATACTCCATACAACAA	This work
41R up	GGCTGC**GAATTC**TACCATTATTTTTTCTACAATCGCATTCTCTATCG	This work
42R up	GGCTGC**GAATTC**AGATGCTTCTGATTTTAAAATATCAGATTCAACAC	This work
43R up	GGCTGC**GAATTC**TATTCAACAAATACTTAAAATCATTGAGTTTTCTT	This work
44R up	GGCTGC**GAATTC**ACCACCTTCTGTATGTATATTTAACTCCTGAACAG	This work
45R up	GGCTGC**GAATTC**CTGTTACTTCGAGAATATCATTATCAGGGACAATA	This work
46R up	GGCTGC**GAATTC**AACAGAATCTGCAGTTATATTTTTTACTTTCCCTG	This work
47R up	GGCTGC**GAATTC**ATTCCGATCAACAGATATAATCACGTTAAATACAA	This work
48R up	GGCTGC**GAATTC**TTCTTTCTCCCTGTATATCATTCCGATCAACAGAT	This work
1R down	CGCCCG**AAGCTT**CCTCCTGGCCAGTTTAGATCTGGAATGTGAGGTAAAATATA	This work
2R down	CGCCCG**AAGCTT**CCTCCTACTTATGTCTGAAGTAAAAAACAGACAAGATTTTAATTTTA	This work
3R down	CGCCCG**AAGCTT**CCTCCTTGGAATAGAAATACACCGTACTGAACGAGATGATA	This work
4R down	CGCCCG**AAGCTT**CCTCCTAGCCGGAACAGTTCTCTCAGCACTACAGAATTTTA	This work
5R down	CGCCCG**AAGCTT**CCTCCTACAACGGGAAGGAGAGGATATAAGTCAGAATGATA	This work
6R down	CGCCCG**AAGCTT**CCTCCTTTATTAATCAGCTTGTAGATACAGTATCAAGTATA	This work
7R down	CGCCCG**AAGCTT**CCTCCTTCTTGGAAATGTTGGTAAAGCTGTTTCGCAATATA	This work
8R down	CGCCCG**AAGCTT**CCTCCTGATAAATTTGAGCGAGCTAAGCAGCTTGAATCATA	This work
9R down	CGCCCG**AAGCTT**CCTCCTTAGAGAAAGAAAACAGAGTGGTAAATATGAATATA	This work
10R down	CGCCCG**AAGCTT**CCTCCTGTGGTAAATATGAATATATGACATCTCTTATTGTA	This work
11R down	CGCCCG**AAGCTT**CCTCCTGGATACATGGTCTGTAAAAGGCATAAAAAATCATA	This work
12R down	CGCCCG**AAGCTT**CCTCCTCTCATCAGAAGTATTTGCTGGTGAAGGTTATGATACCGTATCTTATAATA	This work

13R down	CGCCCG**AAGCTT**CCTCCTTGATGCAACAGGAGCATCAAAACCTGGTGAGTATA	This work
14R down	CGCCCG**AAGCTT**CCTCCTTGATGATGAATTCAAAGGCGGTAAGTTTAATGATA	This work
15R down	CGCCCG**AAGCTT**CCTCCTTGATATATTCCATGGCGCAGATGGGAACGATTATA	This work
16R down	CGCCCG**AAGCTT**CCTCCTTATATCGAAGGTAATTATGGTAATGATCGACTATA	This work
17R down	CGCCCG**AAGCTT**CCTCCTTGATCGACTATACGGCGATGATGGGGATGATTATA	This work
18R down	CGCCCG**AAGCTT**CCTCCTTAATGATGAGCTTCAGGCACACGGAGCTTATAATA	This work
19R down	CGCCCG**AAGCTT**CCTCCTTTATAATATTCTGTCAGGTGGTACTGGTGATGATA	This work
20R down	CGCCCG**AAGCTT**CCTCCTTAATGACTATCTGAATGGTGGTTTTGGTAATGATA	This work
21R down	CGCCCG**AAGCTT**CCTCCTTATTTATGTTTATGGGCAAAACTATGGTCATCATA	This work
22R down	CGCCCG**AAGCTT**CCTCCTTCGTTTGCACTTATCTGATATTAGCTTTGATGATA	This work
23R down	CGCCCG**AAGCTT**CCTCCTTCGAGAGATAAAAGTTGATAAGATACCTCATAATAATA	This work
24R down	CGCCCG**AAGCTT**CCTCCTATGAACGGTCAGGTTATATAAAAGCCAGTAATATA	This work
25R down	CGCCCG**AAGCTT**CCTCCTAAACATGGTTAATATCACCAGTGTTGCCAATGATA	This work
26R down	CGCCCG**AAGCTT**CCTCCTTCGTTATATGATGAGTAAATGTAGTTCTCATAATA	This work
27R down	CGCCCG**AAGCTT**CCTCCTTCTGTATGCACTGATATTGCTTGCACAATATCATAATATA	This work
28R down	CGCCCG**AAGCTT**CCTCCTAATATCATAATATAACTGTCAATGCTGAAACTATAAGGCATCAGTATAATA	This work
29R down	CGCCCG**AAGCTT**CCTCCTTGCGTTGATATGGCGGGATGACGGTAAGCATTATA	This work
30R down	CGCCCG**AAGCTT**CCTCCTTTATCCCCTCTGTTGTGAAATACAGGAGGATTTTA	This work
31R down	CGCCCG**AAGCTT**CCTCCTCAACGTTAAATATTATCACAATAGCATTTATTATA	This work
32R down	CGCCCG**AAGCTT**CCTCCTTATATCACCCTTATTACGTCGACGTCTTGATGATA	This work
33R down	CGCCCG**AAGCTT**CCTCCTATACACCGGTAGAGAAAAAGTCAGGCAGAAATATA	This work
34R down	CGCCCG**AAGCTT**CCTCCTGTATTCTTCTGACGGTAATGTTATTTTGAATAATA	This work
35R down	CGCCCG**AAGCTT**CCTCCTAGCGCTTTTATATACCAGAGACCGGACAGATTTTA	This work
36R down	CGCCCG**AAGCTT**CCTCCTGACGCCAGATTGGTGTTGTATTGCAGGAAAATATA	This work
37R down	CGCCCG**AAGCTT**CCTCCTAAATATACTATTAAATCGTAGTATTATCGATAATA	This work
38R down	CGCCCG**AAGCTT**CCTCCTTGAGGCAGCCAGACTTGCAGGTGCCCATGATTTTA	This work
39R down	CGCCCG**AAGCTT**CCTCCTCAACCAGTGCTCTTGATTATGAGTCGGAAAATATAATA	This work
40R down	CGCCCG**AAGCTT**CCTCCTCCAGACGTTCTCATTTTGTTGTATGGAGTATTTTATTATTTGTAATTATA	This work
41R down	CGCCCG**AAGCTT**CCTCCTGCAAAGAAATAAAACCGATAGAGAATGCGATTGTA	This work
42R down	CGCCCG**AAGCTT**CCTCCTAATTGACGGTGCCTGGTGTTGAATCTGATATTTTA	This work
43R down	CGCCCG**AAGCTT**CCTCCTGGTATCACAGGAAGGAAGAAAACTCAATGATTTTA	This work
44R down	CGCCCG**AAGCTT**CCTCCTCTCCTGTGAGTGGTACTGTTCAGGAGTTAAATATA	This work
45R down	CGCCCG**AAGCTT**CCTCCTAGAAACGCTGATGATTATTGTCCCTGATAATGATA	This work
46R down	CGCCCG**AAGCTT**CCTCCTGTCATGGTTACCTTACAGGGAAAGTAAAAAATATA	This work
47R down	CGCCCG**AAGCTT**CCTCCTTTCCGGATACAGGGCTTGTATTTAACGTGATTATA	This work
48R down	CGCCCG**AAGCTT**CCTCCTATTTAACGTGATTATATCTGTTGATCGGAATGATA	This work

Appropriate oligonucleotide primers (upstream and downstream) were purchased from Invitrogen or Alta Bioscience (University of Birmingham) and used at a final concentration of 1 µM. Deoxynucleoside triphosphates (dNTPs) (Bioline) were used at a final concentration of 2 mM (0.5 mM each). The PCR reaction mixes were made up to 100 µl in sterile distilled water. An example of a typical PCR cycle used is shown in Table 2.4.

2.9.2 DNA restriction digests

Two µl of each restriction enzyme (all from Invitrogen) was added to 50 µl of DNA solution and made up to a final volume of 100 µl with appropriate buffer and dH_2O. The restriction digests were incubated for 2-3 hours at 37 °C. When making plasmid vector for cloning, the digested plasmid DNA was treated with 3 µl of calf alkaline phosphatise (NEB) to remove the terminal 5' phosphate groups. The restriction digest mix was then incubated for a further 30-60 minutes at 37 °C. Where appropriate, phenol-choloroform extraction, or various QIAgen DNA extraction kits, was used to purify digested DNA.

2.9.3 Ligation of DNA fragments into plasmid vectors

One µl T4 DNA ligase (New England Biolabs) was used to ligate 1-10 µl of restriction digested insert DNA and 1-3 µl of restriction digested, and alkaline phosphatase treated, vector DNA. The reaction mixture was made up to a final volume of 20 µl with dH_2O and T4 DNA ligase buffer (New England Biolabs). The ligation mixes were incubated for 1 hour at room temperature. The entire ligation mix was used to transform into chemically competent cells using the technique defined in section 2.8. Nutrient agar or MacConkey-lactose plates, supplemented with appropriate antibiotics, were used to select cells transformed with ligated DNA. Candidate transformants were picked and grown as overnight cultures. The plasmid was then extracted and screened for the presence of the expected DNA insert. Thus, the

Table 2.4 Standard PCR cycle

Annealing temperature (T_A °C) was usually set 3 °C above the melting temperature of the primers. Extension time (X) was calculated based on PCR product length (15-30 seconds/bp)

Temperature	Time		Purpose
98 °C	5 min		Initial Denaturation
94 °C	10 s		Denaturation
T_A °C	30 s	30-35 cycles	Annealing
72 °C	X		Extension
72 °C	5 min		Final extension

plasmids were digested using restriction enzymes to excise the cloned fragment that was visualised on either 7.5 % PAGE gels or 1 % (*v/v*) agarose gels to check for the insert of expected size.

2.10 Sequencing

"Plasmid-to-profile" sequencing was done by the Functional Genomics and Proteomics Laboratory (University of Birmingham) and Research facilities (University of Warwick). Eight µl of plasmid miniprep was mixed with 2 µl of 10 µM sequencing primer for sequencing plasmid templates. Table 2.3 lists the primers used for sequencing inserts in plasmids.

2.10.1 Calibration of sequencing gels

Reagents and chemicals used:

Formic acid: 100 %

Sodium acetate: 3 M, pH 7.0

Piperidine: 1 M

Ethanol: 100 %

Gel loading buffer: 40 % *v/v* deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Method:

Maxam-Gilbert G+A sequencing ladders, derived from the DNA fragment used in the DNA footprinting experiment were used to calibrate DNA sequencing gels. Twelve µl of labelled

DNA fragment was added to 50 µl formic acid and left at room temperature for 2-3 minutes. The DNA was then precipitated using 20 µl of sodium acetate and 700 µl of 100% ethanol. The DNA pellet was dried and resuspended in piperidine and incubated for 30 minutes at 90 °C. The DNA was again precipitated, washed twice with 70 % ethanol and the dried pellet was resuspended in 20 µl of loading buffer.

2.10.2 Sequencing reactions for Primer extension

T7 Sequencing Kit (USB) was used for sequencing reactions. Sequencing reactions were used to calibrate gels for visualising primer extension experiments.

Reagents and chemicals used (provided in the kit):

'A' Mix-Short: 840 µM each dCTP, dGTP and dTTP; 93.5 µM dATP; 14 µM ddATP; 40 mM Tris-HCl (pH 7.5) and 50 mM NaCl.

'C' Mix-Short: 840 µM each dATP, dGTP and dTTP; 93.5 µM dCTP; 17 µM ddCTP; 40 mM Tris-HCl (pH 7.5) and 50 mM NaCl.

'G' Mix-Short: 840 µM each dATP, dCTP and dTTP; 93.5 µM dGTP; 14 µM ddGTP; 40 mM Tris-HCl (pH 7.5) and 50 mM NaCl.

'T' Mix-Short: 840 µM each dATP, dCTP and dGTP; 93.5 µM dTTP; 14 µM ddTTP; 40 mM Tris-HCl (pH 7.5) and 50 mM NaCl.

T7 DNA Polymerase: 8 units/µl in buffered glycerol solution.

Enzyme Dilution Buffer: 25 mM Tris-HCl (pH 7.5), 5 mM DTT, 100 µg BSA/ml and 5% glycerol.

Universal Primer: 5'-d[GTAAAACGACGGCCAGT]-3' in aqueous solution, 0.86 A_{260} units/ml (5 pmol/µl).

Annealing Buffer: 1 M Tris-HCl (pH 7.5), 100 mM MgCl2 and 160 mM DTT.

Labeling Mix-dATP: 1.375 µM each dCTP, dGTP and dTTP and 333.5 mM NaCl.

Stop Solution: 0.3 % each Bromophenol Blue and Xylene Cyanol FF; 10 mM EDTA (pH 7.5) and 97.5 % deionized formamide

Control template: 10 µg of single stranded M13mp18 DNA in 50 µl of Tris-EDTA buffer.

Method:

The control template was used at a concentration of 2 µg (~ 10 µl). Primer (2 µl) and annealing buffer (2 µl) was then added to make a total volume of 14 µl. The tube was vortexed gently, then centrifuge at 20000 g briefly. After incubation at 60 °C for 10 minutes. The tube was left at room temperature for at least 10 minutes, then centrifuge at 20000 g briefly. Labeling Mix-dATP with added [α-^{32}P]dATP was used. Two point five µl of 'A' 'G' 'C' and 'T' mix-shorts were pipetted into four different tubes. To label the reaction, the tube containing the 14 µl annealed template and primer was mixed with 3 µl of [α-^{32}P]dATP and 2 µl of diluted T7 DNA polymerase. This made a total volume 20-21 µl. The components were mixed by gentle pipetting, and the contents collected at the bottom of the tube by a brief centrifugation at 20000 g. The samples were then incubated at room temperature for 5 minutes. While this incubation is in progress, the four sequencing mixes were warmed at 37 °C for at least 1 minute. Immediately, after the 5 minute incubation of the labelling reaction, 4.5 µl of this reaction was pipetted into each of the four pre-warmed sequencing mixes, using a fresh pipette tip for each transfer. The components were mixed by gentle pipetting,

incubated at 37 °C for 5 minutes. The reactions were then stopped by adding 5 µl of Stop Solution to each tube. Spin briefly to collect the contents at the bottom.

2.10.3 DNA sequencing gels

Denaturing sequencing gels were used to analyse the DNA fragments obtained from DNA footprinting and primer extension experiments.

Reagents and chemicals used:

Acrylamide: The 6 % Denaturing acrylamide gel was made using UreaGel concentrate, diluent and buffer supplied by Geneflow

N,N,N′,N′-Tetramethylethylenediamine (TEMED)

Ammonium persulphate (APS) solution: 10 % (*w/v*) (diluted in dH_2O)

Fixing solution: 10 % (*v/v*) Methanol, 10 % (*v/v*) Acetic acid

Method:

The Ureagel concentrate, diluents and buffer were mixed according to the manufacturer's instructions and polymerisation of the gel was initiated by addition of 100 µl APS and 40 µl TEMED. The gel mixture was then poured between two 40 cm x 30 cm glass plates separated by 0.4 mm thick plastic spacers. Prior to sample loading the wells were washed thoroughly and the gels were pre-run at 60 W for 30 minutes. After washing the wells, the gel was loaded. Once run, the gel was soaked in fixing solution for 15 minutes and transferred to a 3 mm Whatman filter paper. Gels were dried under vacuum at 80 °C for 45 minutes and exposed to a Biorad phosphor screen in a light-proof autoradiography cassette overnight.

2.11 RNA polymerase σ factor purification

The protocol given below is for σ^{70} RA451 purification. However, all σ factors used in this work were purified using the same procedure.

Reagents and chemicals used:

Buffer FB: 20 mM Tris, pH 7.5 (3.6g in 1.5L), 1 mM EDTA (0.558g in 1.5L), 10 % Glycerol (150ml in 1.5L). Adjust pH before adding EDTA and Glycerol

Dialysis buffer: FB + 100 mM NaCl

Elution buffer: FB + 1.5 M NaCl

Denaturing buffer: Dialysis buffer (40 ml) + 6 M Guanidine hydrochloric acid

Method:

The expression vector, pET21b/PVRσ274, was constructed by cloning the mutant *E. coli* PVRσ 274A on a *Nhe*I-*Hin*dIII DNA fragment.

Day 1

1. Transform pET21b/PVRσ RA451 into T7 express and plate on agar plate with ampicillin.

Day 2

2. Inoculate 5 ml LB (with ampicillin) with a single colony from transformation plate and incubate overnight at 37 °C.

Day 3

3. Inoculate 500 ml LB (with ampicillin) with 5 ml overnight culture.

4. Incubate with shaking at 37 °C until the culture reaches an OD_{650} of ~ 0.6. Record OD_{650} and take a sample for SDS-PAGE (500 µl), then add IPTG to a final concentration of 1 mM (500 µl of 1 M IPTG) to induce expression of σ^{70}. Continue to grow for a further 3 hours.
5. After 3 hours, record OD_{650} and take sample for SDS-PAGE. Chill cells for 15 min on ice, then pour culture into centrifuge bottles and spin at 4 °C for 10 min at 8000 rpm. Discard supernatant and store cell pellets overnight at -20 °C.

Day 4

6. Resuspend pellet in 20 ml lysis buffer (Buffer FB).
7. Add PMSF (3.5 µl of 100 mg/ml solution in ethanol).
8. Add lysozyme to a final concentration of 0.5 mg/ml (350 µl of 30 mg/ml solution).
9. Add sodium deoxycholate to a final concentration of 0.2 % (400 µl of 10 % stock solution).
10. Incubate 10 min on ice.
11. Sonicate 3 times for 40 s each time with 30 s on ice in between.
12. Centrifuge for 15 min at 17,000 rpm at 4 °C.
13. Discard supernatant and resuspend pellet in 20 ml lysis buffer. The pellet will be chalky, and difficult to resuspend. Transfer to a homogeniser to aid resuspension.
14. Add lysozyme to a final concentration of 0.5 mg/ml (350 µl of 30 mg/ml solution).
15. Add n-octyl β-D glucopyranoside to a final concentration to a final concentration of 0.2 % (400 µl of 10 % solution).
16. Sonicate 3 times for 40 s each time with 30 s on ice in between.
17. Centrifuge for 15 min at 17,000 rpm at 4 °C.
18. Discard supernatant and resuspend pellet in 40 ml denaturing buffer.

19. Transfer to dialysis tubing, and dialyse overnight against 2 litre dialysis buffer at 4 °C.

Day 5

20. After dialysis, a precipitate may form in the dialysis bag. Transfer the contents of the dialysis bag to a centrifuge tube, and spin for 15 min at 17,000 rpm at 4 °C to remove this precipitate.
21. Remove supernatant and filter through 0.2 µm syringe filter.
22. Wash a 5 ml HiTrap Q FF sepharose anion exchange column (GE Healthcare/Amersham) with dialysis buffer.
23. Load the filtered supernatant on to the column.
24. Elute the protein from the column in a linear sodium chloride gradient (from 100 mM to 1 M sodium chloride over 100 min). Collect fractions and measure the OD_{280}.
25. Analyse protein containing fractions by SDS-PAGE (Figure 2.4), and measure protein concentration using Bradford assay.
    a. Blank- 200 µl Biorad Solution + 800 µl dH_2O
    b. Protein as above but subtract protein volume from dH_2O and measure OD_{595}.

Fraction	OD_{595}
F32	1.323
F33	2.196
F34	1.890
F35	1.337

26. Mix protein solution with an equal volume of glycerol to make stock solution. Store at -20 °C. Store remaining protein samples at -80 °C.

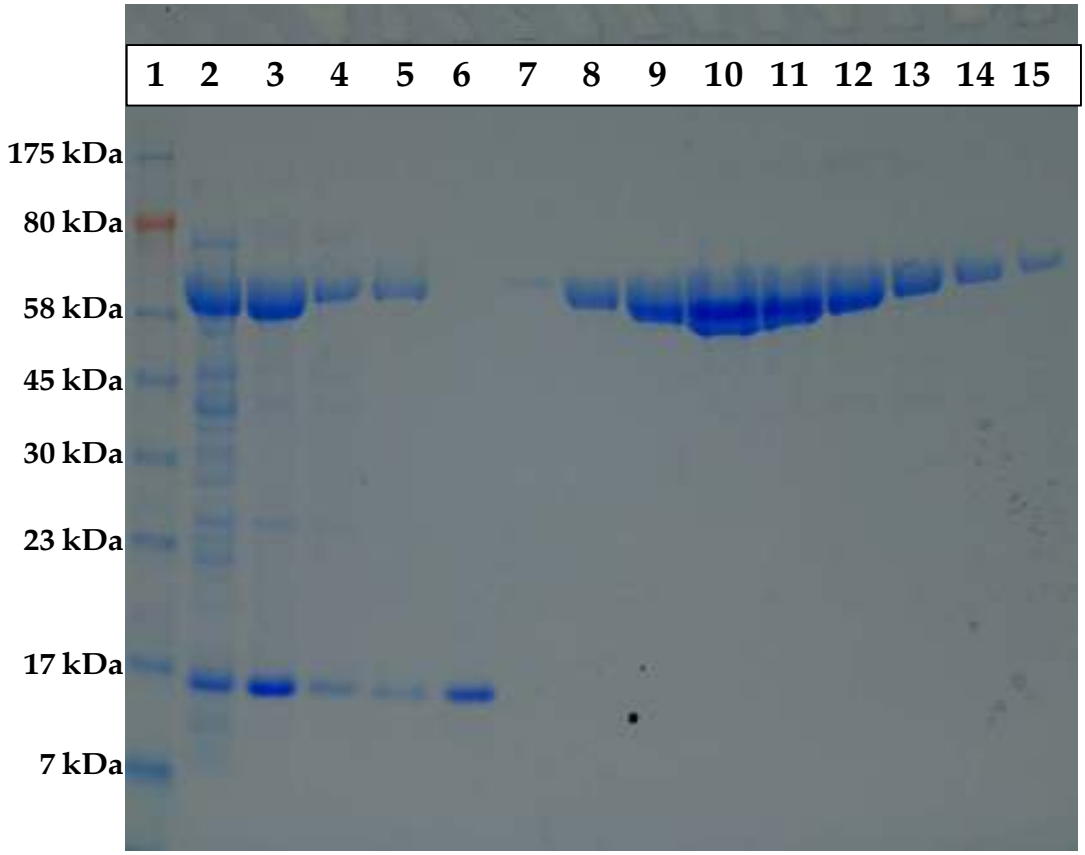


Figure 2.4 SDS-PAGE gel showing protein samples generated during purification of *E. coli* σ^{70} RA451

Lanes

1. Marker*
2. Total protein
3. Insoluble fraction
4. Soluble fraction
5. Guanidine HCl fraction
6. Peak Fraction 4
7. Peak Fraction 30
8. Peak Fraction 31
9. Peak Fraction 32
10. Peak Fraction 33
11. Peak Fraction 34
12. Peak Fraction 35
13. Peak Fraction 36
14. Peak Fraction 37
15. Peak Fraction 38

* Colorplus Prestained Protein marker, Broad range (7-175 kDa), New England Biolabs

Holoenzyme reconstitution: core enzyme ordered from Epicentre and purified σ factor was added in equimolar amounts on the day of the experiment and incubated for 20 min at room temperature.

2.12 Potassium permanganate ($KMnO_4$) footprinting

Potassium permanganate can readily oxidise double bonds between carbon atoms in organic molecules, including nucleic acids. This oxidation does not break DNA strands but makes modified nucleotides susceptible to attack by piperidine. Thus, the location of oxidised nucleotides can be determined by subsequent piperidine treatment. Single stranded DNA is sensitive to $KMnO_4$ mediated oxidation because bases are accessible to $KMnO_4$. However, native double stranded DNA is not sensitive to the $KMnO_4$ treatment. Protocol mentioned in Savery *et*. *al*. (1996a) was followed.

Reagents and chemicals used:

Binding buffer: 60 mM Hepes (pH 8.0), 1.25 mM potassium glutamate, 38 μg/ml Herring sperm DNA.

$KMnO_4$: 200 mM $KMnO_4$

Stop solution: 3 M Ammonium acetate, 0.1 mM EDTA, 1.5 M 2-mercaptoethanol

Piperidine: 1 M piperidine

Gel loading buffer: 40 % *v/v* deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Method:

Purified *Aat*II–*Hin*dIII DNA fragments were derived from maxi preparations (using a Qiagen maxiprep kit) of plasmid pSR carrying the fragment to be analysed. Fragments were labelled at the *Hin*dIII end as described in section 2.2.3. DNA fragments were used at a final concentration of 10–40 nM in binding buffer. RNA polymerase was then added and the volume reactions were incubated for 10 minutes at 37 °C. One µl of 200 nM $KMnO_4$ was added to each reaction and samples were incubated at 37 °C for 4 minutes. Stop solution was then added and the DNA was extracted using phenol/chloroform. The recourse DNA was then precipitated using ethanol, dried under vacuum, and resuspended in 100 µl of 1 M piperidine. After incubation for 30 minutes at 95 °C DNA was again precipitated using ethanol, dried, and resuspended in 10 µl of gel loading buffer. Footprints were analysed on a 6 % DNA sequencing gel (2.10.3). The results of all footprinting experiments were visualized by exposing a Biorad phosphor screen to the dried gel. After exposure the phosphor screen was scanned using a phosphor imager and Quantity One software was used for image analysis.

2.13 FeBABE footprinting

Fe^{2+} chelated Bromoacetamidobenzyl-EDTA (FeBABE) is a specific DNA cleavage reagent that can be attached to specific cysteine side chains in proteins, which then cleave nucleic acids within a 12 Å radius of the attachment site.

Reagents and chemicals used:

Binding buffer: 60 mM Hepes (pH 8.0), 1.25 mM potassium glutamate, 38 µg/ml Herring sperm DNA.

Gel loading buffer: 40 % *v/v* deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Sodium ascorbate: 5 mM

Hydrogen peroxide: 5 mM

Stop solution: 0.15 mM EDTA, 25 mM Thiourea

Method:

Reactions were set up with 40 nM of radiolabelled fragment and protein at the desired concentration in 1 x binding buffer. Reactions were then incubated for 20 minutes at room temperature. 5 µl of H_2O_2 and 5 µl of sodium ascorbate was then added to the reactions and incubated at 37 °C for a further 20 minutes. The reactions were then stopped by adding 21 µl of stop solution. Reactions were diluted to 200 µl with TE buffer and DNA was extracted using phenol/chloroform. The aqueous layer was transferred to a fresh tube and 1 µl of 20 mg/ml glycogen was added along with 400 µl of ice-cold ethanol. The solutions were incubated at -70 °C for 15 minutes to stimulate DNA precipitation. After centrifuging at 20000 g for 15 minutes at 4 °C the DNA pellet was washed with 600 µl of ice-cold 70 % ethanol and recentrifuged at 20000 g for 10 minutes at 4 °C. The supernatant was discarded and pellet was dried. After drying DNA was resuspended in 8 µl of gel loading buffer. Before loading onto a 6 % sequencing gel, the samples were incubated at 90 °C for 2 minutes.

2.14 DNAse I footprinting

Reagents and chemicals used:

Binding buffer: 20 mM Tris pH 7, 10 mM $MgCl_2$, 100 mM EDTA and 120 mM KCl. 12.5 mg/ml Herring sperm DNA

DNAse I: 10 Units/ µl, 1 in 50 dilutions

Stop solution: 1 ml 3 M NaAc pH 7.0 + 200 µl 500 mM EDTA, made up to 10 ml with dH_2O

Phenol/Chloroform/Isoamyl alcohol: pH 6.7/8.0

Ethanol: 100 % *v/v* and 70 % *v/v*

Gel loading buffer: 40 % *v/v* deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Method:

Purified *Aat*II–*Hin*dIII DNA fragments were derived from maxi preparations (using a Qiagen maxiprep kit) of plasmid pSR carrying the fragment to be analysed. Fragments were labelled at the *Hin*dIII end as described in section 2.2.3. DNA fragments were used at a final concentration of 10–40 nM in buffer containing 20 mM Tris pH 7, 10 mM $MgCl_2$, 100 mM EDTA and 120 mM KCl. Herring sperm DNA (12.5 µg/ml) was also added as a nonspecific competitor. H-NS was then added and the reactions were incubated for 20 minutes at room temperature. Two µl DNAse I was added, the reactions were incubated at room temperature for 40 seconds and stopped with 200 µl stop solution. An equal volume of phenol/chloroform was then added, the sample was vortexed and centrifuged at 20000 g for 3 minutes. The aqueous layer was transferred to a fresh tube and 1 µl of 20 mg/ml glycogen, along with 400 µl ice-cold ethanol, was added to the reaction. The solutions were incubated at -70 °C for 15 minutes. After centrifuging at 20000 g for 15 minutes at 4 °C, the DNA pellet was washed with 600 µl ice-cold 70 % ethanol and recentrifuged at 20000 g for 10 minutes at 4 °C. The supernatant was then discarded and the pellet was dried and resuspended in 8 µl gel loading buffer. Before loading, the samples were incubated at 90 °C for 2 minutes and loaded onto a 6 % sequencing gel which was run at 60 W.

2.15 Primer Extension

Primer extension assays were used to map the 5' end of RNA transcripts and thus define transcription start sites. A radiolabelled primer, 37 nucleotides in length, complementary to a region near the 5' end of the gene is annealed to the RNA and reverse transcription is initiated to obtain a complementary DNA (cDNA). The product is analysed on a denaturing polyacrylamide gel, alongside DNA sequencing reactions as size calibrators, and the transcription start site can be mapped.

Chemicals and reagents used:

Sodium acetate: 3 M, pH 7.0

Phenol/Chloroform/Isoamyl alcohol: pH 6.7/8.0

Ethanol: 100 % *v/v* and 70 % *v/v*

Gel loading buffer: 40 % *v/v* deionised formamide, 5 M urea, 5 mM NaOH, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Hybridisation buffer: 20 mM HEPES, 0.4 M NaCl, 80 % formamide

5x reverse transcriptase buffer (Promega)

DTT: 50 mM

dNTPs: 10 mM

AMV reverse transcriptase (Promega)

RNasin (Promega)

Ammonium acetate: 3 M, pH 4.8

RNA was purified from cells containing pRW50 carrying the promoter to be analysed, using a QIAgen RNeasy mini kit. The D49724 primer, that corresponds to sequence downstream of

the *Hin*dIII site in pRW50, was labelled with [γ-^{32}P]-ATP as in section 2.2.3. The RNA extracted was then hybridized to 5' end-labelled D49724. Thus, 1 µl of labelled primer was mixed with 20-30 µg of extracted RNA. Zero point one volumes of 3 M sodium acetate pH 7.0 and 2.5 volumes of ice-cold ethanol was then added to the reaction, which was vortexed, and incubated at -70 °C for 15 minutes. After centrifuging at 20000 g for 15 minutes at 4 °C, the DNA pellet was washed with 600 µl of ice-cold 70 % ethanol and recentrifuged at 20000 g for 10 minutes at 4 °C. The supernatant was discarded and the pellet was dried before being resuspended in 30 µl hybridisation buffer. Reactions were incubated at 50 °C for 5 minutes, 75 °C for 15 minutes, and then at 50 °C for 3 hours. Seventy five µl of ice-cold ethanol was then added to the reaction and incubated at -70 °C for 15 minutes or overnight. After centrifuging at 20000 g for 15 minutes at 4 °C, the DNA was pelleted and washed with 600 µl of ice-cold 70 % ethanol. The DNA was again pelleted by recentrifugation at 20000 g for 10 minutes at 4 °C. The supernatant was discarded and the pellet was dried and resuspended in 31 µl RNAse free water. Ten µl of 5x reverse transcriptase buffer, 1 µl 50 mM DTT, 5 µl 10 mM dNTPs, 2.5 µl AMV reverse transcriptase and 0.6 µl RNasin was added and the reaction was incubated at 37 °C for 1 hour. The reactions were then incubated at 72 °C for 10 minutes to deactivate the enzyme and treated with 1 µl 10 mg/ml RNase for 30 minutes at 37 °C. 6.7 µl 3 M ammonium acetate pH 4.8 and 125 µl ice-cold ethanol were added to the reactions. After centrifuging at 20000 g for 15 minutes at 4 °C, the DNA pellet was washed with 600 µl ice-cold 70 % ethanol and recentrifuged at 20000 g for 10 minutes at 4 °C. The supernatant was discarded and the pellet was dried and resuspended in 4 µl gel loading buffer. Before loading, the samples were incubated at 90 °C for 2 minutes and loaded onto a 6 % sequencing gel which was run at 60 W.

2.16 β-galactosidase assays

The Lac activity was determined for cells carrying different plasmid-encoded promoter::*lacZ* fusions.

Reagents and chemicals used:

Z-buffer: 0.75 g KCl, 0.25 g $MgSO_4.7H_2O$, 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_20$, 2.70 ml β-mercaptoethanol

Toluene

Sodium deoxycholate: 1 % (*w/v*)

ONPG in Z-buffer: 13 mM (made on day of use).

Sodium carbonate: 1 M

Z-buffer: 0.75 g KCl, 0.25 g $MgSO_4.7H_2O$, 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4.2H_20$, 2.70 ml **β-mercaptoethanol**

Method:

A fresh colony of cells carrying the desired plasmid construct was used to inoculate a 5 ml aliquot of LB or minimal M9 salts medium, supplemented with antibiotics/sugars as appropriate. These cultures were grown at 37 °C overnight in a shaking incubator. The next day, ~ 50 µl of the overnight culture was used to inoculate 5 ml of the same. These cultures were then incubated in a shaking incubator at 37 °C until they reach mid- exponential phase i.e. an OD_{650} of 0.3-0.5 (for stationary phase assays this step was omitted). Two drops each of toluene and 1 % sodium deoxycholate solution was then added to each culture and vortexed for 15 seconds to lyse the cells. The lysed cultures were then returned to 37 °C for about 20

minutes to allow the toluene to evaporate. To assay the lysates for β-galactosidase activity, 2.5 ml of 2-Nitrophenyl β-D-galactopyranoside (ONPG) solution was then added to 100 µl of each lysate. The solution was incubated until it turned yellow in colour. The reaction was then stopped by adding 1 ml of 1 M sodium carbonate and the OD_{420} was measured.

The following formula was used to calculate the β-galactosidase activity:

$$\beta\text{-galactosidase activity} = \frac{1000 \times 2.5 \times 3.5 \times OD420}{OD650 \times 4.5 \times t \times v} \text{ nmol/min/mg bacterial mass}$$

Wherein:

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

3.5 = final assay volume (ml)

1000/4.5 = factor for conversion of OD_{420} into nmol o-nitrophenyl (ONPG), based on 1 nmol/ml ONP having an OD_{420} of 0.0045

t = incubation time (min)

v = volume of lysate added (in ml)

In each experiment, each assay was done in triplicate. The mean β-galactosidase activity and standard deviation was calculated and plotted as a graph or presented in a table. A minimum of at least two separate experiments for each strain and set of conditions were done.

2.17 *In vitro* transcription assays

Reagents and chemicals used:

***In vitro* transcription buffer**: 20 mM Tris pH 7.9, 5 mM $MgCl_2$, 500 μM DTT, 50 mM KCl, 100 μg/ml BSA, 200 μM ATP, 200 μM GTP, 200 μM CTP, 10 μM UTP with 5 μCi [α-^{32}P]-UTP.

Stop solution: 40 % *v/v* deionised formamide, 1 mM EDTA, 0.025 % *w/v* bromophenol blue, 0.025 % *w/v* xylene cyanol FF

Method:

The *in vitro* transcription experiments were performed as described previously by Rhodius *et. al*. (2001). A Qiagen maxiprep kit was used to purify supercoiled pSR plasmid carrying the DNA promoter to be analysed. This template (16 μg/ml) was incubated with *in vitro* transcription buffer. The reaction was started by adding purified *E. coli* $E\sigma^{70}$ or $E\sigma^{38}$. Labelled RNA products were analysed on a denaturing polyacrylamide gel.

2.18 DNA bending assays

DNA fragments generated by PCR were separated by electrophoresis on a 7.5 % polyacrylamide non-denaturing gel. Electrophoresis was performed at 4 °C in TBE buffer. After electrophoresis, the gels were stained with ethidium bromide and DNA was visualized by UV illumination.

2.18.1 Modelling of DNA fragments *in silico*

Changes in DNA bending were modelled computationally using the 'model.it' web server (http://hydra.icgeb.trieste.it/dna/model_it.html) using the default parameters (Vlahovicek *et al.,* 2003). Predicted DNA structures were downloaded in pdb format and PyMOL was used to prepare figures.

2.19 Tethered Particle Motion (TPM)

Reagents and buffers:

10 x RE Buffer (40 ml): 1 M Tris pH 7.9, 1 M KAc, dH_2O

Casein Buffer (5 ml): 10 x RE Buffer, 20 mg/ml Casein, 1 M DTT, dH_2O

CbpA Buffer: 20 mM Tris pH 7.0, 10 mM $MgCl_2$, 100 µM EDTA, 120 mM KCl

Method

A glass slide and coverslip were thoroughly cleaned using Acetone and Ethanol and a flow cell was created using a parafilm. The flow cell was filled with 20 µg/ml Anti DIG. The Anti DIG was then washed out and the flow chamber was incubated with DNA tether labelled with biotin or dig at either ends (See Section 4.2). Washed Streptavidin coated polystyrene particles, with diameter 0.46 µm, were then flushed into the flow cell and then incubated with DNA tethers. CbpA buffer was used during the incubation of the tethers and attached beads with CbpA. Figure 4.2 explains the steps involved in preparation of the flow cell for the TPM experiment.

Effect of promoter spacer region sequence on promoter selection

Chapter 3

3.1 Introduction

The -10 and -35 promoter elements are recognized by the σ subunit of RNA polymerase during transcription initiation (Burgess *et. al.,* 1969). As mentioned in Section 1.1.2, multiple σ factors, with different DNA binding specificities, are encoded by *E. coli* (Gruber and Gross, 2003). Interestingly, it is becoming increasingly apparent that, at many gene regulatory regions, promoters with a different σ factor preference overlap (Wade *et. al*., 2006). In this chapter I have investigated an example of such overlapping promoters. In particular, I have focussed on the shared promoter spacer region (the DNA between the -10 and -35 elements) of the overlapping promoters. My hypothesis, based on recent work (Liu *et. al*., 2004), was that altering spacer region sequence might affect overlapping promoter differently. The two overlapping promoters that I have investigated drive expression of a nucleoid associated protein known as Curved-DNA binding protein A (CbpA).

CbpA is a homologue of the co-chaperone protein DnaJ. However, CbpA has additional DNA-binding activity and is considered a nucleoid associated protein. CbpA was originally isolated based on its ability to bind curved DNA (Ueguchi *et. al.,* 1994). Hence, whilst CbpA will bind to any DNA sequence, it binds with a higher affinity to curved DNA than non-curved DNA (Ueguchi *et. al.,* 1994). CbpA homologues are found in many bacteria related to *E. coli* (Chintakayala and Grainger, 2011). Importantly, in *E. coli*, CbpA is expressed in stationary phase. Hence, whilst the *cbpA* regulatory DNA contains overlapping promoters for $E\sigma^{38}$ and $E\sigma^{70}$, the $E\sigma^{38}$ promoter is by far the most active (Yamashino *et. al*., 1994).

3.2 The *cbpA* regulatory region and binding of RNA polymerase

Figure 3.1A shows the *cbpA* regulatory region DNA fragment used in this work. The fragment contains two overlapping promoters separated by 5 bp. Each of the two promoters is recognised by a different RNA polymerase holoenzyme. Thus, one promoter is recognised by $E\sigma^{70}$ and the other by $E\sigma^{38}$ (Yamashino *et. al*., 1994; Cho *et. al*., 2009). In a preliminary experiment, $KMnO_4$ footprinting, which detects open complex formation by RNA polymerase (Chapter 2, Section 2.12), was done to confirm the σ factor preference of the two *cbpA* promoters. Figure 3.1B shows the results obtained. As expected, $E\sigma^{70}$ and $E\sigma^{38}$ produce different DNA opening patterns. Furthermore, the patterns are offset by the expected 5 bp. Note that the concentration of $E\sigma^{70}$ required for open complex formation in this assay is higher than that required for $E\sigma^{38}$. Again, this was expected because CbpA is known to be expressed in stationary phase.

In a further set of confirmatory experiments *in vitro* transcription assays (Chapter 2, Section 2.17) were used to distinguish the two promoters. Hence, the DNA fragment shown Figure 3.1A was cloned into plasmid pSR. This places the two *cbpA* promoters upstream of the factor independent λ*oop* transcription terminator. The pSR plasmid also encodes the 108 base RNAI transcript that functions as an internal control in these assays. Thus, *in vitro*, RNA polymerase generates distinct transcripts using pSR as a template. Figure 3.2 shows the results obtained from *in vitro* transcription assays. Both the $E\sigma^{70}$ and $E\sigma^{38}$ dependent *cbpA* promoters produce transcripts that can be detected after electrophoresis. As expected, $E\sigma^{70}$ produces less transcript than $E\sigma^{38}$. Also, the transcript produced by $E\sigma^{70}$ is longer than that produced by $E\sigma^{38}$. This confirms that the *cbpA* regulatory DNA fragment, shown in Figure 3.1A, contains two promoters with different σ factor specificity.

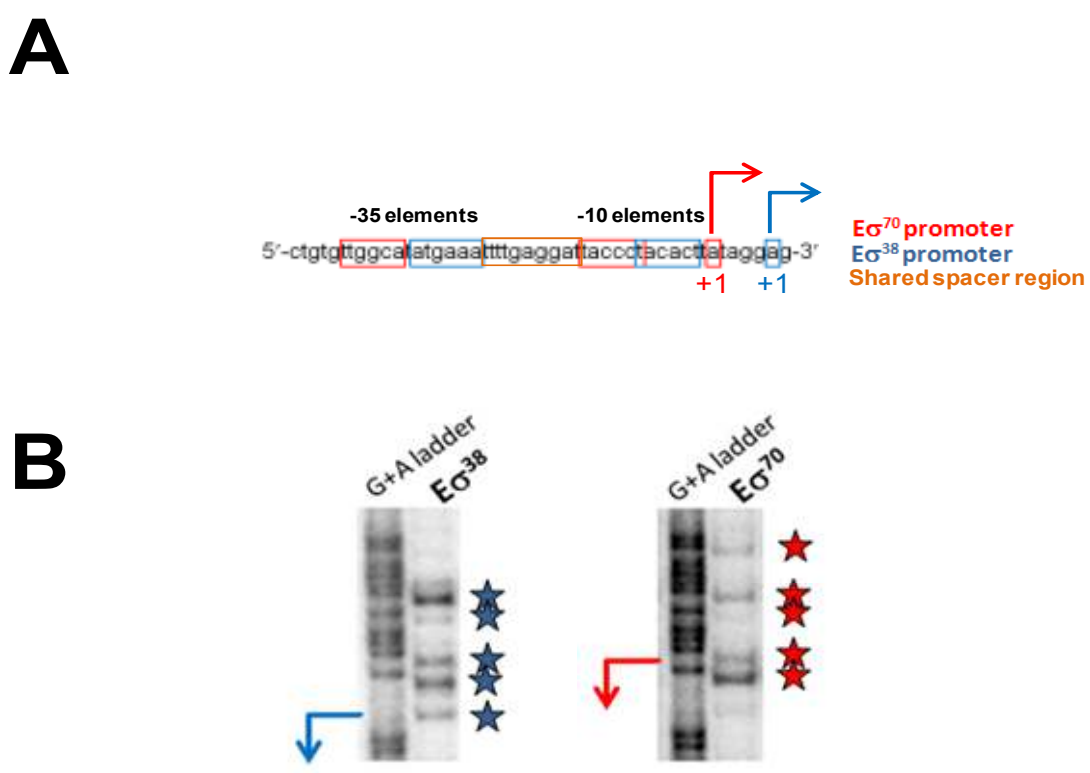


Figure 3.1 Overlapping promoters in the *cbpA* regulatory region.

A. Sequence of the *cbpA* regulatory region. The panel shows the *cbpA* regulatory DNA fragment used in this chapter. Promoters for $E\sigma^{70}$ and $E\sigma^{38}$ are highlighted in red and blue respectively. The transcription start sites are designated as '+1' and labelled with arrows. Please note that, throughout this work, all mutations in the *cbpA* regulatory region have been numbered with respect to the $E\sigma^{70}$ '+1'. The shares spacer region is highlighted by orange box

B. Open complex formation by $E\sigma^{70}$ and $E\sigma^{38}$. Open complex formed by RNA polymerase, at the *cbpA* regulatory region, was detected by $KMnO_4$ footprinting. The positions of $KMnO_4$ sensitivity are highlighted by blue (for $E\sigma^{38}$) and red stars (for $E\sigma^{70}$) respectively. The transcription start sites are shown by arrows. 480 nM $E\sigma^{38}$ was added and $E\sigma^{70}$ was added at three times higher concentration.

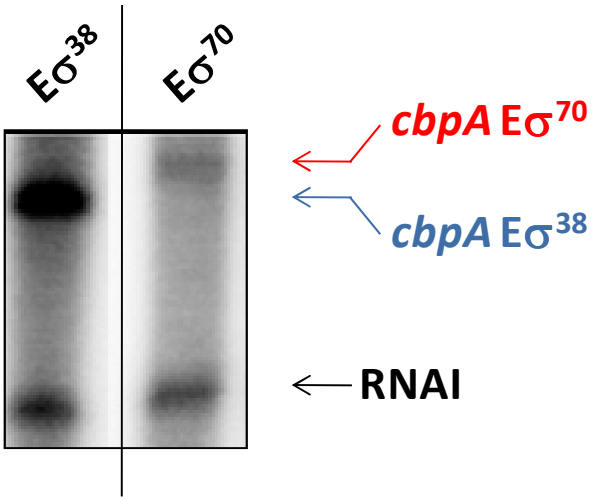


Figure 3.2 *In vitro* transcriptions by Eσ^{38} and Eσ^{70}.

Results of by an *in vitro* transcription assay are shown. The lower band corresponds to the internal control RNAI transcript. The upper bands are transcripts that initiate from the *cbpA* regulatory region and are due to transcription by either Eσ^{38} (blue) or Eσ^{70} (red). Eσ^{38} and Eσ^{70} were added in equal amounts (480 nM).

3.3 Selection of spacer region mutations that influence transcription

Having confirmed the position, and σ factor preference, of the two *cbpA* promoters, our next goal was to understand how the shared promoter spacer region influenced transcription. A library of DNA fragments, carrying derivatives of the *cbpA* regulatory region (Figure 3.1A), was created as a first step. The library (based on the DNA fragment shown in Figure 3.1A) contained single base mutations in the shared portion of the spacer region (highlighted orange in Figure 3.1). Table 2.3 (Chapter 2) lists the oligonucleotides used to introduce these single random base substitutions. The library of DNA fragments was cloned upstream of *lacZ* in plasmid pRW50 to create a library of *lacZ* fusions. The plasmid library was then used to transform *lac*$^-$ JCB387 cells. The transformants were plated on MacConkey agar medium. Note that the wild type *cbpA* regulatory region stimulates very low levels of *lacZ* transcription. Hence, the majority of transformants had a *lac*$^-$ phenotype on the MacConkey indicator plates (i.e. the colonies appear white). A total of about 216 transformants were screened. Of these, 12 transformants had a *lac*$^+$ phenotype (i.e. the colonies were red). The 12 *lac*$^+$ colonies were restreaked on fresh MacConkey indicator plates. Finally, β-galactosidase assays were used to measure *lacZ* expression for each of the *lac*$^+$ candidates. The *cbpA* regulatory region, carried by pRW50 derivative in each of the 12 candidates, was sequenced simultaneously. Table 3.1 shows the data obtained from this analysis. The biggest stimulatory effect on transcription was due to the introduction of either a T at promoter position -18 or a T at promoter position -17. Note that we used the Eσ^{70} transcription start site as the point of reference for numbering mutations in the *cbpA* regulatory region. Transcription was also

Mutation	No. of isolates	β-galactosidase activity
WT	N/A	95
-14G	1	127
-15T	2	126
-17C	1	158
-17T	4	199
-18T	3	249
-18A	1	221

Table 3.1 Spacer DNA mutations that increase transcription from the *cbpA* regulatory region

This table shows β-galactosidase activities obtained from overnight cultures of JCB387 cells carrying different *cbpA::lacZ* fusions in plasmid pRW50. Please note that the mutations are numbered with respect to the σ^{70} dependent *cbpA* transcription start site (Figure 3.1A). β-galactosidase activities were measured at least three times and the standard derivation for each measurement is less than 10 % of the activity value.

stimulated by the -18A and -17C mutations but to a lesser extent. Single base changes in the extended -10 element (-15T and-14G) had the smallest effect.

3.4 Mutations at position -18 stimulate only the σ^{70} dependent *cbpA* promoter

As the substitutions at position -18 had the biggest stimulatory effect, we decided to investigate this position further. In particular, we wondered if mutations at position -18 affected both, or only one, of the two overlapping *cbpA* promoters. Site directed mutagenesis was used to make a -18C mutation so that all base changes, at position -18, could be tested. Thus, *Eco*RI-*Hin*dIII fragments, containing the wild type, -18A, -18C and -18T derivatives of the *cbpA* regulatory region (Figure 3.1) were cloned into plasmid pSR. *In vitro* transcription assays (using either purified $E\sigma^{70}$ or $E\sigma^{38}$) were then done using these pSR based templates. This allowed me to investigate the effects of the mutations on transcription by different holoenzymes. Figure 3.3 shows the data obtained. Remarkably, all of the mutations at position -18 stimulate transcription by $E\sigma^{70}$ (compare lane 2 with lanes 4, 6 and 8) but do not stimulate transcription by $E\sigma^{38}$ (compare lane 1 with lanes 3, 5 and 7). To confirm this observation *in vivo*, the same promoter derivatives were cloned into the *lacZ* expression vector pRW50. The pRW50 derivatives were then used to transform MC4100 and MC4100*rpoS::kan* cells. β-galactosidase expression was then measured in the transformants. Recall that the wild type *cbpA* regulatory region is almost totally dependent on $E\sigma^{38}$ (Figure 3.2, Yamashino *et. al*., 1994). Thus, only background LacZ expression was detectable in MC4100*rpoS::kan* cells carrying the WT *cbpA::lacZ* fusion (Figure 3.4). Strikingly, for all of the regulatory region derivatives with mutations at position -18, the dependence on $E\sigma^{38}$ was

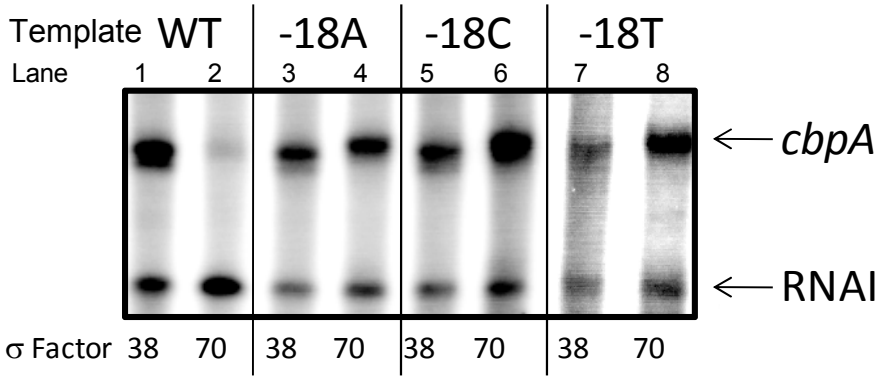


Figure 3.3 Effect of mutations at position -18 on transcription by Eσ^{38} and Eσ^{70} *in vitro.*

The gel shows transcripts produced *in vitro* by Eσ^{38} and Eσ^{70} from the *cbpA* regulatory region. Lanes 1, 3, 5 and 7 shows transcripts produced by Eσ^{38} and the control RNAI transcript. Lanes 2, 4, 6 and 8 shows transcripts produced by Eσ^{70}. The control RNAI transcript is labelled.

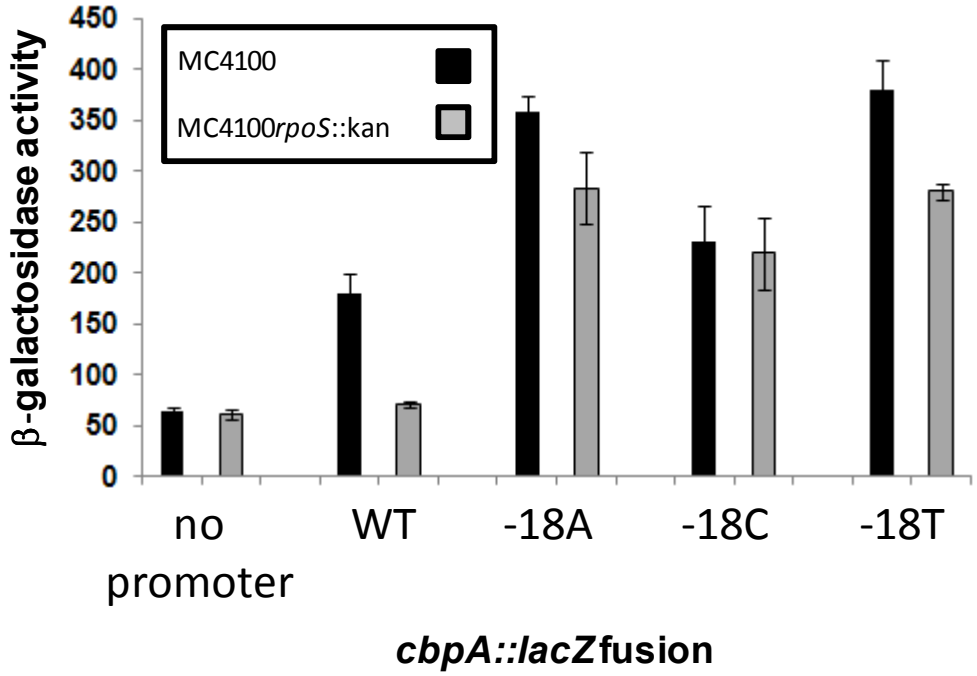


Figure 3.4 Activity of *cbpA* regulatory region derivatives *in vivo.*

The bar chart shows β-galactosidase expression driven by different *cbpA* regulatory region derivatives cloned in plasmid pRW50. β-galactosidase expressions were measured in MC4100 and the *rpoS::kan* derivative. The 'no promoter' value was obtained using pRW50 carrying no promoter insert. Overnight cultures were used to perform the assays.

drastically reduced. Hence, transcription increased substantially in MC4100*rpoS::kan* cells (Figure 3.4). I conclude that this is because the activity of the $E\sigma^{70}$ dependent promoter is stimulated by changes at position -18. To check that mutations at position -18 had not generated a third promoter, of which I was unaware, $KMnO_4$ footprinting was used to monitor the position of RNA polymerase. Figure 3.5 shows the footprinting results. As expected, the -18T substitution stimulated open complex formation. Importantly, the open complex is not repositioned or altered in any way by the -18T substitution (compare Figure 3.1B and 3.5).

3.5 Spacer region mutations alter conformation of the *cbpA* regulatory DNA

Changes in base sequence can cause DNA fragments of equivalent length to migrate differently in native PAGE analysis. This is because DNA conformation is dependent on the base sequence of DNA. Hence, native PAGE analysis was performed with DNA fragments carrying the wild type, -18A, -18C and -18T derivatives of the *cbpA* regulatory region. Figure 3.6A shows that these fragments have different mobility. Thus, changes at position -18 must alter bending of the *cbpA* regulatory DNA. Out of the four derivatives, the wild type DNA fragment was least mobile and the -18T fragment was the most mobile. The -18A and -18C fragments had intermediate mobility. Based on this result, computational modelling was used to predict topology of the different DNA sequences. The predicted changes in conformation of the double helix, based on the mutations at the -18 position, are shown in Figure 3.6B (The Bend.It server was used to generate results) (Vlahovicek *et. al*., 2003). The computational modelling is in agreement with the experimental data. Thus, the -18T substitution causes the biggest change in conformation of the *cbpA* regulatory DNA.

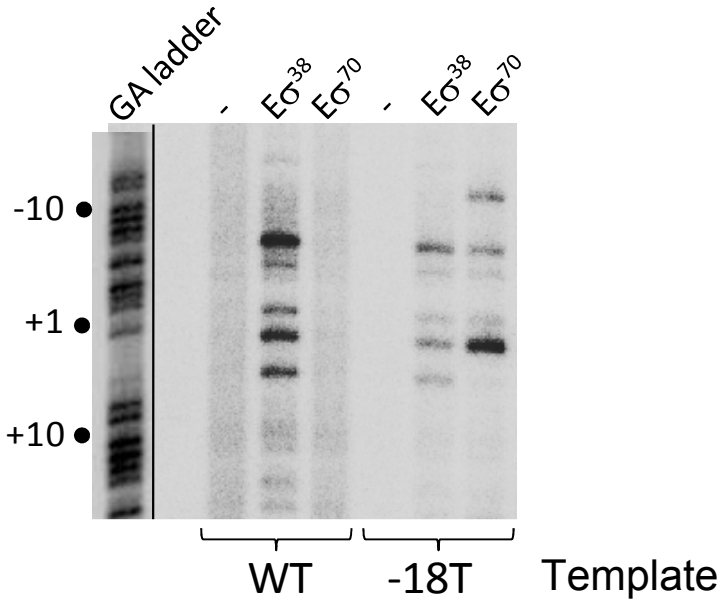


Figure 3.5 Effect of -18T substitution on open complex formation by RNA polymerase

The figure shows the results of $KMnO_4$ footprinting experiments with either the wild type *cbpA* regulatory region or the -18T derivative. $E\sigma^{38}$ and $E\sigma^{70}$ were added in equal amounts (480 nM).

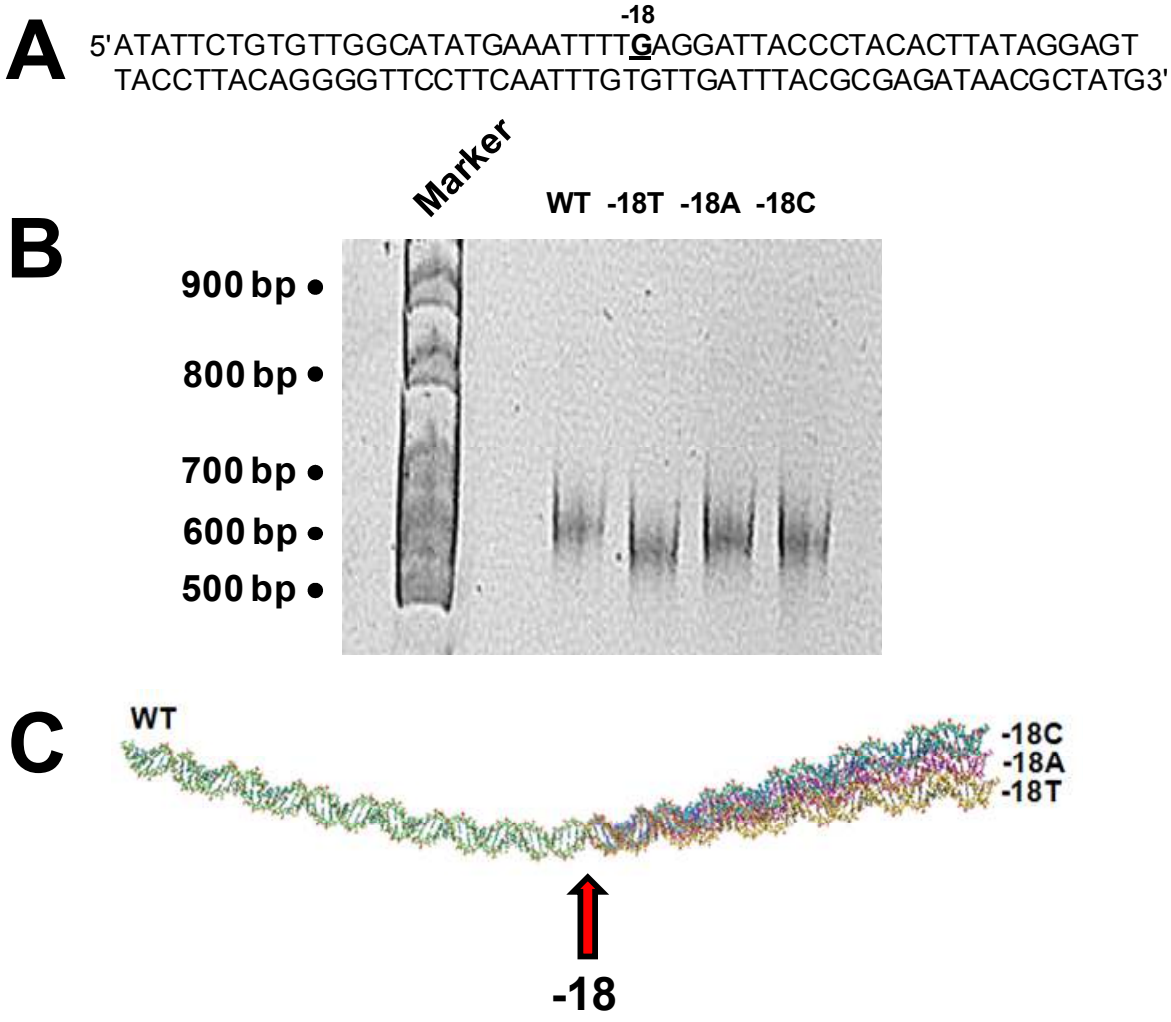


Figure 3.6 DNA conformation changes induced by mutations in the *cbpA* spacer region.

A. 106 bp *cbpA* fragment was used for this analysis. Position -18 is highlighted

B. DNA fragments carrying different mutations in the *cbpA* regulatory region have different mobility on 7.5 % native acrylamide gel.

C. Predicted topology of the wild type, -18C, -18A and -18T derivatives of the *cbpA* regulatory region (Coloured in green, blue, pink and yellow respectively). The arrow indicates the position -18.

3.6 The -18T substitution affects growth phase dependent control of *cbpA* expression

My data suggest that the -18T *cbpA* regulatory region mutation specifically stimulates σ^{70} dependent transcription. Hence, I reasoned that the -18T regulatory DNA derivative should trigger transcription earlier than the wild type DNA fragment during the transition to stationary phase. Figure 3.7 shows β-galactosidase activity stimulated by the WT *cbpA* promoter, and promoter with the -18T substitution, throughout growth. The WT *cbpA* promoter is inactive when the cells are growing rapidly and activity increases once the cells reach stationary phase. However, with the -18T substitution, the promoter is active even during rapid growth. This indicates that substitution at position -18 is capable of affecting the stationary phase dependent expression profile of CbpA.

3.7 Conservation of sequences in the promoter spacer region

Figure 3.8 shows a DNA logo created by aligning the 554 *E. coli* promoters catalogued by Mitchell *et. al.,* 2003. The best conserved promoter element is the -10 sequence. In comparison, the -35 element is poorly conserved. The DNA logo also suggests that some sequences in the 17 bp spacer region are also conserved. Strikingly, some sections of the spacer region are better conserved than parts of the -35 and the extended -10 elements. Moreover, T is the preferred base at both positions -17 and -18 (Figure 3.8). Thus, the alignment in Figure 3.8 is consistent with our genetic screen for mutations that increase activity of the $E\sigma^{70}$ dependent *cbpA* promoter (Table 3.1). The structural model of RNA polymerase holoenzyme bound to DNA (Murakami *et al*., 2002) predicts that the linker between σ^{70} domains 2 and 3 is very close to position of -18 of the promoter spacer region

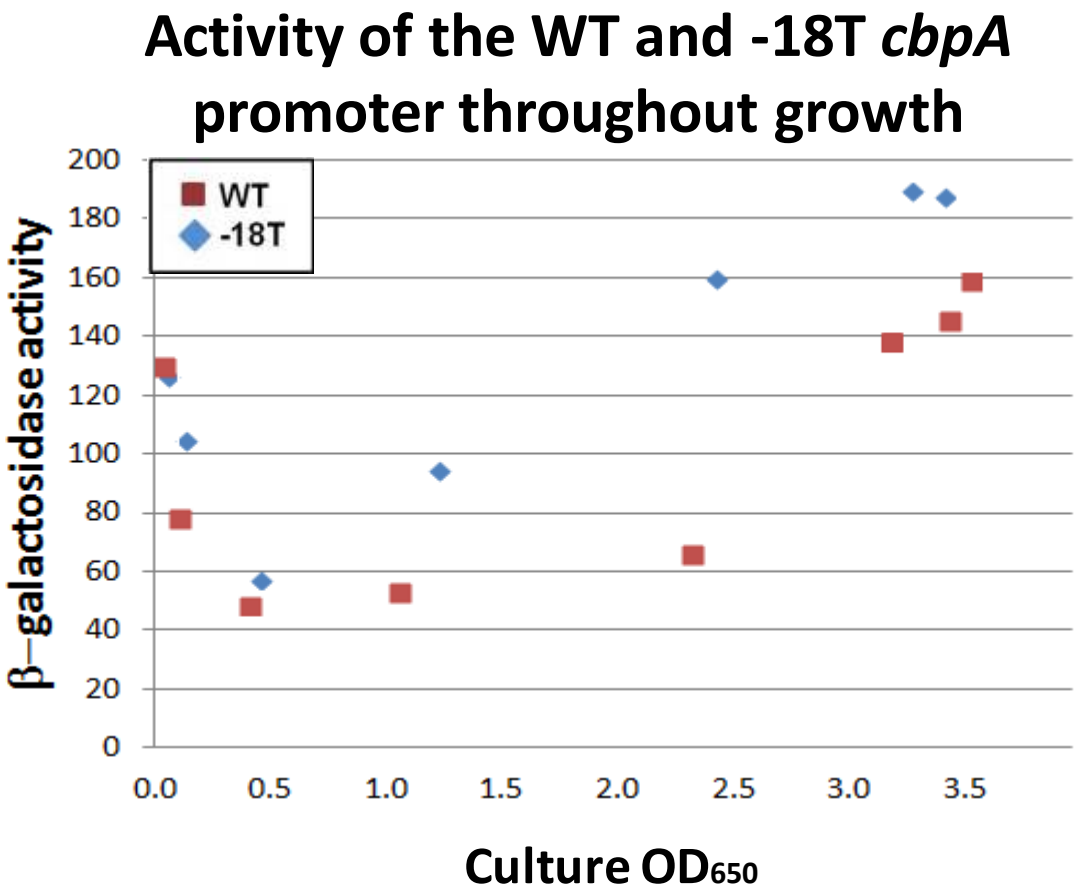


Figure 3.7 Activity of the wild type (WT) and -18T derivative of the *cbpA* promoter

The graph shows β–galactosidase activities, obtained from cultures of JCB387 cells, carrying different *cbpA::lacZ* fusions in plasmid pRW50, throughout growth.

Alignment of 554 *E. coli* promoters

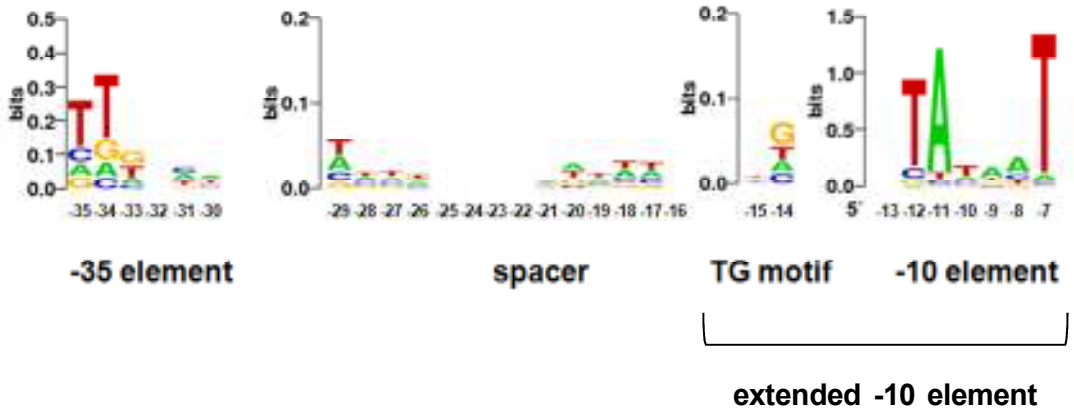


Figure 3.8 DNA sequence logo

DNA sequence logo generated by aligning the 554 *E. coli* promoters documented by Mitchell *et. al*., 2003. All the promoter elements are labelled.

(Figure 3.9). Thus, only 2 Å separate the linker and the non-template DNA strand at this position. This suggests that the promoter spacer region might interact with the linker between σ^{70} domains 2 and 3. Consistent with this, it was previously suggested by Fenton *et al*. (2000) that side chain R451 (located in the linker) contacts the DNA upstream of the -10 element in the spacer region. Thus, we reasoned that changes in promoter conformation, caused by mutations at position -18, might be detected by the σ^{70} domain 2-3 linker. To test this, alanine substitutions were introduced into the *rpoD* gene, encoded by plasmid pVRσ, at the positions highlighted in Figure 3.8. MC4100*rpoS::*kan cells, carrying the wild type and -18T version of the *cbpA* regulatory region in plasmid pRW50, were transformed with the starting pVRσ plasmid and derivatives. Recall that, in this genetic background, the WT *cbpA* regulatory DNA fragment cannot drive LacZ expression. Conversely, because of the increase in dependence on $E\sigma^{70}$, the -18T derivative can drive LacZ expression. Figure 3.10 shows the results obtained from the β-galactosidase assays with these transformants. None of the substitutions in *rpoD* had an effect at the wild type *cbpA* regulatory region (Figure 3.10, A). This is not surprising since there is hardly any LacZ expression in this experiment (compare 'no promoter' and wild type promoter, Figure 3.4). However, activity of the -18T promoter was drastically reduced by the RA451 substitution (Figure 3.10, B). A derivative of the *cbpA* regulatory region, with an improved -10 element for $E\sigma^{70}$, was used as a control. The improved -10 element was generated by substituting the 'C' at positions -9 and -10 with an 'A' and 'T' respectively (mutant named -9A-10T). This creates an $E\sigma^{70}$ dependent *cbpA* promoter with a consensus -10 hexamer. The data obtained show that the RA451 $E\sigma^{70}$ derivative is functional at the -9A-10T version of the *cbpA* regulatory region (Figure 3.10, C). Transcription was also measured *in vitro* with purified $E\sigma^{70}$, and the RA451 derivative, to

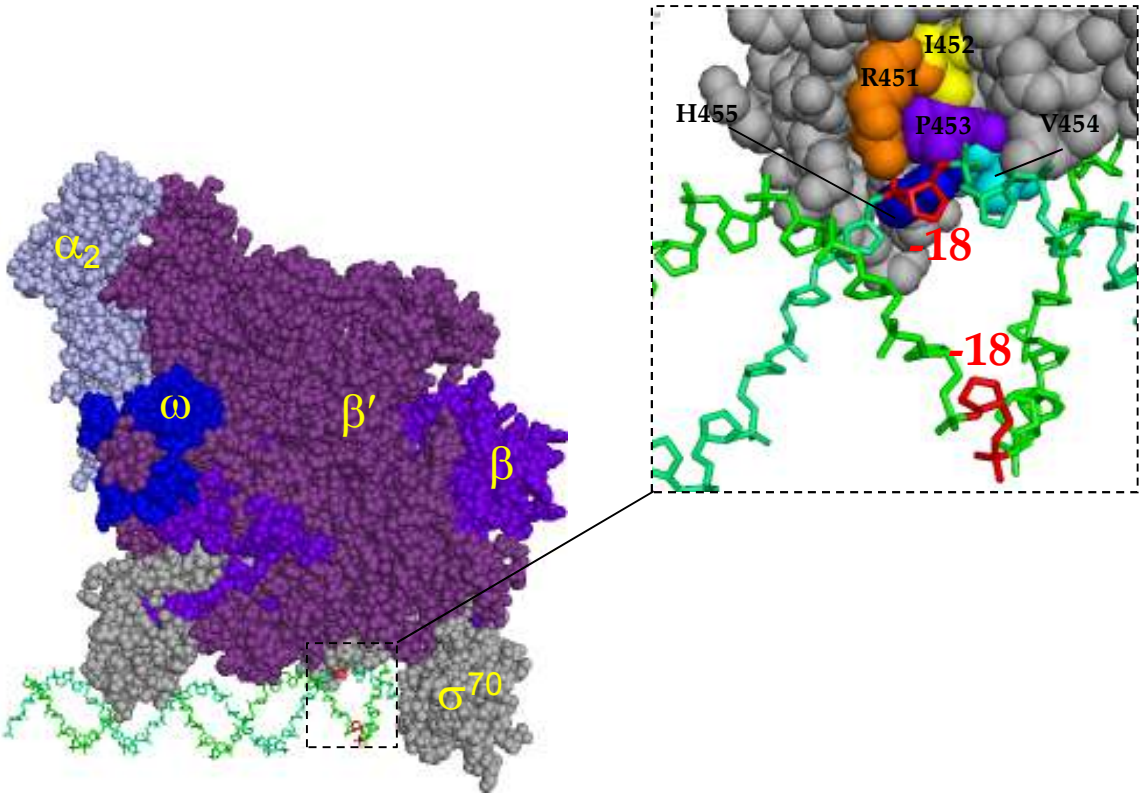


Figure 3.9 Structural model of RNA holoenzyme-DNA complex

Each RNA polymerase component protein is shown in a different colour (except the two α subunits) and labelled. The sugar phosphate backbone is shown in green and the position -18 is highlighted in red. The bases are not shown. The expansion shows the close proximity of the loop between σ^{70} domains 2 and 3 and the promoter non-template strand. Residues mutated for alanine scan analysis are highlighted and labelled.

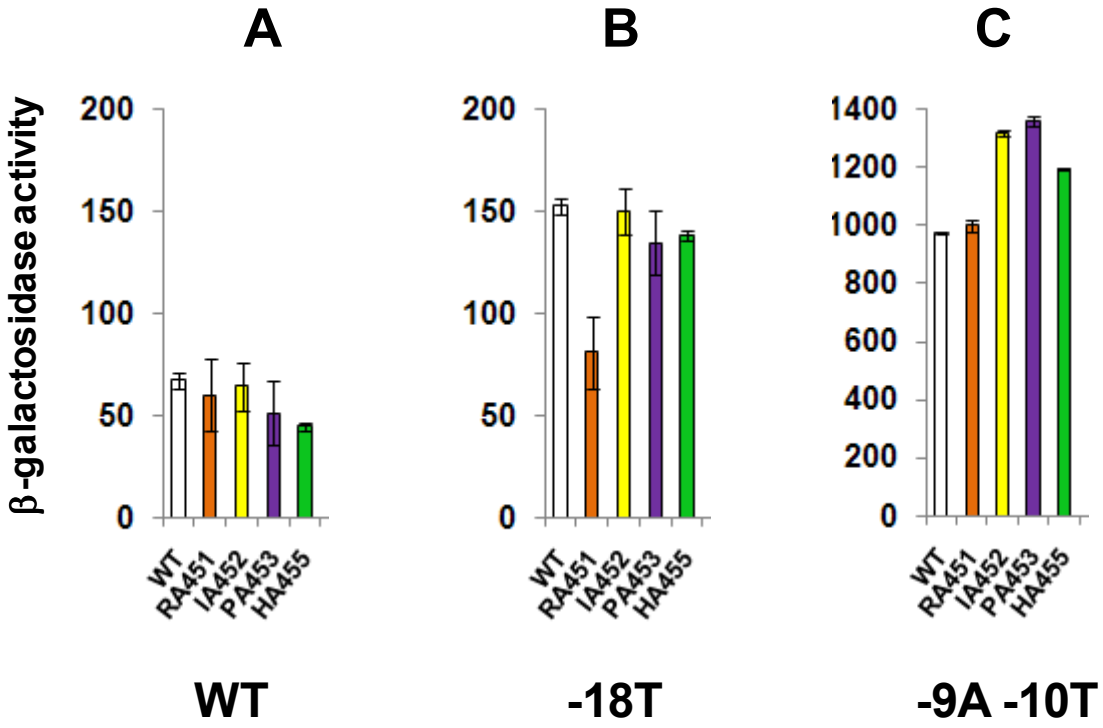


Figure 3.10 Stimulatory effects of the -18T substitution requires σ^{70} side chain R451

The bar chart shows β-galactosidase expression driven by (A) the wild type *cbpA* regulatory region (B) the -18T regulatory region and (C) -9A-10T derivative of the regulatory region. In each case the regulatory DNA is cloned in plasmid pRW50. Assays were done in MC4100*rpos*::kan cells also carrying pVRσ or pVRσ derivatives. The bars are coloured with respect to the residues highlighted in Figure 3.9. Assays were done using overnight cultures.

confirm the results obtained *in vivo*. The data confirms that RA451 $E\sigma^{70}$ is functional at -9A-10T promoter but not at -18T promoter (Figure 3.11).

3.7 Promoter position -18, and σ^{70} side chain R451, are required at other regulatory regions

3.7.1 The *dps* promoter

To test whether effects of base substitutions at position -18 are observed at other regulatory regions, the *dps* promoter was examined. The *dps* promoter, which also has a G at position -18, can be shared by $E\sigma^{70}$ and $E\sigma^{38}$. However, $E\sigma^{38}$ is more efficient at initiating transcription. Thus, the 'G' at position -18 was substituted with 'C', 'T' and 'A' (Figure 3.12A). *In vitro* transcription analysis was then used to examine the effects on $E\sigma^{38}$ and $E\sigma^{70}$ dependent transcription. The base substitutions at position -18 did not affect transcription by $E\sigma^{38}$ (Figure 3.12B; compare lane 1 with lanes 3, 5 and 7). Strikingly, $E\sigma^{70}$ dependent transcription was stimulated, particularly by the -18T substitution (Figure 3.12B; compare lane 2 with lanes 4, 6 and 8).

3.7.2 The LEE1 promoter

E. coli O157 LEE1 promoter drives expression of genes in the locus for enterocyte effacement (LEE). The GrlA transcriptional activator is required to achieve maximal levels of transcription. However, substantial basal levels of transcription are observed in *E. coli* K-12 cells that do not encode *grlA*. A LEE1 promoter, with increased GrlA-independent activity, was isolated recently by Islam *et. al.*, 2011. This increase in activity was due to a G to A substitution at promoter position -18 (Figure 3.13A). To investigate whether the σ^{70} side chain

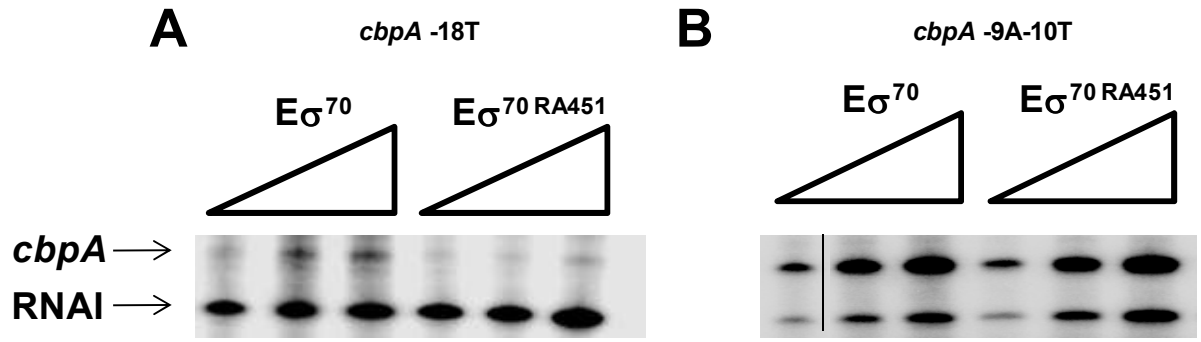


Figure 3.11 *In vitro* transcription assays showing the effect of -18T substitution requires σ^{70} side chain R451

The gels show transcripts produced *in vitro* from the -18T (A) and -9A-10T (B) derivatives of the *cbpA* regulatory region. RNA polymerase was added at a concentration of 80, 160 or 240 nM.

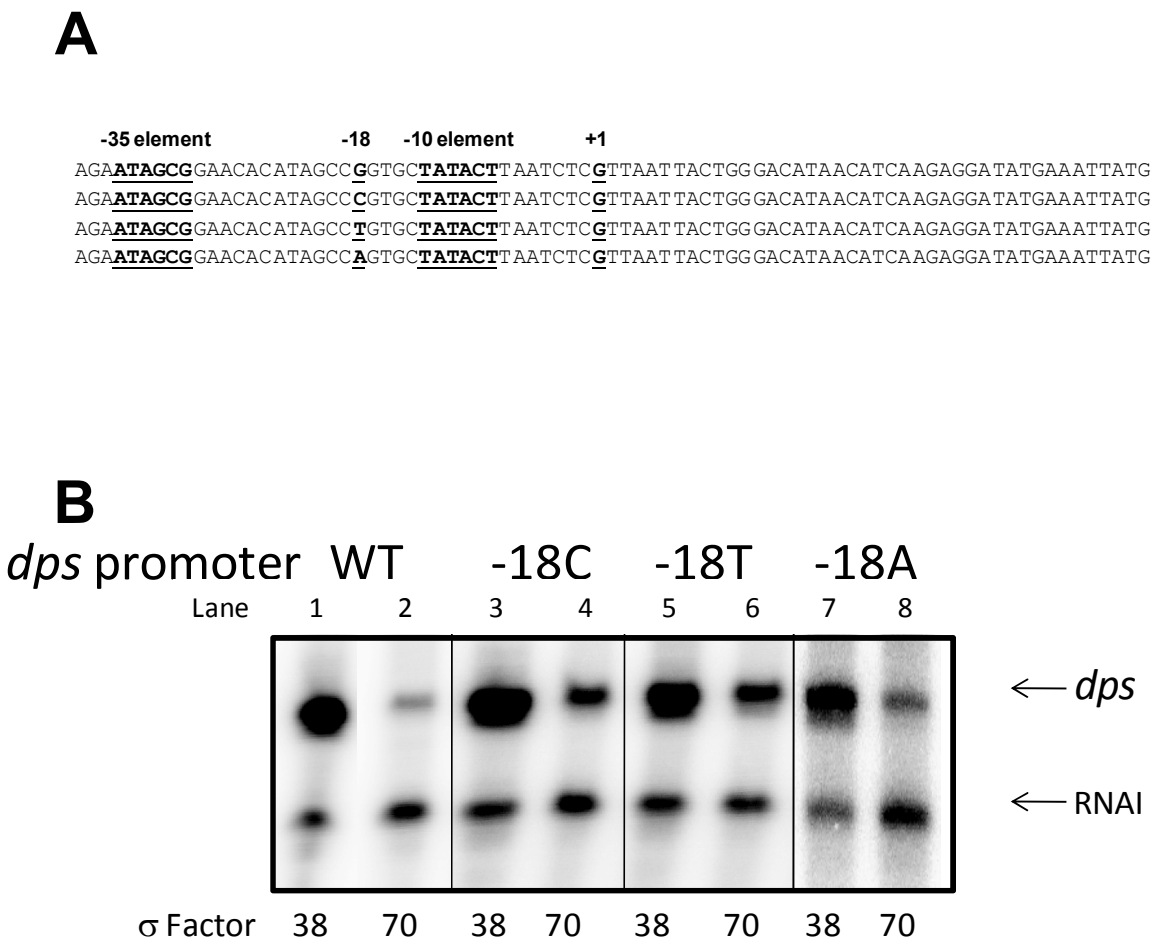


Figure 3.12 Eσ^{70} dependent transcription at the *dps* promoter is stimulated by substitutions at position -18

A. Sequence of *dps* promoter and derivatives used in this work (all are derivatives of the *dps*100 fragment described by Grainger *et al*., 2008).

B. *In vitro* transcription assay results using a *dps* promoter fragment and derivatives cloned in pSR plasmid as a template.

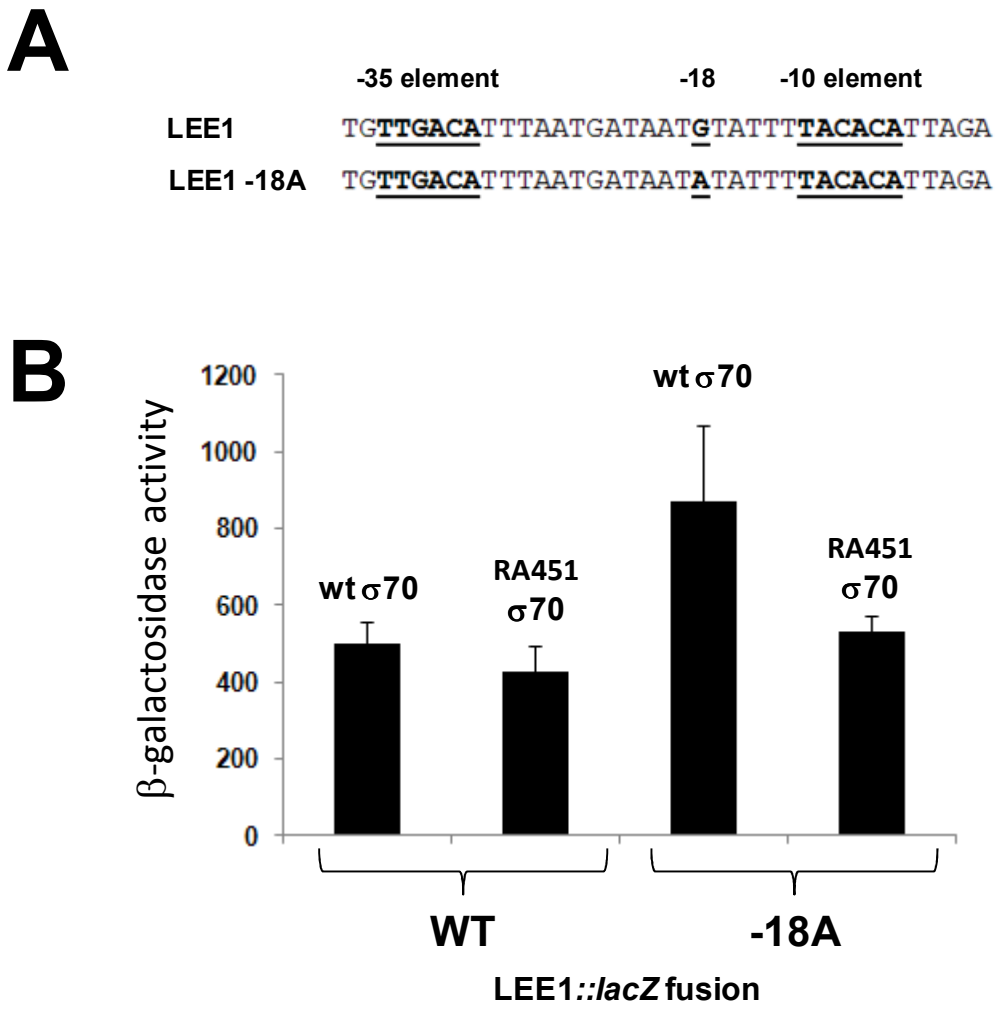


Figure 3.13 The LEE1 promoter responds to changes in the spacer region at position -18

A. Sequence of the LEE1 promoter and -18A derivative. The -10 and -13 promoter elements and position -18 are highlighted

B. Bar chart displaying β-galactosidase expression driven by the wild type and the -18A derivatives of the LEE1 regulatory region in JCB387 cells transformed with either pVRσ or pVRσ RA451. Measurements were taken in mid-log phase using the LEE promoter derivatives described in panel A.

R451 is important for mediating this effect, β-galactosidase assays were done. The results showed that RA451 $E\sigma^{70}$ is not defective with the wild type LEE1 promoter. However, the -18A LEE1 promoter had reduced activity with RA451 $E\sigma^{70}$ (Figure 3.13B). Computational and native PAGE analysis was done to confirm alterations in DNA bending due to base substitutions at position -18 of the LEE1 promoter (Figure 3.14).

3.9 The σ^{70} side chain R451 is conserved and essential for optimal growth

It has been shown that R451 is highly conserved in RNA polymerase σ factors from different bacteria (Campbell *et al*., 2002). Thus, growth rate analysis was done, over a 5 hour period, to compare the fitness of *E. coli* cells transformed with pVRσ carrying either the wild type *rpoD* allele or the RA451 *rpoD* allele. A growth defect was detected in *E. coli* cells transformed with pVRσ carrying the RA451 *rpoD* allele (Figure 3.15).

3.10 Discussion

In this chapter, my aim was to understand how specificity is achieved at the *cbpA* regulatory region that has two overlapping promoters. The results obtained show that alterations in the shared spacer region of the two promoters can affect the two promoters differently. Thus, a mutation (-18T) was isolated that specifically increases promoter activity due to σ^{70}. The σ^{70} side chain R451 is important for this effect. Interestingly, several recent studies have focussed on promoter spacer sequence and the role it plays in controlling transcription (Hook-Barnard and Hinton, 2009; Liu *et al*., 2004). For example, Liu *et. al*., showed that an 8 bp sequence upstream of the -10 element can stimulate transcription from the *lac* promoter in the absence

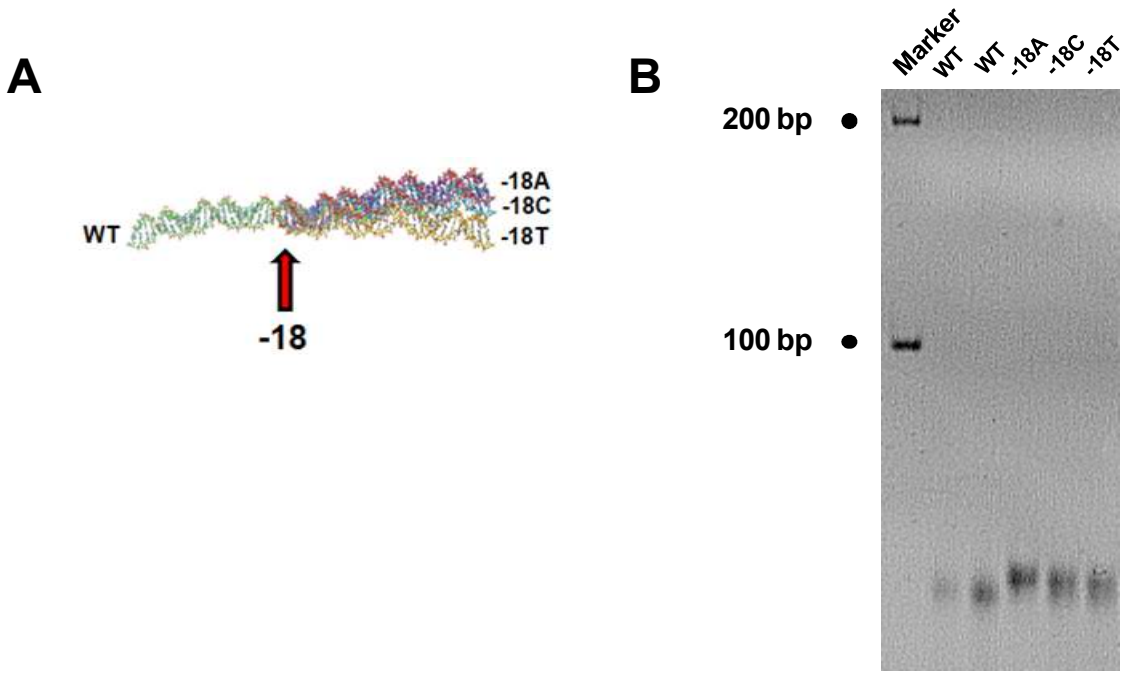


Figure 3.14 Changes in DNA conformation induced by mutations at position -18 of the LEE1 regulatory region

A. Predicted topology of the wild type (green), -18C (blue), -18A (pink) and -18T (yellow) derivatives of the LEE1 regulatory region. Arrow indicates the position -18.

B. LEE1 regulatory region derivatives, with different base substitution at position -18, have different mobility on a 7.5% native acrylamide gel.

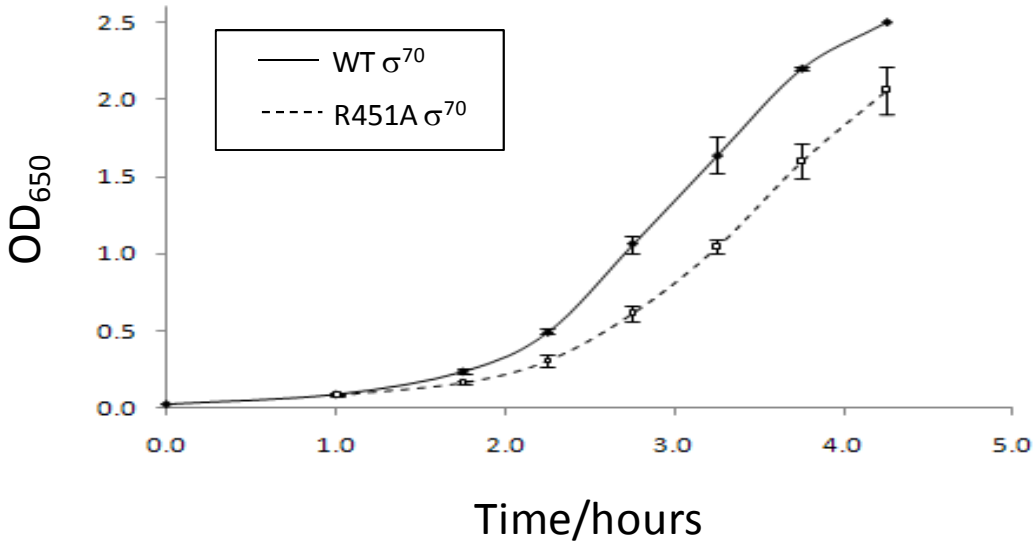


Figure 3.15 σ^{70} side chain RA451 mutation induces growth defects

The graph shows growth of JCB387 cells transformed with either pVRσ (solid line) or pVRσ RA451 (dashed line). Cells were grown in LB medium with vigorous aeration at 37 °C. The experiment was repeated three times and error bars show the standard deviation of the recorded OD_{650} values.

of CRP (Liu *et al*., 2004). Similarly, Yuzenkova *et*. *al*., also showed that the sequence upstream of the -10 element is crucial (Yuzenkova *et*. *al*., 2011). The results in this chapter confirm that there is a clear T>A>C>G preference at promoter position -18, consistent with promoter sequence alignments (Figure 3.8). This conclusion is supported by biochemical and genetic experiments done with three different *E. coli* promoters (Table 3.1, Figure 3.12 and Figure 3.13). Moreover, similar observations exist in the literature. Thus, Busby *et*. *al*., (1984) showed that transcription from a modified *gal* P2 promoter is stimulated by a G>A substitution at position -18. Similarly, *gap*A1 promoter activity is decreased almost 3-fold, both *in vivo* and *in vitro*, by a T>G substitution at position -18 (Thouvenot *et al*., 2004). The 8 bp motif identified by Liu *et al*. (2004) has a T at position -18 that is essential for activity. In addition, a preference for 5'-TTT-3' trinucleotide sequences centered at position -18 is shown by Mitchell *et. al.* (2003). Similar observations have been made at the *rrn*B P1 promoter by Gaal *et al*. (1989). It is likely that the effect of base sequence at position -18 is dependent on overall promoter strength. Thus, although base sequence at promoter position -18 plays an important role at some promoters, it is unlikely that it plays a central role at all promoters.

Previous structural and biochemical studies have shown that position -18 of the promoter non-template strand, and the loop between σ^{70} domains 2 and 3, of are in very close proximity (Murakami *et al*., 2002; Colland *et al*., 1999). Hence, I propose that a contact might occur between the two structures. Consistent with this proposition, 'extensive' RNA polymerase contacts with DNA in the spacer region of the phage T7 A3 promoter have been reported (Siebenlist and Gilbert, 1980). Similar observations have been made at the *lac*UV5 and the λc17 promoters (Siebenlist *et al*., 1980; Rosenberg and Court, 1979). It has also been shown that the promoter DNA binding activity of RNA polymerase *in vitro* is disrupted by an RS451

substitution (Fenton *et al*., 2000). The results obtained in this work suggests that the R451-DNA backbone contact is modulated by changes in DNA conformation at promoter position -18 (Figure 3.6 and Figure 3.14) giving rise to artificial sequence specificity. The introduction of a polar serine side chain adjacent to the DNA backbone could be the reason for the extreme deleterious effect of the RS451 substitution seen by Fenton *et. al.* (2000) as σ^{70} subunits with an RA451 substitution retain some activity (Figure 3.11 and Figure 3.13). In summary, a mechanism, via which changes in sequence and conformation of the promoter sequence region are sensed, is described in the chapter. My results show that, at overlapping promoters, the sequence of the spacer region can play a key role. Hence, at the *cbpA* regulatory region, the spacer region is important to prevent high levels of $E\sigma^{70}$ dependent transcription (Figure 3.3 and Figure 3.4).

Single molecule analysis of CbpA-DNA complexes

Chapter 4

4.1 Introduction

In the previous chapter I examined the *cbpA* regulatory region, which drives expression of CbpA in starved *E. coli*. I next turned my attention to the CbpA protein and its DNA binding properties. CbpA is a stationary phase nucleoid associated protein and is believed to aggregate with DNA (Cosgriff *et. al*., 2010). Hence, in late stationary phase, CbpA represents almost 3 % of the total cellular protein (15,000 molecules per cell) (Ali Azam *et. al*., 1999). Figure 4.1 shows the domain organisation of CbpA defined by Bird *et. al*., 2006. The J-domain possesses co-chaperone activity and domain CTD II is required for dimerisation. DNA binding activity locates to the linker region and CTD I. However, the precise region of CbpA required for DNA binding is unknown. Proteins often contact DNA via ionic interactions. Hence, positively charged amino acids frequently interact with the negatively charged DNA backbone. Presuming this, all arginine (R), lysine (K) and histidine (H) residues in the linker and CTD I were substituted with alanine (Chintakayala and Grainger, *in preparation*). Out of the twenty alanine substitutions made, one (R116A) resulted in defective DNA binding by CbpA. The side chain R116 is situated at the end of the linker region, close to CTD I, and is highly conserved in CbpA proteins from other enterobacteria. Thus, the corresponding mutation in *Y. enterocolitica* CbpA also renders CbpA unable to bind DNA (Chintakayala and Grainger, *in preparation*). Hence, CbpA residue R116 is an important DNA binding determinant. In this chapter I have used single molecule tethered particle motion (TPM) analysis to test the ability of CbpA, and the R116A derivative, to bind and compact DNA.

4.2 Tethered Particle Motion (TPM)

Single molecule approaches are proving to be essential for understanding biological processes. They allow researchers to examine individual molecules, as opposed to their

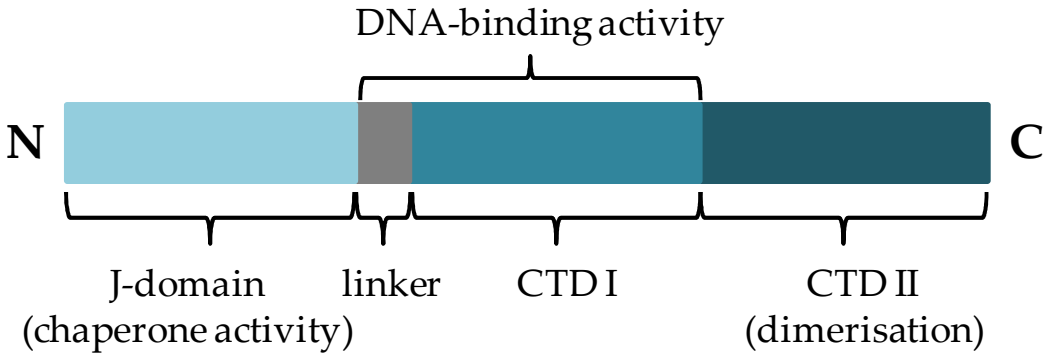


Figure 4.1 Domain organisation of CbpA

Individual domains of CbpA are highlighted using different coloured boxes and the function of each domain is described.

ensemble-averaged behaviour, in real time. Protein-DNA interactions are central to gene regulation, DNA replication and DNA repair. However, to investigate protein-DNA conformation, the resolution required is at the nanometre-scale. This poses an impossible challenge for conventional techniques. Thus, single molecule techniques are being increasingly applied to these questions. In TPM experiments, a biopolymer (i.e. DNA) is tethered between a stationary substrate (glass slide) and a micrometer scale sphere (bead) which is large enough to be imaged with conventional optical microscopy. The underlying macromolecular dynamics are reported by the constrained Brownian motion of the bead in solution. Hence, any changes in the conformation of the tethered DNA molecule (looping, bending etc.) prompt changes Brownian motion (Towles *et. al*., 2009; Han *et. al*., 2008). Figure 4.2 shows a schematic of the procedure. The sequence of the AT-rich DNA tether used as a template for CbpA binding is shown in Figure 4.3. The beads used in this work were Streptavidin coated polystyrene particles with a diameter 0.46 mm (Kisker Biotech GmbH & Co. KG). The data obtained by tracking the Brownian motion of the DNA tether is plotted on a graph where the X axis shows the Root Mean Square (RMS) motion of the particles. Briefly, this is the square root of the sum of the variances of particle position (x, y) along two orthogonal image-plane axes. The Y axis shows the number of beads counted (the values are written as n/100). If CbpA binding to the DNA tether compacts the DNA tether, the RMS value would decrease (Figure 4.4).

4.3 CbpA binds and compacts DNA

Figure 4.5 shows the data obtained by tracking the Brownian motion of the DNA tethered beads in the absence of CbpA. The results show that, in the absence of any protein, the beads have a RMS value of about 160. Next, we examined the effect of CbpA on particle motion. A

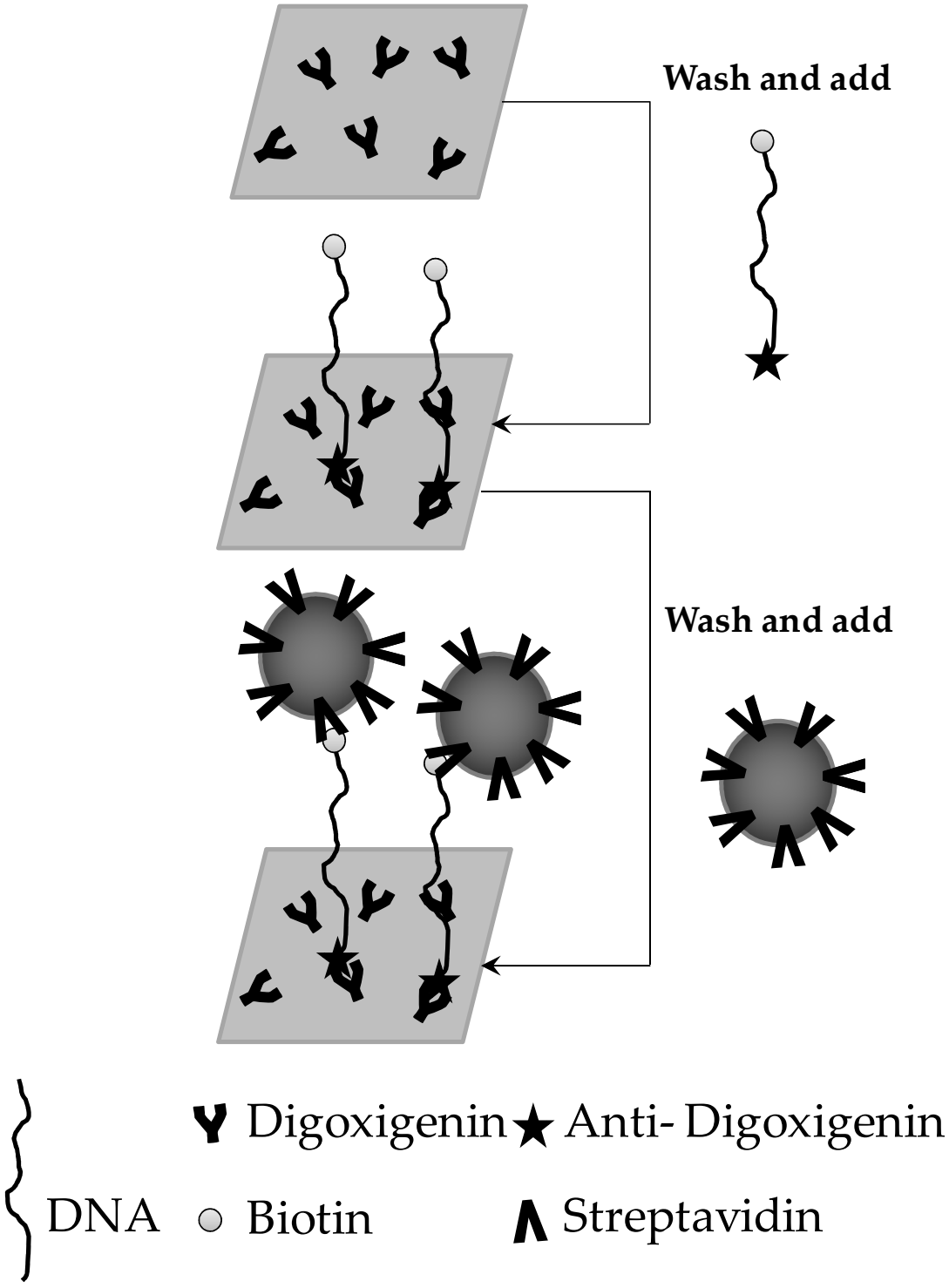


Figure 4.2 Steps in TPM

Simplified cartoon to explain the steps involved in preparation of a slide for TPM

URA3 685 bp

U4-DIG
CGGTAATCTCCGAGCAGAAGGAAGAACGAAGGAAGGAGCACAGACTTAGATTGGTATATATACGCATATGTGGTG
TTGAAGAAACATGAAATTGCCCAGTATTCTTAACCCAACTGCACAGAACAAAAACCTGCAGGAAACGAAGATAAA
TCATGTCGAAAGCTACATATAAGGAACGTGCTGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCA
TGCACGAAAAGCAAACAAACTTGTGTGCTTCATTGGATGTTCGTACCACCAAGGAATTACTGGAGTTAGTTGAAG
CATTAGGTCCCAAAATTTGTTTACTAAAAACACATGTGGATATCTTGACTGATTTTTCCATGGAGGGCACAGTTA
AGCCGCTAAAGGCATTATCCGCCAAGTACAATTTTTTACTCTTCGAAGACAGAAAATTTGCTGACATTGGTAATA
CAGTCAAATTGCAGTACTCTGCGGGTGTATACAGAATAGCAGAATGGGCAGACATTACGAATGCACACGGTGTGG
TGGGCCCAGGTATTGTTAGCGGTTTGAAGCAGGCGGCGGAAGAAGTAACAAAGGAACCTAGAGGCCTTTTGATGT
TAGCAGAATTGTCATGCAAGGGCTCCCTAGCTACTGGAGAATATACTAAGGGTACTGTTGACATTGCGAAGAGCG
ACCCGGGAAA U5-bio
SmaI

Figure 4.3 DNA tether sequence

The sequence of the AT-rich DNA fragment used for TPM experiments with CbpA. The DNA was labelled with digoxigenin (DIG) at one end and biotin at the other.

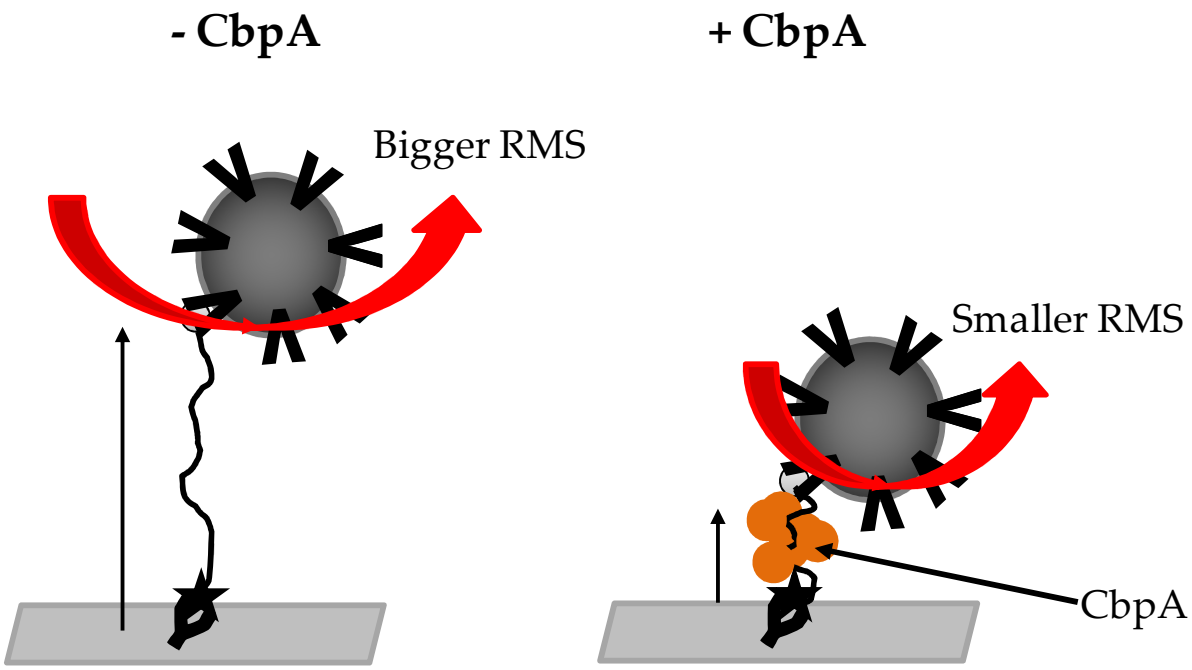


Figure 4.4 DNA compaction by CbpA

The figure gives a schematic representation of the result expected. When CbpA (orange circles) is absent, the DNA tether is not compacted and hence gives a higher RMS value. In the presence of CbpA, the DNA is compacted. This results in a shortened tether length and, hence, a lower RMS value.

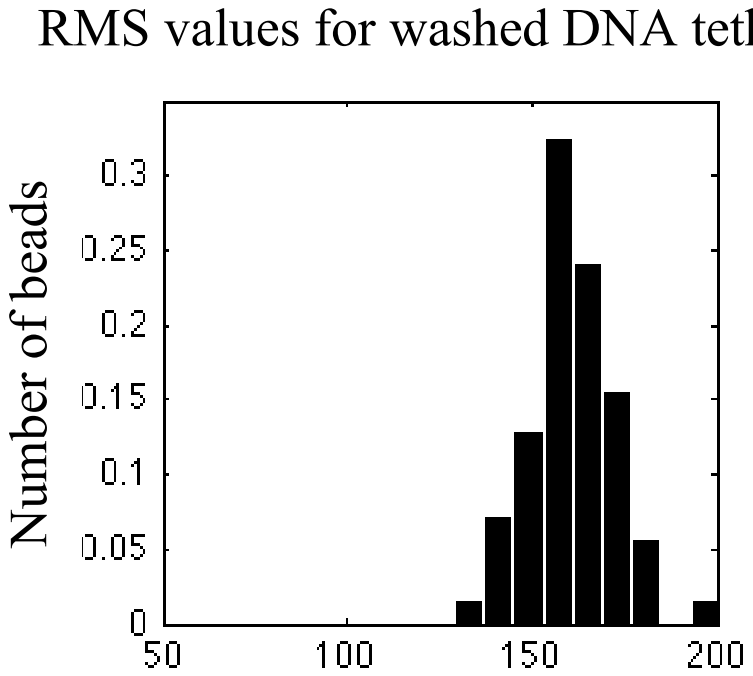


Figure 4.5 Control data obtained for TPM experiments

The flow cell was prepared as mentioned in Section 2.19 (Chapter 2). The X-axis shows RMS and the Y-axis shows number of beads written as n/100. The readings were taken in CbpA buffer Section 2.19 (Chapter 2) in the absence of protein.

range of CbpA concentrations were used. Figure 4.6 shows the data obtained using 100 nM and 750 nM CbpA. The figure shows a clear decrease in RMS values with increase in CbpA concentrations. This indicates that the persistence length of the DNA tether is reduced due to CbpA binding. Hence, CbpA must compact the DNA. Note that beads with RMS values comparable to the "no CbpA" control are still present. This is expected because the experiments could not be done at saturating concentrations of CbpA because the solution becomes too viscous.

4.4 R116A disrupts DNA compaction by CbpA

We next examined CbpA carrying the R116A substitution. The data show that R116A CbpA does not have any effect on the length of the tether at 100 nM or 750 nM concentrations (Figure 4.7). This confirms that i) the effects seen with wild type CbpA are not indirect due to effects on viscosity of solution and ii) that CbpA R116A cannot bind DNA. Figure 4.8 shows a direct comparison of results obtained with wild type CbpA (750 nM) and R116A CbpA (750 nM) alongside a control (no CbpA). These data confirm that, at 750 nM CbpA, RMS values indicate two populations (i.e. bound and unbound DNA) (Figure 4.8, blue bars). Conversely, 750 nM R116A CbpA had no effect, displaying a trend similar to control DNA (Figure 4.8, compare red and black bars).

4.5 Discussion

For performing single-molecule analysis of DNA-protein complexes, the TPM assay is the simple and highly accurate. TPM allows direct observations of molecular dynamics and not averaged behaviour. CbpA is known to bind to AT-rich DNA and form aggregates (Cosgriff *et. al*., 2010). Results in this chapter show that, when CbpA forms aggregates with DNA, it

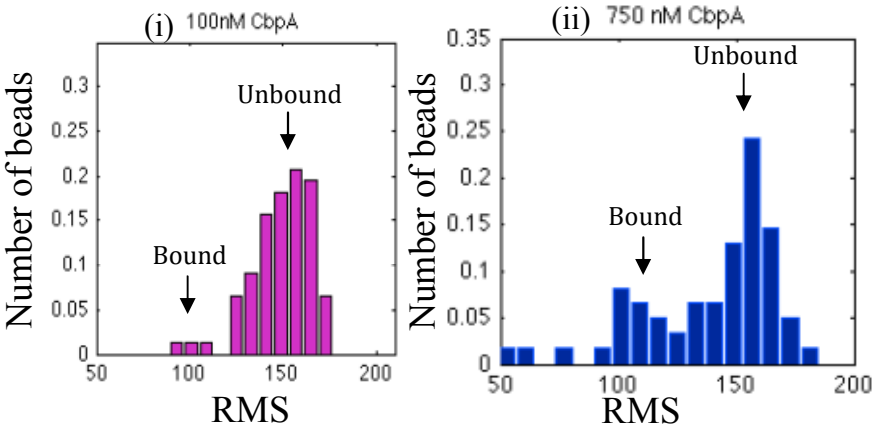


Figure 4.6 TPM experiments at different CbpA concentrations

The flow cell was prepared as mentioned in Section 2.19 (Chapter 2). A new flow cell was used for CbpA concentrations (i) 100 nM and (ii) 750 nM. The X-axis shows RMS and the Y-axis shows number of beads written as n/100

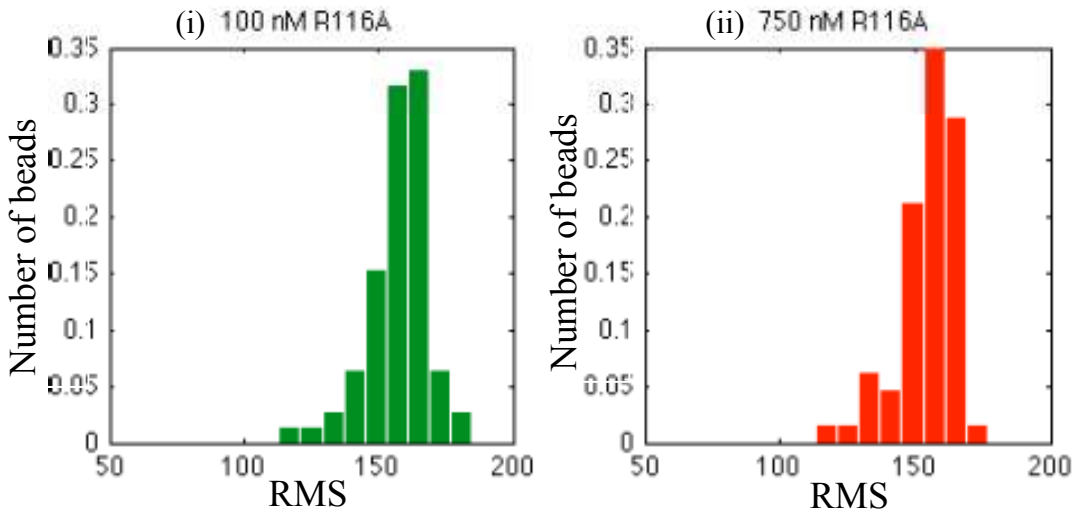


Figure 4.7 TPM experiments at different R116A CbpA concentrations

The flow cell was prepared as mentioned in Section 2.19 (Chapter 2). A new flow cell was used for R116A CbpA concentrations (i) 100 nM (ii) 750 nM. The X-axis shows RMS and the Y-axis shows number of beads written as n/100

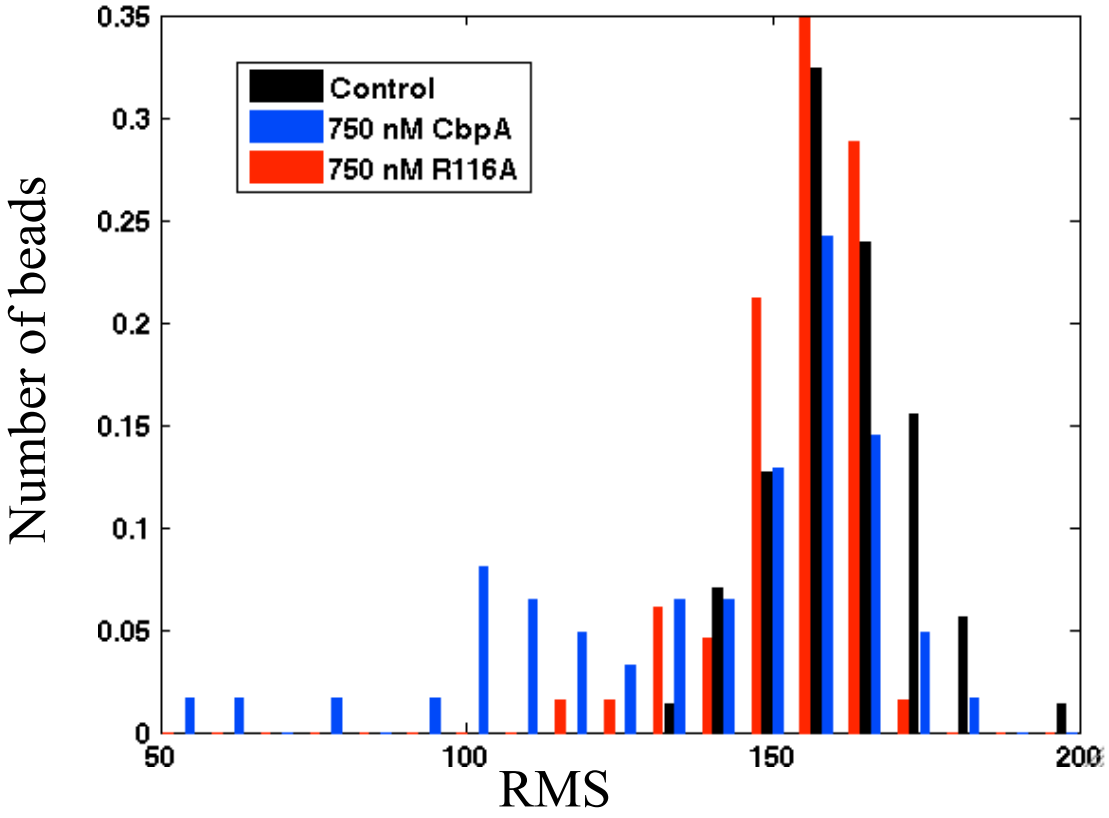


Figure 4.8 Combined data

Graph showing the data obtained from Control (black), CbpA (750 nM) (blue) and R116A (750 nM) (red) experiments. The X-axis shows RMS and the Y-axis shows number of beads written as n/100

also drives DNA compaction (Figure 4.4). As expected, the mutant R116A CbpA was unable to bind to DNA (Figure 4.6 and Figure 4.7).

H-NS improves DNA-binding specificity by RNA polymerase at an AT-rich gene regulatory region

Chapter 5

5.1 Introduction

In the previous two Chapters I have discussed: i) complicated gene regulatory regions with multiple promoters; ii) effects of chromosome folding proteins on DNA organisation. In this final chapter, I have studied the effects of chromosome folding on DNA binding specificity by RNA polymerase at an AT-rich regulatory DNA region. Recall that RNA polymerase recognition of promoter DNA elements is the first step in transcription initiation. Promoter recognition is followed by DNA unwinding. This unwinding is facilitated by weak interactions associated with A:T base pairing. Hence, the -10 promoter element (5'-TATAAT-3'), which is AT-rich, is unwound during transcription initiation. Interestingly, because RNA polymerase recognises AT-rich DNA elements, regulatory regions that have a high AT-content are difficult for RNA polymerase to recognise specifically. This can lead to unusual, sometimes uncontrolled patterns of transcription (Section 1.3.2, iii). Here, I have studied a regulatory region that has an AT content of 71%. In particular, I have investigated how DNA folding factors aid RNA polymerase binding at this complex regulatory region.

5.1.1 The *ehxCABD* operon

The *ehxCABD* operon, consisting of four genes, is located on the pO157 plasmid of *E. coli* O157:H7. The *ehxCABD* operon is horizontally acquired and is therefore very AT-rich. However, despite being unusually AT-rich, this foreign operon is recognised by RNA polymerase and expressed. The organisation of the *ehxCABD* operon is shown in Figure 5.1. The genes encode enterohaemolysin toxin and proteins used for its post-translational modification and transport (Schmidt *et. al.*, 1996). The *ehxA* gene product, EhxA, is a non toxic prohaemolysin. Thus, EhxA is converted into a mature toxin by acylation. Acylation is mediated by the *ehxC* gene product (Trent *et. al.*, 1999). The regulatory region, of 284 bp,

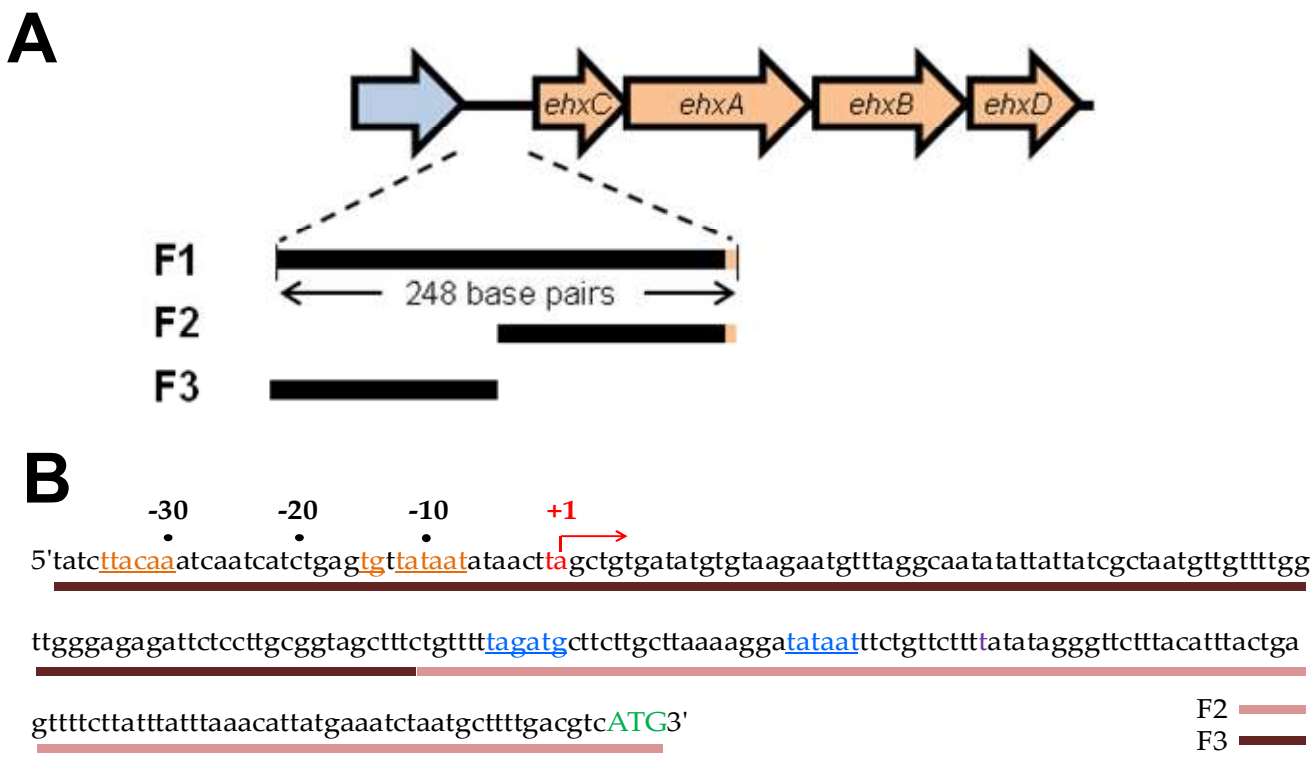


Figure 5.1 The *ehxCABD* operon and corresponding gene regulatory region.

A. Block orange arrows represent the genes within the *ehxCABD* operon. The blue block arrow represents the adjacent open reading frame. Solid black lines labelled F1 through F3 represent the *ehxCABD* regulatory region fragments made. The F1 fragment is 248 bp long and contains the *ehxC* start codon along with the regulatory DNA upstream of it. The F1 fragment was divided into upstream and downstream parts (the F3 and F2 fragments) respectively.

B. The sequence of the F1 fragment. The -10, -35 and the TG motif identified in this work are highlighted in orange, transcription start site in red and start codon in green. The promoter elements proposed in literature are highlighted in blue and purple. The F2 and F3 fragments are highlighted in pink and brown respectively.

immediately upstream of the *ehxCABD* operon, has an AT content of 71%. As a result, the regulatory region contains many sequences that resemble promoter -10 elements. Previously, Rogers *et. al*. (2009), identified a potential promoter in the *ehxCABD* regulatory region. The starting point for this work was to validate this promoter and understand how it is recognised in the context of extremely AT-rich DNA.

5.2 Identification of promoter activity in the *ehxCABD* regulatory region

In initial experiments DNA fragments, carrying discrete sections of the *ehxCABD* regulatory region, were cloned into one of two reporter plasmids; pRW50 and pLux. The different DNA fragments are shown in Figure 5.1. Note that the promoter proposed by Rogers *et. al*. resides in the F2 fragment. The pRW50 reporter encodes LacZ and was used to infer promoter activity in *E. coli* K-12. The pLux reporter encodes luciferase and was used to measure promoter activity in *E. coli* O157. After cloning, the ability of each DNA fragment to drive transcription was measured. The results obtained for each fragment, in both of the plasmids, are shown in Figure 5.2. The data show that gene expression driven by fragments F1 and F3 is similar. Conversely, the fragment F2, proposed to contain a promoter by Rogers *et. al*., was not able to drive any transcription in *E. coli* K-12 (Figure 5.2A). However, in *E.coli* O157 fragment F2 drives some expression which could be due to the presence of *E.coli* O157 specific regulator (Figure 5.2B). Hence, the *ehxCABD* promoter must be located in the upstream fragment, F3, rather than the F2 fragment.

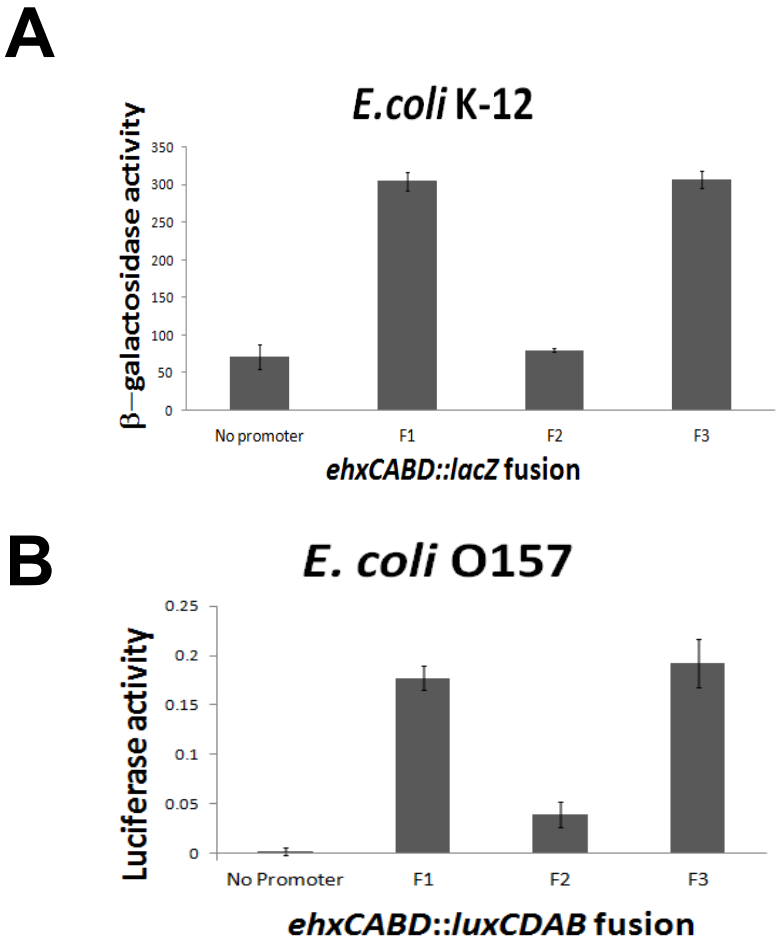


Figure 5.2 Promoter activities of different *ehxCABD* regulatory DNA fragments.

A. Data from β-galactosidase assays using the different constructs shown in Figure 5.1 fused with *lacZ* in pRW50 in *E. coli* strain JCB387

B. Luciferase assays data using the different constructs shown in Figure 5.1 fused with pLux in *E. coli* strain O157:H7

5.2.1 Identification of promoter in *ehxCABD* obtained from different *E. coli* strains

The sequence of the *ehxCABD* F3 fragment from *E. coli* O157 aligns well with *ehxCABD* regulatory regions obtained from *E. coli* TW15838 (Goat isolate) and *E. coli* 1.2264 (fresh water isolate). Thus, equivalent sections of DNA from these organisms were cloned upstream of LacZ in the pRW50 reporter plasmid to see if they too stimulated transcription. Figure 5.3 shows the result of the analysis. As expected, the DNA fragments equivalent to the F3 region from *E. coli* O157 were able to drive gene expression

5.3 Identification of *ehxCABD* promoter and transcription start site

To identify the transcription start site in the F3 fragment, mRNA primer extension assays were used. Hence, RNA was extracted from *E. coli* JCB387 cells, carrying the F3 fragment cloned in pRW50. An oligonucleotide, binding within the *lacZ* gene of plasmid pRW50, was then used to prime reverse transcription. The primer extension products were analysed on a denaturing sequencing gel. Two extension products of 155 and 154 nucleotides (nt) in length, were generated (Figure 5.4A). The transcription start site corresponding to the more abundant 154 nt extension product is labelled +1 (Figure 5.4B). A consensus extended -10 promoter element, and a four out of six match to a -35 promoter element, were found at the expected positions upstream of this transcription start site (Figure 5.4B, highlighted in green). In parallel experiments, with the F1 fragment, we identified the same transcription start site but not the transcription start site proposed by Rogers *et. al.* (2009) (Figure 5.5). Thus, from here

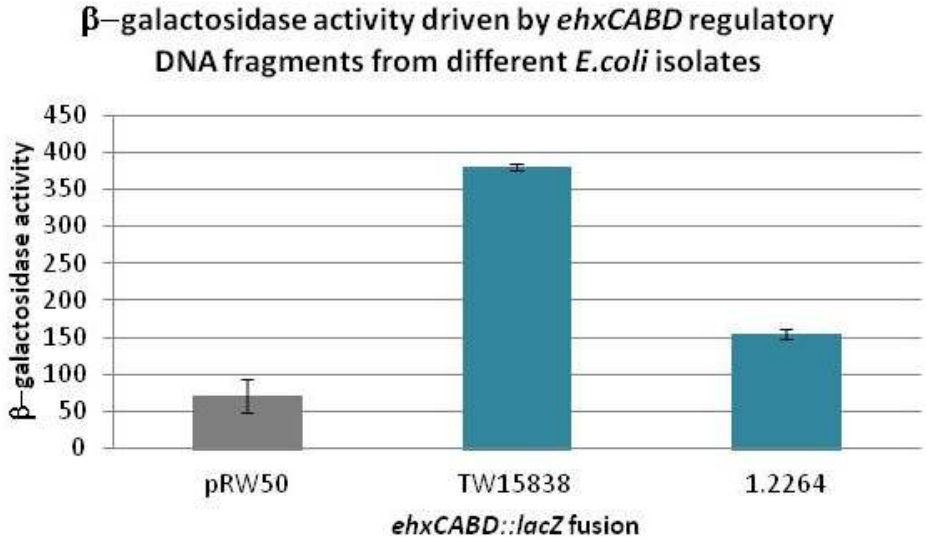


Figure 5.3 Activities of different *ehxCABD* regulatory DNA fragments obtained from different *E. coli* isolates.

Data from β-galactosidase assays, using regulatory DNA fragments obtained from *E. coli* TW15838 (Goat isolate) and *E.coli* 1.2264 (fresh water isolate), fused to *lacZ* in pRW50. Assays were done using *E. coli* strain JCB387.

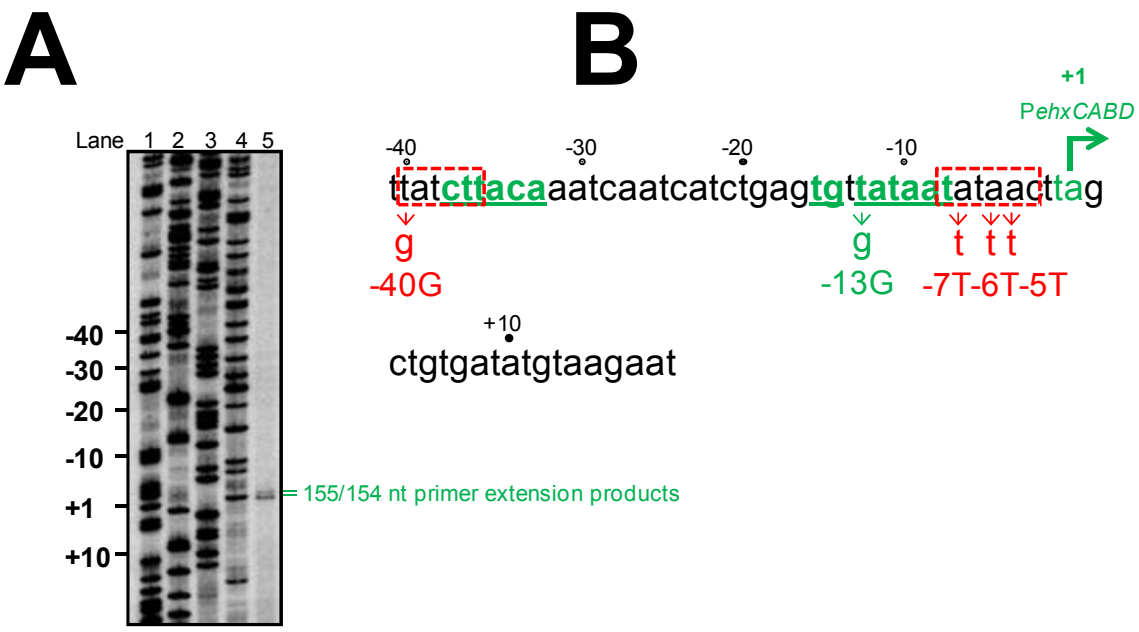


Figure 5.4 Location of the *ehxCABD* transcription start site and promoter.

A. The gel shows products from an mRNA primer extension analysis of the F3 fragment (Lane 5). The gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1–4)

B. Base sequence of the non-template strand. Identified transcript start sites are highlighted in green with the most abundant start site labelled as ‘‘+1’’. The proposed extended -10 and -35 elements are also in green and underlined. Two sequences resembling promoter -10 elements are boxed by a dashed red line. The positions of mutations designed to disrupt the various RNA polymerase binding elements are also shown. The -40G mutation disrupts the highly conserved ‘‘T’’ that occurs in the first position of -10 elements.

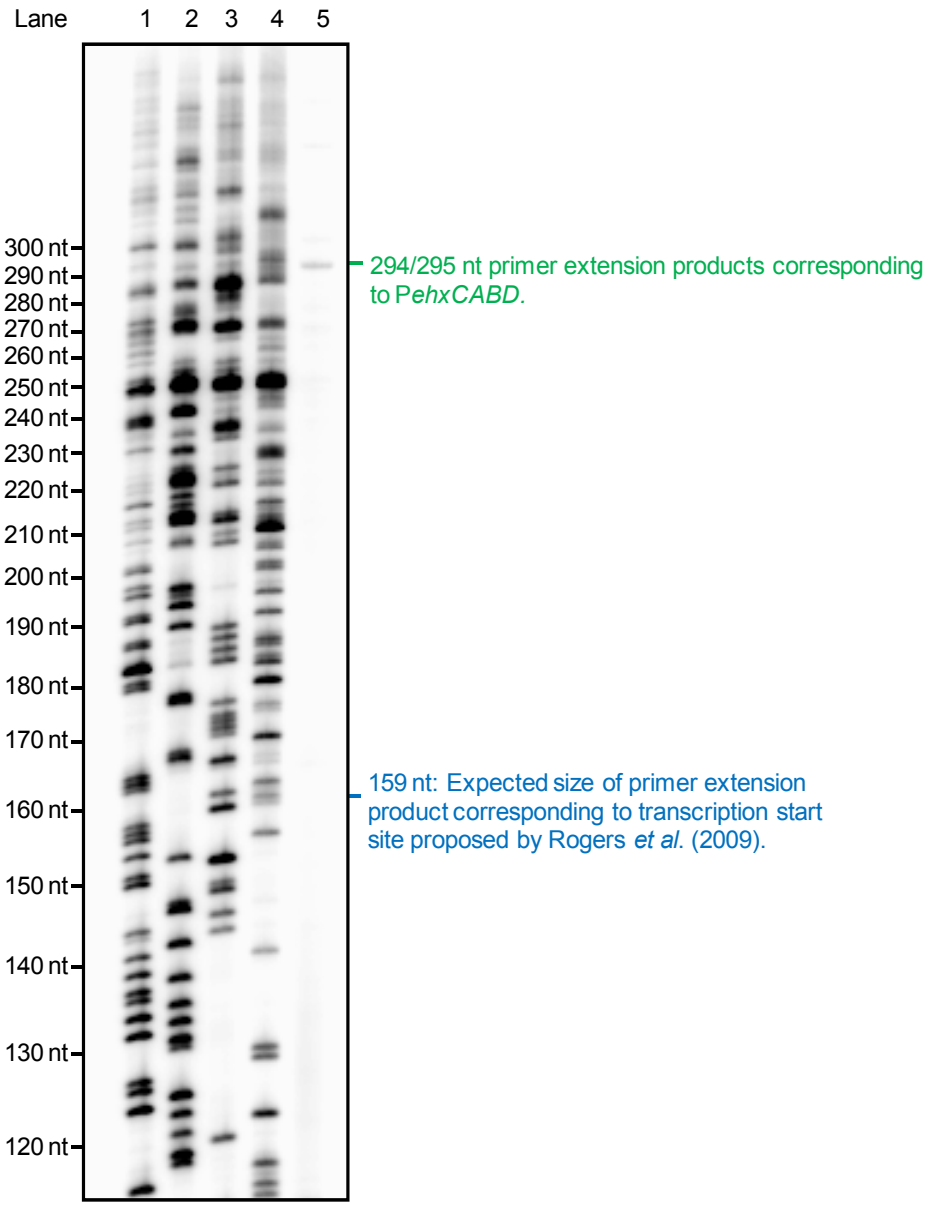


Figure 5.5 Location of the *ehxCABD* transcription start site in the context of the F1 fragment.

Products obtained from an mRNA primer extension analysis of the F1 fragment is shown (Lane 5). Gel was calibrated using arbitrary size standards (A, C, G and T in Lanes 1-4). P*ehxCABD* transcription start site is highlighted in green and the position of the transcription start site proposed by *Rogers et. al.* (2009) is highlighted in blue.

on, we refer to the promoter in the F3 fragment as P*ehxCABD*. We conclude that Rogers *et.al*. incorrectly identified the *ehxCABD* promoter.

To further validate P*ehxCABD*, site directed mutagenesis was used to disrupt the promoter -10 element. Hence, a single point mutation (T to G) was made at promoter position -13 and the resulting DNA fragment was cloned into pRW50 to check promoter activity. As expected, the -13G mutation causes a decrease in P*ehxCABD* activity (Figure 5.6A). Similar results were obtained *in vitro* using the same promoter derivatives cloned in plasmid pSR (Figure 5.6B).

5.4 Binding of RNA polymerase to P*ehxCABD in vitro*

RNA polymerase initiates *ehxCABD* transcription with precision *in vivo* from P*ehxCABD* (Figures 5.4 and 5.5). This is surprising as the *ehxCABD* regulatory region is AT-rich and has many potential -10 promoter elements (two such elements are shown in Figure 5.4B, highlighted in red). Thus, I next analysed the naked F3 fragment to better understand how specificity is achieved by RNA polymerase. I used two *in vitro* DNA footprinting techniques. First, Fe^{2+} chelated Bromoacetamidobenzyl-EDTA (FeBABE) was exploited. FeBABE is a specific DNA cleavage reagent that can be attached to specific cysteine side chains in proteins. FeBABE then cleaves nucleic acids within a 12Å radius of the attachment site. Thus, FeBABE conjugated with the RC461 derivative of *E. coli* σ^{70} cleaves promoter -10 elements (Brown *et. al*., 1999). FeBABE cleavage of the F3 fragment was performed and the result is shown in Figure 5.7A. As expected, the P*ehxCABD* -10 element was cleaved (highlighted by the green box). However, DNA cleavage at additional sites, overlapping P*ehxCABD*, also occurred (highlighted by red stars, Figure 5.7, A). These sites of additional cleavage align with the pseudo -10 elements highlighted red in Figure 5.4B.

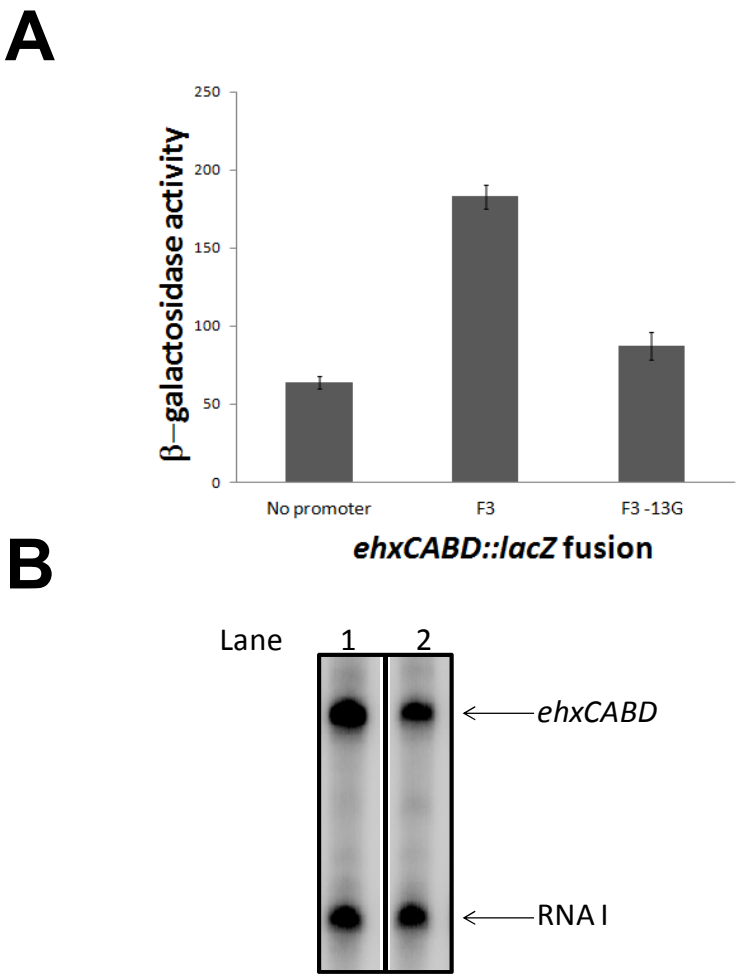


Figure 5.6 Effect of disrupted -10 on the fragment F3 activity *in vitro* and *in vivo*

A. The graph shows LacZ activity data for *E. coli* JCB387 cells carrying different F3::lacZ fusions in pRW50.

B. The figure shows the results of an *in vitro* transcription reaction. Lane 1 shows results with the F3 fragment and lane 2 shows F3 with the -13G mutation. The control RNA I transcript is highlighted.

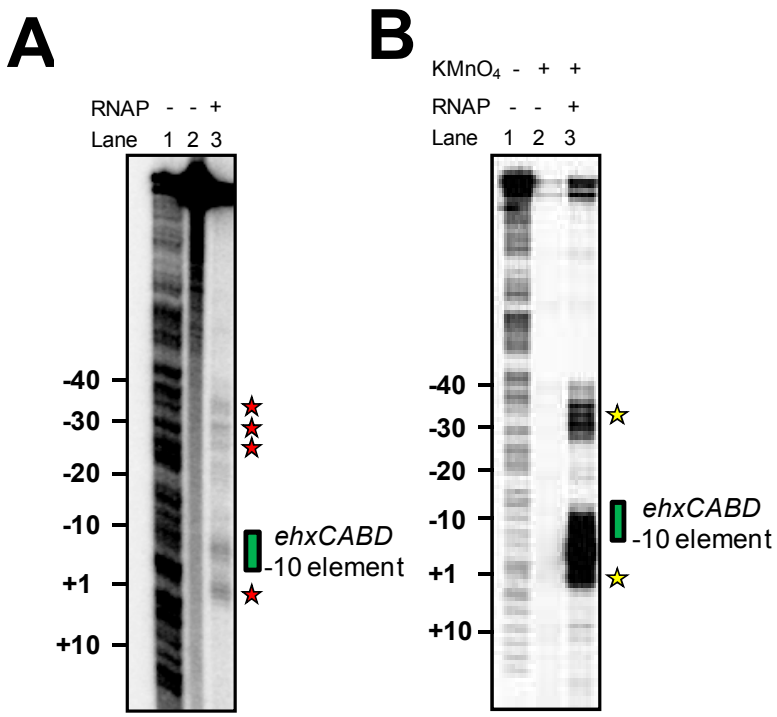


Figure 5.7 RNA polymerase binds multiple sites in the *ehxCABD* gene regulatory region

A. FeBABE Footprint of RNA polymerase (σ^{70} RC461-FeBABE) binding with the *ehxCABD* F3 fragment. The gel shows DNA cleavage products resulting from cleavage of the F3 fragment by RNA polymerase containing σ^{70} RC461-FeBABE (640 nM). The P*ehxCABD* -10 element is indicated by a green box. Additional sites at which the DNA is cleaved are highlighted by red stars. The gel was calibrated with a G+A sequencing ladder (Lane 1)

B. $KMnO_4$ reactivity pattern of the *ehxCABD* promoter in the presence and absence of RNA polymerase. The panel shows DNA cleavage products resulting from $KMnO_4$ treatment of a complex formed between RNA polymerase (320 nM) and the *ehxCABD* F3 fragment. Thus, the sites of DNA cleavage correspond to DNA unwinding by RNA polymerase at -10 hexamers. The P*ehxCABD* -10 element is indicated by a green box. Additional sites at which the DNA is cleaved are highlighted by yellow stars.

Second, $KMnO_4$ footprinting, which detects DNA unwinding by RNA polymerase, was used to investigate binding of RNA polymerase with the F3 fragment. As expected DNA melting at the P*ehxCABD* -10 element was observed (highlighted by the green box, Figure 5.7B). However, DNA opening was also observed at additional sites (highlighted by yellow stars, Figure 5.7B). The additional sites of FeBABE and $KMnO_4$ reactivity align with each other and with the -10 like hexamer sequence highlighted, in red Figure 5.4B. To eliminate the concern that additional FeBABE and $KMnO_4$ reactivity signals might originate from RNA polymerase bound at P*ehxCABD*, identical reactions, with an unrelated control promoter from the *yccE* gene, were done. No DNA cleavage at sites other than the promoter -10 element were observed in these controls (Figure 5.8). Thus, the naked P*ehxCABD* F3 fragment must contain overlapping binding sites for RNA polymerase.

5.5 Effects of H-NS on binding of RNA polymerase to the *ehxCABD* regulatory region

RNA polymerase is capable of initiating transcription specifically from P*ehxCABD in vivo* (Figures 5.4 and 5.5). Conversely, *in vitro*, additional RNA polymerase binding sites are observed (Figure 5.7). Thus, factors present *in vivo* must influence RNA polymerase-DNA interactions to induce specificity. The nucleoid associated protein H-NS is known to bind to curved AT-rich DNA and influence DNA-RNA polymerase interactions (Navarre *et. al*., 2006; Nagarajavel *et al.,* 2007). Interestingly, the 248 bp regulatory region upstream of the *ehxCABD* operon is known to bind H-NS but a detailed molecular analysis of H-NS action at this locus has not been completed (Li *et. al*., 2008; Rogers *et. al*., 2009). Thus, DNAse I footprinting was used to locate the H-NS binding site at the *ehxCABD* F3 fragment. Figure 5.9 shows that H-NS recognises and binds to the same AT-rich region, extending from +10 to

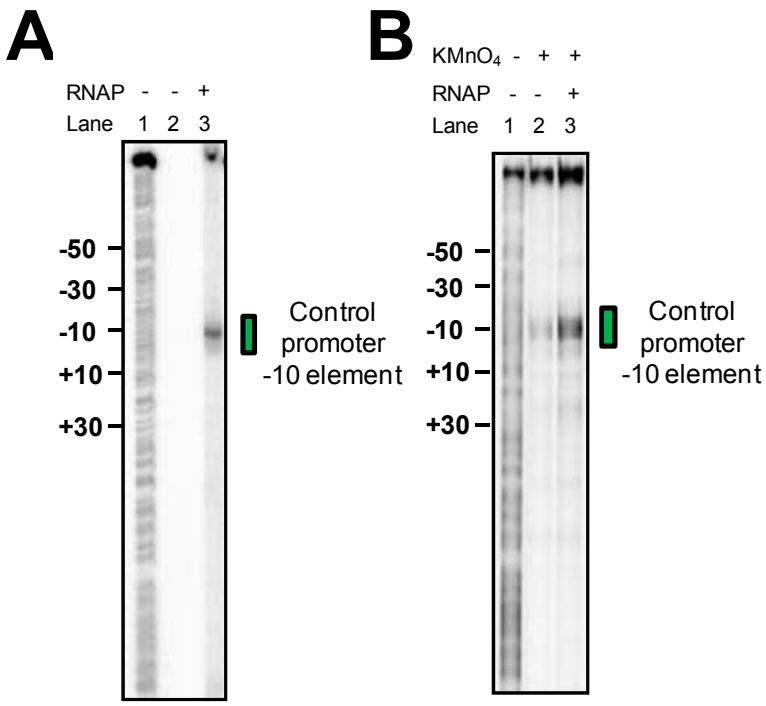


Figure 5.8 RNA polymerase binds specifically to the -10 element at a control promoter

A. Footprint of RNA polymerase (σ^{70} RC461-FeBABE) interactions with -10 elements in a control promoter from the *yccE* gene. The gel shows DNA cleavage products resulting from cleavage of the control promoter by RNA polymerase containing σ^{70} RC461-FeBABE (640 nM). The -10 hexamer is highlighted by a green box.

B. $KMnO_4$ reactivity pattern of a control promoter from the *yccE* gene in the presence and absence of RNA polymerase. The panel shows DNA cleavage products resulting from $KMnO_4$ treatment of a complex formed between RNA polymerase (320 nM) and the control promoter. The control -10 hexamer is highlighted by a green box.

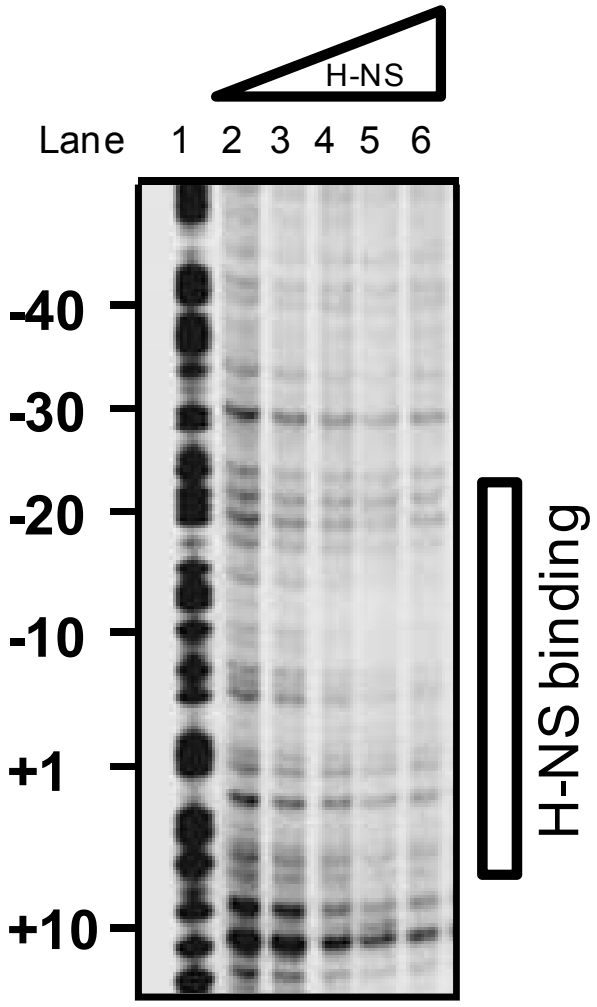


Figure 5.9 Binding of H-NS to the *ehxCABD* F3 fragment.

The panel shows the result of a DNAse I footprint to monitor binding of H-NS to the *ehxCABD* DNA fragment. The gel is calibrated with a Maxim-Gilbert DNA sequencing reaction. H-NS was added at concentrations of 235 nM, 470 nM, 940 nM, 1645 nM or 2350 nM.

-30, as RNA polymerase. We wondered if competition between RNA polymerase and H-NS for binding may stimulate specific recognition of P*ehxCABD* by RNA polymerase. To test this, σ^{70}RC461-FeBABE analysis was repeated with increasing H-NS concentrations. In the presence of H-NS, the signal obtained for RNA polymerase binding at P*ehxCABD* is specific and RNA polymerase binding at additional sites is lost (Figure 5.10, compare peak 6 with other peaks).

We next assessed the effect of H-NS on RNA polymerase binding with P*ehxCABD in vivo*. Hence, we repeated our primer extension analysis with RNA extracted from wild type *E. coli* K-12 cells and cells lacking *hns*. As mentioned above, two primer extension products, of 155 and 154 nt in length, were observed (Figure 5.4, lane 5). Strikingly, RNA from Δ*hns* cells yielded an additional 9 extension products ranging from 138 and 194 nt in length (Figure 5.11). Note that the additional sites of RNA polymerase binding observed in Figure 5.6 align with the additional primer extension products (Figure 5.11, lane 6). Interestingly, to observe primer extension products in the Δ*hns* experiment the gel had to be overloaded. This implies that the overall effect of reduced RNA polymerase binding specificity is a reduction in transcription. Consistent with this, LacZ expression driven by the F3 fragment was reduced in cells lacking H-NS (Figure 5.12).

5.6 H-NS stimulates P*ehxCABD* activity

The data imply that H-NS stimulates specific recognition of P*ehxCABD* by RNA polymerase. To test this model further, *in vitro* transcription assays were used. The F3 fragment was cloned upstream of the λ*oop* terminator in plasmid pSR. P*ehxCABD* is expected to produce transcripts of 178/179 nt in length. However, additional transcripts, corresponding to additional primer extension products in Figure 5.11, should also be observed. Based on this,

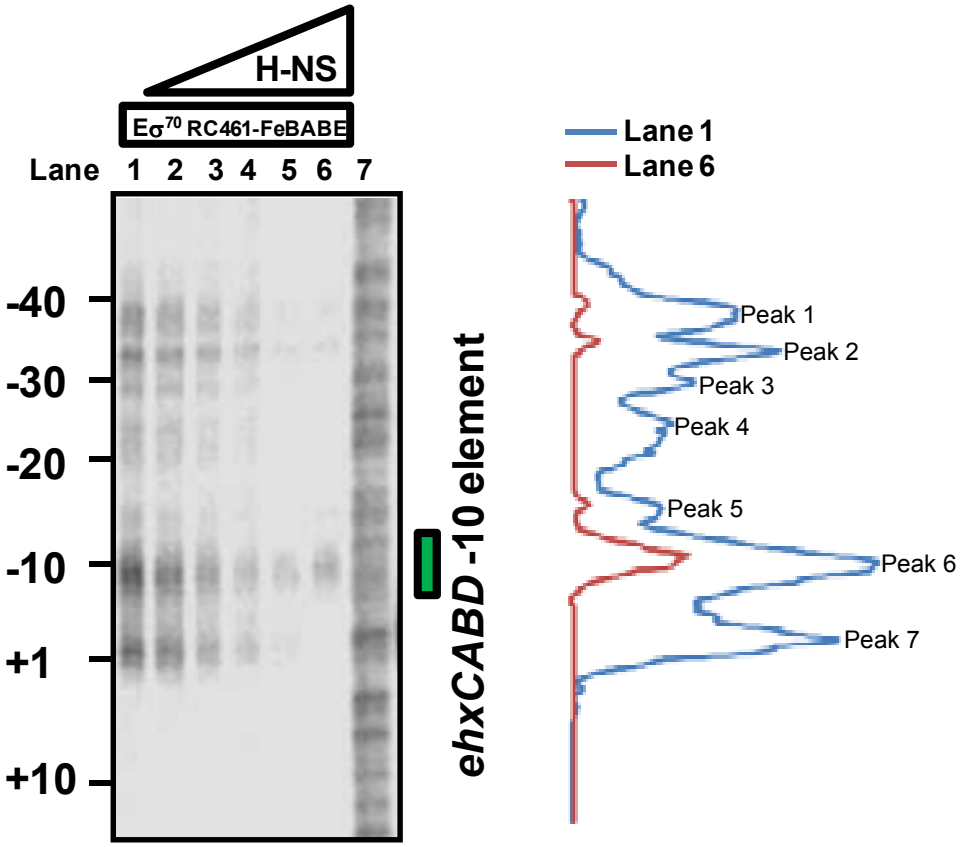


Figure 5.10 Footprint of RNA polymerase (σ^{70} RC461-FeBABE) interactions the *ehxCABD* regulatory region in the presence of H-NS.

The panel shows an image of *ehxCABD* DNA cleavage products separated by electrophoresis on a denaturing acrylamide gel. DNA cleavage was mediated by 640 nM RNA polymerase associated with the σ^{70} RC461-FeBABE derivative that cleaves -10 hexamer sequences. Where present, H-NS was pre-incubated with the DNA at concentrations of 235 nM, 470 nM, 940 nM, 1645 nM or 2350 nM. The position of the *ehxCABD* promoter -10 hexamer is indicated by a green box. The graph indicates band intensities of Lane 1 (blue) compared to Lane 6 (red).

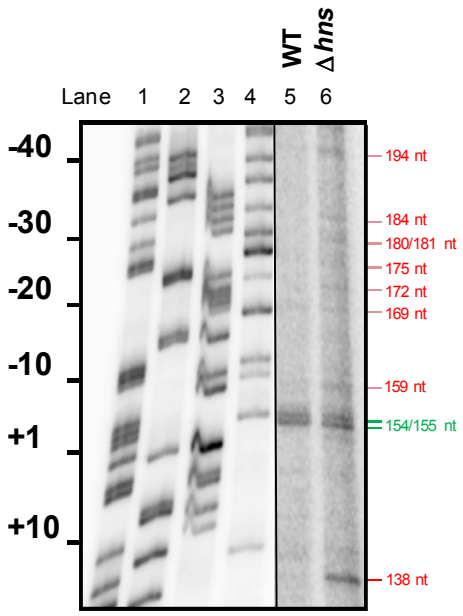


Figure 5.11 Effect of H-NS on transcription start site selection at the *ehxCABD* regulatory region

The panel shows the result of primer extension analysis using RNA extracted from strain M182 or the Δhns derivative, carrying the *ehxCABD* F3 fragment cloned in pRW50, grown aerobically to midexponential phase (OD650 0.4–0.6) in LB medium. The sizes of primer extension products were determined by calibration against size standards (A, C, G and T in Lanes 1–4). The brightness and contrast have been set differently for lanes 1–4 and 5–6 so that the primer extension products can be more easily compared to the marker lanes. The image otherwise represents a single continuous gel.

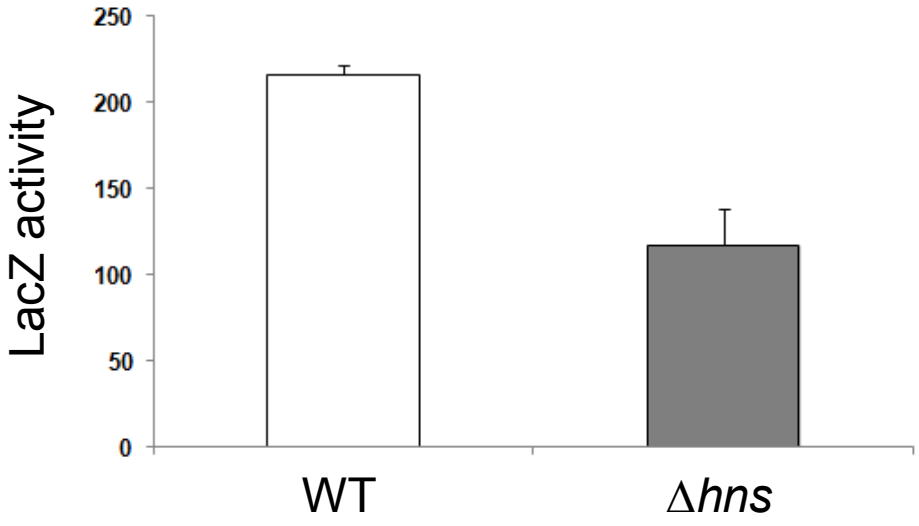


Figure 5.12 H-NS stimulates transcription from the F3 fragment.

The graph shows LacZ activity data for *E. coli* M182 cells, and the Δ*hns* derivative, carrying the F3::*lacZ* fusion in pRW50.

we expected a 162 nt transcript (corresponding to the 138 nt extension product in Figure 5.11) and scarce transcripts sized between 183 nt and 218 nt (equivalent to the primer extension products in the 159-194 nt range). Figure 5.13 shows the result of the *in vitro* transcription assay alongside a set of marker transcripts (Lane 1). As expected, two bands corresponding to the 178/179 and 162 nt products are seen in absence of H-NS (Figure 5.13, Lane 2). However, bands in the 183-218 nt range, which were expected to be less abundant and poorly resolved, were not visible. Even so, increasing the concentration of H-NS added to the reactions stimulated transcription from P*ehxCABD* (Figure 5.13, Lanes 2-5).

5.7 The -10 like sequences overlapping the P*ehxCABD* are inhibitory

In our model, P*ehxCABD* is flanked by at least two overlapping elements that can bind RNA polymerase. Hence, there should be competition between RNA polymerase molecules for binding the various targets. This would lead to reduced transcription from P*ehxCABD*. To test this, the spurious overlapping RNA polymerase binding elements were disrupted (Figure 5.4B, mutations highlighted in red). LacZ activity data for *E. coli* cells carrying the various promoter::*lacZ* fusions are shown in Figure 5.14A. LacZ expression is increased by the -40G mutation, which disrupts the upstream -10 like sequence. Incorporation of the -7T-5T-4T mutations, which disrupt the downstream -10 like sequence, further increases LacZ expression. Thus, removing spurious RNA polymerase binding sites leads to an increase in P*ehxCABD* activity. An *in vitro* transcription analysis was used to confirm the effect of -40G mutation (Figure 5.14B). As expected, the -40G mutation stimulates P*ehxCABD in vivo* as well as *in vitro*.

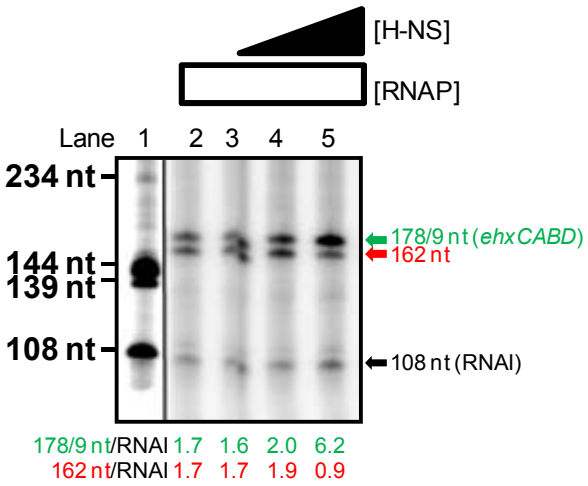


Figure 5.13 Stimulation of P*ehxCABD* by H-NS in vitro

The figure shows the results of an in vitro transcription reaction calibrated with transcripts of known size from a regulatory region. The 178 nt transcript initiates from P*ehxCABD* and the 108 nt RNAI transcript is an internal control.

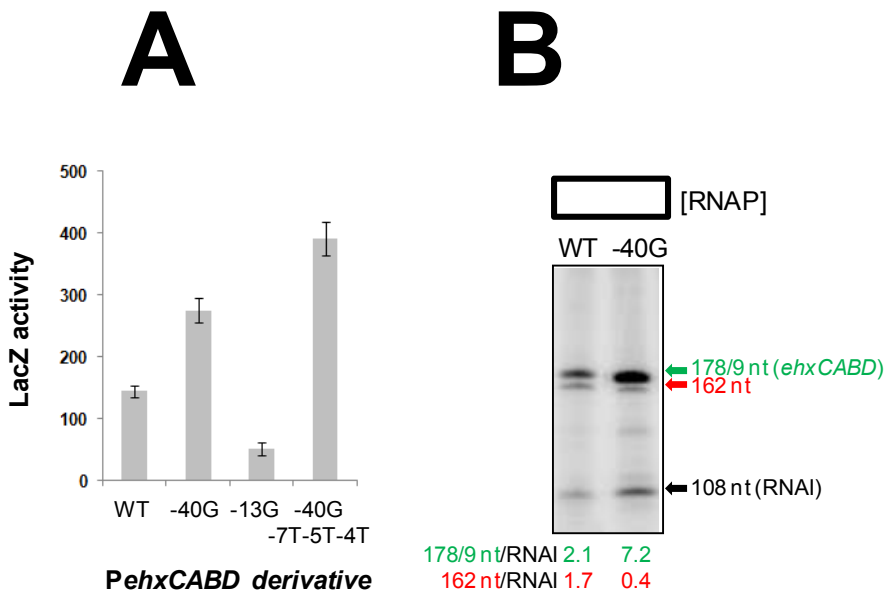


Figure 5.14 Effects of mutations in P*ehxCABD*, and overlapping RNA polymerase binding sites.

A. The graph shows LacZ activity data for *E. coli* JCB387 cells carrying different F3::*lacZ* fusions in pRW50.

B. Stimulation of P*ehxCABD* by the -40G mutation in vitro. The figure shows the results of an in vitro transcription assay comparing the wild type *ehxCABD* F3 fragment with a derivative carrying a mutation at promoter position -40.

5.8 Identification of promoter activity in the *ehxCABD* coding region

Our data show that the AT-rich *ehxCABD* regulatory DNA contains a single promoter but that RNA polymerase can recognise multiple "decoy" -10 elements close to this promoter. As the *ehxCABD* operon itself is AT-rich (62%), we reasoned that the entire coding region could be enriched for RNA polymerase binding sites and, perhaps, promoters. The DNA logo created by aligning the 554 *E. coli* promoters (Mitchell *et al.,* 2003; Figure 3.8) suggests that the most conserved bases in the -10 promoter element are the T at position -12, the A at position -11 and the T at position -7 (First, second and sixth positions in 5'-TATAAT-3') (Figure 3.1). Thus, we scanned *ehxCABD* for -10 like sequences with only one mismatch, at one of the least conserved positions of the -10 element, allowed. The scan identified a total of 94 -10 like hexamers. Of these, 39 were found in the 5'-3' orientation and 55 in the 3'-5' orientation. The different -10 like sequences are listed in Tables 5.1 and 5.2. The tables show the sequence of the ~ 62 bp DNA fragments that were subsequently fused to *lacZ*.

5.9 Activity of the putative intragenic *ehxCABD* operon promoters

A total of 83 ~ 62 bp DNA fragments, carrying the sequences shown in Tables 5.1 and 5.2, were made and cloned into the pRW50 plasmid. We then measured LacZ expression in JCB387 cells carrying these different plasmids. Our aim was to identify active intragenic promoters. Note that some of the -10 like hexamers overlapped and were thus cloned in the same DNA fragment. LacZ expression, driven by each fragment, is shown in Figure 5.15. Results show that, of the 83 constructs, 20 stimulated LacZ activity at least 2 fold.

Table 5.1 Constructs made from the -10 hexamers identified within the *ehxCABD* operon (Forward). The –10 like hexamer is highlighted in red and the shine dalgarno sequence in blue

Construct Name	Sequence	Orientation	Bases
1F	CTGGGCCAGTTTAGATCTGGAATGTGAGGTAAAATATATAAATGATGTAAAGGAGG	5'-3'	62
2F	TTGAGAAAGAAAACTAATATGACAGTAAATAAAATAAAGAACATTTTCAAAGGAGG	5'-3'	62
3F	AAGCGTACGTTCCGCTGGAAAAAAACTCATATTATTAATACCTGATAATTAGGAGG	5'-3'	62
4F	CTTCTGCAGAAGTATCAGAAAGTTGGGAGTAAAATAGGAGGAACCGCTGAAGGAGG	5'-3'	62
5F	TTGGAAATGTTGGTAAAGCTGTTTCGCAATATATTCTGGCTCAGAGAATGAGGAGG	5'-3'	62
6F	ATGCTGATCGCAGTCAGAGTGGTAAGGCATATATTAATTATCTGGAAAATAGGAGG	5'-3'	62
7F	TAAAAGGCATAAAAAATCATAAAGGTGTATATGATTATTCAAAATTGATTAGGAGG	5'-3'	62
8F	GCTCATCAGAAGTATTTGCTGGTGAAGGTTATGATACCGTATCTTATAATAAGACGGATGTTGGTAGGAGG	5'-3'	77
9F	GCAACAGGAGCATCAAAACCTGGTGAGTATATAGTTTCAAAAAATATGTAAGGAGG	5'-3'	62
10F	GGTGTCAGTAGGGAAGCGAACAGAGAAAATACAATATCGTGATTTTGAATAGGAGG	5'-3'	62
11F	ATTTTGAATTCAGAACCGGTGGAATTCCTTATGATGTAATAGATAATCTTAGGAGG	5'-3'	62
12F	GATGAATTCAAAGGCGGTAAGTTTAATGATATATTCCATGGCGCAGATGGAGGAGG	5'-3'	62
13F	GTAATGATGAGCTTCAGGCACACGGAGCTTATAATATTCTGTCAGGTGGTAGGAGG	5'-3'	62
14F	TATGTTTATGGGCAAAACTATGGTCATCATACAATTGCAGATGAAGGAGGAGGAGG	5'-3'	62
15F	ATCGCATTTAAGAGAGTTGGAAATGATCTTATCATGAATAAAGCCATTAAAGGAGG	5'-3'	62
16F	ATTCAGGTGATGAACGATTAGCATCTTTATATAATTTATCCTTACATCAAAGGAGG	5'-3'	62
17F	GGCCATTCCAGGAGGAATGGTCATCGTTATATGATGAGTAAATGTAGTTCAGGAGG	5'-3'	62
18F	GTCTGTATGCACTGATATTGCTTGCACAATATCATAATATAACTGTCAATAGGAGG	5'-3'	62
19F	CTGTCAATGCTGAAACTATAAGGCATCAGTATAATACCCACACACAAGATAGGAGG	5'-3'	62
20F	TTGATATGGCGGGATGACGGTAAGCATTATATATTGTCTCGTATTACTAAAGGAGG	5'-3'	62
21F	TTACTAAAGATTCATCACGCTATCTTGTTTATGATCCAGAACAACATCAGAGGAGG	5'-3'	62
22F	TTAGTTCACCGGGGGTTTTCAACGTTAAATATTATCACAATAGCATTTATAGGAGG	5'-3'	62
23F	ACGTTAAATATTATCACAATAGCATTTATTATAGTGATACTTTTTGAAGTAGGAGG	5'-3'	62
24F	TTGAAGTGATATTAACCGGAGCCAGAACTTATATTTTCTCTCATACTACAAGGAGG	5'-3'	62
25F	TGGTATTACAGCCCTCAATTAACACTGGTTATATTATTGTCACTACCTTGTTAGGAGG	5'-3'	64
26F	TTTCTTGTCGAAACGGTAACAGCAATAAATACAATCAAATCCATGGCAGTAGGAGG	5'-3'	62
27F	ACACCGGTAGAGAAAAAGTCAGGCAGAAATATACTGCCGGAAATTCAGGGAGGAGG	5'-3'	62
28F	TTATACATATCAAAGGGGGATGTTATCGGTATAGTTGGCCGTTCTGGTTCAGGAGG	5'-3'	62
29F	CGCCAGATTGGTGTTGTATTGCAGGAAAATATACTATTAAATCGTAGTATAGGAGG	5'-3'	62
30F	CGCCAGATTGGTGTTGTATTGCAGGAAAATATACTATTAAATCGTAGTATTATCGATAATATTACATTAGGAGG	5'-3'	80
31F	ATGATTTTATTAGAGAACTAAAAGAAGGGTACAATACTATTGTTGGAGAAAGGAGG	5'-3'	62
32F	ACCAGTGCTCTTGATTATGAGTCGGAAAATATAATAATGAAAAATATGTCAGGAGG	5'-3'	62
33F	TCAAGAATATGTAAGAACAGAACCGTAATTATTATTGCACACAGGTTGTCAGGAGG	5'-3'	62
34F	ACGTTCTCATTTTGTTGTATGGAGTATTTTATTATTTGTAATTATATCTCAGGAGG	5'-3'	62
35F	GACGGTGCCTGGTGTTGAATCTGATATTTTAAAATCAGAAGCATCTTTGTAGGAGG	5'-3'	62
36F	AACATTCAGTTATGGAGCAGGAAAATAGCTATATTCAGGCAAAAAATGAAAGGAGG	5'-3'	62

Table 5.2 Constructs made from the -10 hexamers identified within the *ehxCABD* operon (Reverse). The –10 like hexamer is highlighted in red and the shine dalgarno sequence in blue

Construct Name	Sequence	Orientation	Bases
1R	ATGTTACATCATTTATATATTTTACCTCACATTCCAGATCTAAACTGGCCAGGAGG	3'-5'	62
2R	TTTCTTTCTCAATATTAAAATTAAAATCTTGTCTGTTTTTTACTTCAGACATAAGTAGGAGG	3'-5'	68
3R	GGTTTGCAATCGCTGTATCATCTCGTTCAGTACGGTGTATTTCTATTCCAAGGAGG	3'-5'	62
4R	ATAAAGCAATCCCCGTAAAATTCTGTAGTGCTGAGAGAACTGTTCCGGCTAGGAGG	3'-5'	62
5R	TACTACTTTTGGCAATATCATTCTGACTTATATCCTCTCCTTCCCGTTGTAGGAGG	3'-5'	62
6R	ATCAACGGTACTGTTTATACTTGATACTGTATCTACAAGCTGATTAATAAAGGAGG	3'-5'	62
7R	TTCTCTGAGCCAGAATATATTGCGAAACAGCTTTACCAACATTTCCAAGAAGGAGG	3'-5'	62
8R	TTAAATCGTTCAGAATATGATTCAAGCTGCTTAGCTCGCTCAAATTTATCAGGAGG	3'-5'	77
9R	CAATAAGAGATGTCATATATTCATATTTACCACTCTGTTTTCTTTCTCTAAGGAGG	3'-5'	62
10R	TGTATCCTTACCATTTACAATAAGAGATGTCATATATTCATATTTACCACAGGAGG	3'-5'	62
11R	AATCATATACACCTTTATGATTTTTTATGCCTTTTACAGACCATGTATCCAGGAGG	3'-5'	62
12R	TACCAACATCCGTCTTATTATAAGATACGGTATCATAACCTTCACCAGCAAATACTTCTGATGAGAGGAGG	3'-5'	77
13R	TATTTTTTGAAACTATATACTCACCAGGTTTTGATGCTCCTGTTGCATCAAGGAGG	3'-5'	62
14R	CTGCGCCATGGAATATATCATTAAACTTACCGCCTTTGAATTCATCATCAAGGAGG	3'-5'	62
15R	CATAATTACCTTCGATATAATCGTTCCCATCTGCGCCATGGAATATATCAAGGAGG	3'-5'	62
16R	TCCCCATCATCGCCGTATAGTCGATCATTACCATAATTACCTTCGATATAAGGAGG	3'-5'	62
17R	CCTGTCCTCCGGATATATAATCATCCCCATCATCGCCGTATAGTCGATCAAGGAGG	3'-5'	62
18R	TACCACCTGACAGAATATTATAAGCTCCGTGTGCCTGAAGCTCATCATTAAGGAGG	3'-5'	62
19R	CACCACCATAAAGTTTATCATCACCAGTACCACCTGACAGAATATTATAAAGGAGG	3'-5'	62
20R	GCCCATAAACATAAATATCATTACCAAAACCACCATTCAGATAGTCATTAAGGAGG	3'-5'	62
21R	CTTCATCTGCAATTGTATGATGACCATAGTTTTGCCCATAAACATAAATAAGGAGG	3'-5'	62
22R	CTCTCTTAAATGCGATATCATCAAAGCTAATATCAGATAAGTGCAAACGAAGGAGG	3'-5'	62
23R	AACCTGACCGTTCATTATTATTATGAGGTATCTTATCAACTTTTATCTCTCGAAGGAGG	3'-5'	65
24R	GTTTTTTTCAGATGCTATATTACTGGCTTTTATATAACCTGACCGTTCATAGGAGG	3'-5'	62
25R	AAATAATCTTATTAATATCATTGGCAACACTGGTGATATTAACCATGTTTAGGAGG	3'-5'	64
26R	TCAGTGCATACAGACTATTATGAGAACTACATTTACTCATCATATAACGAAGGAGG	3'-5'	62
27R	TTCAGCATTGACAGTTATATTATGATATTGTGCAAGCAATATCAGTGCATACAGAAGGAGG	3'-5'	67
28R	AATCTTGTGTGTGGGTATTATACTGATGCCTTATAGTTTCAGCATTGACAGTTATATTATGATATTAGGAGG	3'-5'	78
29R	TAATACGAGACAATATATAATGCTTACCGTCATCCCGCCATATCAACGCAAGGAGG	3'-5'	62
30R	AGTTAACACCTCAAGTAAAATCCTCCTGTATTTCACAACAGAGGGGATAAAGGAGG	3'-5'	80
31R	TTCAAAAAGTATCACTATAATAAATGCTATTGTGATAATATTTAACGTTGAGGAGG	3'-5'	62
32R	CATTCCTGAGAAACTTATCATCAAGACGTCGACGTAATAAGGGTGATATAAGGAGG	3'-5'	62
33R	CTGAATTTCCGGCAGTATATTTCTGCCTGACTTTTTCTCTACCGGTGTATAGGAGG	3'-5'	62
34R	ATATGTATAAATTAATATTATTCAAAATAACATTACCGTCAGAAGAATACAGGAGG	3'-5'	62
35R	ATCATGCCCATCAATTAAAAATCTGTCCGGTCTCTGGTATATAAAAGCGCTAGGAGG	3'-5'	62
36R	ACTACGATTTAATAGTATATTTTCCTGCAATACAACACCAATCTGGCGTCAGGAGG	3'-5'	62
37R	GAGAAGCTAATGTAATATTATCGATAATACTACGATTTAATAGTATATTTAGGAGG	3'-5'	62
38R	CTTTTAGTTCTCTAATAAAAATCATGGGCACCTGCAAGTCTGGCTGCCTCAAGGAGG	3'-5'	62
39R	TGACATATTTTTCATTATTATATTTTCCGACTCATAATCAAGAGCACTGGTTGAGGAGG	3'-5'	65
40R	AGATAAGAGAAGAGATATAATTACAAATAATAAAATACTCCATACAACAAAATGAGAACGTCTGGAGGAGG	3'-5'	77
41R	TACCATTATTTTTTCTACAATCGCATTCTCTATCGGTTTTATTTCTTTGCAGGAGG	3'-5'	62
42R	AGATGCTTCTGATTTTAAAAATATCAGATTCAACACCAGGCACCGTCAATTAGGAGG	3'-5'	62
43R	TATTCAACAAATACTTAAAAATCATTGAGTTTTCTTCCTTCCTGTGATACCAGGAGG	3'-5'	62
44R	ACCACCTTCTGTATGTATATTTAACTCCTGAACAGTACCACTCACAGGAGAGGAGG	3'-5'	62
45R	CTGTTACTTCGAGAATATCATTATCAGGGACAATAATCATCAGCGTTTCTAGGAGG	3'-5'	62
46R	AACAGAATCTGCAGTTATATTTTTTACTTTCCCTGTAAGGTAACCATGACAGGAGG	3'-5'	62
47R	ATTCCGATCAACAGATATAATCACGTTAAATACAAGCCCTGTATCCGGAAAGGAGG	3'-5'	62
48R	TTCTTTCTCCCTGTATATCATTCCGATCAACAGATATAATCACGTTAAATAGGAGG	3'-5'	62

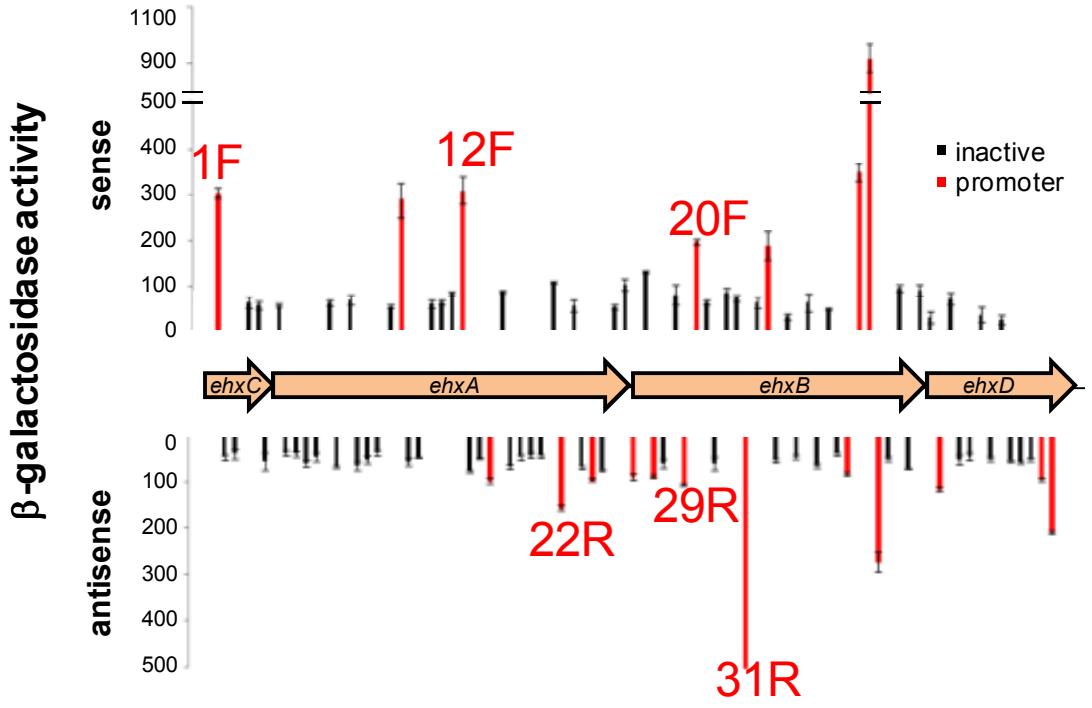


Figure 5.15 Intragenic promoters identified in the *ehxCABD* operon

Graph shows the location of each promoter identified within the *ehxCABD* operon and their activities measured *in vivo*. Promoters with activities that increased the activity at least 2-fold are highlighted in red.

5.10 Intragenic promoters require a -10 element

To test whether LacZ activity, driven by the different 62 bp fragments, was due to the predicted -10 hexamers, we measured the effect of disrupting these sequences. Thus, for 6 out of the 20 active fragments, the predicted -10 element was altered. Hence, the T at position -12 and the A at position -11 were both replaced with G. The LacZ expression properties of these fragments are shown (Figure 5.16). All of the 6 fragments examined showed a significant reduction in their ability to drive LacZ expression when the -10 element was disrupted. Hence, the observed intragenic transcriptional activity is indeed driven by the predicted promoters.

5.11 Discussion

This chapter demonstrates that the background DNA sequence of a regulatory region is an important factor. Thus, at extremely AT-rich regulatory regions, RNA polymerase must ensure binding to the correct promoter -10 element and ignore the background sequence. The *ehxCABD* regulatory region has an AT content of 71% and is full of sequences that resemble promoter -10 elements. I propose that competition with H-NS for binding these elements induces specificity in RNA polymerase binding. Interestingly, the P*ehxCABD* has a consensus extended -10 element which is incredibly rare; it is found in only 3 of the 554 documented *E.coli* promoters (Mitchell *et. al*., 2003). This close match to the consensus sequence could be beneficial for specific recognition by RNA polymerase in very AT-rich regulatory regions. In particular, in the presence of H-NS, overlapping pseudo -10 elements are ignored whilst the close match to the -10 consensus element is not (Figure 5.10 and 5.11). H-NS ensures correct positioning of RNA polymerase and has a positive effect on transcription from P*ehxCABD* (Figure 5.13). Recently, H-NS has been found to positively regulate *mal*T expression (Park *et*.

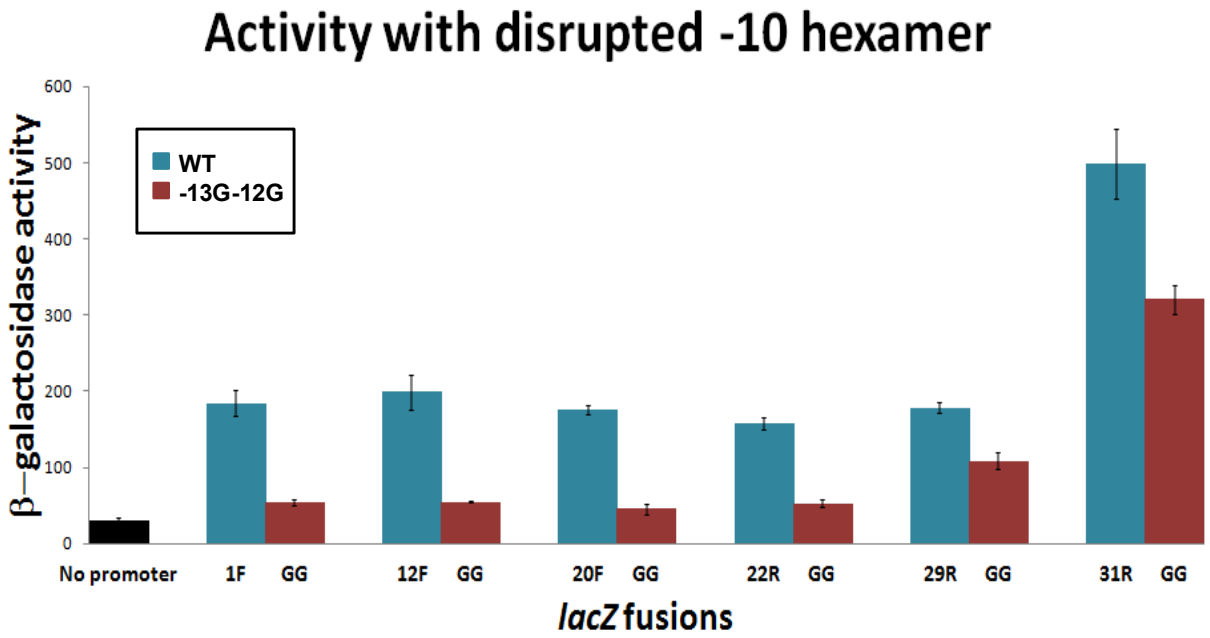


Figure 5.16 LacZ expression driven by of six *ehxCABD* fragments with disrupted -10 elements

The graph shows the data obtained from six fragments- 1F, 12F, 20F, 22R, 29R, and 31R with and without the disrupted -10 element. The blue and red bars represent the activity obtained with the wild type (WT) and the mutated fragments respectively.

al., 2010). Here, H-NS binds to *malT* mRNA at the suboptimal ribosome binding site. This ensures correct positioning of the ribosome to initiate translation. At the *acs*P2 promoter, CRP has an effect similar to that of H-NS on P*ehxCABD*, although the mechanism is different. Hence, at *acs*P2, CRP makes contacts with RNA polymerase that ensure RNA polymerase is bound to the promoter specifically (Beatty *et*. *al*., 2003).

Previously, Rogers *et. al.* (2009) analysed a 1338 bp fragment that included 126 bp of the *ehxCABD* gene regulatory region, the *ehxC* gene and 695 bp of the *ehxA*. The 1338 bp fragment showed LacZ expression and the conclusion was made that a promoter was located within the 126 bp of the *ehxCABD* gene regulatory region. However, in this chapter, I show that the 126 bp of *ehxCABD* regulatory DNA upstream of *ehxC* (fragment F2, Figure 5.1) is unable to drive any transcription in *E. coli* K-12 (Figure 5.2A). In *E.coli* O157, the 126 bp fragment F2 drives some expression which could be due to the presence of *E.coli* O157 specific regulator. Saitoh *et. al.* (2008) showed that GrlA induces the expression of the *ehxCABD* operon. Iyoda *et. al.* (2011) analysed the entire *ehxCABD* gene regulatory region (similar to fragment F1) and found that deleting the upstream part of the regulatory region reduced transcription. Combined with the findings from Rogers *et. al.*, they concluded that the reduced transcription was due to deletion of an activator binding site. The speculated activator binding site aligns with the consensus extended -10 hexamer of the P*ehxCABD*. Thus, the reduction in transcriptional activity, caused by deleting the upstream part of the *ehxCABD* regulatory region, is probably due to deletion of P*ehxCABD*. The transcriptional activity seen by Rogers *et. al.* (2009) likely arose from promoters in the *ehxCABD* coding sequence. Interestingly, transcription from the 1338 bp fragment examined by Rogers *et*. *al*. (2009) was repressed by H-NS. This hints that H-NS not only plays a role in ensuring correct RNA

polymerase positioning at P*ehxCABD* but also controls transcription that initiates within the *ehxCABD* coding region.

Final conclusions

Chapter 6

RNA polymerase σ factor recognition of DNA elements is the first step in transcription initiation. The housekeeping sigma factor of *Escherichia coli*, σ^{70}, recognises two conserved DNA elements; the -10 promoter element (5'-TATAAT-3') and the -35 promoter element (5'-TTGACA-3'). In particular, the -10 promoter element plays an important role in DNA melting to form open complex. However, recognition by RNA polymerase is not always straight forward, especially at promoters with multiple binding sites for one or more than one σ factor. The DNA elements recognised by σ^{70}, and the stationary phase sigma factor, σ^{38} are almost the same. Thus, this work contributes to understanding how specificity is achieved at regulatory regions that have binding sites for both of these σ factors. For example, the *cbpA* regulatory region has overlapping binding sites for σ^{70} and σ^{38}. At this regulatory region, the spacer region sequence (between the -10 and -35 promoter elements) plays a role in determining which promoter is most active and hence the overall σ factor preference of the regulatory DNA. This is surprising because the spacer region sequence is generally considered unimportant. Interestingly, σ^{70} senses the sequence and conformation of the spacer region and has an effect on promoter activity. The σ^{70} side chain R451 and promoter position -18 of the non-template strand are 2 Å apart. Thus base changes at promoter position -18 is sensed by the σ^{70} side chain R451, which alters promoter activity. However, at an improved -10 element for, R451 is not essential to mediate transcriptional stimulation. Similarly, mutation of σ^{70} side chain R451 results in reduced growth rate which is expected as it plays an important role in promoter recognition.

AT-rich DNA is a difficult template for RNA polymerase to recognise and promoter elements as the -10 DNA element, which plays a very important role in recognition, is AT-rich. Thus, it is interesting to know the mechanism behind specific recognition by RNA polymerase at

DNA target sites that are AT-rich. This work finds that Histone- like Nucleoid Structuring protein (H-NS), a nucleoprotein that binds to AT rich DNA, plays an important role in correct recognition of promoter by RNA polymerase at such gene regulatory regions. *ehxCABD* operon regulatory region has a high AT content (71 % AT) and RNA polymerase is not capable of discriminating the promoter -10 element from similar overlapping sequence in the absence of H-NS.

CbpA is known to bind to curved DNA and form aggregates. Single molecule technique, tethered particle motion (TPM), was used to detect the compaction properties of CbpA. It was found that CbpA, when bound, is capable of compacting DNA and a mutant R116A was unable to do so. This was expected as previous work has found that the residue R116 plays an important role in CbpA-DNA binding.

References

- Aiyar S, Mcleod S, Ross W, Hirvonen C, Thomas M, Johnson R and Gourse R (2002) Architecture of Fis-activated transcription complexes at the Escherichia coli rrnB P1 and rrnE P1 promoters. J. Mol. Biol. 316: 501-516.
- Ali Azam T, Iwata A, Nishimura A, Ueda S and Ishihama A (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J. Bacteriol. 181: 6361-6370.
- Ali B, Amit R, Braslavsky I, Oppenheim AB, Gileadi O and Stavans J (2001). Compaction of single DNA molecules induced by binding of integration host factor (IHF). Proc Natl Acad Sci USA 98(19): 10658-10663.
- Almirón M, Link A, Furlong D and Kolter R (1992) A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. Genes Dev. 6(12B):2646-54.
- Arnosti D and Chamberlin M (1989) Secondary sigma factor controls transcription of flagellar and chemotaxis genes in Escherichia coli. Proc Natl Acad Sci USA. 86(3):830-4.
- Barnard A, Wolfe A and Busby S (2004). Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Curr Opin Microbiol. 7(2): 102-108.
- Beatty CM, Browning DF, Busby SJW, Wolfe AJ (2003) Cyclic AMP Receptor Protein-Dependent Activation of the Escherichia coli acsP2 Promoter by a Synergistic Class III Mechanism. J Bacteriol 185: 5148–5157.
- Bell C and Lewis M (2001) The Lac repressor: A second generation of structural and functional studies. Curr. Opin. Struct. Biol. 11:19–25.

- Bertoni G, Fujita N, de Ishihama A, and Lorenzo V (1998) Active recruitment of σ54-RNA polymerase to the Pu promoter of Pseudomonas putida: role of IHF and aCTD. EMBO J 17:5120-5128.
- Bird JG, Sharma S, Roshwalb SC, Hoskins JR and Wickner S (2006) Functional analysis of CbpA, a DnaJ homolog and nucleoid-associated DNA-binding protein. J. Biol. Chem. 281: 34349-34356.
- Blake T, Barnard A, Busby SJW and Green J (2002) Transcription Activation by FNR: Evidence for a Functional Activating Region 2. J Bacteriol 184: 5855-5861.
- Blatter E, Ross W, Tang H, Gourse R and Ebright R (1994) Domain organization of RNA polymerase alpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell. 78:889-896.
- Blattner F, Plunkett G, Bloch C, Perna N, Burland V, Riley M, Collado-Vides J, Glasner J, Rode C, Mayhew G, Gregor J, Davis N, Kirkpatrick H, Goeden M, Rose D, Mau B and Shao Y (1997). The complete genome sequence of Escherichia coli K-12. Science 277, 1453–1474.
- Brown JA, Owens JT, Meares CF, Fujita N, Ishihama A, et al. (1999) Organization of open complexes at Escherichia coli promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem 274: 2263–2270.
- Browning D and Busby S (2004) The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. 2: 57–65.
- Browning D, Grainger D and Busby S (2010) Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. Curr Op Microbiol. 13: 773-780.
- Browning D, Lee D, Green J and Busby S (2002) Secrets of bacterial transcription initiation taught by the Escherichia coli FNR protein. In “Signals, Switches, Regulons and

Cascades: Control of bacterial Gene Expression" (D. Hodgson and C. Thomas, Eds.), pp 127- 142. Cambridge University Press, Cambridge, UK.

- Burgess R, Travers A, Dunn J and Bautz E (1969) Factor stimulating transcription by RNA polymerase. Nature. 221: 43–46.
- Burr T, Mitchell J, Kolb A, Minchin S and Busby S (2000) DNA sequence elements located immediately upstream of the -10 hexamer in Escherichia coli promoters: a systematic study. Nucleic Acids Res 28: 1864-1870.
- Burton NA, Johnson MD, Antczak P, Robinson A, Lund PA (2010) Novel aspects of the acid response network of E. coli K-12 are revealed by a study of transcriptional dynamics. J. Mol. Biol. 401:726-742.
- Busby S and Ebright R (1999) Transcription Activation by Catabolite Activator Protein (CAP). J Mol Biol 293: 199-213.
- Busby S, Kotlarz D, and Buc, H (1983) Deletion mutagenesis of the Escherichia coli galactose operon promoter region. J. Mol. Biol. 167:259–274.
- Busby S, Truelle N, Spassky A, Dreyfus M and Buc H (1984) The selection and characterisation of two novel mutations in the overlapping promoters of the Escherichia coli galactose operon. Gene. 28: 201–209.
- Campbell E, Muzzin O, Chlenov M, Sun J, Olson C, Weinman O, Trester-Zedlitz M and Darst S (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. Mol. Cell. 9: 527–539.
- Chintakayala K and Grainger D (2011) A conserved acidic amino acid mediates the interaction between modulators and co-chaperones in enterobacteria. J Mol Biol. 2011 Aug 12;411(2):313-20. doi: 10.1016/j.jmb.2011.05.043. Epub 2011 Jun 12.

- Colland F, Fujita N, Kotlarz D, Bown J, Meares C, Ishihama A and Kolb A (1999) Positioning of sigma (38), the stationary phase sigma factor, in Escherichia coli RNA polymerase-promoter open complexes. EMBO J. 18: 4049–4059.
- Cosgriff S, Chintakayala K, Chim YT, Chen X, Allen S, Lovering AL, Grainger DC. (2010) Dimerization and DNA-dependent aggregation of the Escherichia coli nucleoid protein and chaperone CbpA. Mol Microbiol. 77(5):1289-300.
- Cukier-Kahn R, Jacquet M, and Gros F (1972) Two Heat-Resistant, Low Molecular Weight Proteins from Escherichia coli That Stimulate DNA-Directed RNA Synthesis. Proc Natl Acad Sci USA. 69(12): 3643–3647.
- Dame R (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. Mol. Microbiol. 56: 858-870.
- Dame R, Noom M and Wuite G (2006). Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. Nature 444(7117): 387-390.
- Dame R, Wyman C, Wurm R, Wagner R and Goosen N (2002). Structural basis for H-NS-mediated trapping of RNA polymerase in the open initiation complex at the rrnB P1. J Biol Chem. 277(3): 2146-2150.
- Dillon S and Dorman C (2010) Bacterial nucleoid-associated protein, nucleoid structure and gene expression. Nat Rev Microbiol. 8: 185-195.
- Dorman C (2013) Genome architecture and global gene regulation in bacteria: making progress towards a unified model? Nature Reviews Microbiology 11, 349–355 (2013)
- Dornenburg J, DeVita A, Palumbo M and Wade J (2010) Widespread Antisense Transcription in Escherichia coli. mBio 1(1):e00024-10. doi:10.1128/mBio.00024-10.
- Dorobantu L and Gray M (2010) Applications of atomic force microscopy in bacterial research. Scanning. 32: 74–96.

- Dove S, Darst S and Hochschild A (2003) Region 4 of σ as a target for transcription regulation. Mol. Microbiol. 48: 863–874.
- Drlica K and Bendich A (2009) Chromosome, Bacterial. Encyclopedia of Microbiology, M. Schaechter, Ed. 507-516.
- Erickson J, Vaughn V, Walter W, Neidhardt F and Gross C (1987) Regulation of the promoters and transcripts of rpoH, the Escherichia coli heat shock regulatory gene. Genes Dev. 1(5):419-32.
- Feklistov A and Darst S (2011) Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. Cell. 147: 1257-1269.
- Fenton M, Lee S and Gralla J (2000) Escherichia coli promoter opening and -10 recognition: mutational analysis of sigma70. EMBO J. 19: 1130–1137.
- Finzi L and Dunlap D (2010) Single-molecule approaches to probe the structure, kinetics, and thermodynamics of nucleoprotein complexes that regulate transcription. The Journal of Biological Chemistry. 285: 18973-18978.
- Gaal T, Barkei J, Dickson R, deBoer H, deHaseth P, Alavi H and Gourse R (1989) Saturation mutagenesis of an Escherichia coli rRNA promoter and initial characterization of promoter variants. J. Bacteriol. 171: 4852–4861.
- Giladi H, Koby1 S, Prag G, Engelhorn M, Geiselmann J and Oppenheim A (1998) Participation of IHF and a distant UP element in the stimulation of the phage λ PL promoter. Molecular Microbiology 30(2): 443–451
- Goosen N and Vandeputte P (1995) The Regulation of Transcription Initiation by Integration Host Factor. Molecular Microbiology (16)1:1-7

- Grainger D, Belyaeva T, Lee D, Hyde E and Busby S (2003) Binding of the Escherichia coli MelR protein to the melAB promoter: orientation of MelR subunits and investigation of MelR–DNA contacts. Molecular Microbiology 48(2): 335–348
- Grainger D, Belyaeva T, Lee D, Hyde E and Busby S (2004) Transcription activation at the Escherichia coli melAB promoter: interactions of MelR with the C-terminal domain of the RNA polymerase alpha subunit. Mol. Microbiol. 51: 1311-1320.
- Grainger D, Hurd D, Harrison M, Holdstock J and Busby S (2005). Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. Proc Natl Acad Sci USA. 102(49): 17693-17698.
- Grainger D and Busby S (2008 a) Global Regulators of Transcription in Escherichia coli: Mechanism of Action and Methods for Study. Advances in Applied Microbiology. 65: 93-113.
- Grainger D and Busby S (2008 b) Methods for studying global patterns of DNA binding by bacterial transcription factors and RNA polymerase. Biochem Soc Trans. 36 (Pt 4): 754-757.
- Grainger D, Goldberg M, Lee D and Busby S (2008) Selective repression by Fis and H-NS at the Escherichia coli dps promoter. Mol. Microbiol. 68:1366-77.
- Griffiths A, Wessler S, Lewontin R, Gelbart W, Suzuki D and Miller J (2005) RNA: Transcription and Processing in “Introduction to genetic analysis”. 259. W. H. Freeman and Company.
- Grossman A, Erickson J and Gross C (1984) The htpR gene product of E. coli is a sigma factor for heat-shock promoters. Cell. 38(2):383-90.
- Gruber T and Gross C (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol. 57: 441-466.

- Hampsey M (2001) Omega meets its match. Trends Genet. 17: 190–191.
- Han L, Lui B, Blumberg S, Beausang J, Nelson P and Phillips R (2008) Calibration of Tethered Particle Motion Experiments. Quantitative Biology - Quantitative Methods. eprint arXiv:0810.2284
- Hancock S, Ghane T, Cascio D, Rohs R, Di Felice R and Johnson R (2013). Control of DNA minor groove width and Fis protein binding by the purine 2-amino group. Nucleic Acids Res. 41(13): 6750-6760.
- Herring C, Raffaelle M, Allen T, Kanin E, Landick R, Ansari A and Palsson B (2005). Immobilization of Escherichia coli RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. J Bacteriol 187(17): 6166-6174.
- Hook-Barnard,I.G. and Hinton,D.M. The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase. Proc. Natl Acad. Sci. USA, 106, 737–742.
- Hunt T and Magasanik B (1985) Transcription of glnA by purified Escherichia coli components: core RNA polymerase and the products of glnF, glnG, and glnL. Proc Natl Acad Sci USA. 82(24):8453-7.
- Igarashi K, Fujita N, Ishihama A (1991) Identification of a subunit assembly domain in the alpha subunit of Escherichia coli RNA polymerase. J Mol Biol. 218:1-6.
- Islam M, Bingle L, Pallen M and Busby S (2011) Organisation of the LEE1 operon regulatory region of enterohaemorrhagic Escherichia coli O157:H7 and activation by GrlA. Mol. Microbiol. 79: 468–483.
- Iyoda S, Honda N, Saitoh T, Shimuta K, Terajima J, et al. (2011) Coordinate control of the locus of enterocyte effacement and enterohemolysin genes by multiple common virulence regulators in enterohemorrhagic Escherichia coli. Infect Immun 79: 4628–4637.

- Jafri S, Evoy S, Cho K, Craighead HG and Winans SC. (1999) An Lrp-type transcriptional regulator from Agrobacterium tumefaciens condenses more than 100 nucleotides of DNA into globular nucleoprotein complexes. J Mol Biol. 21;288(5):811-24.
- Kimura M, Fujita N, Ishihama A (1994) Functional map of the alpha subunit of Escherichia coli RNA polymerase. J Mol Biol. 242:107-115.
- Koch C and Kahmann R (1986) Purification and properties of the Escherichia coli host factor required for inversion of the G segment in bacteriophage Mu. J Biol Chem. 261(33):15673-8.
- Kolb A, Kotlarz D, Kusano S and Ishihama A (1995) Selectivity of the Escherichia coli RNA polymerase Esigma38 for overlapping promoters and ability to support CRP activation. Nucleic Acids Res. 23: 819–826.
- Lamberg K and Kiley P (2000) FNR-dependent activation of the class II dmsA and narG promoters of Escherichia coli requires FNR-activating regions 1 and 3. Molecular Microbiology. 38(4): 817-827.
- Leela J, Syeda A, Anupama K and Gowrishankar J (2013) Rho-dependent transcription termination is essential to prevent excessive genome-wide R-loops in Escherichia coli. Proc Natl Acad Sci USA 110(1): 258-263.
- Li H, Granat A, Stewart V, Gillespie JR (2008) RpoS, H-NS, and DsrA influence EHEC hemolysin operon (ehxCABD) transcription in Escherichia coli O157:H7 strain EDL933. FEMS Microbiol Lett 285: 257–262.
- Liu M, Tolstorukov M, Zhurkin V, Garges S and Adhya S (2004) A mutant spacer sequence between -35 and -10 elements makes the Plac promoter hyperactive and cAMP receptor protein-independent. Proc. Natl Acad. Sci. USA. 101: 6911–6916.

- Liu S and Wang Y (2010) Application of AFM in microbiology: a review. Scanning. 32: 61–73.
- Luijsterburg M, Noom M, Wuite G and Dame, R (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. J Struct Biol 156: 262-272.
- Mathew R and Chatterji D (2006) The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol. 14: 450–455.
- McClure W and Cech C (1978) On the mechanism of rifampicin inhibition of RNA synthesis. J. Biol. Chem. 253:8949–8956.
- Mcleod S and Johnson R (2001). Control of transcription by nucleoid proteins. Curr. Opin. Microbiol. 4: 152-159.
- Mcleod S, Aiyar S, Gourse R and Johnson R (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: 517-529.
- Mechold U, Potrykus K, Murphy H, Murakami K and Cashel M (2013). Differential regulation by ppGpp versus pppGpp in Escherichia coli. Nucleic Acids Res 41(12): 6175-6189.
- Meyer A and Grainger D (2013) The Escherichia coli Nucleoid in Stationary Phase. Adv Appl Microbiol. 2013;83:69-86. doi: 10.1016/B978-0-12-407678-5.00002-7
- Miller H and Friedman D (1980). An E. coli gene product required for lambda site-specific recombination. Cell. 20(3):711-9.
- Mitchell J, Zheng D, Busby S and Minchin S (2003) Identification and analysis of ‘extended -10’ promoters in Escherichia coli. Nucleic Acids Res. 31: 4689–4695.

- Müller-Hill B (1996) The Lac Operon: A Short History of a Genetic Paradigm. Walter de Gruyter, New York.
- Mulvey M, Switala J, Borys A and Loewen P (1990) Regulation of transcription of katE and katF in Escherichia coli. J Bacteriol. 172(12):6713-20.
- Murakami K, Masuda S, Campbell E, Muzzin O and Darst S (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science. 17: 1285–1290.
- Nagarajavel V, Madhusudan S, Dole S, Rahmouni A and Schnetz K (2007) Repression by binding of H-NS within the transcription unit. J. Biol. Chem. 282: 23622-23630.
- Navarre W, Porwollik S, Wang Y, McClelland M, Rosen H, Libby S and Fang F (2006) Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. Science. 313: 236-238.
- Newlands J, Josaitis C, Ross W and Gourse R (1992) Both fis-dependent and factor-independent upstream activation of the rrnB P1 promoter are face of the helix dependent. Nucleic Acids Res. 20:719-726.
- Owen-Hughes T, Pavitt G, Santos D, Sidebotham J, Hulton C, Hinton J and Higgins C (1992) The chromatin-associated protein H-NS interacts with curved DNA to influence DNA topology and gene expression. Cell. 71: 255-265.
- Park H, Ostberg Y, Johansson J, Wagner E and Uhlin B (2010) Novel role for a bacterial nucleoid protein in translation of mRNAs with suboptimal ribosome binding sites. Genes Dev 24: 1345–1350.
- Paul S, Zhang X and Hulett F (2004) Autoinduction of Bacillus subtilis phoPR operon transcription results from enhanced transcription from EsigmaA- and EsigmaE responsive promoters by phosphorylated PhoP. J. Bacteriol. 186, 4262–4275

- Peters J, Mooney R, Grass J, Jessen E, Tran F and Landick R (2012). Rho and NusG suppress pervasive antisense transcription in Escherichia coli. Genes Dev 26(23): 2621-2633.
- Peters J, Vangeloff A and Landick R (2011). Bacterial transcription terminators: the RNA 3'-end chronicles. J Mol Biol 412(5): 793-813.
- Popham D and Setlow P (1993) Cloning, nucleotide sequence, and regulation of the Bacillus subtilis pbpF gene, which codes for a putative class A high-molecular-weight penicillin-binding protein. J. Bacteriol. 175, 4870–4876
- Pouget N, Dennis C, Turlan C, Grigoriev M, Chandler M and Salomé L (2004) Single-particle tracking for DNA tether length monitoring. Nucleic Acids Res. 32 (9).
- Pressler U, Staudenmaier H, Zimmermann L and Braun V (1988) Genetics of the iron dicitrate transport system of Escherichia coli. J Bacteriol. 170(6):2716-24.
- Raghavan R, Sloan D and Ochman H (2012) Antisense transcription is pervasive but rarely conserved in enteric bacteria. mBio 3(4):pii e00156-12.
- Ramirez M, Castellanos-Juarez F, Yasbin R & Pedraza-Reyes M (2004) The ytkD (mutTA) gene of Bacillus subtilis encodes a functional antimutator 8-Oxo-(dGTP/GTP)ase and is under dual control of Sigma A and Sigma F RNA polymerases. J. Bacteriol. 186, 1050–1059
- Rao L, Ross W, Appleman J, Gaal T, Leirmo S, Schlax P, Record M, and Gourse R (1994) Factor- independent activation of rrnB P1- an extended promoter with an upstream element that dramatically increases promoter strength. J. mol. Boil. 235: 1421-1435.
- Reppas N, Wade J, Church G and Struhl K (2006). The transition between transcriptional initiation and elongation in E. coli is highly variable and often rate limiting. Mol Cell 24(5): 747-757.

- Reyes-Lamothe; R, Wang X and Sherratt D (2008) Escherichia coli and its chromosome. Trends Microbiol. 16(5): 238-45.
- Rhodius V, Savery N, Kolb A and Busby S (2001) Assays for Transcription Factor Activity. DNA-Protein Interactions. Methods in Molecular Biology. Volume 148, 2001, pp 451-464
- Rimsky S and Travers A (2011) Pervasive regulation of nucleoid structure and function by nucleoid-associated proteins. Curr Opin Microbiol. 14(2): 136-41.
- Rogers M, Zimmerman R and Scott M (2009) Histone-like nucleoid-structuring protein represses transcription of the ehx operon carried by locus of enterocyte effacement-negative Shiga toxin-expressing Escherichia coli. Microb Pathog 47:202–211.
- Rosenberg M and Court D (1979) Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13: 319–353.
- Ross W, Ernst A and Gourse R (2001). Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove. Genes Dev 15(5): 491-506.
- Ross W, Vrentas C, Sanchez-Vazquez P, Gaal T and Gourse R (2013). The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. Mol Cell 50(3): 420-429.
- Rouvière-Yaniv J and Gros F(1975) Characterization of a novel, low-molecular-weight DNA-binding protein from Escherichia coli. Proc Natl Acad Sci USA. 72(9): 3428–3432.
- Saitoh T, Iyoda S, Yamamoto S, Lu Y, Shimuta K, Ohnishi M, Terajima J, Watanabe H (2008) Transcription of the ehx Enterohemolysin Gene Is Positively Regulated by GrlA, a Global Regulator Encoded within the Locus of Enterocyte Effacement in Enterohemorrhagic Escherichia coli. J Bacteriol. 190(14): 4822–4830.

- Sanderson A, Mitchell J, Minchin S, Busby S (2003) Substitutions in the Escherichia coli RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. FEBS Lett 544(1-3):199-205
- Savery N, Belyaeva T and Busby S (1996a) Introduction to protein: DNA interactions, DNase I foot printing, hydroxyl radical foot printing, permanganate foot printing and supplementary protocols. In Essential Techniques: Gene Transcription. Docherty K. (ed.). Oxford: BIOS Scientific Publishers, pp. 1–5 and 21–33.
- Savery N, Rhodius V and Busby S (1996b) Protein- protein interactions during transcription activation: The case of Escherichia coli cAMP receptor protein. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 351: 543-550.
- Schmidt H, Kernbach C, Karch H (1996) Analysis of the EHEC hly operon and its location in the physical map of the large plasmid of enterohaemorrhagic Escherichia coli O157:H7. Microbiology 142: 907–914.
- Schröder O and Wagner R. (2000) The bacterial DNA-binding protein H-NS represses ribosomal RNA transcription by trapping RNA polymerase in the initiation complex. J. Mol. Biol. 298, 737–748
- Selinger D, et al. (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. Nat Biotechnol 18(12):1262–1268.
- Sharma C, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler P and Vogel J (2010). The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464(7286): 250-255.
- Shin M, Song M, Rhee J, Hong Y, Kim Y, Seok Y, Ha K, Jung S and Choy H (2005) DNA looping-mediated repression by histone-like protein H-NS: specific requirement of $E\sigma^{70}$ as a cofactor for looping. Genes Dev. 19: 2388-98.

- Siebenlist U and Gilbert W (1980) Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. Proc. Natl Acad. Sci. USA. 77: 122–126.
- Siebenlist U, Simpson RB and Gilbert W (1980) E. coli RNA polymerase interacts homologously with two different promoters. Cell. 20: 269–281.
- Tagami H and Aiba H (1998) A common role of CRP in transcription activation: CRP acts transiently to stimulate events leading to open complex formation at a diverse set of promoters. The EMBO Journal 17:1759 – 1767.
- Tagami H & Aiba H (1999). An inactive open complex mediated by an UP element at Escherichia coli promoters. Proc. Natl Acad. Sci. USA, 96, 7202-7207.
- Tang H, Severinov K, Goldfarb A, Fenyo D, Chait B and Ebright R (1994) Location, structure, and function of the target of a transcriptional activator protein. Genes Dev 8:3058-3067
- Thanbichler M, Viollier P and Shapiro L (2005) The structure and function of the bacterial chromosome. Curr Opin Genet & Dev 15: 153-162.
- Thomason M and Storz G (2010) Bacterial antisense RNAs: How many are there, and what are they doing? Annu Rev Genet 44:167–188.
- Thouvenot B, Charpentier B and Branlant C (2004) The strong efficiency of the Escherichia coli gapA P1 promoter depends on a complex combination of functional determinants. Biochem. J. 383: 371–382.
- Towles K, Beausang J, Garcia H, Phillips R and Nelson P (2009) First-principles calculation of DNA looping in tethered particle experiments. Phys Biol. 6(2).
- Trent, M. S., Worsham, L. M., & Ernst-Fonberg, M. L. (1999). HlyC, the internal protein acyltransferase that activates hemolysin toxin: role of conserved histidine, serine, and

cysteine residues in enzymatic activity as probed by chemical modification and site-directed mutagenesis. Biochemistry, 38(11), 3433-3439.

- Tucker G (1996) Biotechnology and enzymes in the food industry. British Food Journal. 98(4/5)14-19
- Typas A and Hengge R (2006) Role of the spacer between the -35 and -10 regions in sigma38 promoter selectivity in Escherichia coli. Mol. Microbiol. 59: 1037–1051.
- Ueguchi C, Kakeda M, Yamada H and Mizuno T (1994) An analogue of the DnaJ molecular chaperone in Escherichia coli. Proc. Natl. Acad. Sci. USA. 91:1054-1058.
- Ussery D, Larsen T, Wilkes T, Friis C, Worning P, Krogh A and Brunak S (2001) Genome organisation and chromatin structure in Escherichia coli. Biochimie 83 (2001) 201−212
- Vlahovicek K, Kaja´n L and Pongor S (2003) DNA analysis servers: plot.it, bend.it, model. It and IS. Nucleic Acids Res.31: 3686‑3687.
- Wade J, Belyaeva T, Hyde E and Busby S (2001). A simple mechanism for co-dependence on two activators at an Escherichia coli promoter. EMBO J. 20, 7160-7.
- Wade J, Castro Roa D, Grainger D, Hurd D, Busby S, Struhl K and Nudler E (2006). Extensive functional overlap between sigma factors in Escherichia coli. Nat Struct Mol Biol 13(9): 806-814.
- Willins D, Ryan C, Platko J and Calvo J (1991) Characterization of Lrp, and Escherichia coli regulatory protein that mediates a global response to leucine. J Biol Chem. 266(17):10768-74.
- Wilson C, Zhan H, Swint-Kruse L and Matthews K (2007) The lactose repressor system: paradigms for regulation, allosteric behaviour and protein folding. Cell Mol Life Sci 64: 2-16.
- Wosten M (1998) Eubacterial sigma-factors. FEMS Microbiol.Rev. 22: 127–150.

- Wyborn N, Stapleton M, Norte V, Roberts R, Grafton J, and Green J (2004) Regulation of Escherichia coli hemolysin E expression by H-NS and Salmonella SlyA. J Bacteriol 186:1620–1628.
- Yamashino T, Kakeda M, Ueguchi C and Mizuno T (1994) An analogue of the DnaJ molecular chaperone whose expression is controlled by sigma s during the stationary phase and phosphate starvation in Escherichia coli. Mol. Microbiol., 13, 475–483.
- Young B, Anthony L, Gruber T, Arthur T, Heyduk E, Lu C, Sharp M, Heyduk T, Burgess R and Gross C (2001). A Coiled-Coil from the RNA Polymerase β′ Subunit Allosterically Induces Selective Nontemplate Strand Binding by σ70. Cell 105(7): 935-944.
- Yuzenkova Y, Tadigotla V, Severinov K and Zenkin N (2011) A new basal promoter element recognized by RNA polymerase core enzyme. EMBO J. 30(18):3766-75.
- Zhang A and Belfort M (1992) Nucleotide sequence of a newly-identified Escherichia coli gene, stpA, encoding an H-NS-like protein. Nucleic Acids Res. 25;20(24):6735.
- Zhou Y, Kusukawa N, Erickson J, Gross C and Yura T (1988) Isolation and characterization of Escherichia coli mutants that lack the heat shock sigma factor Sigma 32. J. Bacteriol. 170, 3640–3649.
- Zhou Y, Pendergrast P, Bell A, Williams R, Busby S and Ebright R (1994) The functional subunit of a dimeric transcription activator protein depends on promoter architecture. EMBO J. 13:4549-4557.
- Zuo Y, Wang Y and Steitz T (2013). The mechanism of E. coli RNA polymerase regulation by ppGpp is suggested by the structure of their complex. Mol Cell 50(3): 430-436.

Appendix: Abbreviations

A	adenine
A (Ala)	Alanine
AFM	Atomic force microscopy
APS	Ammonium persulphate
AR	Activating Region
aRNA	Antisense ribonucleic acid
Arg, R	arginine
Å	Ångstrom
B. subtilis	*Bacillus subtilis*
bp	Base pair
BSA	Bovine serum albumin
C	cytosine
CbpA	Curved DNA binding protein
ChIP	Chromatin Immunoprecipitation
CIAP	Calf intestine alkaline phosphatase
CRP	Cyclic-AMP receptor protein
cAMP	3'-5'-cyclic adenosine monophosphate
CTD	Carboxy-terminal domain
D (Asp)	Aspartic acid
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate (N = A, C, G, T)

Dps	DNA- binding protein from starved cells
E	RNA polymerase core enzyme
E (Glu)	Glutamic acid
E. coli	*Escerichia coli*
EDTA	diaminoethanetetra-acetic acid
EMSA	electrophoretic mobility shift assays
FeBABE	*p*-bromoacetamidobenzyl-EDTA-Fe
Fis	Factor for Inversion Stimulation
FNR	fumarate and nitrate reductase
G	guanosine
H. pylori	*Helicobacter pylori*
H (His)	Histidine
H-NS	Histone-like Nucleoid Structuring protein
HU	Histone-like protein from *E. coli* strain U93
I (Ile)	Isoleucine
IHF	Integration host factor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
$KMnO_4$	Potassium permanganate
L (Leu)	Leucine
LB	Lennox Broth
Lrp	Leucine responsive protein
Mbp	Mega base pairs
M (Met)	Methionine

MES	2-(*N*-morpholino)ethanesulfonic acid
Mg	Magnesium
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	Messenger ribonucleic acid
MT	Magnetic tweezers
N (Asn)	Asparagine
NTD	Amino-terminal domain
O	Operator
OD	Optical Density
ONPG	*o*-nitrophenyl-β-D-galactopyranoside
PCR	Polymerase chain reaction
ppGpp	Guanosine tetraphosphate
Q (Gln)	Glutamine
R (Arg)	Arginine
$RbCl_2$	Rubidium chloride
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNase	Ribonuclease
SDS	Sodium dodecyl sulphate
StpA	Suppression of td^- phenotype
T	Thymine
T (Thr)	Threonine
T. aquaticus	*Thermus aquaticus*
TEMED	N,N,N',N'-tetramethylethylene diamine

T_m	Melting temperature
TPM	Tethered particle motion
Tris	Tris (hydroxymethyl) aminoethane
TSS	Transcription start site
U	Uracil
V (Val)	Valine
W (Trp)	Tryptophan
WT	Wild type
Y (Tyr)	Tyrosine