A Study of Intragenic Promoters in *E. coli*

by

EMILY WARMAN

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

DOCTOR OF PHILOSOPHY

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

June 2020

Abstract

RNA polymerases initiate transcription at DNA sequences called promoters. In bacteria, the best conserved promoter feature is the AT-rich -10 element; a sequence essential for DNA unwinding. Sequences that resemble the -10 element occur frequently in AT-rich DNA. As a result, horizontally-acquired genes are predisposed to spurious transcription. However, it is not clear how RNA polymerase initially recognises such sequences.

This work explores the properties of promoters located inside AT-rich genes. Intragenic promoter sequences frequently contain a short AT-tract which resides upstream of sub-optimal -10 elements. The AT-tract stimulates transcription by altering DNA conformation and enhancing contacts between the DNA backbone and RNA polymerase.

Intragenic promoter regions are also frequently divergent; able to drive transcription in both directions. We show that this is an inherent property of all promoters. Hence, whilst promoters are traditionally believed to drive transcription in a single direction, they are often bidirectional. Mechanistically, this occurs because -10 elements have inherent symmetry and often coincide on opposite DNA strands.

Finally, this work shows that RNAP-associated proteins σ^{38}, Rho and NusA are recruited to AT-rich DNA. The alternative σ factor σ^{38} frequently recognises intragenic promoters and plays a role in initiating spurious transcription. Rho and NusA are involved in termination of non-coding transcripts and so are also recruited to sites of spurious transcription. Titration of these proteins away from sites of normal function is expected to be toxic to the cell.

Acknowledgments

I am very grateful for the support and excellent supervision from Professor David Grainger. Thank you for giving me the opportunity to pursue a PhD when I wasn't sure it was possible, restoring my confidence as a scientist and academic. I am also thankful for the insight and feedback offered by my secondary supervisor, Dr Pawel Grzechnik, and my internal assessor, Dr Pete Lund. Thank you to all the other staff members at the University of Birmininghам who taught me as an undergrad and helped my research as a postgrad. Outside of Birmingham, I am very grateful to Bob Landick and his lab group in Madison, Wisconsin who hosted me on a research visit. The research might not have been a great success but I made new friends and learnt a lot.

I am lucky to have been part of a wonderful and supportive lab group. Drs Lisa Lamberte, Shivani Singh, James Haycocks and Prateek Sharma have been excellent mentors and patient teachers. Dr Gemma Warren has been an enormous help and I am particularly grateful for all the media, sterilised tips and pipettes I have borrowed from her over the years. Thank you to Dr David Forrest, Ali Trigg, Lucas Walker and Alistair Middlemiss for being good colleagues and good friends, who make the office a fun place to be. Special thanks to Dr Rachel Kettles and Tom Guest for always being up for a coffee break.

Thank you to all my friends and family. My partner and biggest fan, Thomas McClay, has cheered me on every step of the way. Thank you for listening to me complain when an experiment hasn't worked and for celebrating with me when it has. Thank you to my amazing mum, Ana, and Jes and Wilf for always being there for me. I am so lucky to

have such a kind, supportive and badass family. Finally thanks to the feline part of my family Salem, Pepper and Tulip; they never fail to cheer me up.

Contents

i. List of Figures

ii. List of Tables

iii. Abbreviations

°C	Degrees Celsius
A	Alanine
APS	Ammonium persulfate
bp	Base pair
BSA	Bovine serum albumin
B. subtilis	*Bacillus subtilis*
ChIP	Chromatin Immunoprecipitation
ChIP-seq	ChIP and sequencing
Ci	Curie
CIP	Calf intestine alkaline phosphatase
CTD	C-terminal domain
dH_2O	Distilled water
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	2'-deoxyribonucleoside 5'-triphosphate
dRNA-seq	Differential RNA sequencing
dsDNA	Double stranded DNA
DTT	Dithiothreitol
E. coli	*Escherichia coli*

EDTA	Ethylenediaminetetraacetic acid
Fis	Factor for inversion stimulation
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
H-NS	Histone-like nucleoid structuring protein
IHF	Integration Host Factor
LB	Lennox Broth
LEE	Locus of enterocyte effacement
Mg^{2+}	Magnesium ion
mRNA	Messenger RNA
NAP	Nucleoid Associated Protein
nt	Nucleotide
NTD	N-terminal domain
OD	Optical density
Oligo	Oligonucleotide
ONPG	Ortho-nitrophenyl-β-galactosidase
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PPP-seq	RNA 5' polyphosphatase sequencing
R	Arginine
RNA	Ribonucleic Acid

RNAP	RNA polymerase
RNase	Ribonuclease
RNA-seq	RNA extraction and sequencing
SDS	Sodium dodecyl sulphate
S. Typhimurium	*Salmonella enterica*, serovar Typhimurium
sRNA	small RNA
TEMED	Tetramethylethylenediamine
Tris	Tris (hydroxymethyl) aminoethane
TSS	Transcription start site
UP element	Upstream promoter element
V	Volts
WT	Wild type

Chapter 1. **Introduction**

1.1. RNA polymerase

All life forms depend on the information encoded in their DNA. Transcription is the first step required to use this information. Briefly, RNA polymerase (RNAP) enzymes recognise specific sequences of DNA called promoters. The enzyme then translocates along the DNA, transcribing the sequence to make a complementary RNA molecule. RNA messages, which provide information required by ribosomes for protein synthesis, are known as mRNA. Other RNA molecules are not translated. For example, tRNA is used to transfer amino acids and rRNA makes up part of the ribosome. In bacteria, transcription requires a single RNAP enzyme (Browning and Busby, 2004). The core enzyme consists of six subunits; β, β′, 2 α subunits, and ω. These associate with an additional subunit, σ, to form a holoenzyme with a distinctive 'crab claw' shape (Murakami, 2013) (Figure 1). The holoenzyme binds to promoter DNA to form a closed complex. The DNA then unwinds, and RNAP undergoes a conformational change, to form an open complex. Transcription then initiates.

The two largest subunits of RNAP are β and β'. These form the 2 'pincers' of the claw structure, with the β' subunit acting as a clamp which can move between an open and closed conformation (Murakami, 2013). Within the channel created by these subunits is the active site, which contains an essential catalytic Mg^{2+} ion (Murakami and Darst, 2003). The α subunits have two distinct N and C terminal domains, separated by a flexible linker. The C-terminal domain (CTD) recognises sequences upstream of the

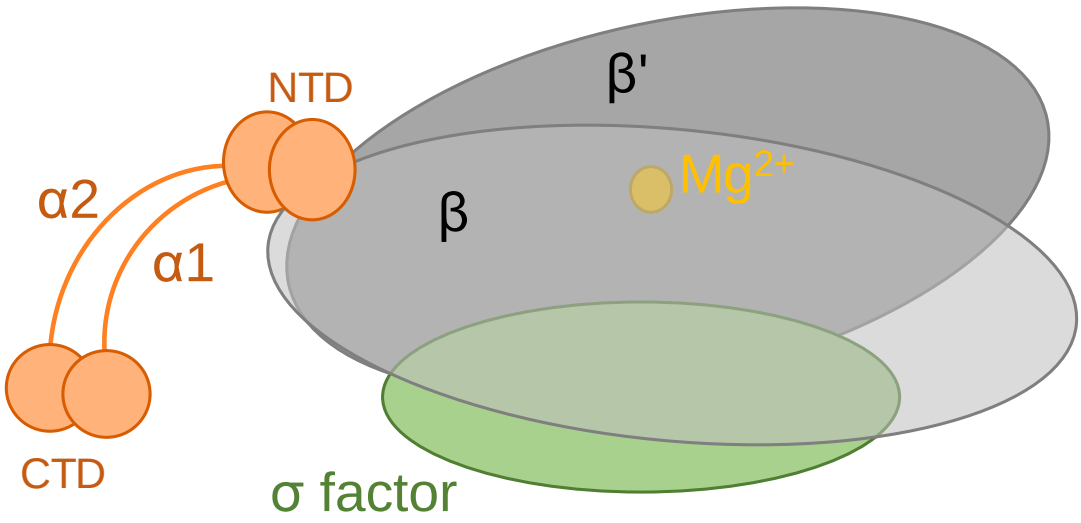


Figure 1: Structure of the RNA polymerase holoenzyme.

Cartoon illustrating the structure of bacterial RNA polymerase. The β and β' subunits which form the 'crab claw' shape are shown in grey. The active site in the channel between β and β' is represented by an Mg^{2+} ion in yellow. The σ subunit is shown in green. The α subunits are shown in orange, with the N-terminal and C-terminal domains (NTD and CTD) connected by a linker (Browning and Busby, 2004).

promoter (UP elements) and interacts with transcription factors (Browning and Busby, 2004; Murakami, 2015).The N-terminal domain (NTD) dimerises as the first step of RNAP assembly and acts as a platform for β and then β' binding (Zhang and Darst, 1998). The fifth subunit, ω, is not essential for cell viability or RNAP function. However, ω is involved in RNAP assembly, acting as a chaperone for the β' subunit (Ghosh *et al.*, 2001). There is evidence to suggest that ω is involved in the stringent response and recognition of prophage genes (Yamamoto *et al.*, 2018).

1.2. The σ factors

The RNAP σ subunit recognises and binds to promoters, which are positioned just upstream of transcription start sites (TSSs). Hence, σ directs RNAP to sites of transcription. The σ factor also facilitates promoter unwinding which triggers open complex formation (Browning and Busby, 2004). There are many different bacterial σ factors that respond to different cellular conditions. Each σ factor recognises a different subset of promoters, known as its regulon, enabling widespread changes in gene expression in response to specific stimuli (reviewed in Paget, 2015). Bacterial σ factors are divided in to two classes depending on whether they are structurally similar to *Escherichia coli* σ^{70} or σ^{54}. The σ^{54} factor is distinct in that it requires energy from ATP hydrolysis, via ATPase transcription factors, to form a transcriptionally active RNAP open complex (Rappas *et al.*, 2006). In contrast, the σ^{70}/RNAP holoenzyme is able to spontaneously isomerise into the open complex state (Feklistov and Darst, 2011).

The σ^{70} family of σ factors is subdivided into 4 groups based on domain structure as illustrated in Figure 2. Most gene expression at log phase is controlled by Group 1 σ

factors such as σ^{70}, which contain a unique domain σ1.1 (Murakami and Darst, 2003). This domain is required for autoinhibition; it compacts free σ factor to stop binding to the DNA before association with core RNAP (Schwartz *et al.*, 2008). Once σ binds the core enzyme, σ1.1 sits in the DNA binding channel of RNAP and must be removed for open complexes to form (Murakami, 2013).

Domain σ2 is highly conserved and is present in all 4 σ factor groups. This domain contains several key regions.

- Region 2.2, which forms an interface with the β' clamp (Murakami *et al.*, 2002).
- Region 2.4, which recognises a promoter sequence approximately 10 bases upstream of the transcription start site, known as the -10 element (Murakami and Darst, 2003).
- Region 2.3 facilitates promoter melting at the -10 element (Feklistov and Darst, 2011).
- Region 1.2, present in domain σ2 of the Group 1 and Group 2 σ factors, recognises a 'discriminator sequence with a '5′-GGG-3" motif downstream of the -10 element (Feklistov *et al.*, 2006; Paget, 2015; Zhang *et al.*, 2012).

Domain σ3 is missing from Group 4 σ factors. In Groups 1 to 3, this domain interacts with the DNA bases at positions -13, -14 and/or -15. Promoter sequences which interact with σ3 are known as extended -10 elements (Becker and Hengge-Aronis, 2001; Mitchell *et al.*, 2003). The σ4 domain contains a helix-turn-helix motif in region 4.2 which interacts with the DNA between positions -30 and -38, at a site known as the -35 element (Campbell *et al.*, 2002).

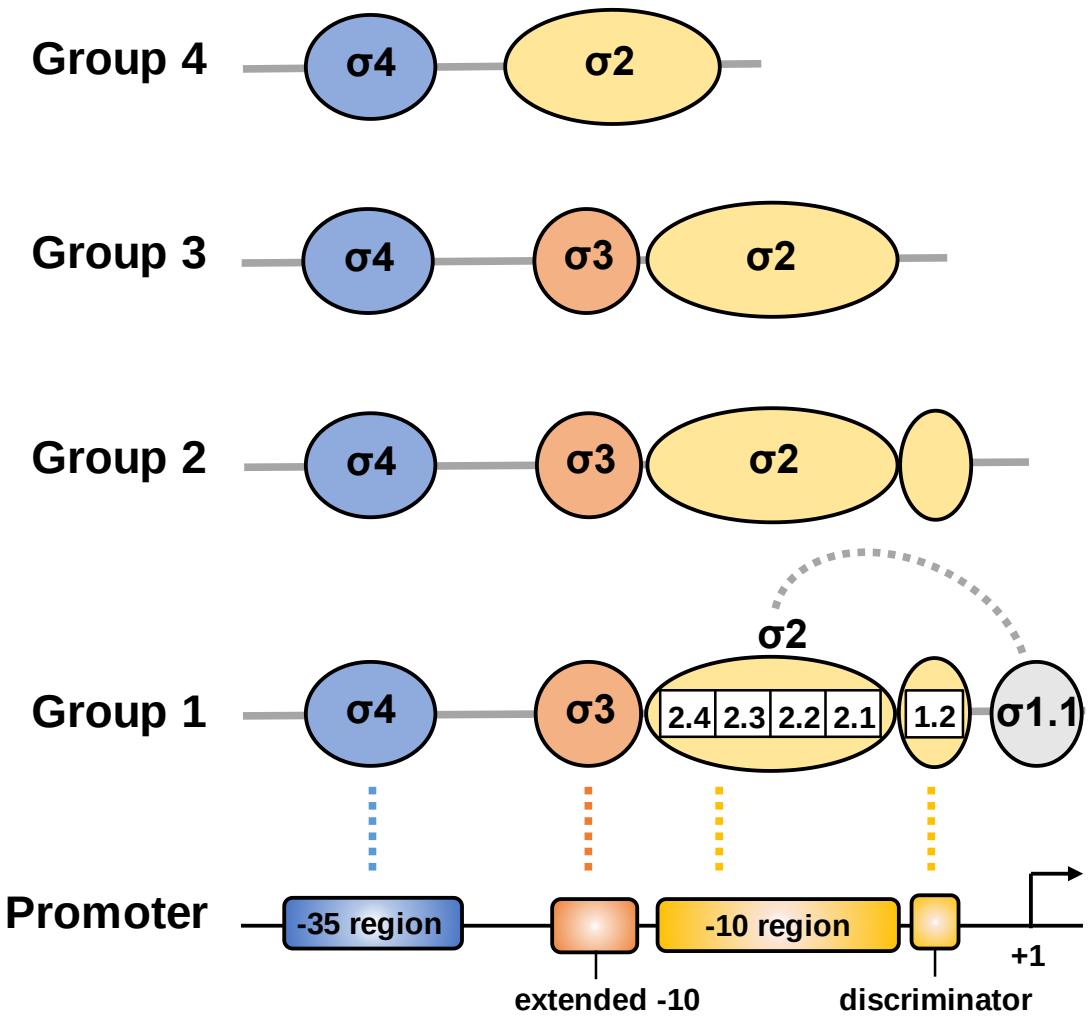


Figure 2: Comparison of σ factor domains.

Diagram showing which protein domains are present in each group of σ factors. Domains are represented as ovals. Target promoter sequences are shown as coloured rectangles. Domain σ4 (blue) interacts with the promoter -35 region. Domain σ3 (orange) interacts with an extended -10 element. The σ2 domain (yellow) is separated into regions, shown in rectangles. Region 2.4 recognises the -10 element. Bases in the promoter -10 element interact with σ2 region 2.3 during promoter melting. Region 1.2 interacts with a downstream discriminator sequence. σ1.1 (grey) interacts with other domains to inhibit σ binding to DNA (Paget, 2005).

1.2.1. RpoS (σ38)

Group 2 σ factors are structurally similar to Group 1 factors but lack the σ1.1 domain and are non-essential (Paget, 2015). This group includes σ^{38}, the general stress response σ factor, encoded by the gene *rpoS*. At high protein levels, σ^{38} outcompetes σ^{70} and binds to RNAP to induce the stress response. The RNAP/σ^{38} complex is responsible for the expression of up to 10 % of all *E. coli* genes, some of which are only expressed in specific conditions (Weber *et al.*, 2005). For example, the gene *csiD* is only transcribed during nutrient starvation when both σ^{38} and the activator cAMP receptor protein (CRP) are present (Metzner *et al.*, 2004).

Levels of σ^{38} are influenced by stimuli including cell density, pH, temperature and osmolarity (Hengge-Aronis, 2002). A summary of σ^{38} regulation is illustrated in Figure 3. During log phase *rpoS* is expressed at low levels, with expression increasing at slower growth rates and as cells enter stationary phase (Lange and Hengge-Aronis, 1994). However, the *rpoS* mRNA is self-inhibitory, due to formation of a stem loop in the 5' region which blocks the ribosome binding site. Removal of the stem loop requires the action of one of three sRNAs (DsrA, RprA, ArcZ) which are each expressed in response to a different stress. A cofactor Hfq recruits the sRNA to the *rpoS* mRNA, opening the stem loop and initiating translation (Soper *et al.*, 2010; Soper and Woodson, 2008).

Proteolysis of σ^{38} is also regulated, via the two-component phosphorelay system ArcB/RssB (Mika and Hengge, 2005). The response regulator RssB binds σ^{38} and targets it for degradation by the protease ClpXP (Zhou *et al.*, 2001). ArcB phosphorylates RssB in response to changes in cellular energy levels, but RssB is still able to promote σ^{38} degradation independently of its phosphorylation status (Peterson

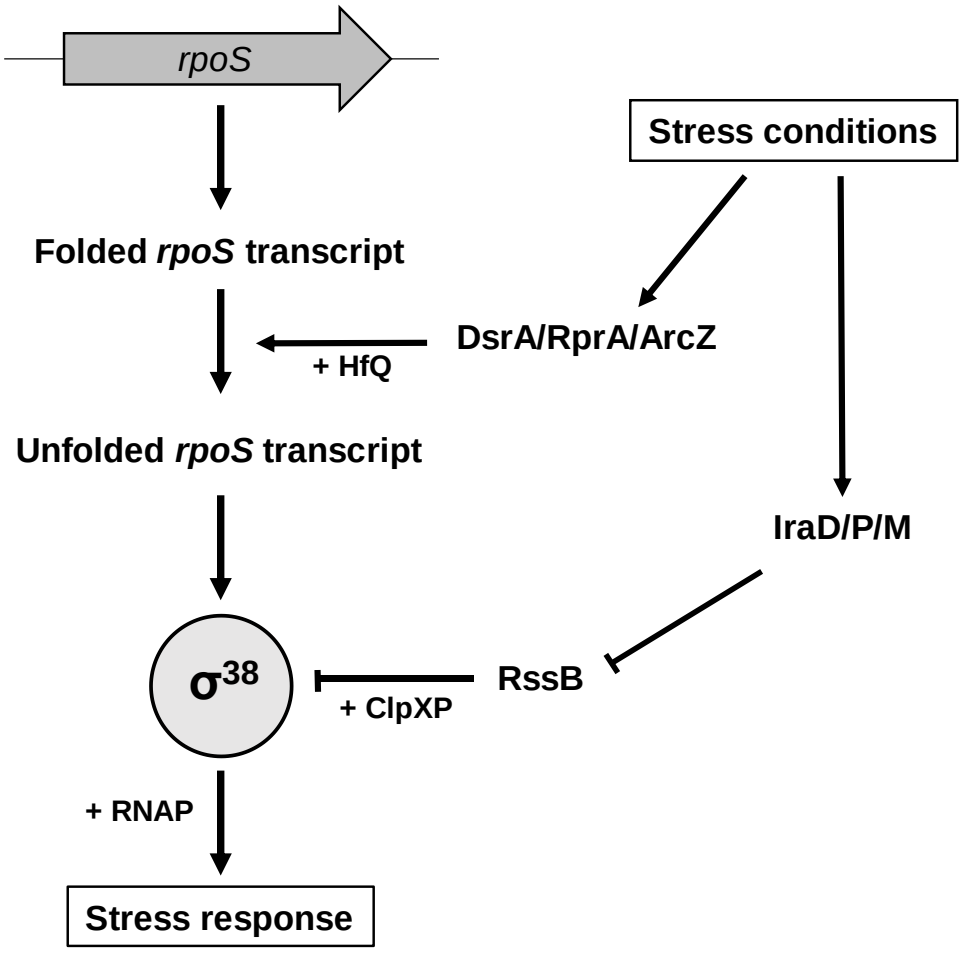


Figure 3: Regulation of σ^{38} protein levels.

Schematic showing the pathways responsible for controlling σ^{38} protein levels in response to stress. The σ^{38} protein is shown as a grey circle. Pointed arrows indicate activation and flat arrows show an inhibitory interaction. The *rpos* mRNA is folded to inhibit translation. Protein which is translated is bound by RssB and directed to the ClpXP protease for degradation. When the cells undergo stress, the expression of sRNAs DsrA, RprA and ArcZ is increased. These interact with Hfq and unfold the mRNA to allow translation. Stress also induces expression of anti-RssB proteins IraD, IraP or IraM which rescue σ^{38} from degradation (Hengge-Aronis, 2002).

et al., 2004). One of three 'anti-adaptor' proteins, IraP, IraD or IraM, are required to stop RssB-mediated degradation of σ^{38}. These proteins sequester RssB and block binding to σ^{38}. Each adaptor is induced in response to a different stress; IraP responds to phosphate starvation (Bougdour *et al.*, 2006), and IraM is induced in low magnesium conditions (Bougdour *et al.*, 2008). IraD responds to DNA damage and is induced in stationary phase (Merrikh *et al.*, 2009).

1.2.2. Group 3 and 4 σ factors

Groups 3-4 contain further alternative σ factors. One of the most widespread Group 3 σ factors is σ^{28}, which has related σ factors present in all motile bacteria (Paget, 2015). These σ factors are required for the transcription of genes encoding flagellar proteins (Helmann, 1991). This is because σ^{28} acts as a gatekeeper between formation flagellum base and the flagellin filament. Before the base is made, a protein FlgM maintains low levels of σ^{28}. The completed flagellar base is able to export FlgM out of the cell, stimulating σ^{28} expression, which initiates transcription of the flagellin gene (Brown and Hughes, 1995). Another Group 3 σ factor, σ^{32}, responds to heat shock, when cellular proteins become damaged or misfolded (Storz and Hengge, 2010, pp.93–94). Subsequently, σ^{32} induces expression of heat shock proteins. Such proteins include the protein chaperones DnaK and GroE which assist in reassembly of misfolded proteins (Yura *et al.*, 1993). DnaK also negatively regulates the heat shock response, by binding to free σ^{32} and promoting degradation (Liberek *et al*., 1992). At high temperatures, this interaction is abolished and σ^{32} is free to associate with RNAP (Chattopadhyay and Roy, 2002). Other genes in the σ^{32} regulon encode proteases such as Lon, which degrade irreparable proteins (Yura *et al.*, 1993).

The Group 4 σ factors are a large and diverse group of small proteins, consisting of only the σ2 and σ4 regions. They are also known as extracytoplasmic function (ECF) σ factors due to their ability to act as effector molecules in response to signals from outside the cell (Lonetto *et al.*, 1994). Group 4 σ factors recognise unique promoter elements with conserved 'AAC' and 'CGT' motifs in the -35 and -10 regions respectively (Staroń *et al.*, 2009). *E. coli* σ^{24} is also involved in the heat shock response, like σ^{32}, but specifically reacting to misfolded proteins in the cell membrane (Mecsas *et al.*, 1993). Another *E. coli* ECF σ factor is FecI, which is required for the Fec iron transport system (Braun and Mahren, 2005). In other organisms, ECF σ factors are involved in diverse stress responses (Staroń *et al.*, 2009).

1.3. Promoter sequences

1.3.1. Promoter elements

The DNA sequences recognised by σ factors are known as promoter elements. As the majority of *E. coli* promoters are σ^{70} dependent, the sequences recognised by σ^{70} are the best defined. This σ factor recognises a highly conserved -10 element with the consensus sequence 5'-T_{-12}ATAAT$_{-7}$-3' and a -35 element (consensus 5'-T_{-35}TGACA$_{-30}$ -3') (Figure 4) (deHaseth *et al.*, 1998). The bases T_{-7}, T_{-12} and A_{-11} are the best conserved positions in the -10 element and are key for promoter melting (Feklistov and Darst, 2011). The -35 sequence is poorly conserved compared to the -10 element, with the positions T_{-35} T_{-34} being present most frequently. Approximately 20 % of *E. coli* σ^{70} promoters have extended -10 elements; a 5′-T_{-15} G_{-14} -3′ motif (Mitchell *et al.*, 2003). Promoters with an extended -10 element are often more active and form stable complexes with RNAP even in the absence of a -35 element (Campbell *et al.*, 2002).

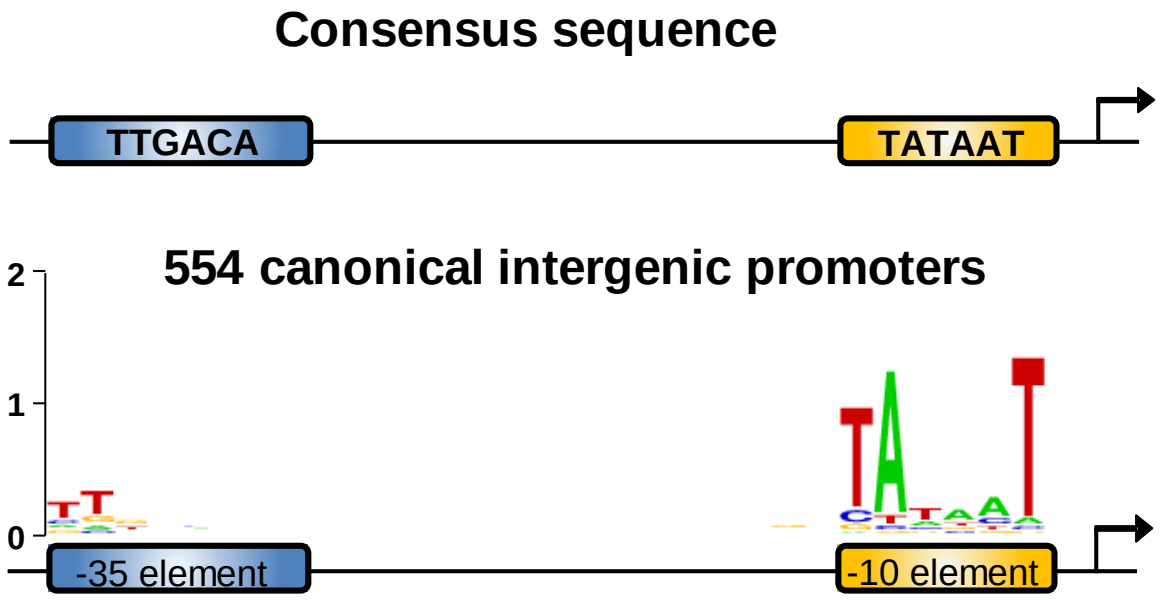


Figure 4: Sequence logo from *E. coli* σ^{70} promoters.

554 canonical σ^{70} promoters, identified by Mitchell *et al*. (2003), were aligned to create a sequence logo. The logo shows a highly conserved -10 element (yellow box) and a less well conserved -35 element (blue box) (adapted from Singh *et al*., 2011).

Promoters also frequently contain UP elements; an AT rich sequence upstream of the -35 element, which interacts with the RNAP α-CTD (Busby and Ebright, 1994; Gourse *et al.*, 2000). Another common feature of promoter sequences is the base at the TSS (position +1), which is an A or a G at 60 % of *E. coli* promoters (Salgado *et al.*, 2013).

Group 2 σ factors share similar structures with σ^{70}, particularly in domain σ2, and thus recognise very similar promoter elements (Paget, 2015). Factor σ^{38} targets the same -35 consensus sequence as σ^{70}, although this element is less conserved at σ^{38}-dependent promoters. The σ^{38} protein also recognises an almost identical -10 element (5'-TATACT-3') to σ^{70} (5'-TATAAT-3') (Becker and Hengge-Aronis, 2001). An additional determinate for σ^{38} binding is a C at position -13, which interacts with a σ^{38}-specific charged residue in domain σ3 (Becker and Hengge-Aronis, 2001). In general, σ^{38} is also better able to tolerate both deviations from promoter element consensus and different spacer lengths. These features decrease overall promoter activity, but selects against σ^{70} binding (Typas *et al.*, 2007; Typas and Hengge, 2006).

1.3.2. The spacer sequence

The promoter -10 and -35 elements are usually spaced 17 bp apart, which creates the best configuration for optimum promoter activity (Aoyama *et al.*, 1983; Mulligan *et al.*, 1985). However, spacer length can vary between 15 and 19 bp whilst maintaining promoter functionality (Mitchell *et al.*, 2003). Despite the general view that spacer sequence is not important for promoter activity, there is a degree of sequence conservation in this region (Singh *et al.*, 2011). Furthermore, a growing body of evidence suggests that spacers with a high AT-content increase promoter activity (Singh *et al*., 2011). This could be because promoters with AT-rich spacer sequences have a higher degree of bending, influencing their interaction with the σ factor (Hook-

Barnard and Hinton, 2009; Singh *et al.*, 2011). When promoter DNA is bound to RNAP, the spacer region is positioned to a cluster of amino acids which make up part of the linker between domains σ2 and σ3. This cluster includes an arginine R451. Singh *et al.* (2011) showed that R451 is responsible for the stimulatory effect of a T at position -18 in the *cbpA* promoter, suggesting a previously undescribed σ/promoter interaction.

1.4. Transcription factors

Transcription factors are DNA binding proteins which target promoters to stimulate or hinder transcription initiation and thus regulate gene expression. There are over 300 transcription factors in *E. coli* (Pérez-Rueda and Collado-Vides, 2000). Each recognises and binds to a specific DNA sequence found close to the promoters of target genes. These genes form the transcription factor's regulon, which can include one to over 100 genes (Martínez-Antonio and Collado-Vides, 2003). Some transcription factors can act as both activators and repressors depending on the promoter to which it is bound (Browning and Busby, 2004). Additionally more than half of genes are regulated by multiple transcription factors, resulting in a complex and interlinked regulatory network that tailors gene expression to the cell's requirements (Martínez-Antonio and Collado-Vides, 2003). This network responds to environmental conditions. For example, some transcription factors are response regulators; part of two-component phosphotransferase systems. Response regulators only function when phosphorylated by their corresponding histidine protein kinase, in response to external stimuli (Gao *et al.*, 2007). Other regulators are bound directly by small ligands acting as signals (Browning and Busby, 2004). Mechanisms of transcription factor activation are summarised in Table 1 and illustrated in Figure 5 (adapted from Browning and

Busby, 2004; Lee *et al.*, 2012). Mechanisms of repression are in Table 2 and Figure 6 (Browning and Busby, 2004; Rojo, 1999).

1.4.1. Nucleoid Associated Proteins

Nucleoid associated proteins (NAPs) are involved in maintaining chromatin structure and can also function as transcription factors (Dorman *et al.*, 2020). The spatial distribution of NAP binding during different growth phases alters the structure and superhelicity of the chromosome to modulate gene expression (Blot *et al.*, 2006; Sobetzko *et al.*, 2012). In this way, NAPs function as global regulators. Some NAPs act in a similar way to eukaryotic histones; they bind and condense DNA, restricting access of RNAP to the promoter and thus inhibiting transcription initiation (Arold *et al.*, 2010). NAPs can also have a positive regulatory effect on transcription. For instance, via the DNA distortion method of transcription activation (Table 1) (Dame, 2005; Dorman and Dillon, 2010). The main *E. coli* NAPs IHF, HU, and Fis are described briefly here. The histone-like nucleoid structuring protein (H-NS) is discussed in more detail in sections 1.7. -1.9.

IHF (integration host factor) and HU (heat-unstable nucleoid protein) are related DNA binding proteins with similar functions. IHF recognises an AT-rich consensus sequence and bends the DNA ~160° upon binding. Hence, IHF regulates gene expression by bringing RNAP/promoter complexes closer to distally bound transcription factors (Arfin *et al.*, 2000; Santero *et al.*, 1992). IHF can also directly recruit σ^{54}/RNAP holoenzyme to promoters (Macchi *et al.*, 2003). Furthermore, bending induced by IHF can introduce DNA loops and compact the chromosome (Swinger and Rice, 2004). HU does not recognise specific sequences but is preferentially recruited to distorted or damaged

Table 1: Mechanisms of transcription activation

Type	Mechanism	Example
Class I	The transcription factor binds upstream of the -35 element and recruits RNAP via the α-CTD	CRP at the *lac* promoter (Busby and Ebright, 1999)
Class II	The transcription factor binding site overlaps the -35 element, interacts with RNAP at domain σ4, and/or the α subunit	CRP at the *galP1* promoter (Busby and Ebright, 1999), the AraC family (Martin and Rosner, 2001)
Class III	More than one transcription factor interacts independently with RNAP	CRP at the *acs* promoter (Beatty *et al.*, 2003)
DNA distortion	The transcription factor changes conformation of the promoter to facilitate promoter element/σ factor interactions	The MerR family (Brown *et al.*, 2003)

Table 2: Mechanisms of transcription repression

Type	Mechanism	Examples
Steric hindrance	Binds core promoter elements to occlude RNAP, or prevent promoter clearance	LacI at the *lac* promoter (Schlax *et al.*, 1995)
DNA looping	Binds upstream of promoters and causes the DNA to loop, blocking access to the promoter	AraC at the *araB* promoter in the absence of arabinose (Lobell and Schleif, 1990)
Anti-activation	Directly interacts with activators to stop activation	CytR (Valentin- Hansen *et al.*, 1996)
Anti-σ	Binds σ factors (usually alternative σ factors) to block them from binding core RNAP	anti-σ^{70} Rsd, anti-σ^{24} RseA (Treviño-Quintanilla *et al.*, 2013)

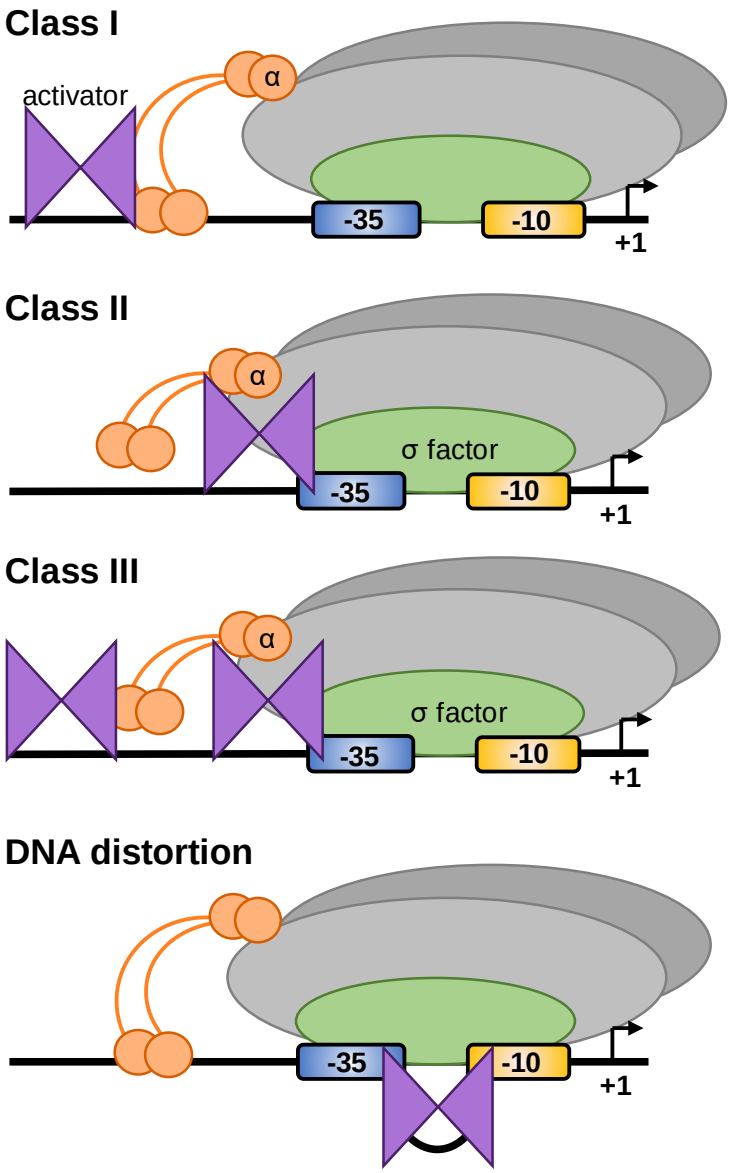


Figure 5: Methods of activation at bacterial promoters.

Activators are shown in purple. The -35 and -10 elements are shown as blue and yellow boxes. Class I activators bind upstream of the promoter and interact with the α subunits of RNAP (orange) to stimulate transcription. Class II activators bind at sites overlapping the -35 element and interact with the α subunits and/or the σ factor (green). Class III activation occurs when two activators function independently at the same promoter. DNA distortion activators bend the DNA (black lines) to optimise the positioning of the promoter elements and RNAP (Browning and Busby, 2002).

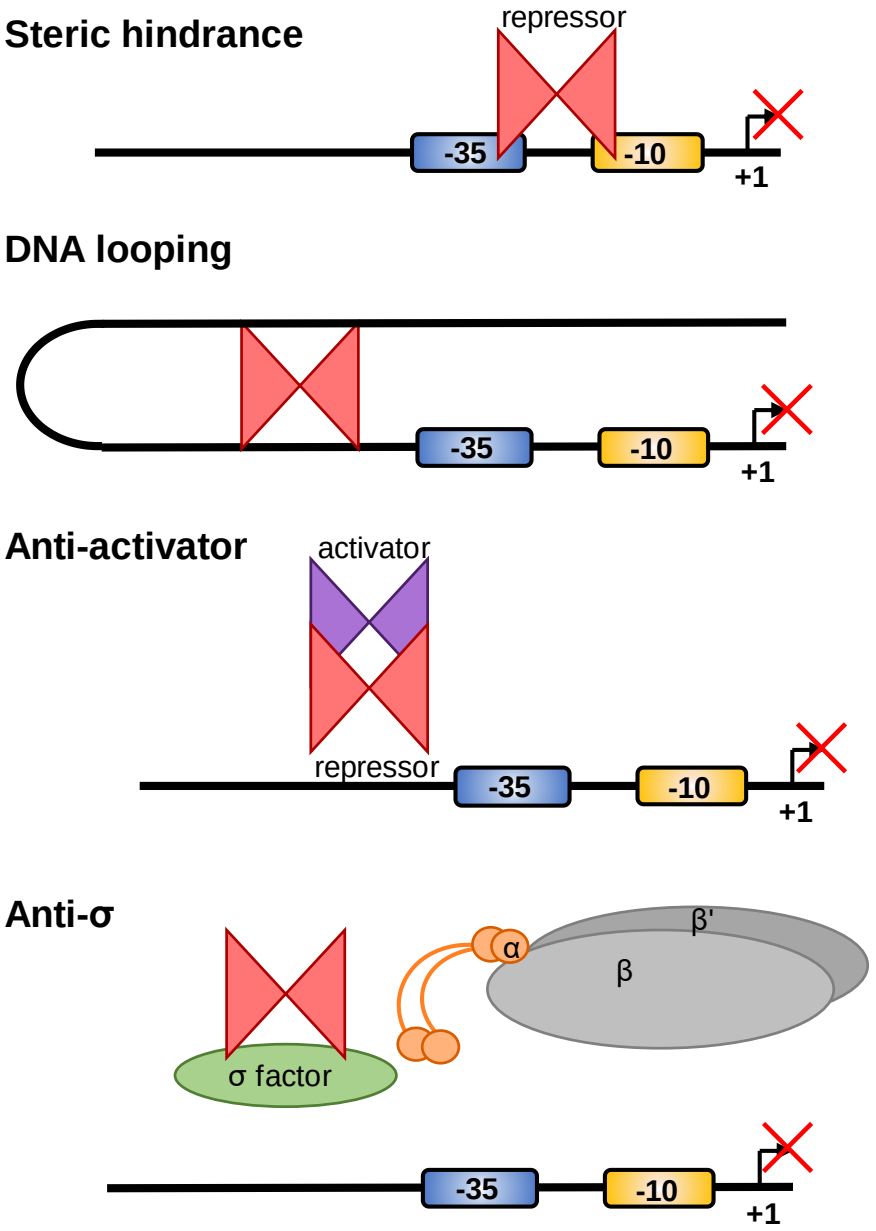


Figure 6: Methods of repression at bacterial promoters.

Repressor proteins are shown in red and activators in purple. The -35 and -10 elements are shown as blue and yellow boxes. Repressors bind to promoter elements (steric hindrance) or introduce DNA loops to occlude RNAP. Anti-activators bind directly to essential activators, preventing them from binding to promoters. Anti-σ factors bind to the σ factor to block binding to core RNAP, inhibiting expression of all genes in the regulon of that σ factor (Browning and Busby 2002)

DNA, suggesting a role in DNA repair. The DNA bending introduced by HU ranges from ~105° to 140° (Swinger *et al.*, 2003). Via this mechanism HU has a role in compacting the chromosome. HU also interacts with topoisomerase to induce negative supercoils (Rouvière-Yaniv *et al.*, 1979). As a transcription factor, HU is a global regulator of genes involved in the SOS, osmolarity, aerobic and acid stress responses (Oberto *et al.*, 2009).

The Fis protein binds to 17 bp AT-tracts with a G and a C at positions 2 and 16 respectively (Cho *et al.*, 2008). Again, this NAP can bend the DNA to promote chromosome compaction and regulate transcription by a number of mechanisms. At the *dps* promoter, Fis traps RNAP at the promoter to repress *dps* expression (Grainger *et al.*, 2008). At the *proP* promoter, Fis functions as a Class II activator; it binds to a site overlapping the -35 element and interacts with the α-CTD (McLeod *et al.*, 2002).

1.5. Transcription initiation

Transcription initiation occurs when RNAP is bound at a promoter and moves from a closed complex to an open complex. This conformational change depends on a stable closed complex first being formed. This involves RNAP interacting with both the -10 and -35 elements. An exception to this is when a promoter has an extended -10 element, which facilitates sufficient complex stability to negate the requirement for a -35 sequence (Campbell *et al.*, 2002). UP elements and transcription factors can also assist RNAP binding (Busby and Ebright, 1994). Once RNAP is bound, promoter melting occurs, where the dsDNA unwinds at the -10 element. This triggers the isomerisation of RNAP from the closed to open complex (Figure 7). The first base to

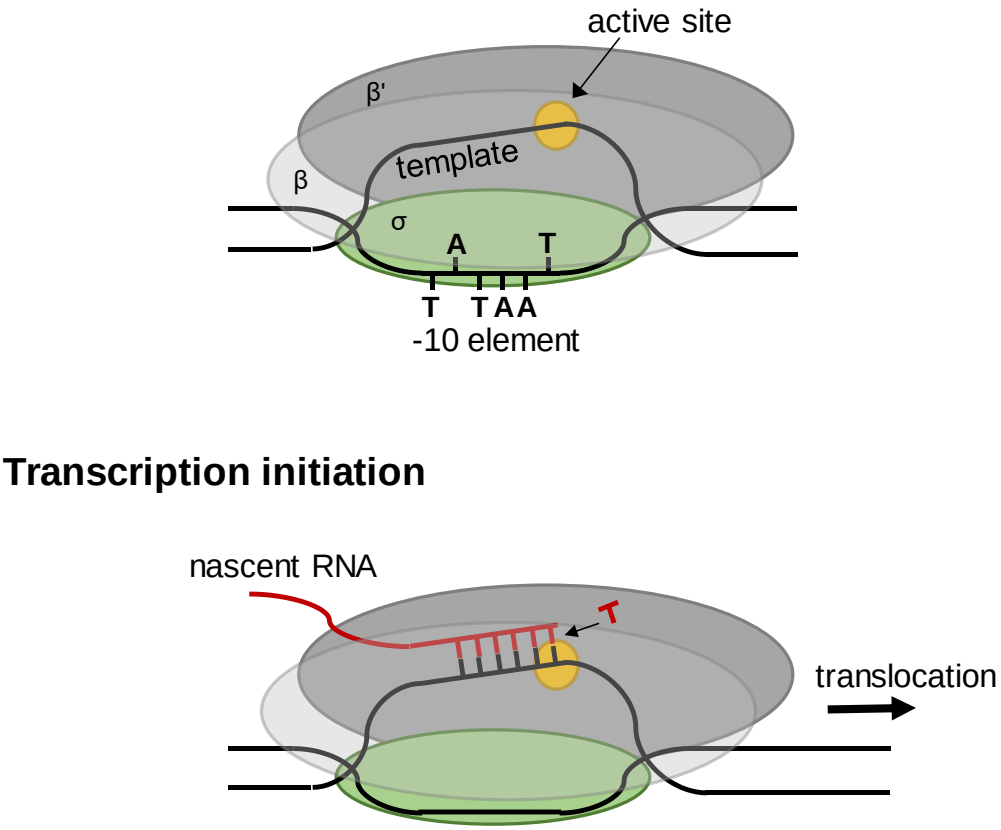


Figure 7: Processes of transcription initiation.

Diagrams showing how RNAP moves from a closed to open complex and then initiates transcription. RNAP adopts an open conformation when the promoter DNA (black line) melts and forms a transcription bubble. The A_{-11} and T_{-7} of the promoter non-template strand flip out of the DNA into the σ factor (green). The template strand moves into the active site (yellow). Transcription initiates and the RNAP translocates down the DNA. Meanwhile, nucleotides are recruited to the growing RNA chain (red line) at the active site (Feklistov and Darst, 2011).

unwind is at position -11, then each base of the -10 element follows to form a transcription bubble (Chen *et al.*, 2020, 2010). The open complex is stabilised by interactions between σ region 2.3 and bases A_{-11} and T_{-7} on the non-template strand. These nucleotides flip out of the DNA backbone into protein pockets containing aromatic amino acids (Chen *et al.*, 2020; Feklistov and Darst, 2011; Zhang *et al.*, 2012). T_{-12} does not unwind but is important for promoter recognition, making extensive van der Waals interactions, polar bonds and hydrogen bonds with σ (Feklistov and Darst, 2011). As the transcription bubble continues to unwind up to the +2 position, the DNA template strand moves into the RNAP channel towards the active site and RNA synthesis begins. The first rounds of transcription are abortive; downstream DNA is pulled into the channel via a 'scrunching' mechanism and accumulates as RNA is synthesised (Kapanidis *et al.*, 2006). Once 8-15 nt of RNA is made, transcription is either aborted, with RNAP returning to the promoter to repeat the cycle, or moves to the elongation phase (Goldman *et al.*, 2009). During promoter escape, RNAP breaks its contacts with the promoter DNA. Then σ factor generally dissociates from RNAP to allow space for the DNA/RNA hybrid (Vassylyev *et al.*, 2007). However in some cases σ can be maintained during elongation (Hsu, 2002). RNAP facilitates elongation by translocating down the DNA 1 bp at a time, while synthesising RNA at the active site. The nascent RNA leaves the active site via an RNA exit channel, which is separated from the main channel by an RNAP lid domain (Vassylyev *et al.*, 2007).

1.6. Transcription termination

Termination of transcription occurs when the RNAP elongation complex stops translocating along the DNA, halts RNA synthesis and dissociates from both the DNA

and RNA (reviewed in Peters *et al*., 2011; Ray-Soni *et al*., 2016; Roberts, 2019). This is an important regulatory step; uncontrolled elongation would result in unnecessary transcription of downstream, non-coding DNA, and an increased risk of collisions with DNA replication machinery. It is also important that RNAP is recycled promptly as the number of available RNAP enzymes is limited (Piper *et al.*, 2009). There are two pathways of transcription termination, intrinsic and factor-dependent (Figure 8). Intrinsic termination occurs when the RNA in the exit channel forms a hairpin secondary structure. Factor-dependent termination relies on a protein factor, usually Rho, which binds nascent RNA. Both pathways are more effective with paused RNAP complexes, which form when the RNAP encounters a consensus pause site in the DNA sequence (Larson *et al.*, 2014; Ray-Soni *et al.*, 2016).

1.6.1. Intrinsic termination

To terminate transcription independently of Rho, a specific sequence of RNA forms a hairpin in the RNAP exit channel. The sequence consists of a symmetrical GC-rich region followed by tract of 7-9 uracils (Carafa *et al.*, 1990). As the last 3' uracil is incorporated into the nascent RNA, the RNAP pauses as the hairpin forms (Ray-Soni *et al.*, 2016). Hairpin formation destabilises the elongation complex, breaking RNAP/RNA contacts and unwinding part of the DNA/RNA hybrid (Gusarov and Nudler, 1999). Complexes are further destabilised, either by hypertransclocation (RNAP continues to move down the DNA but does not add new RNA NTPs), or by hybrid shearing (RNA pulls away from the DNA/RNA complex) (Larson *et al.*, 2008). Both pathways open the RNAP channel resulting in the collapse of the transcription bubble. RNAP then dissociates from the DNA, which returns to a duplex (Roberts, 2019).

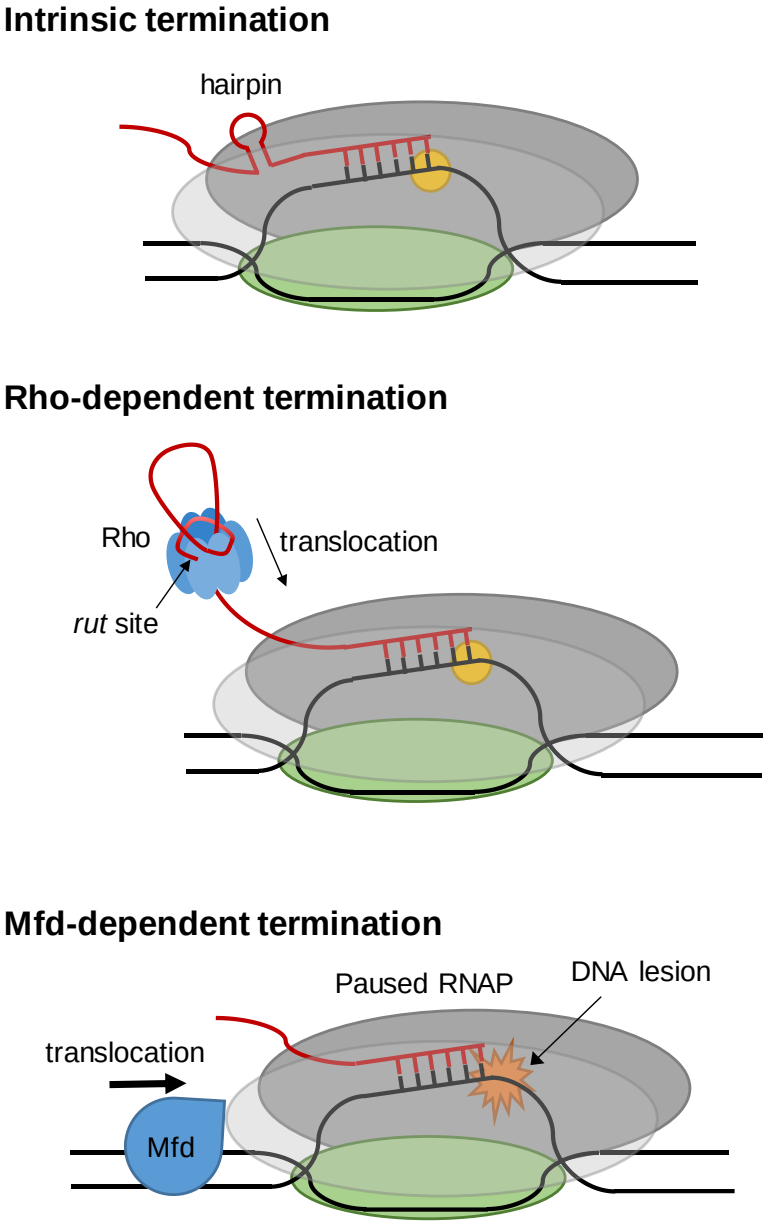


Figure 8: Mechanisms of transcription termination.

In intrinsic termination, a hairpin is introduced into the nascent RNA (red line). Rho-dependent termination occurs when the protein (blue hexamer) is recruited to an RNA *rut* site. Rho translocates down the DNA toward RNAP. Mfd (shown in blue) facilitates termination as part of the transcription-coupled repair pathway. It binds and moves along the DNA. When it encounters paused RNAP elongation complexes it can promote termination (Ray-Soni *et al*., 2016).

1.6.1. Factor-dependent termination

At termination sites lacking an intrinsic terminator sequence, a protein called Rho is required. Rho-dependent termination accounts for at least 20 % of termination events in *E. coli*. Rho is particularly important for terminating non-canonical transcripts, for example those derived from antisense (Peters *et al.*, 2012) or horizontally acquired promoters (Cardinale *et al.*, 2008). The Rho protein oligomerises into a ring-shaped hexamer with ATPase activity (Mitra *et al.*, 2017). Rho has 3 functional domains, a primary RNA binding site at the N terminal, a secondary RNA binding site and the ATP-binding catalytic site (Mitra *et al.*, 2017). The primary RNA binding site is responsible for Rho recruitment to the nascent RNA as it leaves the RNAP exit channel. Rho targets a C-rich, single stranded RNA sequence known as a *rut* (*R*ho *ut*ilisation) site (Chen and Richardson, 1987; Morgan *et al.*, 1985). However, the primary feature of RNA recognised by Rho is the absence of a ribosome. In bacteria, transcription and translation are coupled, so nascent RNA is usually immediately bound by the ribosome. Actively translated RNA is therefore protected from termination by Rho (Richardson, 2003). Hence, Rho functions to terminate transcription when the ribosome has stopped translating, e.g. at the end of an operon, or when the transcript is non-coding. Transcription which is not coupled to translation must be quickly terminated because free RNA is able to interact with its template DNA. This results in the formation of toxic R-loops, a structure which blocks DNA replication and promotes dsDNA breaks (Raghunathan *et al.*, 2018).

Once the *rut* site is bound to the Rho primary binding domain, the RNA is pulled through the core of the hexamer and binds to the secondary domain. At this point ATP is recruited to the active site and the ring structure isomerises to a closed state

(Thomsen *et al.*, 2016). The energy from ATP allows Rho to translocate down the RNA towards RNAP. Throughout translocation Rho remains tethered to the *rut* site, and the upstream RNA is pulled through the Rho channel (Figure 8) (Koslover *et al.*, 2012). When Rho reaches RNAP, it triggers transcription termination by an unknown mechanism. The force of the translocation may result in RNA being pulled out of the RNAP complex (Roberts, 2019).

A DNA repair protein, Mfd, can also act as a termination factor (Figure 8). Mfd binds DNA and acts as a translocase, using ATP for energy. As it moves along the DNA Mfd encounters paused RNAP complexes and either rescues elongation or terminates transcription (Le *et al.*, 2018). This is an example of transcription-coupled repair, a mechanism by which elongating RNAP can scan the DNA for damage. By pausing at DNA lesions, RNAP acts as a marker for sites requiring repair (Savery, 2007). The mechanism of termination by Mfd is unclear; Mfd translocation force may be sufficient to dislodge RNAP (Roberts, 2019), or Mfd might act in parallel with Rho (Jain *et al.*, 2019).

1.6.2. NusG and NusA

NusG is a protein factor involved in transcription elongation and termination. The NusG/Spt5 family is universally conserved in all life forms (Tomar and Artsimovitch, 2013). NusG has an important role in promoting forward translocation of the elongating RNAP complex and reducing pauses, by binding RNAP via a domain at the NusG N terminus (Herbert *et al.*, 2010). The NusG CTD interacts with the ribosome to facilitate transcription/translation coupling and occlude Rho from the *rut* sites (Burmann *et al.*,

2010). However, the NusG CTD can bind Rho instead of the ribosome to positively regulate Rho-dependent termination. This is particularly helpful at sites with a less C-rich *rut* sequence, where NusG helps Rho encircle the RNA and isomerise into a closed ring state (Lawson *et al.*, 2018).

NusA can also act as both a terminator and an anti-terminator. This protein binds to the RNA exit channel to stimulate pausing and intrinsic termination, by promoting hairpin folding and then stabilising its interaction with the RNAP flap domain (Guo *et al.*, 2018; Ray-Soni *et al.*, 2016). Conversely, NusA also inhibits Rho-dependent termination. The RNA sequences targeted by NusA often overlap *rut* sites, resulting in a direct competition for binding (Qayyum *et al.*, 2016).

1.7. Structure of H-NS

H-NS is a NAP which binds selectively to AT-rich DNA. The protein is 137 amino acids in length and contains two functional domains; the NTD, responsible for oligomerisation, and the CTD required for DNA binding (Figure 9) (Grainger, 2016a). The domains are connected by a positively charged flexible linker which facilitates optimum binding to DNA (Gao *et al.*, 2017). H-NS molecules can dimerise via interactions involving two different sections of the N-terminal domain. Head-to-head interactions occur at the N-terminal end of the domain and form stable coiled coils. In parallel, tail-to-tail interactions at the opposite end of the domain form a dimer by a helix-turn-helix interface. When both contacts occur simultaneously H-NS self-associates into oligomers, which bind long tracts of DNA (Arold *et al.*, 2010). The H-NS CTD contains the conserved DNA binding motif ‘TWTXGRXP’. The XGR forms an

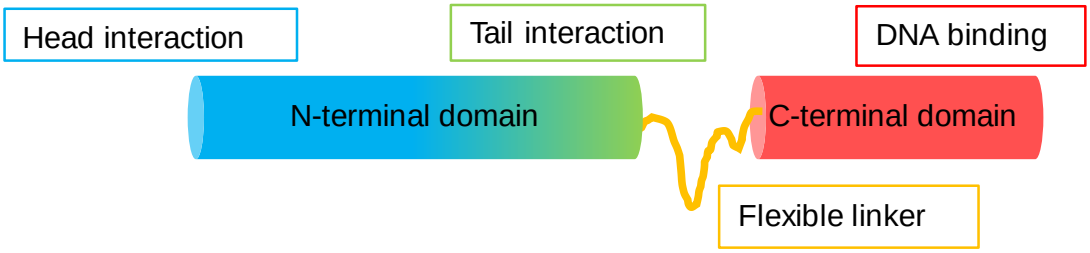


Figure 9: The structure of H-NS.

Schematic of the two H-NS protein domains. H-NS has an N-terminal domain (blue-green cylinder) for oligomerisation and a C-terminal domain (red cylinder) for DNA binding. The area of the N-terminal domain that forms head-to-head interactions is shown in blue and the tail-to-tail region is shown in green. The two domains are connected by a flexible linker (orange line) (Grainger, 2016).

AT hook that binds the DNA minor groove (Gordon *et al.*, 2011). H-NS specifically binds to AT-rich DNA because the minor groove is narrower and deeper in A-tract sequences. This favours electronegative interactions between the phosphate backbone of the DNA and the charged amino acid of the AT hook (Singh *et al.*, 2016). As such, H-NS does not recognise a specific DNA sequence but is instead recruited to a short AT-rich nucleation site with high affinity. Then, as H-NS oligomerises, it binds secondary low affinity sites along the DNA. The nucleation site often contains a central T-A base step that is necessary to increase flexibility of the DNA, for accommodation of the AT hook (Lang *et al.*, 2007). H-NS is also commonly believed to prefer curved DNA, however several studies analysing H-NS binding sites have failed to find this correlation (Singh *et al.*, 2016).

1.7.1. H-NS/DNA binding modes

The structure of H-NS in complex with DNA has not been solved, nor has the position of the CTD relative to the H-NS oligomer. However, the oligomer exists in such a way that the CTDs could be directed outwards and in opposite directions (Figure 10A) (Arold *et al.*, 2010). This would allow H-NS to simultaneously bind two DNA molecules. *In vitro* studies confirm that H-NS can form a bridged complex, linking two molecules of DNA together (Dame *et al.*, 2000). Other studies suggest that H-NS can also form linear complexes with just one DNA molecule as illustrated in Figure 10B (Amit *et al.*, 2003). Both modes are predicted to exist *in vivo,* depending on conditions within the cell. Linear binding is dominant when the concentration of Mg^{2+}, the temperature and the pH is low (<5 mM, <37 °C and <7 respectively). However this mode is sensitive to changes in temperature and pH, whereas the bridging mode is not. Therefore when cellular conditions change, in particular if Mg^{2+} concentrations rise, the complexes

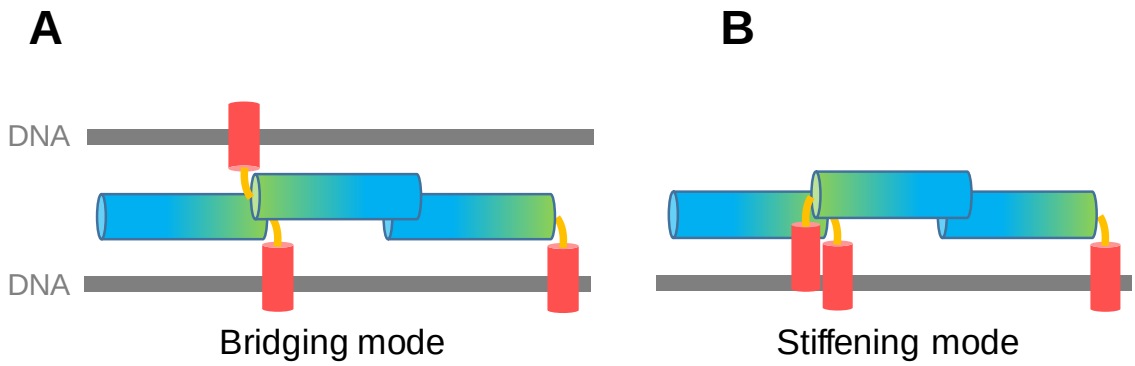


Figure 10: Structures of the H-NS/DNA complex.

Diagrams of the two possible conformations adopted by H-NS/DNA complexes. The NTD of H-NS is blue-green, CTD is red and linker is orange. DNA is shown as grey lines. A) Bridging occurs in high Mg^{2+} concentrations and involves two DNA molecules wrapped around one H-NS oligomer. B) Stiffening occurs in low Mg^{2+} concentrations and involves one linear DNA molecule that becomes stiffened and extended (Grainger, 2016).

switch to the bridging mode (Liu *et al.*, 2010). Mg^{2+} alters the affinity of H-NS for DNA in a linear fashion. This suggests that H-NS oligomers move continuously between different binding modes without dissociating from the DNA (Will *et al.*, 2018).

1.8. Functions of H-NS

1.8.1. Chromosome structure

There is a general consensus that H-NS helps to compact the genome (Dorman, 2004). Evidence from *in vivo* experiments is limited, but cells overexpressing H-NS contain highly condensed chromosomes (Spurio *et al.*, 1992). Additionally, the distance between H-NS binding patches is consistent with the boundaries between DNA loop domains, suggesting that H-NS has a role in loop formation (Noom *et al.*, 2007). An atomic force microscopy experiment by Dame *et al*. (2000) generated 3D topographic images of DNA/H-NS complexes. These images showed that H-NS condensed DNA by bridging two dsDNA molecules, which they proposed then led to the formation of highly condensed DNA foci. However, evidence from a magnetic tweezer experiment by Amit *et al*., (2003) suggests that H-NS actually makes DNA strands less compact. In this study, a magnetic bead is tethered to a surface with a strand of DNA. A magnetic field is applied to stretch the DNA molecule and the force is measured to determine the rigidity of the DNA. DNA/H-NS complexes were more extended (and therefore less compacted) than naked DNA. However, this experiment used an excess of H-NS compared to relatively low concentrations of DNA, which may not be representative of *in vivo* conditions (Dame and Wuite, 2003). Subsequently, an adapted optical tweezers study was designed with two DNA strands to account for DNA bridging (Dame *et al.*, 2006). This confirmed that H-NS interacts with two DNA

molecules, further supporting the case that H-NS bridges DNA to compact the chromosome. As both bridged and linear complexes are predicted to exist *in vivo,* it is possible that H-NS can both compact and extend the chromosome depending on cellular conditions (Liu *et al.*, 2010; Will *et al.*, 2018). Protein factors which associate with H-NS may also affect its ability to compact the DNA (see 1.9.3. ; Wang *et al*., 2014)

1.8.2. Regulation of gene expression

H-NS is a global repressor of transcription initiation (Dorman, 2004). The ability of H-NS to drastically alter the structure of DNA, into linear or bridged filaments, influences RNAP/promoter interactions in multiple ways. These mechanisms are illustrated in Figure 11. Firstly, H-NS bridging can create a DNA loop which traps RNAP at the promoter, as seen at the *rrnB* promoter (Dame *et al.*, 2002). Secondly, H-NS binding in either mode may occlude RNAP or transcription factors from the promoter (Grainger, 2016a). At the LEE5 promoter, H-NS interacts directly with the RNAP α-CTD to prevent open complex formation (Shin *et al.*, 2012). H-NS repression can also discriminate between σ factors. At the *dps* promoter, H-NS selectively displaces RNAP in complex with σ^{70}, but not σ^{38} (Grainger *et al.*, 2008). Similarly, at the *hdeAB* promoter, H-NS forms a repression loop with σ^{70} RNAP but not σ^{38} RNAP (Shin *et al.*, 2005). H-NS can also impact elongation. Bridged H-NS/DNA filaments slow the rate of elongation and stimulate RNAP pausing (Kotlajich *et al.*, 2015). As previously discussed, paused RNAP complexes are more susceptible to termination (Larson *et al.*, 2014). Post-transcription, H-NS binds to mRNA molecules with sub-optimal ribosome binding sites, and stimulates translation by correctly repositioning the ribosome (Park *et al.*, 2010). This is a rare example of H-NS activating gene expression.

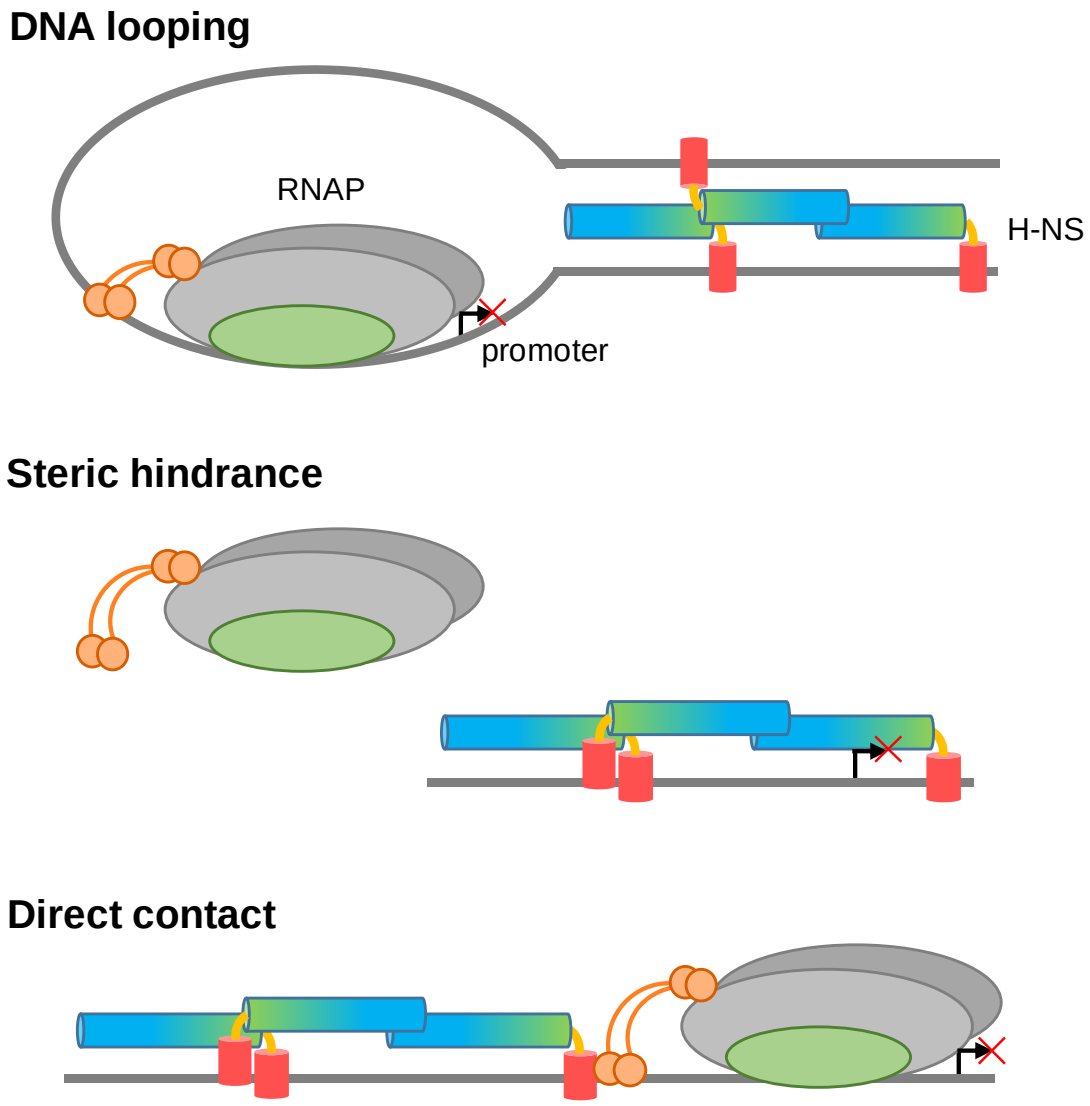


Figure 11: Mechanisms of repression by H-NS.

RNAP and H-NS molecules are represented as described earlier. Repressed promoters are shown as arrows with red crosses. Bridged DNA/H-NS complexes create DNA loops which trap RNAP at promoters and prevent elongation. Linear or bridged complexes at promoters occlude RNAP. AT some promoters, H-NS interacts directly with the RNAP to repress transcription (Grainger, 2016).

1.8.3. H-NS as a xenogeneic silencer

Many of the AT-rich genes targeted by H-NS were acquired by horizontal gene transfer (Baños *et al.*, 2009). By binding such genes, H-NS helps to integrate foreign DNA into existing gene regulatory networks. In doing so, H-NS protects the cell from expression of unnecessary or damaging gene products. H-NS also represses intragenic promoters found frequently in AT-rich DNA (Singh *et al.*, 2014). H-NS binding to these non-canonical promoters blocks RNAP recruitment and repositions RNAP to canonical promoters (Singh and Grainger, 2013). This ensures that transcriptional machinery is reserved for the expression of housekeeping genes; in Δ*hns* cells, intragenic promoters sequester RNAP resulting in a decrease in global transcription (Lamberte *et al.*, 2017). This is toxic to the cell and provides an explanation for the dramatic loss of fitness in bacteria lacking H-NS (Singh *et al.*, 2016). Thus, H-NS acts as a silencer of both horizontally acquired genes and spurious transcription while maintaining canonical gene expression.

1.9. Regulators of H-NS

Many other proteins are involved in enhancing or relieving H-NS binding at genes, to regulate H-NS repression in response to changing cellular conditions. This is important for pathogenic bacteria such as *Shigella*, where expression of virulence genes is only required during infection. Outside the host, H-NS binds and represses multiple virulence genes, including *virF* which encodes a virulence regulator. As temperature, pH and salt concentration increases upon entry to host cells the H-NS/DNA complex destabilises at the *virF* promoter. VirF is then expressed, which in turn promotes

expression of VirB, an activator at multiple virulence genes (Dorman, 2004; Tobe *et al.*, 1993).

1.9.1. Anti-silencers

VirB is not a traditional transcriptional activator as it does not interact with RNAP, but instead functions to relieve H-NS repression (Stoebel *et al.*, 2008). Many other 'anti-silencers' similarly displace H-NS at promoters. Genes for anti-silencers tend to locate to regions subject to H-NS repression; AT-rich pathogenicity islands, mobile genetic elements or bacteriophages. One example is the gene 5.5 protein of bacteriophage T7, which binds one of the H-NS dimerization sites and prevents oligomerisation. This relieves H-NS repression at other T7 bacteriophage genes (Sabrina *et al.*, 2011). Other transcription factors can be anti-silencers, particularly those with winged helix-turn-helix domains, such as ToxR in *Vibrio cholerae* and HilA/OmpR in *Salmonella enterica*. These regulators bind the DNA minor groove so are well placed to block H-NS binding (Dorman and Dorman, 2017). AraC-like transcription factors, including VirF and ToxT, are also implicated in both anti-silencing and direct activation at H-NS-bound promoters (Stoebel *et al.*, 2008).

Other H-NS anti-silencers are NAPs themselves. Ler is a protein encoded on the locus of enterocyte effacement (LEE) pathogenicity island in pathogenic *E. coli.* As a member of the H-NS family, Ler has a VGR hook motif similar to the H-NS DNA binding domain (Cordeiro *et al.*, 2011). Ler competitively binds to the same target DNA as H-NS, remodelling the DNA structure to displace H-NS and positively regulating LEE gene expression (Winardhi *et al.*, 2014). Some pathogenic *E. coli* also encode a truncated H-NS protein (H-NST) that forms hetero-oligomers with H-NS. This interaction relieves H-NS repression of LEE genes, promoting Ler-mediated

displacement of H-NS via an unknown mechanism (Levine *et al.*, 2014; Williamson and Free, 2005).

1.9.2. StpA

StpA is a paralogue of H-NS, sharing a highly similar DNA binding domain with a QGR hook motif. The exact function of StpA is unclear; it acts as an RNA chaperone (Zhang *et al.*, 1995), but also forms heteromers with H-NS (Johansson *et al.*, 2001). This association is essential for StpA survival as, without H-NS binding, a Lon protease site is exposed on StpA and StpA is rapidly degraded (Johansson and Uhlin, 1999). Chromatin immunoprecipitation (ChIP) experiments show that StpA and H-NS bind to the same AT-rich genomic sites, but the expression of most target genes is not impacted by the loss of StpA (Uyar *et al.*, 2009). Additionally, StpA does not complement loss of *hns*, but a mutation in StpA which prevented Lon protease degradation was able to restore viability in a Δ*hns* strain (Johansson and Uhlin, 1999). This suggests that StpA can perform some of the functions of H-NS if not degraded. Interestingly, StpA protein rapidly gains compensatory mutations in *Salmonella* cells lacking H-NS (Ali *et al.*, 2014). The gain-of-function mutations either affect protease degradation or the ability to silence gene expression.

1.9.3. Hha

Hha is a homologue of H-NS unique to enteric bacteria. Originally identified as a regulator of hemolysin gene expression, Hha targets multiple promoters in response to changes in osmolarity (Balsalobre *et al.*, 1999; Nieto *et al.*, 1991). Hha represses gene expression by forming complexes with H-NS (Nieto *et al.*, 2000). Heteromers assemble due to the strong electrostatic interactions which form between charged residues in Hha and the H-NS NTD. These interactions stabilise longer H-NS

oligomers, which frequently occur to repress horizontally acquired DNA (Cordeiro *et al.*, 2011). Consequently, Hha is required for H-NS to function as a xenogeneic silencer, but not for H-NS binding at core promoters (Baños *et al.*, 2009). Hha/H-NS heteromers may also contribute to chromosome condensation. Singh *et al.* (2016) propose an 'intrabridging' structure, where H-NS and Hha form an extended polymer in a corkscrew shape that is able to bind a single DNA molecule on both sides. Evidence from an optical tweezer experiment supports this as a potential mode of DNA binding, which is unaffected by changing Mg^{2+} concentration (Wang *et al.*, 2014).

1.9.4. Control of σ^{38} levels by H-NS

H-NS promotes RssB-mediated degradation of σ^{38} (Zhou and Gottesman, 2006). This is because H-NS represses expression of the anti-adaptor proteins IraD and IraM, which inhibit RssB (Battesti *et al.*, 2012; Bougdour *et al.*, 2008). Hence, σ^{38} accumulates in a Δ*hns* strain (Yamashino *et al.*, 1995). An excess of σ^{38} may contribute to the loss of fitness in *hns* mutants. Deleting the gene for σ^{38} (*rpoS*) alongside *hns* partially restores growth, and spontaneous *rpoS* mutants occur frequently in evolution experiments with Δ*hns* strains (Barth *et al.*, 1995; Singh *et al.*, 2016; Srinivasan *et al.*, 2015).

1.10. Spurious transcription

The model of transcription discussed above applies to canonical gene expression, where transcription initiates at promoters upstream of discrete genes. This process is tightly regulated by multiple protein factors during each step of the transcription cycle. Therefore, resulting mRNAs have a defined length and are usually translated into functional protein (Jacob and Monod, 1961). However, whole genome arrays in *E. coli* suggest that most of the genome is transcribed at a detectable level (Selinger *et al.*,

2000; Grainger, 2016b). TSS mapping experiments have identified thousands of examples where transcription initiates from non-canonical sites, such as inside genes or antisense to coding sequences (Figure 12) (Singh *et al.*, 2014; Thomason *et al.*, 2015; Ettwiller *et al.*, 2016; Ju *et al.*, 2019; Dornenburg *et al.*, 2010). Termed 'spurious transcription', this observation is widespread across all domains of life, even though spurious transcripts are non-coding and are therefore not predicted to benefit the cell (Wade and Grainger, 2014). Some non-coding RNAs are functional and act as regulators of their overlapping gene by complementary base pairing (Dornenburg *et al.*, 2010; Georg and Hess, 2011; Storz *et al.*, 2011). Spurious transcription may also function to protect non-coding DNA from mutations, by transcription-coupled repair. Highly transcribed regions are scanned for damage by elongating RNAP, which senses DNA lesions and promotes repair via factors like Mfd (Le *et al.*, 2018). Consequently, genomic regions which have low gene expression are more prone to mutations (Martincorena *et al.*, 2012). Spurious transcription would therefore enable RNAP to periodically check non-coding regions for DNA damage (Savery, 2007).

The low DNA sequence specificity of RNAP is partly responsible for the prevalence of spurious transcription in *E. coli*. In particular, RNAP requires just a partial match to the -10 hexamer to initiate transcription (Browning and Busby, 2004). As such, sequences which can act as promoters appear frequently in the *E. coli* genome, particularly inside AT-rich genes (Singh *et al.*, 2014). In other organisms, the number of antisense RNAs increases when the genome AT-content is higher (Lloréns-Rico *et al.*, 2016). Thus, spurious transcription is a consequence of AT-rich DNA. Terminator read-through also contributes to spurious transcription; intrinsic termination occasionally fails resulting in the transcription of downstream DNA (Nicolas *et al.*, 2012; Wade and Grainger, 2014).

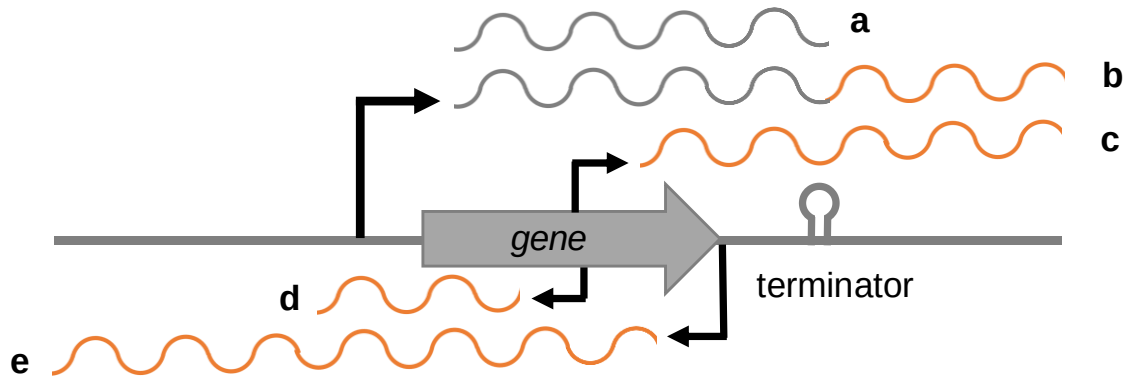


Figure 12: Pervasive transcription in prokaryotes.

Schematic of the different transcripts made during pervasive transcription. Canonical mRNA transcripts are shown in grey and spurious transcripts in orange. Transcripts are derived from: a) Canonical promoters upstream of genes; b) terminator readthrough; c) intragenic promoters positioned in the sense direction of a coding sequence d) intragenic promoters positioned antisense to a coding sequence e) antisense promoters in the 3' untranslated region of a gene (Wade and Grainger, 2014)

Rho therefore has a key role in ensuring these untranslated transcripts are quickly terminated (Cardinale *et al.*, 2008; Peters *et al.*, 2012). *E. coli* also utilises the H-NS protein to control spurious transcription. H-NS selectively silences intragenic promotersin AT-rich genes (Singh *et al.*, 2014). In an *hns* mutant, these cryptic promoters sequester RNAP. As cellular levels of RNAP are limited, this leaves less RNAP available for transcription of housekeeping genes. As such, an *hns* mutant displays a decrease in canonical transcription but increased spurious transcription. This corresponds to a loss of fitness. Removing intragenic promoters partially restores fitness, showing that spurious transcription is toxic if unregulated (Lamberte *et al.*, 2017).

One of the challenges of understanding spurious transcription is the difficulty of identifying non-canonical promoters. In *E. coli*, a variety of methods have been used to map novel start sites, including RNA 5' polyphosphatase sequencing (PPP-seq) (Singh *et al.*, 2014), differential RNA sequencing (dRNA-seq) (Thomason *et al.*, 2015), SEnd-seq (Ju *et al.*, 2019) and cappable-seq (Ettwiller *et al.*, 2016). All of these processes are variations of RNA sequencing (RNA-seq), modified to select for full length transcripts (which have a 5' triphosphate) and against processed RNA (which have a 5' monophosphate). RNA-seq involves extracting the total cell RNA, adding an adaptor (or biotin) to RNA ends, reverse transcribing the RNA and sequencing the resulting cDNA library. The 5' end of a transcript can then be mapped back to the genome to identify the TSS (Sharma and Vogel, 2014). SEnd-seq is the most recently developed method and unique in that it maps both 3' and 5' ends of the same transcript. It therefore defines the whole transcription unit from the site of initiation to termination. Using this method, Ju *et al*. identified 3,578 unique transcription units. Of these, 22 %

initiated from inside a gene, although more than half of these intragenic TSSs drove transcription of a downstream gene. They also observed that ~15 % of transcripts were antisense RNAs, originating from the 3' end of a gene. However, the proportion of predicted antisense and intragenic transcripts vary between different RNA-seq experiments. A dRNA-seq experiment identified 14,868 potential TSSs, 43 % of which were antisense and 37 % of which were intragenic (Thomason *et al.*, 2015). Ettwiller *et al*. (2016) found that 63 % of 16,539 predicted TSSs were intragenic, with two-thirds of these initiating transcription in the sense direction

RNA-seq has limitations which may explain the lack of a consistent measure of spurious transcripts between experiments. Firstly, non-coding transcription is quickly terminated by Rho, thus the resulting transcripts will be short (Panyukov and Ozoline, 2013; Peters *et al.*, 2009). Spurious transcripts are also predicted to be degraded by RNases so that they can be recycled via metabolic pathways (Durand *et al.*, 2012; Lybecker *et al.*, 2014). It is less likely that short, unstable RNA are purified during RNA-seq experiments, and so spurious TSSs will be underrepresented. Non-canonical promoters are also often silenced by NAPs such as H-NS (Singh *et al.*, 2016, 2014). Promoters repressed by NAPs in normal lab conditions would not produce RNA and so their TSSs would not be identified. Singh at al. (2014) used an *E. coli hns* mutant strain to identify spurious promoters which would usually be repressed by H-NS. They also used a chromatin immunoprecipitation assay with sequencing (ChIP-seq) to identify σ^{70} binding sites, which correspond to promoter regions. Combining the data sets revealed only 668 promoter regions with both a σ^{70} binding site and a TSS. However, there were many more σ^{70} binding sites than TSSs mapped to the genome,

supporting the hypothesis that RNA-seq experiments do not efficiently capture spurious transcripts.

1.10.1. Eukaryotic spurious transcription

Eukaryotic transcription initiation is more complicated than transcription in bacteria, involving multiple different RNAPs, transcription factors and promoter elements. RNAP II is responsible for transcribing messenger RNA and is structurally similar to bacterial RNAP (Ebright, 2000; Kornberg, 1999). Both enzymes recognise multiple AT-rich promoter elements. For RNAP II, these include the TATA box (consensus $T_{-31}ATA_{-28}$) (Decker and Hinton, 2013). Widespread spurious transcription has been observed in many eukaryotic organisms including *Drosophila* (Manak *et al.*, 2006), human (Gingeras, 2007; Kapranov *et al.*, 2007), mouse (Okazaki *et al.*, 2002) and yeast (David *et al.*, 2006; Lu and Lin, 2019). In all examples, the majority of the genome is transcribed when only a small fraction is required for functional mRNA and protein production.

The nucleosome is an essential modulator of eukaryotic transcription and acts to silence promoters by condensing the DNA and occluding RNAP access. In this way the key nucleosome proteins, histones, are analogous to bacterial NAPs. DNA is released from the nucleosome due to the activity of histone remodelling activators. RNAP II is then recruited to a promoter to form a pre-initiation complex (Jensen *et al.*, 2013). Frequently, two RNAP complexes, orientated in opposite directions, can bind to a single region of nucleosome free DNA (Rhee and Pugh, 2012). Consequently, many eukaryotic promoter regions are divergent and antisense transcription is prevalent (Seila *et al*., 2008) (Figure 13). This phenomenon has been linked to promoter evolution; promoters in foreign DNA or those which have evolved more recently are

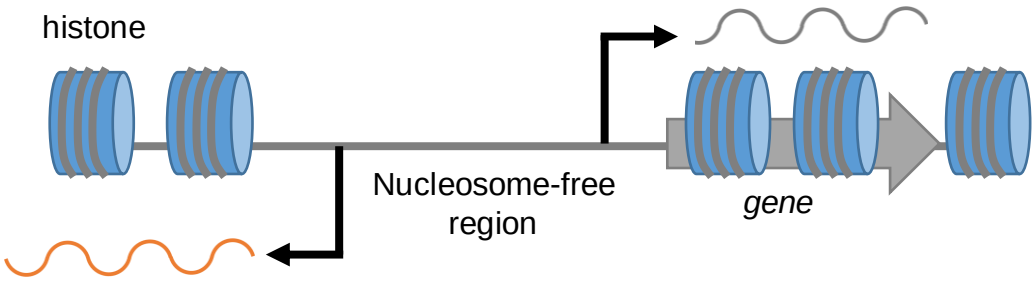


Figure 13 Bidirectional transcription in eukaryotes.

Transcription frequently initiates from two distinct promoters in a single nucleosome-free region. One direction of transcription typically results in the expression of a gene (grey transcript) whereas the antisense transcript (orange) has no known function (Rhee and Pugh, 2012).

more likely to be bidirectional (Jin *et al.*, 2017). This suggests that eukaryotic promoters are inherently bidirectional, and directional promoters are a consequence of evolution.

1.11. Objectives

Many questions remain about the origin of non-coding transcripts in the *E. coli* genome. This work explores the molecular basis of spurious transcription in *E. coli*, particularly transcription which initiates from inside H-NS repressed genes. The directionality of promoters and the role of RNAP-associated proteins Rho and NusA are also discussed.

Chapter 2. **Materials and Methods**

2.1. Strains and plasmids

Bacterial strains used in this work are listed in Table 3. In general, strains were streaked on solid agar and grown at 37 °C overnight. Strains used to inoculate liquid culture were grown at 37 °C, shaking until an appropriate optical density (OD) was reached. Liquid cultures grown overnight were prepared for storage at -80 °C by mixing with 50 % glycerol in a 1:3 ratio.

Plasmids are listed in Table 4. Purified plasmids were stored in dH_2O at -20 °C.

2.2. Oligonucleotides

Table 5 lists oligonucleotides used in this thesis.

2.3. Media

All media was purchased from Sigma and dissolved in dH_2O before autoclaving.

Lennox broth (LB) – 20 g/L

LB agar – 35 g/L

MacConkey agar – 52 g/L

Table 3. Bacterial strains

Strain	Genotype	Source
MG1655		(Keseler *et al.*, 2012)
MG1655Δ*hns*	Δ*hns*, Kan^R	(Singh *et al.*, 2014)
JCB387	Δ*nirB* Δ*lac*	(Page *et al.*, 1990)
RPB104	*rpoS* with SPA tag, from MG1655 background	(Wong *et al.*, 2017)
RPB104Δ*hns*	*rpoS* with SPA tag, Δ*hns*, Kan^R	
SL1344	*Salmonella enterica* serovar Typhimurium	(Kröger *et al.*, 2012)

Table 4. Plasmids

Plasmid	Description	Source
pRW50	Encodes Tet^R. Features a cloning site upstream of a *lacZ* fusion. Used for LacZ assays. Low copy number.	Lodge *et al*., 1992
pDOC-*hns::kan*	Encodes Amp^R and SacB, derived from pEX100T. Features a kanamycin resistance cassette between two large regions homologous to upstream and downstream of the *hns* gene. This section is flanked by two I-SceI recognition sites.	Lamberte *et al*., 2017
pACBSR	Encodes Cam^R. Recombination plasmid for gene doctoring; carries arabinose inducible λ-Red and I-SceI endonuclease genes.	Herring *et al*., 2003
pSR	Encodes Amp^R, derived from pBR322. Features a cloning site upstream of a λ*oop* terminator site. Used for *in vitro* transcription.	Kolb *et al*., 1995
pSR -9A -10T	pSR carrying a derivative of the *cbpA* promoter	Singh *et al*., 2011
pET-21b	T7 expression vector	Novagen

Table 5: Oligonucleotides

Name	Sequence
Promoter DNA sequences made by Shivani Singh prior to this work	
-10 element only	GAATTCgaccggcgagcttcgcagtcagctgactataattgccgcgcgcA GGAGGATGCGGACTATGAAGCTT
-10 element, -35 TT	GAATTCgaccttcgagcttcgcagtcagctgactataattgccgcgcgcA GGAGGATGCGGACTATGAAGCTT
-10 element, AT-i	GAATTCgaccggcgagcttcgctatttattgactataattgccgcgcgcA GGAGGATGCGGACTATGAAGCTT
-10 element, AT-i, -35 TT	GAATTCgaccttcgagcttcgctatttattgactataattgccgcgcgcA GGAGGATGCGGACTATGAAGCTT
Primers for synthesising promoter DNA with AT-ii or a random AT-tract	
AT-ii F	GGCTGCGAATTCgaccggcgagcttcgcagaattttgactataattgccg cgcgcAGG
AT-ii, -35TT F	GGCTGCGAATTCgaccttcgagcttcgcagaattttgactataattgccg cgcgcAGG
Random AT-tract F	GGCTGCGAATTCgaccggcgagcttcgcwwwwwwwtgactataattgccg cgcgcAGG[1]
Random AT-tract, -35TT F	GGCTGCGAATTCgaccttcgagcttcgcwwwwwwwtgactataattgccg cgcgcAGG[1]
Promoter R	GCCCGAAGCTTCATAGTCCGCATCCTCCTgcgcgcggcaattatagtc
Primers for amplifying intragenic σ^{70} binding sites in a and b orientations	
wzxB 1a F	GGCTGCGAATTCacgttactttatctttactatctgc
wzxB 1a R	GCCCGAAGCTTCCTCCTttgtaagaacacttggtcctgaaaa
wzxB 1b F	GGCTGCGAATTCttgtaagaacacttggtcctgaaaa
wzxB 1b R	GCCCGAAGCTTCCTCCTacgttactttatctttactatctgc
yigG 1a F	GGCTGCGAATTCcattgcctgaacaggcaaaatcttc
yigG 1a R	GCCCGAAGCTTCCTCCTtactccattatctcgtcatcaacatg
yigG 1b F	GGCTGCGAATTCtactccattatctcgtcatcaacatg
yigG 1b R	GCCCGAAGCTTCCTCCTcattgcctgaacaggcaaaatcttc
yqiI 2a F	GGCTGCGAATTCgaatattttatgaatgttttctg
yqiI 2a R	GCCCGAAGCTTCCTCCTataagttacaccgaaagtataagag
yqiI 2b F	GGCTGCGAATTCataagttacaccgaaagtataagag
yqiI 2b R	GCCCGAAGCTTCCTCCTgaatattttatgaatgttttctg

[1]*w is A or T, with an equal likelihood of either base being incorporated.*

ygaQ 1a F	GGCTGCGAATTCcggttacacaatactaacttatttaac
ygaQ 1a R	GCCCGAAGCTTCCTCCTtgaaaaatcaatggcgcttaaatcatc
ygaQ 1b F	GGCTGCGAATTCtgaaaaatcaatggcgcttaaatcatc
ygaQ 1b R	GCCCGAAGCTTCCTCCTcggttacacaatactaacttatttaac

Primers for amplifying DNA surrounding TSSs identified by Singh *et al.* (2014).

trmA F	GGCTGCGAATTCTctgcgccacaagttgttccag
trmA R	GCCCGAAGCTTCATAGTCCGCATCCTCCTttcttatggtatagcagggaaacc
wcaD F	GGCTGCGAATTCTcaaacagtttggtatcaaaacg
wcaD R	GCCCGAAGCTTCATAGTCCGCATCCTCCTcccctgaaaacgatccgg
ybvC F	GGCTGCGAATTCtgccaacaccccggtgctgacacg
ybvC R	GCCCGAAGCTTCATAGTCCGCATCCTCCTagcattcagaacactacac
yhcD F	GGCTGCGAATTCacgcctctgggcagtttttgc
yhcD R	GCCCGAAGCTTCATAGTCCGCATCCTCCTagtcggattcagtgcgataac
yqiI 1 F	GGCTGCGAATTCtaatacaagtttttatttac
yqiI 1 R	GCCCGAAGCTTCATAGTCCGCATCCTCCTtcagtaacttttccagtacg
lpxD f	GGCTGCGAATTCAccagtgccagattgcacataacg
lpxD r	GCCCGAAGCTTCATAGTCCGCATCCTCCTtcaggctgcccgccataatgacg
ribB F	GGCTGCGAATTCGtcacgagcgccatattgtg
ribB R	GCCCGAAGCTTCATAGTCCGCATCCTCCTgctgactaatgacgatggcacg
stfP F	GGCTGCGAATTCagtatattttcatcttaatacc
stfP R	GCCCGAAGCTTCATAGTCCGCATCCTCCTgtttttattacggtgcgcagc

Primers and GeneStrings for replacing AT-rich spacer sequences with GC-rich spacers[2]

wzxB 1a GC spacer	CTTGAGTCCACGCTAGATCTGGCTGCGAATTCacgttactttatctttactatctgctgctttggcaatactctgagttgctgtgagattgaaacttcgcagtcagctgactatcatatatagcatagtcgcttggcaaaaaccgaatataccgaaattttcaggaccaagtgttcttacaaAGGAGGAAGCTTCGGGCTTGTCAGTGCGCAAAAAGAT
yigG 1b GC spacer	CTTGAGTCCACGCTAGATCTGGCTGCGAATTCtactccattatctcgtcatcaacatgaattgccagcgactccgtgatagtggtttcatctatatacttcgcagtcagcttggtacattagcagtatatatcatctctatcatcacaatgatagccgaagattttgcctgttcaggcaatgAGGAGGAAGCTTCGGGCTTGTCAGTGCGCAAAAAGAT
yqiI 2b GC spacer	CTTGAGTCCACGCTAGATCTGGCTGCGAATTCataagttacaccgaaagtataagagttttgattataaaagtcttgacctcttcgcagtcagctgactatatttgcccatgcagatgggtattcttctcctggagatgggcctggtagt

	gcattattacagaaaacattcataaaatattcAGGAGGAAGCTTCGGGCT TGTCAGTGCGCAAAAAGAT
ygaQ 1a GC spacer	CTTGAGTCCACGCTAGATCTGGCTGCGAATTCcggttacacaatactaac ttatttaacccaaaatatcataaaaaagccgttatgaatttcgcagtcag cttggtaacttgtcagttggatgaacaacaaatgtcatcactgctttatg aaagagatgatttaagcgccattgatttttcaAGGAGGAAGCTTCGGGCT TGTCAGTGCGCAAAAAGAT
wcaD GC spacer F	GGCTGCGAATTCtcaaacagtttggtatcaaacttcgcagtcagcttgct atgat
wcaD GC spacer R	AGCCCGAAGCTTcCTCCTcccctgaaaacgatccggataatattatccct gcgagaatcatagcaagctgactgcgaagtttgat
yqiI 2 GC spacer F	GGCTGCGAATTCTaatacaagtttttttatttacttcttcgcagtcagctga ctagctt
yqiI 2 GC spacer R	AGCCCGAAGCTTCCTCCTtcagtaacttttccagtacgagtgcagctata tattaagctagtcagctgactgcgaagagtaaa
lpxD GC spacer F	GGCTGCGAATTCaccagtgccagattgcaccttcgcagtcagctgacgac aat
lpxD GC spacer R	AGCCCGAAGCTTCCTCCTtcaggctgcccgccataatgacgccaccggca accgccgtattgtcgtcagctgactgcgaaggtgcaa

[2]*underlined bases show new spacer sequence*

Primers for amplifying additional intragenic σ^{70} binding sites in a and b orientation[3]

yeeL 1a F	GGCTGCGAATTCcttgccatatgtaatattgaggtg
yeeL 1a R	GCCCGAAGCTTCCTCCTcaaaggtgaagataaagccagggc
yeeL 1b F	GGCTGCGAATTCcaaaggtgaagataaagccagggc
yeeL 1b R	GCCCGAAGCTTCCTCCTcttgccatatgtaatattgaggtg
yigF 1a F	GGCTGCGAATTCtttctcatagaaccatttgttcgtg
yigF 1a R	GCCCGAAGCTTCCTCCTtttgtcctggggttatggaaaaaag
yigF 1b F	GGCTGCGAATTCtttgtcctggggttatggaaaaaag
yigF 1b R	GCCCGAAGCTTCCTCCTtttctcatagaaccatttgttcgtg
gadE 1a F	GGCTGCGAATTCtaaacccttcaaggttatcattgatac
gadE 1a R	GCCCGAAGCTTCCTCCTaaaataaataggcgctttagcttttag
gadE 1b F	GGCTGCGAATTCaaaataaataggcgctttagcttttag
gadE 1b R	GCCCGAAGCTTCCTCCTtaaacccttcaaggttatcattgatac
ybdO 1a F	GGCTGCGAATTCtcaaccctgctgatttagatattaattg
ybdO 1a R	GCCCGAAGCTTCCTCCTcagccgttccgtcatttgcatgcc
ybdO 1b F	GGCTGCGAATTCcagccgttccgtcatttgcatgcc
ybdO 1b R	GCCCGAAGCTTCCTCCTtcaaccctgctgatttagatattaattg
ybdO 2a F	GGCTGCGAATTCtcagcattctctgctgacatgagg
ybdO 2a R	GCCCGAAGCTTCCTCCTcaataagtccgaactaaagaaaaac

ybdO 2b F	GGCTGCGAATTCcaataagtccgaactaaagaaaaac
ybdO 2b R	GCCCGAAGCTTCCTCCTtcagcattctctgctgacatgagg
yehA 1a F	GGCTGCGAATTCtttaattttatggtatcgttataaag
yehA 1a R	GCCCGAAGCTTCCTCCTttgttcaggggattttgcacttac
yehA 1b F	GGCTGCGAATTCttgttcaggggattttgcacttac
yehA 1b R	GCCCGAAGCTTCCTCCTtttaattttatggtatcgttataaag
fepE 1a F	GGCTGCGAATTCgtggtgaaagagtcgctagaaaac
fepE 1a R	GCCCGAAGCTTCCTCCTgttggcaatgtcgagtgaataattg
fepE 1b F	GGCTGCGAATTCgttggcaatgtcgagtgaataattg
fepE 1b R	GCCCGAAGCTTCCTCCTgtggtgaaagagtcgctagaaaac
leuO 1a F	GGCTGCGAATTCtcagaacactgaacatcagctgcg
leuO 1a R	GCCCGAAGCTTCCTCCTaattcccgggttgttggatgatttttg
leuO 1b F	GGCTGCGAATTCaattcccgggttgttggatgatttttg
leuO 1b R	GCCCGAAGCTTCCTCCTtcagaacactgaacatcagctgcg
mcrC 1a F	GGCTGCGAATTCagtgtcctttgttttgacctggaatag
mcrC 1a R	GCCCGAAGCTTCCTCCTgaaaattaccgggcattagcactcttc
mcrC 1b F	GGCTGCGAATTCgaaaattaccgggcattagcactcttc
mcrC 1b R	GCCCGAAGCTTCCTCCTagtgtcctttgttttgacctggaatag
mcrC 2a F	GGCTGCGAATTCataaagtgaacgagcttcatctctg
mcrC 2a R	GCCCGAAGCTTCCTCCTttccatcttaatcatgggaaaaccg
mcrC 2b F	GGCTGCGAATTCttccatcttaatcatgggaaaaccg
mcrC 2b R	GCCCGAAGCTTCCTCCTataaagtgaacgagcttcatctctg
mcrB 1a F	GGCTGCGAATTCgtatgattcagtttttgacataggtg
mcrB 1a R	GCCCGAAGCTTCCTCCTcagtcctattacgcctgttcccaaaaag
mcrB 1b F	GGCTGCGAATTCcagtcctattacgcctgttcccaaaaag
mcrB 1b R	GCCCGAAGCTTCCTCCTgtatgattcagtttttgacataggtg
mcrB 2a F	GGCTGCGAATTCctccaagaaatgcaaaccagggaatag
mcrB 2a R	GCCCGAAGCTTCCTCCTaaccctggattgaaaaatttattaag
mcrB 2b F	GGCTGCGAATTCaaccctggattgaaaaatttattaag
mcrB 2b R	GCCCGAAGCTTCCTCCTctccaagaaatgcaaaccagggaatag
trkG 1a F	GGCTGCGAATTCttctttataactttcgttatatttttttg
trkG 1a R	GCCCGAAGCTTCCTCCTtaaagggaatgcactaataacagaaaac
trkG 1b F	GGCTGCGAATTCtaaagggaatgcactaataacagaaaac
trkG 1b R	GCCCGAAGCTTCCTCCTttctttataactttcgttatatttttttg
trkG 2a F	GGCTGCGAATTCtatcggatattttaataactatttg
trkG 2a R	GCCCGAAGCTTCCTCCTaacagaaagaaacgaagttcaatatc
trkG 2b F	GGCTGCGAATTCaacagaaagaaacgaagttcaatatc

trkG 2b R	GCCCGAAGCTTCCTCCTtatcggatattttaataactatttg
trkG 3a F	GGCTGCGAATTCaggtctgtatggagtttctttttttc
trkG 3a R	GCCCGAAGCTTCCTCCTtgcagtagccccaaaacctaatccc
trkG 3b F	GGCTGCGAATTCtgcagtagccccaaaacctaatccc
trkG 3b R	GCCCGAAGCTTCCTCCTaggtctgtatggagtttctttttttc
idnK 1a F	GGCTGCGAATTCggagtggtaaaacattaattggtag
idnK 1a R	GCCCGAAGCTTCCTCCTgttccagccagggaagtcgatcttc
idnK 1b F	GGCTGCGAATTCgttccagccagggaagtcgatcttc
idnK 1b R	GCCCGAAGCTTCCTCCTggagtggtaaaacattaattggtag
idnK 2a F	GGCTGCGAATTCgtctttataaaaagaatgaaacagg
idnK 2a R	GCCCGAAGCTTCCTCCTcccgacgctgcattcgcgcgagaatag
idnK 2b F	GGCTGCGAATTCcccgacgctgcattcgcgcgagaatag
idnK 2b R	GCCCGAAGCTTCCTCCTgtctttataaaaagaatgaaacagg
yqiI 1a F	GGCTGCGAATTCcgcagtctgtagtggcggtcctgaac
yqiI 1a R	GCCCGAAGCTTCCTCCTatacctttaaattcaagtgctatatattc
yqiI 1b F	GGCTGCGAATTCatacctttaaattcaagtgctatatattc
yqiI 1b R	GCCCGAAGCTTTCCTCCTcgcagtctgtagtggcggtcctgaac
ygaQ 2a F	GGCTGCGAATTCttaaagatccagtaacaaaagaacg
ygaQ 2a R	GCCCGAAGCTTCCTCCTgcattccatttaaacgcttttc
ygaQ 2b F	GGCTGCGAATTCgcattccatttaaacgcttttc
ygaQ 2b R	GCCCGAAGCTTCCTCCTttaaagatccagtaacaaaagaacg
evgS 1a F	GGCTGCGAATTCcctaatgaactttatcattttcttattc
evgS 1a R	GCCCGAAGCTTCCTCCTtttttccgtcaggcgcagaacttcg
evgS 1b F	GGCTGCGAATTCtttttccgtcaggcgcagaacttcg
evgS 1b R	GCCCGAAGCTTCCTCCTcctaatgaactttatcattttcttattc
evgS 2a F	GGCTGCGAATTCaaagcactctcggattccttaccg
evgS 2a R	GCCCGAAGCTTCCTCCTaaagggtgagtcactgttttctaatg
evgS 2b F	GGCTGCGAATTCaaagggtgagtcactgttttctaatg
evgS 2b R	GCCCGAAGCTTCCTCCTaaagcactctcggattccttaccg
evgS 3a F	GGCTGCGAATTCatatacacacaggtatttgaaattg
evgS 3a R	GCCCGAAGCTTCCTCCTtgcattaattagatcacgcgtttcag
evgS 3b F	GGCTGCGAATTCtgcattaattagatcacgcgtttcag
evgS 3b R	GCCCGAAGCTTCCTCCTatatacacacaggtatttgaaattg
yibA 1a F	GGCTGCGAATTCgatatcggagcatttatactcgggc
yibA 1a R	GCCCGAAGCTTCCTCCTtttcttgcatcgctgagccgttgac
yibA 1b F	GGCTGCGAATTCtttcttgcatcgctgagccgttgac
yibA 1b R	GCCCGAAGCTTCCTCCTgatatcggagcatttatactcgggc

yibA 2a F	GGCTGCGAATTCgggttttatctgttttatgcgatgag
yibA 2a R	GCCCGAAGCTTCCTCCTcggaagttataatttcattgtcatc
yibA 2b F	GGCTGCGAATTCcggaagttataatttcattgtcatc
yibA 2b R	GCCCGAAGCTTCCTCCTgggttttatctgttttatgcgatgag
elaD 1a F	GGCTGCGAATTCgccgaatgaagtccagttattcccc
elaD 1a R	GCCCGAAGCTTCCTCCTttttctttatcatagcctagtgcac
elaD 1b F	GGCTGCGAATTCttttctttatcatagcctagtgcac
elaD 1b R	GCCCGAAGCTTCCTCCTgccgaatgaagtccagttattcccc
sfmD 1a F	GGCTGCGAATTCcggcaatacaggaagtgatattttc
sfmD 1a R	GCCCGAAGCTTCCTCCTtgccgaatcgttagctgggccgccg
sfmD 1b F	GGCTGCGAATTCtgccgaatcgttagctgggccgccg
sfmD 1b R	GCCCGAAGCTTCCTCCTcggcaatacaggaagtgatattttc
yeaI 1a F	GGCTGCGAATTCcgattagccagcgaactatggccg
yeaI 1a R	GCCCGAAGCTTCCTCCTcacaacaaacagtttactggaaacttc
yeaI 1b F	GGCTGCGAATTCcacaacaaacagtttactggaaacttc
yeaI 1b R	GCCCGAAGCTTCCTCCTcgattagccagcgaactatggccg
ycjW 1a F	GGCTGCGAATTCgaataaacatgggcatattgaccttc
ycjW 1a R	GCCCGAAGCTTCCTCCTatgcgaaagcaaaattaagcagaaaatg
ycjW 1b F	GGCTGCGAATTCatgcgaaagcaaaattaagcagaaaatg
ycjW 1b R	GCCCGAAGCTTCCTCCTgaataaacatgggcatattgaccttc
yagM 1a F	GGCTGCGAATTCcctcaaaggttgttctatgaataag
yagM 1a R	GCCCGAAGCTTCCTCCTcagttatacgtgaaaggctatcctc
yagM 1b F	GGCTGCGAATTCcagttatacgtgaaaggctatcctc
yagM 1b R	GCCCGAAGCTTCCTCCTcctcaaaggttgttctatgaataag

[3]*sequences are numbered to distinguish between multiple binding peaks in the same gene*

Primers for amplifying canonical promoters in a and b orientations[4]

tsK 2b R	GCCCGAAGCTTCCTCCTGTATGCCACTGTTTGAAAATCCC
tsK 2b F	GGCTGCGAATTCGCAATCGATTACGTAAATGATAGAAC
tsK 2a R	GCCCGAAGCTTCCTCCTGCAATCGATTACGTAAATGATAGAAC
tsK 2a F	GGCTGCGAATTCGTATGCCACTGTTTGAAAATCCC
tdK 8b R	GCCCGAAGCTTCCTCCTaagggagaacgcatataccc
tdK 8b F	GGCTGCGAATTCcccgcattcattgcggaatag
tdK 8a R	GCCCGAAGCTTCCTCCTcccgcattcattgcggaatag
tdK 8a F	GGCTGCGAATTCaagggagaacgcatataccc
ssuE 1b R	GCCCGAAGCTTCCTCCTTACCCGCCAGGGTGATGACACGCATAC

ssuE 1b F	GGCTGCGAATTCctttagtttattttcagaaaaagatacac
ssuE 1a R	GCCCGAAGCTTCCTCCTctttagtttattttcagaaaaagatacac
ssuE 1a F	GGCTGCGAATTCTACCCGCCAGGGTGATGACACGCATAC
rpsU 2b R	GCCCGAAGCTTCCTCCTctcctcacccttataaaagtc
rpsU 2b F	GGCTGCGAATTCcggcatgtgcctctcacctttg
rpsU 2a R	GCCCGAAGCTTCCTCCTcggcatgtgcctctcacctttg
rpsU 2a F	GGCTGCGAATTCctcctcacccttataaaagtc
rpsJ 1b R	GCCCGAAGCTTCCTCCTgattgggagcattgttaggtag
rpsJ 1b F	GGCTGCGAATTCgagagataacccgaaggctg
rpsJ 1a R	GCCCGAAGCTTCCTCCTgagagataacccgaaggctg
rpsJ 1a F	GGCTGCGAATTCgattgggagcattgttaggtag
puT 1b R	GCCCGAAGCTTCCTCCTgcaaaaaatgtgagagagtgcaacc
puT 1b F	GGCTGCGAATTCacggggtaacagagtttatgttttacc
puT 1a R	GCCCGAAGCTTCCTCCTacggggtaacagagtttatgttttacc
puT 1a F	GGCTGCGAATTCgcaaaaaatgtgagagagtgcaacc
puR 1b R	GCCCGAAGCTTCCTCCTaatctcccgtcatttataatgataag
puR 1b F	GGCTGCGAATTCaaaagtgttgcggtacgccgg
puR 1a R	GCCCGAAGCTTCCTCCTaaaagtgttgcggtacgccgg
puR 1a F	GGCTGCGAATTCaatctcccgtcatttataatgataag
pheL 1b R	GCCCGAAGCTTCCTCCTattgagtgtatcgccaacgc
pheL 1b F	GGCTGCGAATTCctcccattcaggggaaggtaaaaaag
pheL 1a R	GCCCGAAGCTTCCTCCTctcccattcaggggaaggtaaaaaag
pheL 1a F	GGCTGCGAATTCattgagtgtatcgccaacgc
osmB 1b R	GCCCGAAGCTTCCTCCTacagccgcggtcatttttttg
osmB 1b F	GGCTGCGAATTCcgtgatataaccctgcgcgcgag
osmB 1a R	GCCCGAAGCTTCCTCCTcgtgatataaccctgcgcgcgag
osmB 1a F	GGCTGCGAATTCacagccgcggtcatttttttg
ompA 2b R	GCCCGAAGCTTCCTCCTgttaaatccttcaccggggg
ompA 2b F	GGCTGCGAATTCatacaagactttttttcatatg
ompA 2a R	GCCCGAAGCTTCCTCCTatacaagactttttttcatatg
ompA 2a F	GGCTGCGAATTCgttaaatccttcaccggggg
mtlA 1b R	GCCCGAAGCTTCCTCCTctatatttatgtgattgatatcacac
mtlA 1b F	GGCTGCGAATTCgtttgctgtcgcgcagg
mtlA 1a R	GCCCGAAGCTTCCTCCTgtttgctgtcgcgcagg
mtlA 1a F	GGCTGCGAATTCctatatttatgtgattgatatcacac
ivbL 1b R	GCCCGAAGCTTCCTCCTgcagttggtagtagttttgc
ivbL 1b F	GGCTGCGAATTCaaacgtgatcaacccctcaattttcc

ivbL 1a R	GCCCGAAGCTTCCTCCTaaacgtgatcaacccctcaattttcc
ivbL 1a F	GGCTGCGAATTCgcagttggtagtagttttgc
ilvIH 1b R	GCCCGAAGCTTCCTCCTattcttattaccccgtgtttatg
ilvIH 1b F	GGCTGCGAATTCgataagcgatcggacgaccatc
ilvIH 1a R	GCCCGAAGCTTCCTCCTgataagcgatcggacgaccatc
ilvIH 1a F	GGCTGCGAATTCtattcttattaccccgtgtttatg
hisB 1b R	GCCCGAAGCTTCCTCCTaaccaactacattctggcgc
hisB 1b F	GGCTGCGAATTCcgctggctttcttcacgggttc
hisB 1a R	GCCCGAAGCTTCCTCCTcgctggctttcttcacgggttc
hisB 1a F	GGCTGCGAATTCaaccaactacattctggcgc
guaB 1b R	GCCCGAAGCTTCCTCCTgtcgtcaaacgtcagagcttc
guaB 1b F	GGCTGCGAATTCcgcccttcggggatagcaag
guaB 1a R	GCCCGAAGCTTCCTCCTcgcccttcggggatagcaag
guaB 1a F	GGCTGCGAATTCgtcgtcaaacgtcagagcttc
gdhA 1b R	GCCCGAAGCTTCCTCCTctcgttattaatttgctttcctggg
gdhA 1b F	GGCTGCGAATTCtgtctgatccatagatataaaaccc
gdhA 1a R	GCCCGAAGCTTCCTCCTgtctgatccatagatataaaaccc
gdhA 1a F	GGCTGCGAATTCctcgttattaatttgctttcctggg
fixA 1b R	GCCCGAAGCTTCCTCCTgaggaacttaacaatattg
fixA 1b F	GGCTGCGAATTCatctccagaaatcatgaagg
fixA 1a R	GCCCGAAGCTTCCTCCTatctccagaaatcatgaagg
fixA 1a F	GGCTGCGAATTCtgaggaacttaacaatattg
fepB 1b R	GCCCGAAGCTTCCTCCTagccctcaccctggaagggag
fepB 1b F	GGCTGCGAATTCatgtcaactcttgaggtaacgc
fepB 1a R	GCCCGAAGCTTCCTCCTatgtcaactcttgaggtaacgc
fepB 1a F	GGCTGCGAATTCagccctcaccctggaagggag
dnaN 3b R	GCCCGAAGCTTCCTCCTcgccagcgccatcgccatc
dnaN 3b F	GGCTGCGAATTCacttgctggcattgcaggaaaaac
dnaN 3a R	GCCCGAAGCTTCCTCCTacttgctggcattgcaggaaaaac
dnaN 3a F	GGCTGCGAATTCtcgccagcgccatcgccatc
cysJ 1b R	GCCCGAAGCTTCCTCCTgtgtcgtcatgcgtcgttatg
cysJ 1b F	GGCTGCGAATTCaggttagtcgatttggttattag
cysJ 1a R	GCCCGAAGCTTCCTCCTaggttagtcgatttggttattag
cysJ 1a F	GGCTGCGAATTCgtgtcgtcatgcgtcgttatg
cstA 3b R	GCCCGAAGCTTCCTCCTactccgatttacatggttgc
cstA 3b F	GGCTGCGAATTCtgctcccattacagagagcac
cstA 3a R	GCCCGAAGCTTCCTCCTtgctcccattacagagagcac

cstA 3a F	GGCTGCGAATTCactccgatttacatggttgc
cirA 2b R	GCCCGAAGCTTCCTCCTttttatgcaggtgatcatcc
cirA 2b F	GGCTGCGAATTCcaattccatttccctgacaaatc
cirA 2a R	GCCCGAAGCTTCCTCCTcaattccatttccctgacaaatc
cirA 2a F	GGCTGCGAATTCttttatgcaggtgatcatcc
bcp 1b R	GCCCGAAGCTTCCTCCTgcggacgcagcaaatattgag
bcp 1b F	GGCTGCGAATTCtcggtgcgatatcaccggctttc
bcp 1a R	GCCCGAAGCTTCCTCCTtcggtgcgatatcaccggctttc
bcp 1a F	GGCTGCGAATTCgcggacgcagcaaatattgag
asnB 1b R	GCCCGAAGCTTCCTCCTtgatatcgaatacgccaaaaattg
asnB 1b F	GGCTGCGAATTCtcaccattacgtttttttatttttc
asnB 1a R	GCCCGAAGCTTCCTCCTtcaccattacgtttttttatttttc
asnB 1a F	GGCTGCGAATTCtgatatcgaatacgccaaaaattg
aroK 1b R	GCCCGAAGCTTCCTCCTaagacagcaaaatgccgcctgaatg
aroK 1b F	GGCTGCGAATTCaagcggtaatgtttttacgctgaacg
aroK 1a R	GCCCGAAGCTTCCTCCTaagcggtaatgtttttacgctgaacg
aroK 1a F	GGCTGCGAATTCaagacagcaaaatgccgcctgaatg
argU 1b R	GCCCGAAGCTTCCTCCTggtcgttcacttgttcagcaac
argU 1b F	GGCTGCGAATTCacctgcggcccacgacttag
argU 1a R	GCCCGAAGCTTCCTCCTacctgcggcccacgacttag
argU 1a F	GGCTGCGAATTCggtcgttcacttgttcagcaac
argE 2b R	GCCCGAAGCTTCCTCCTgttgaatacgctgattgtgg
argE 2b F	GGCTGCGAATTCtgcggatgcaaatcggagattaac
argE 2a R	GCCCGAAGCTTCCTCCTtgcggatgcaaatcggagattaac
argE 2a F	GGCTGCGAATTCgttgaatacgctgattgtgg
araC 1b R	GCCCGAAGCTTCCTCCTcgccgtgcaaataatcaatg
araC 1b F	GGCTGCGAATTCtcttttactggctcttctcg
araC 1a R	GCCCGAAGCTTCCTCCTtcttttactggctcttctcg
araC 1a F	GGCTGCGAATTCcgccgtgcaaataatcaatg
apt 1b R	GCCCGAAGCTTCCTCCTtgatgaaaagcaagaaaagc
apt 1b F	GGCTGCGAATTCactcacggcgcgttttaaacg
apt 1a R	GCCCGAAGCTTCCTCCTactcacggcgcgttttaaacg
apt 1a F	GGCTGCGAATTCtgatgaaaagcaagaaaagc
adiY 1b R	GCCCGAAGCTTCCTCCTgctaaagcaaagcgataccg
adiY 1b F	GGCTGCGAATTCtttttttgcctgttatttatc
adiY 1a R	GCCCGAAGCTTCCTCCTtttttttgcctgttatttatc
adiY 1a F	GGCTGCGAATTCgctaaagcaaagcgataccg

[4]*Promoters are numbered to differentiate between multiple start sites for the same gene*

Primers for introducing mutations into intragenic bidirectional promoters[5]

wzxB 1a mutant1 R	GCCCGAAGCTTcctcctttgtaagaacacttggtcctgaaaatttcggta tattcggtttttgccaagcgactatgctata**ccc**gataatgtttgttg
wzxB 1a mutant2 R	GCCCGAAGCTTcctcctttgtaagaacacttggtcctgaaaatttcggta tattcggtttttgccaagcgactat**cg**tatatatgataatgtttgttg
wzxB 1b mutant1 F	GGCTGCGAATTCttgtaagaacacttggtcctgaaaatttcggtatattc ggtttttgccaagcgactatgctata**ccc**gataatgtttgttgaatatgg
wzxB 1b mutant2 F	GGCTGCGAATTCttgtaagaacacttggtcctgaaaatttcggtatattc ggtttttgccaagcgactat**cg**tatatatgataatgtttgttgaatatgg
yigG 1a mutant F	GGCTGCGAATTCcattgcctgaacaggcaaaatcttcggctatcattgtg atgatagagatgatatatactgct**cc**tgtaccaaaaacataag
yigG 1b mutant R	GCCCGAAGCTTcctcctcattgcctgaacaggcaaaatcttcggctatca ttgtgatgatagagatgatatatactgct**cc**tgtaccaaaaacataag
yqiI 2a mutant R	GCCCGAAGCTTcctcctataagttacaccgaaagtataagagttttgatt ataaaa**Δ**acctgatgctaacaacatcattatatttgc[6]
yqiI 2b mutant F	GGCTGCGAATTCataagttacaccgaaagtataagagttttgattataaa a**Δ**acctgatgctaacaacatcattatatttgcccatgc[6]

[5]*bases in bold show mutations*

[6]Δ *shows where bases have been deleted*

GeneStrings for synthetic bidirectional promoter in orientations a and b[7]

bidirectional promoter i	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcacatcgcg ctattatagcgcg<u>a</u>ctcgccctcccgccagctctaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT
bidirectional promoter ii	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcacatcgcg ctataatagcgcg<u>a</u>ctcagccgctgctggctcccaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT
bidirectional promoter iii	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcacatcgcg ctattatagcgcg**<u>c</u>**ctcgccctcccgccagctctaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT
bidirectional promoter iv	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcaca**g**cgcg ctataatagcgcg<u>a</u>ctcagccgctgctggctcccaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT
bidirectional promoter v	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcaca**g**cgcg ctattatagcgcg<u>a</u>ctcgccctcccgccagctctaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT
bidirectional promoter vi	GAGGCCCTTTCGTCTTCAAGAATTCtgacggtcctttggcgcacatcgcg ctataatagcgcg**<u>c</u>**ctcagccgctgctggctcccaggaggatgcggacta tgaAAGCTTACTCCCCATCCCCTCCAGT

[7]*Start sites are underlined. Mutated bases are shown in bold type.*

Primers for introducing R166A mutation into σ38.

rpoS R166A	gaaccaaacccgtactatt**gct**ttgccgattcacatcgtaaagg[8,9]
rpoS HindIII R	GGGGAAGCTTttactcgcggaacagcgcttcgatattcagc[9]
rpoS NheI F	AGCTCAGCTAGCagtcagaatacgctgaaag

[8]*Bold type shows mutated bases*

[9]*Primer pair used to synthesise a megaprimer*

2.4. Antibiotics

All antibiotics were purchased from Sigma. Stocks were sterilised by filter sterilisation and stored at -20 °C.

Stock concentrations were as follows:

Tetraycline – 35 mg/ml dissolved in methanol

Kanamycin – 50 mg/ml dissolved in dH_2O

Ampicillin – 100 mg/ml dissolved in dH_2O

Chloramphenicol – 100 mg/ml dissolved in ethanol

Antibiotic stocks were added to sterile media at a 1/1000 dilution.

2.5. Gel electrophoresis

2.5.1. Acrylamide gels

Non-denaturing 7.5 % polyacrylamide gels, comprised 0.5 X TBE (using 5 X TBE from National Diagnostics) and 7.5 % (w/v) acrylamide, were used for the DNA bending experiment. To polymerise the gel, 1 mg/ml ammonium persulfate (APS) and 0.2 % (v/v) tetramethylethylenediamine (TEMED) was added. DNA was separated for 7 hrs at 200 V. DNA fragments were visualised with ethidium bromide and UV light on a Bio-Rad Gel Doc transilluminator. For *in vitro* transcription and primer extension assays, reactions were loaded on a denaturing 6 % polyacrylamide gel made using a National Diagnostic SequaGel UreaGel System. 24 ml of concentrate, 66 ml of diluent and 10

ml of buffer was mixed. 0.08 g APS and 50 µl TEMED were added for 1 hour to polymerise the gel. The gel was pre-warmed by electrophoresis for one hour at 60 W. Reactions were loaded and separated for 2 hrs at 60 W. P^{32} labelled DNA was exposed to a Biorad phosphorscreen and imaged with a Biorad FX phosphoimager and Quantity One software.

2.5.2. Agarose gel

1 % (w/v) agarose gels were used for separation and purification of DNA fragments. Gels were made by dissolving 1 g of agarose in 100 ml 1X TBE. SYBR safe DNA stain from Invitrogen was added before the gel set. Electrophoresis was done at 110 V in 1X TBE. A Bio-Rad Gel Doc transilluminator was used to visualise stained DNA.

2.5.3. SDS-PAGE

Proteins were mixed with 1 X SDS loading buffer and separated on NuPAGE 4-12 % Bis-Tris Gels from Invitrogen alongside prestained protein marker. Electrophoresis was done at 150 – 180 V in 1 X NuPAGE MES SDS running buffer from Invitrogen. Proteins were stained with Coomassie blue dye.

2.6. DNA purification

2.6.1. Gel extraction

DNA was purified by gel extraction as required. After visualising the stained DNA, separated by electrophoresis on a 1 % agarose gel, the correct band was excised and DNA was extracted using the Qiagen QIAquick Gel Extraction kit. The gel slice was weighed and dissolved in 3 volumes of buffer QG at 50 °C. 10 µl 3M sodium acetate

pH 5.2 was added to adjust the pH along with 1 volume isopropanol. The mixture was applied to a QIAquick spin column and subjected to centrifugation at 17,900 x g for 1 minute. Flow through was discarded. The column was washed with 500 µl buffer QG and 750 µl buffer PE by sequential centrifugation steps. Further centrifugation removed any residual ethanol. DNA was eluted in 50 µl dH_2O into a 1.5 ml microcentrifuge tube.

2.6.2. PCR purification

PCR (polymerase chain reaction) products were purified using a QIAquick PCR purification kit (QIAGEN). 5 volumes of buffer PB was added to the PCR reaction and mixed. The mixture was applied to a QIAquick spin column and subjected to 17,900 x g for one minute. The flow through was removed and the column was washed with 750 µl buffer PE, again by centrifugation. DNA was eluted in 50 µl nuclease-free dH_2O.

2.6.3. Miniprep

Plasmids were purified using a QIAGEN QIAprep spin miniprep kit. 5 ml of overnight culture was pelleted by centrifugation (1,600 x g) for 5 minutes at 4 °C. The pellet was resuspended in 250 µl buffer P1. 250 µl buffer P2 was added to initiate cell lysis. Lysis was stopped after 4 minutes by addition of 350 µl chilled buffer N3. The lysate was cleared by centrifugation at 17,900 x g for 10 minutes and the supernatant was added to a QIAprep spin column. After centrifugation for 1 minute at 17,900 x g the flow through was discarded. The column was then washed with 500 ul PB buffer and 750 µl PE buffer, with centrifugation between washes. Residual ethanol was removed by centrifugation, and plasmid was eluted in 50 µl nuclease-free dH_2O.

2.6.4. Maxiprep

Higher concentrations of pRW50 and pSR plasmid were required for primer extension and *in vitro* transcription assays respectively. For this, and to generate plasmid stocks, DNA was isolated using a QIAGEN maxiprep kit. For pRW50 purification, 400 ml of culture was used, 100 ml was used for pSR purification. Cells were grown overnight in liquid and pelleted by centrifugation for 15 minutes at 1,600 x g and 4 °C. The pellet was resuspended in 10 ml buffer P1. 10 ml buffer P2 was added and the mixture was incubated at room temperature for 5 minutes. Lysis was stopped with the addition of 10 ml chilled buffer P3. The lysate was incubated on ice for 20 minutes and then centrifuged at 1,600 x g for 10 minute. The supernatant was applied to a QIAGEN-tip 500 which had been equilibrated with 10 ml buffer QBT. After the supernatant had passed through the tip by gravity flow, the tip was washed twice with 30 ml buffer QC. Plasmid was eluted with 15 ml buffer QF. Eluted DNA was precipitated with isopropanol (see 2.6.5.). The final pellet was resuspended in 100 µl dH_2O.

2.6.5. Isopropanol precipitation

0.6-0.7 volumes of room temperature isopropanol was added to DNA. This was immediately centrifuged at 15,000 x g for 45 minutes at 4 °C. The supernatant was removed and the pellet was washed with 750 µl 70 % (v/v) ice-cold ethanol. After centrifugation for 15 minutes, the ethanol was removed and the pellet was air dried for at least 10 minutes, before resuspension in dH_2O.

2.6.6. Ethanol precipitation

DNA in solution was mixed with 2 volumes of cold 100 % (v/v) ethanol. This was incubated at -80 °C for at least 15 minutes, before centrifugation (17,900 x g, 4 °C, 45

minutes). Supernatant was discarded and the pellet was washed with 750 µl cold 70 % (v/v) ethanol, then centrifuged again (17,900 x g, 4 °C, 10-15 minutes). The supernatant was removed and the pellet was air dried for at least 10 minutes.

2.6.7. Phenol chloroform extraction

One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the DNA-containing mixture and mixed vigorously for 15 seconds using a vortex mixer. The tubes were centrifuged for 5 minutes at 17,900 x g. The upper aqueous phase was transferred to a new tube. DNA was recovered by ethanol precipitation (see 2.6.6.).

2.7. Cloning

2.7.1. PCR

To amplify DNA sequences, PCR mixtures were set up in 50 µl with 1 unit Velocity polymerase (Bioline), 1x Hi-Fi Reaction Buffer, 1 mM dNTP mix, 300 ng template, 200 µM oligonucleotide primer and nuclease-free dH_2O. Recommended thermocycling parameters were used as shown in

Table 6.

2.7.2. Digestion

DNA was digested by restriction enzymes from NEB. 50 µl reactions contained 1x CutSmart buffer, up to 1 µg DNA, 20 units of each restriction enzyme and nuclease-free dH_2O. Reactions were incubated at 37° C for an hour. Digested plasmid ends

were dephosphorylated with 2 units of Calf Intestinal Alkaline Phosphatase (CIP) at 37 ° C for an hour. Most digested DNA was purified with a QIAGEN QIAquick PCR

Table 6: PCR cycling parameters

Step	Temperature	Time	Cycles
Initial denaturation	98 °C	2 minutes	1
Denaturation	98 °C	30 seconds	35
Annealing	55-70 °C	30 seconds	
Extension	72 °C	30 seconds/kb	
Final extension	72 °C	4 minutes	1

purification kit, except for pRW50 which, due to its size, was purified by phenol-chloroform extraction (2.6.7.).

2.7.3. Ligation

Digested DNA fragments were mixed with plasmid in a 20 µl ligation reaction with 1 µl T4 DNA ligase (NEB), 1x T4 DNA ligase buffer and nuclease free water. The amounts of insert and vector used was calculated using the NEBioCalculator at a 3:1 molar ratio. The ligation reaction was incubated at 16 °C overnight and then either stored at -20 °C or used immediately to transform *E. coli* cells.

2.7.4. Assembly

The 18 base pair (bp) bidirectional promoter sequences were synthesised as High-Q Strings DNA fragments by Eurofins. The vector was prepared by amplifying pSR divergently from the HindIII site to the EcoRI site, creating a linear double stranded vector. Plasmids were assembled with an NEB HiFi DNA Assembly kit. Double stranded DNA insert and vector were mixed with a 5:1 molar ratio along with 10 µl of HiFi Assembly master mix and nuclease free water to 20 µl. Reactions were incubated at 50 °C for 15 minutes and then stored on ice until transformation.

2.7.5. Preparation of competent cells

E. coli cultures were grown to mid-log phase and pelleted by centrifugation at 1,600 x g at 4 °C. Cells were resuspended in ice-cold 0.1M calcium chloride and incubated on ice for minimum 20 minutes. After pelleting and resuspending in 3.3 ml calcium chloride, cells were incubated on ice overnight. The next morning, 1.2 ml of 50 % (v/v) glycerol was added and cells were aliquoted in 200 µl for storage at -80 °C.

2.7.6. Transformation

100 µl of calcium chloride-competent *E. coli* cells were transformed with 1 µl of purified plasmid DNA or a 20 µl ligation reaction. Cells were incubated on ice with DNA for at least one hour before heat shock at 42 °C for 90 seconds. Cells were recovered for one hour at 37 °C in 750 µl LB. Cells were then isolated by centrifugation for 2 minutes at 2,400 x g and the supernatant was removed. The remaining cells were spread on LB or MacConkey agar supplemented with appropriate antibiotics. Agar plates were incubated at 37 °C overnight or until colonies were visible.

2.7.7. Colony PCR

Correct transformants were identified by colony PCR with MyTaq Red Mix from Bioline. 12.5 µl of polymerase mix was diluted with 10.5 µl nuclease-free dH_2O. 1µl each 10 µM primer was added and a colony was added with a sterile pipette tip. The recommended thermocycling protocol was followed as shown in Table 7.

2.7.8. Sequencing

Plasmid sequences or PCR products were confirmed by Sanger sequencing done by the Functional Genomics Facility at the University of Birmingham. 3 ng of purified PCR product or 300 ng purified plasmid was diluted to 9 µl in nuclease-free dH_2O. 1 µl of 10 µM primer was added.

2.8. LacZ assay

Z-buffer was made as follows:

Table 7: Cycling parameters for colony PCR

Step	Temperature	Time	Cycles
Initial denaturation	95 °C	1 minute	1
Denaturation	95 °C	15 seconds	35
Annealing	65 °C	15 seconds	
Extension	72 °C	10 seconds	

Z-buffer: 8.53 g Na_2HPO_4, 4.87 g $NaH_2PO_4 \cdot 2H_2O$, 0.75 g KCl, 0.25 g $MgSO_4$ dissolved in 1 L of dH_2O

Cultures of JCB387 carrying pRW50-promoter constructs were grown overnight in LB + 35 µg/ml tetracycline. In the morning, 200 µl of overnight culture was used to inoculate 5 ml fresh LB + 35 µg/ml tetracycline. Cells were grown to mid-log (OD_{650} 0.3-0.6) and the optical density (OD_{650}) was recorded. Cultures were lysed by mixing with 2 drops of 10 % (w/v) sodium deoxycholate and 2 drops of toluene on a vortex mixer. The toluene was evaporated by incubating at 37 °C for at least half an hour. Z-buffer was completed with the addition of 8 mg/ml 2-nitrophenyl-β-D-galactopyranoside (ONPG) and 0.27% (v/v) β-mercaptoethanol. 100 µl lysate was mixed with 2.5 ml Z-buffer to start the assay. Reactions were incubated at 37 °C for at between 20 and 40 minutes. The assay was stopped with the addition of 1 ml 0.1M sodium carbonate (Na_2CO_3). The OD_{420} was measured. Activity was calculated in Miller Units with the following equation:

$$Activity\ (Miller\ Units) = \frac{1000 \times 2.5 \times A \times OD_{420}}{4.5 \times T \times V \times OD_{650}}$$

Where:

1000/4.5 converts OD_{650} into nmol ONPG, assuming 1nmol/ml ONPG has an OD_{420} of 0.0045.

2.5 converts OD_{650} into dry protein mass (mg) assuming an OD_{650} of 1 is equivalent to 0.4 mg/ml bacteria.

A = final assay volume (ml)

T = time (minutes)

V = volume of lysate used (ml)

2.9. RNAP σ factor purification

The plasmid pET21b carrying *rpoS* with the R166A mutation was transformed into T7 express cells and selected for on LB agar plates + 100 µg/ml ampicillin.

Buffers were made as follows

Lysis buffer: 20 mM Tris, pH 7.5, 1 mM EDTA, 10 % (v/v) Glycerol, 0.5 mg/ml lysozyme

Dialysis buffer: 20 mM Tris, pH 7.5, 1 mM EDTA, 10 % (v/v) Glycerol, 100 mM NaCl

Elution buffer: 20 mM Tris, pH 7.5, 1 mM EDTA, 10 % (v/v) Glycerol, 1.0 M NaCl

Denaturing buffer: 20 mM Tris, pH 7.5, 1 mM EDTA, 10 % (v/v) Glycerol, 100 mM NaCl, 6 M Guanidine hydrochloric acid

Overnight culture was used to inoculate 500 ml LB with 100 µg/ml ampicillin and incubated at 37 °C until an OD_{650} of 0.6. IPTG was added to a concentration of 1 mM to induce over-expression of RpoS. Cultures were incubated for a further 3 hours, then pelleted by centrifugation (1,600 x g) at 4 °C for 10 minutes. Pellets were stored at -80 °C until use.

Cells were resuspended in 20 ml lysis buffer with 0.2 % (w/v) sodium deoxycholate and sonicated for 3 X 40 s, chilling on ice between steps. Lysate was centrifuged for 15 minutes at 17,900 x g, 4 °C. The supernatant was discarded and the pellet was

resuspended in 20 ml lysis buffer with 0.2 % (w/v) n-octyl β-D glucopyranoside using a homogeniser. This was again sonicated 3 times for 40s and pelleted by centrifugation. Supernatant was discarded and the pellet was resuspended in 40 ml denaturing buffer. This was dialysed overnight in 2 litres of dialysis buffer, kept at 4 °C.

The dialysed lysate was centrifuged at 17,900 x g, for 15 minutes at 4 °C, then filtered with a 0.2 µm syringe filter. Filtered lysate was loaded onto a HiTrap Q FF sepharose anion exchange column with an AKTA liquid chromatography system. The protein was eluted from the column with a linear gradient of elution buffer from 100 mM to 1.0 M sodium chloride over 100 minutes. Protein-containing fractions were collected and analysed on by SDS-PAGE. Suitable fractions were pooled and glycerol was added to a concentration of 25% for storage at -20 °C. Protein concentration was measure by Bradford assay.

2.9.1. Bradford assay

Bradford assays were done with Alfa Aesar Dye Reagent. First, a calibration curve was generated with dilutions of bovine serum albumin (BSA). 100 µl of 0, 0.2, 0.6, 0.9 and 1.2 mg/ml BSA was added to 3 ml Bradford Dye Reagent. 100 µl of undiluted RpoS, and 10x, 20x and 40x dilutions were also mixed with reagent. The OD_{595} was measured and RpoS concentration was calculated using the equation obtained from the BSA standard curve.

2.10. *In vitro* transcription assay

Buffers were made as follows:

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

NTP mix: 1.25 mM ATP, 1.25 mM CTP, 1.25 mM GTP, 62.5 µM UTP

STOP solution: Deionised formamide with 2 % (v/v) EDTA, 0.05 % (v/v) bromophenol blue, 0.05 % (v/v)

For each reaction, 325 ng of pSR template in 5.6 µl of dH_2O was mixed with 2 µl 10x transcription buffer, 0.1 µl of 20 mg/ml BSA (for a final concentration of 18 µg/ml), 3.25 µl NTP mix and 0.2 µl α^{32}P-UTP. Equivalent molar amounts of different σ factors were added in at least 5 X excess to RNAP core enzyme (NEB). 4 µl of the σ factor:RNAP mixture was added to the template mixture and mixed with a pipette tip before incubation at 37 °C for 10 minutes. Reactions were stopped by mixing in 20 µl of STOP solution. 4 µl of each reaction was loaded on to a pre-warmed 6 % denaturing polyacrylamide gel. The RNAI transcript was used as a loading control.

2.11. Primer extension

2.11.1. M13 sequencing

M13mp18 phage DNA sequencing reactions were used as a ladder for calculating length of the primer extension products. All buffers, templates, primers and enzymes were provided in the kit except for sodium hydroxide and sodium acetate. 1.5 µg of template DNA in 32 µl dH_2O was mixed with 8 µl 2M NaOH. This was incubated at room temperature for 10 minutes, then mixed with 7 µl 3M sodium acetate pH 5.1, 4 µl dH_2O and 120 µl ice cold 100 % ethanol. DNA was purified by ethanol precipitation (see 2.6.6.) and dissolved in 10 µl dH_2O. 2 µl of Universal Primer and 2 µl annealing

buffer was added. This was incubated at 65 °C for 5 minutes, then 37 °C for 10 minutes. Before starting the sequencing reaction, T7 DNA polymerase was prepared by diluting 1 in 5 in dilution buffer. The annealed primer-template mix was combined with 3 µl Labelling Mix-dATP, 1 µl α^{32}P dATP and 2µl diluted T7 DNA polymerase. This was incubated for 5 minutes at room temperature before 4.5 µl was taken and added to 4 pre-warmed tubes containing either A, C, G or T Mix-Short. Reactions were incubated at 37 °C for 5 minutes. They were terminated with 5 µl STOP solution. Before gel electrophoresis, they were heated to 72 °C.

2.11.2. RNA purification

RNA was extracted from cultures of JCB387 carrying pRW50 constructs using a QIAGEN RNeasy Mini kit. Cells were grown overnight in LB with 35 µg/ml tetracycline. The next morning, 200 µl of culture was used to inoculate 10 ml LB with 35 µg/ml tetracycline and this was grown to an OD_{650} of 0.4 – 0.5. Cultures were mixed with 4 µl of 19:1 ethanol:acid phenol. This was centrifuged for 10 minutes at 1,600 x g. Supernatant was removed and the pellet was resuspended in 200 µL TE with 40 µg/ml lysozyme and incubated for 15 minutes at room temperature. Buffer RLT was prepared by adding 10 µl/ml β-mercaptoethanol. 700 µl of this Buffer RLT was added to the lysates and mixed with a vortex mixer. 500 µL of 100 % ethanol was then added and the mixture was split between two RNeasy Mini spin columns. Columns were centrifuged for 30 s at 17,900 x g. Flow through was discarded. The columns were washed with 700 µl Buffer RW1 before centrifugation again and flow through was discarded. Wash steps were repeated twice with 500 µl Buffer RPE. The columns were placed into new 1.5 ml microcentrifuge tubes. RNA was eluted in 30 µl RNase-free water.

2.11.3. Removing DNA from RNA preparations

DNA was degraded by mixing the purified RNA with 1 µl TURBO DNase (Ambion) and 3.5 µl 10X TURBO DNase buffer. This was incubated for 30 minutes at 37 °C. To stop the reaction, 6.1 µl DNase Inactivation Reagent was added before incubation at room temperature for 2 minutes. Samples were centrifuged for 1.5 minutes at 8,000 x g and supernatant was collected in a new clean microcentrifuge tube.

2.11.4. Labelling of primer

Before beginning primer extension reactions, primer was labelled with γ^{32}P-ATP using T4 polynucleotide kinase (PNK) from NEB. 50 pg of the pRW50 primer D49724 (0.5 µl of a 100 µM stock) was combined with 2 µl 10X PNK buffer, 1 µl γ^{32}P-ATP, 1 µl T4 PNK and 15.5 µl RNase-free water. This was incubated at 37 °C for 30 minutes and then 68 °C for 10 minutes.

2.11.5. Reverse transcription

Hybridisation buffer was prepared as follows:

Hybridisation buffer: 20mM HEPES, 0.4 M NaCl, 80 % (v/v) formamide

The labelled primer was annealed to the pRW50-derived RNA by mixing 1 µl labelled primer with 30 µg RNA, 0.1 volumes of 3M sodium acetate (pH 7) and 2.5 volumes cold 100 % ethanol. This mixture was purified by ethanol precipitation (see 2.6.6.). The pellet was resuspended in 30 µl hybridisation buffer and incubated for 5 minutes at 50 °C, then 15 minutes at 75 °C, then 3 hours at 50 °C. The annealed primer-RNA mixture was purified again by ethanol precipitation (see 2.6.6.). The pellet was resuspended in 31 µl RNase-free water then mixed with 2.5 µl AMV reverse transcriptase (Promega), 10 µl 5X reverse transcriptase buffer, 1 µl 50 mM DTT, 5 µl

10 mM dNTPs and 0.6 µl RNasin (Promega) and incubated for 1 hour at 37 °C. The transcriptase was inactivated by heating to 72 °C for 10 minutes. To remove the template RNA, 1 µl 10 mg/ml RNase was added and reactions were incubated for 30 minutes at 37 °C. Finally these were purified by ethanol precipitation (see 2.6.6.) and the pellet was resuspended in 4 µl stop solution.

2.11.6. Transcription start site mapping

Primer extension reactions were separated by gel electrophoresis on a 6% polyacrylamide gel (2.5.1.) alongside 4 µl of M13 sequencing reaction A, C, T and G mix. The length of each extension product was determined by comparing to the ladder made by the M13 sequencing reaction. As the primer binds 89 bp from the end of the cloned promoter region, the distance between the TSS and the end of the fragment was calculated by subtracting 89 from the product length. TSSs mapped by *in vitro* transcription (see section 2.10.) follow a similar protocol; the terminator is positioned 94 bp from the end of the fragment, so TSS position is determined by subtracting 94 from the length of the RNA product.

2.12. Gene doctoring

To replace the *hns* gene in strain RPB104 with a kanamycin resistance cassette, the Gene Doctoring method was used (Lee *et al.*, 2009). A pDOC-C plasmid had been made previously carrying a kanamycin resistance cassette, flanked by regions homologous to the upstream and downstream sequences surrounding the *hns* gene. This plasmid was used to co-transform competent RPB104 cells with pACBSR. The transformed cells were grown on LB agar with 100 µg/ml chloramphenicol and 100

µg/ml ampicillin. Colonies were patched onto LB agar with kanamycin and 5 % (w/v) sucrose to confirm that the *sacB* gene was still present. To initiate recombination, a co-transformed colony was inoculated into LB with 100 µg/ml chloramphenicol and 100 µg/ml ampicillin and grown at 37 °C shaking for 3 hours. Cells were harvested by centrifugation for 10 minutes at 1,600 x g. The cell pellet was washed 3 times with 0.1X LB then finally resuspended in 0.5 ml 0.1X LB with 0.4 % (w/v) arabinose. This was incubated at 37 °C shaking for 2.5 hours. 125 µl of culture was spread onto four plates of LB agar with kanamycin and 5 % (w/v) sucrose. These plates were incubated at room temperature for 72 hours. Candidate recombinants were patched onto LB plates with 100 µg/ml chloramphenicol and 100 µg/ml ampicillin to confirm loss of pDOC-C and pACBSR. Colony PCR was used with primers flanking the *hns* gene to confirm loss of the gene. These cPCR reactions were sequenced using the same primers to check the sequence of the kanamycin resistance cassette.

2.13. Chromatin immunoprecipitation and sequencing (ChIP-seq)

Buffers were prepared in advance as follows:

1X TBS: 20 mM Tris-HCl pH 7.4, 0.9 % (w/v) NaCl

FA lysis buffer 150: 50 mM HEPES-KOH pH 7.0, 150 mM NaCl, 1 mM EDTA, 1 % (w/v) Triton-X-100, 0.1 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS.

FA lysis buffer 500: 50 mM HEPES-KOH pH 7.0, 500 mM NaCl, 1 mM EDTA, 1 % (w/v) Triton-X-100, 0.1 % (w/v) sodium deoxycholate, 0.1 % (w/v) SDS.

ChIP wash buffer: 10 mM Tris-HCl pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5 % (w/v) Nonidet-P40, 0.5 % sodium deoxycholate

ChIP elution buffer 50 mM Tris-HCl pH 7.5, 10 mM EDTA, 1 % (w/v) SDS

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

Protein A sepharose beads (GE healthcare) were washed with ddH2O and stored in a 50 % (v/v) slurry with 1X TBS.

The antibodies used were as follows

Anti-FLAG: from Sigma, used to immunoprecipitate SPA-tagged σ^{38}

Anti-NusA: from Neoclone, used to immunoprecipitate NusA

Anti-Rho: provided by Jeffrey Roberts (Cornell University), used to immunoprecipitate Rho.

2.13.1. Lysate preparation

LB broth was inoculated with Δ*hns* cells directly from a glycerol stock to avoid the development of compensatory mutations during overnight growth. 40 ml of culture was incubated at 37 °C, shaking until an OD_{600} of 1.4 (NusA and Rho) or 3.0 (σ^{38}). The NusA and Rho ChIP assay included a spiked-in control for normalisation and better comparison between strains. In parallel, cultures of *Salmonella enterica* serovar Typhimurium (strain SL1344) were grown to log phase. The OD_{600} measurements were used as an approximation of cell number, and cultures were mixed at a 1:4 *S.* Typhimurium:*E. coli* cell ratio. Proteins were crosslinked to DNA with the addition of 1 ml 37 % (v/v) formaldehyde and incubation at room temperature for 20 minutes. This

reaction was quenched by adding 10 ml 2.5 M glycine. Cells were pelleted by centrifugation for 5 minutes at 1,600 x g then washed with 25 ml 1 X TBS and centrifuged again. Cells were resuspended in 1 ml 1 X TBS and transferred to a microcentrifuge tube, then centrifuged for 1 minute at 17,900 x g. The supernatant was discarded and the cells were resuspended in 1 ml FA lysis buffer 150 with 4 mg/ml lysozyme, then incubated at 37 °C for 30 minutes. Lysed cells were sonicated in a Bioruptor sonicator for 3 10 x cycle runs, each cycle consisting of 30 seconds on, 30 seconds off. 250 µl lysate was diluted to 800 µl with FA lysis buffer 150 to make one ChIP lysate.

2.13.2. Immunoprecipitation and library processing

Each ChIP lysate was incubated for 90 minutes at room temperature with 25 µl Protein A beads and 2 µl of the appropriate antibody. For each immunoprecipitation, a mock was done with 25 µl beads but no antibody. Beads were gently pelleted by centrifugation at 1,600 x g for 1 minute. Supernatant was removed, beads were suspended in 700 µl FA lysis buffer 150 and transferred to a Spin-X column (VWR). For all the following washes, beads were centrifuged at 1,600 x g for 1 minute and resuspended in 700 µl buffer by rotating for 3 minutes. Wash steps used FA lysis buffer 150 twice, then 10 mM Tris-HCl pH 7.5 twice. DNA ends were blunted using an NEB quick blunting kit; beads were mixed with 10 µl 10 X quick blunting buffer, 10 µl dNTP mix, 80 µl dH_2O and 2 µl blunt enzyme mix. The blunting reaction was incubated at room temperature for 30 minutes on a rotator. The supernatant was separated by centrifugation and discarded before washes resumed. Beads were washed twice with FA lysis buffer 150 and twice with 10 mM Tris-HCl pH 8. A-tail DNA ends were added with NEB Klenow (3' -> 5' exo-); beads were mixed with 10 µl 10 X NEB buffer 2µl 100

mM dATP, 88 µl dH_2O and 2 µl Klenow. The Klenow reaction was incubated at 37 °C for 30 minutes on a rotator. Columns were centrifuged and the supernatant discarded. Beads were washed twice with FA lysis 150 and twice with 10 mM Tris-HCl pH 8. A unique barcoded adaptor per IP from BioO Scientific was ligated to the DNA ends using NEB quick ligation mix; 100 µl 1 X ligase buffer was added with 1 µl adaptor and 4 µl quick ligase. The ligase reaction was incubated at room temperature for 15 minutes on a rotator. Columns were centrifuged and the supernatant discarded. Beads were washed twice with FA lysis buffer 150, then once each with FA lysis buffer 500, ChIP wash buffer and TE. 100 µl elution buffer was added and the column was incubated at 65 °C for 20 minutes. DNA-protein complexes were eluted by centrifugation at 1,600 x g for 1 minute. Proteins were de-crosslinked from DNA by boiling for 10 minutes.

2.13.3. Purification and size selection

Ampure magnetic beads were used at a 1.1 X volume to remove protein after de-crosslinking or a 0.7 X volume for size selection. The appropriate volume of beads were added to the eluate and resuspended gently, then incubated at room temperature for 5 minutes. Tubes were placed in a magnetic rack for 5 minutes. The supernatant was removed and the beads were washed twice with 70 % (v/v) ethanol. Off the magnet, beads were resuspended in 13 µl dH_2O. This was returned to the magnetic rack and left for 2 minutes, then the supernatant was removed and stored.

2.13.4. Library amplification

To determine the optimum number of PCR cycles for library amplication, PCR reactions were done with 2 µl library, 2 µl NEXTflex primer mix, 1 µl 100 mM dNTPs, 10 µl Velocity 5 X buffer, 34 µl dH_2O and 1 µl Velocity polymerase (Bioline). After an

initial denaturation step at 98 °C for 2 minutes, cycles were as follows: 98 °C for 30 seconds, 65 °C for 30 seconds, 72 °C for 30 seconds. 5 µl of reaction was removed every 3 cycles from 18 to 33 cycles. The optimum cycle number was determined by viewing the reactions on an acrylamide gel and selecting the number of cycles where the library was visible but not adapter dimer. The final amplification was done with the optimum number of cycles and the library was purified with 0.7 X Ampure magnetic beads. Library concentration was quantified with a Qubit dsDNA HS Assay Kit; 1 µl of sample was diluted in 199 µl diluted reagent (1:199 reagent:dilution buffer). This was quantified on a Qubit 2.0 Fluorometer. Equimolar amounts of library were pooled and sequenced with an Illumina MiSeq.

2.13.5. Data processing

FastQ files were converted to Sanger format using FastqGroomer. For the NusA and Rho ChIP with a spiked-in control, reads were first mapped to the *E. coli* MG1655 genome then the remaining reads were mapped to the *S.* Typhimurium SL1344 genome using Bowtie for Illumina. This gave a read number which was used as a correction factor to normalise *E. coli* reads. Next, the original groomed FastQ files were mapped to the SL1344 genome, to remove all *S.* Typhimurium DNA, and the remaining reads were mapped to the MG1655 genome. SAM files were converted to BAM files for assembly into a MultiBam summary, which outputs read counts for each base position on the genome. Each library was normalised using the *S.* Typhimurium correction factor, or for the σ^{38} ChIP, reads were normalised to average read count.

2.14. Western blots

Buffers were prepared as follows:

Grinding buffer: 50 mM Tris pH 8, 200 mM NaCl, 1 protease cocktail inhibitor

Transfer buffer: 50 mM Tris pH 8.3, 40 mM glycine, 5 % (v/v) methanol

Blot buffer: 1 X PBS, 0.1 % (v/v) Tween

Overnight cultures of *E. coli* MG1655, WT (wild type) and Δ*hns*, were used to inoculate 50 ml LB and incubated at 37 °C. 10 ml of culture was taken when the OD_{600} reached 0.4, 0.9 and 1.8. Cells were centrifuged for 5 minutes, at 1,600 x g, 4 °C. The cell pellet was resuspended in 1 ml grinding buffer. Cells were lysed by sonication in a Bioruptor sonicator; 10 cycles, 30 seconds on/30 seconds off each cycle. Total protein concentration was measured by Bradford Assay (see 2.9.1.). 5 µg protein was separated by SDS PAGE (see 2.5.3.). Proteins were transferred from the gel to a nitrocellulose membrane, by wet electroblotting in transfer buffer at 30 V for 1 hour. The membrane was washed in blot buffer then incubated in 25 ml blot buffer with 5 % (w/v) milk powder at 4°C overnight. The next day, the membrane washed and incubated in 10 ml blot buffer with 2 % (w/v) milk powder with 10 µl antibody (anti-NusA or anti-Rho) for 1 hour. The membrane was washed thoroughly and incubated in 10 ml blot buffer with 2 % (w/v) milk powder with 2 µl HRP-conjugated secondary antibody (anti-mouse or anti-rabbit) for 1 hour. The membrane was washed then treated with ECL detection reagent (mixed equal parts of A and B) for 5 minutes. The membrane was exposed with a chemiluminescence camera.

Chapter 3. **The Role of the AT-Tract in Promoter Activity**

3.1. Introduction

Spurious promoters often appear within horizontally acquired AT-rich genes (Singh *et al.*, 2014). It is possible that such promoters occur by chance; frequent A and T bases easily combine to generate promoter -10 elements. To quantify the relationship between DNA AT-content and chance promoter occurrence, 8 separate DNA fragment libraries were created (Warman *et al.*, 2020). In each library, the 43 bp DNA fragments had random sequence but the AT-content was fixed at different percentages. Fragment libraries were fused to *lacZ* gene so that active promoters could be identified. For each library, the percentage of all DNA fragments with promoter activity was determined. These data are plotted against library % AT-content in Figure 1A. Overall, there was a positive correlation between library AT-content and the number of promoters identified. However, few promoters were generated in DNA fragments with an AT-content below 50 %. LacZ activity was also quantified, allowing the average activity of all promoters in each library to be determined (Figure 14). Whilst a positive correlation was evident, there was no increase in average promoter activity when the AT-content exceeded 60 %. Taken together, the data suggest that active promoters can occur by chance and are more prevalent in AT-rich DNA.

All randomly generated active promoters were sequenced and a DNA sequence logo was created (Figure 15A). For comparison, we also generated sequence logos for intragenic promoters, previously identified inside H-NS bound genes (Figure 15B) (Singh *et al.*, 2014), and canonical promoters (Figure 15C) (Mitchell *et al.*, 2003). In all

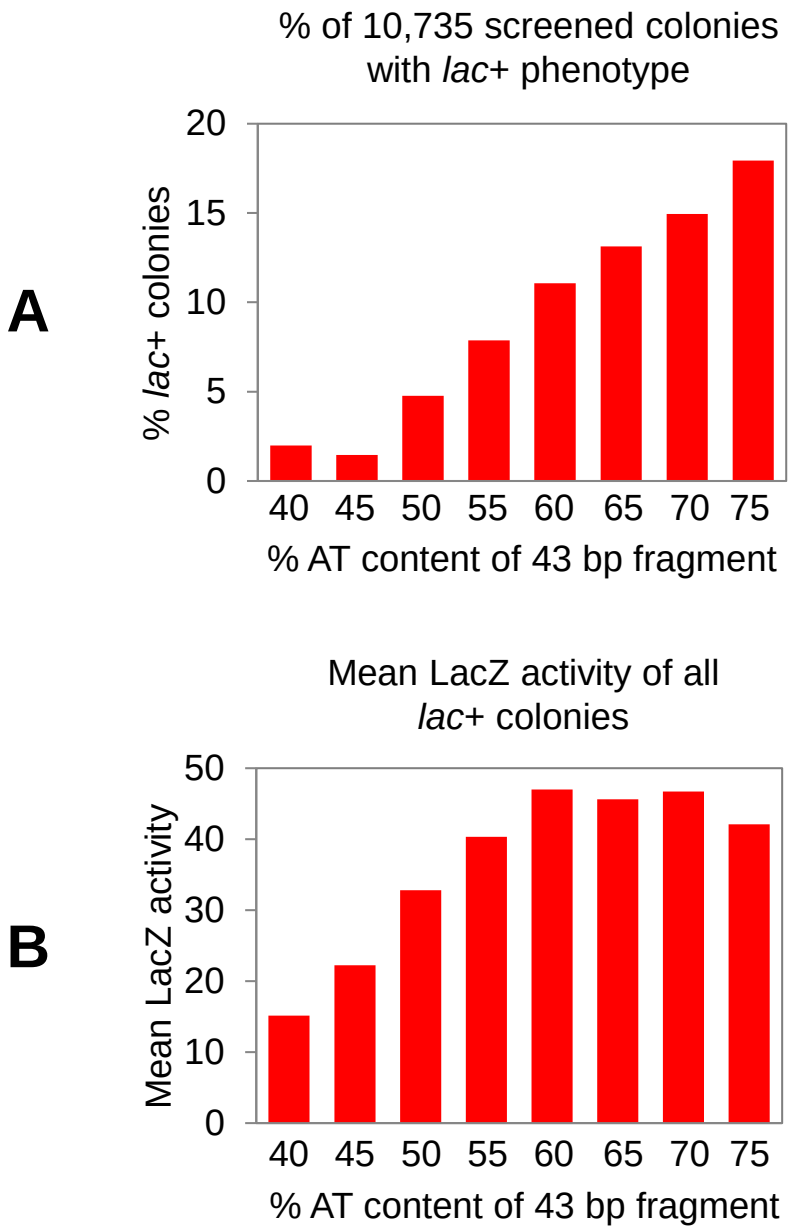


Figure 14: Promoter activity of randomly generated DNA sequences.

A) The number of *lac*+ colonies increases as % AT content increases. 10,735 total colonies were counted and the percentage of pink or red (*lac*+) colonies is plotted against the % AT content of the DNA fragment. B) *lacZ* expression increases as AT content increases. Activity of the *lac*+ colonies was measured by LacZ assay and is plotted against the % AT content of the DNA fragment (Warman *et al*., 2020).

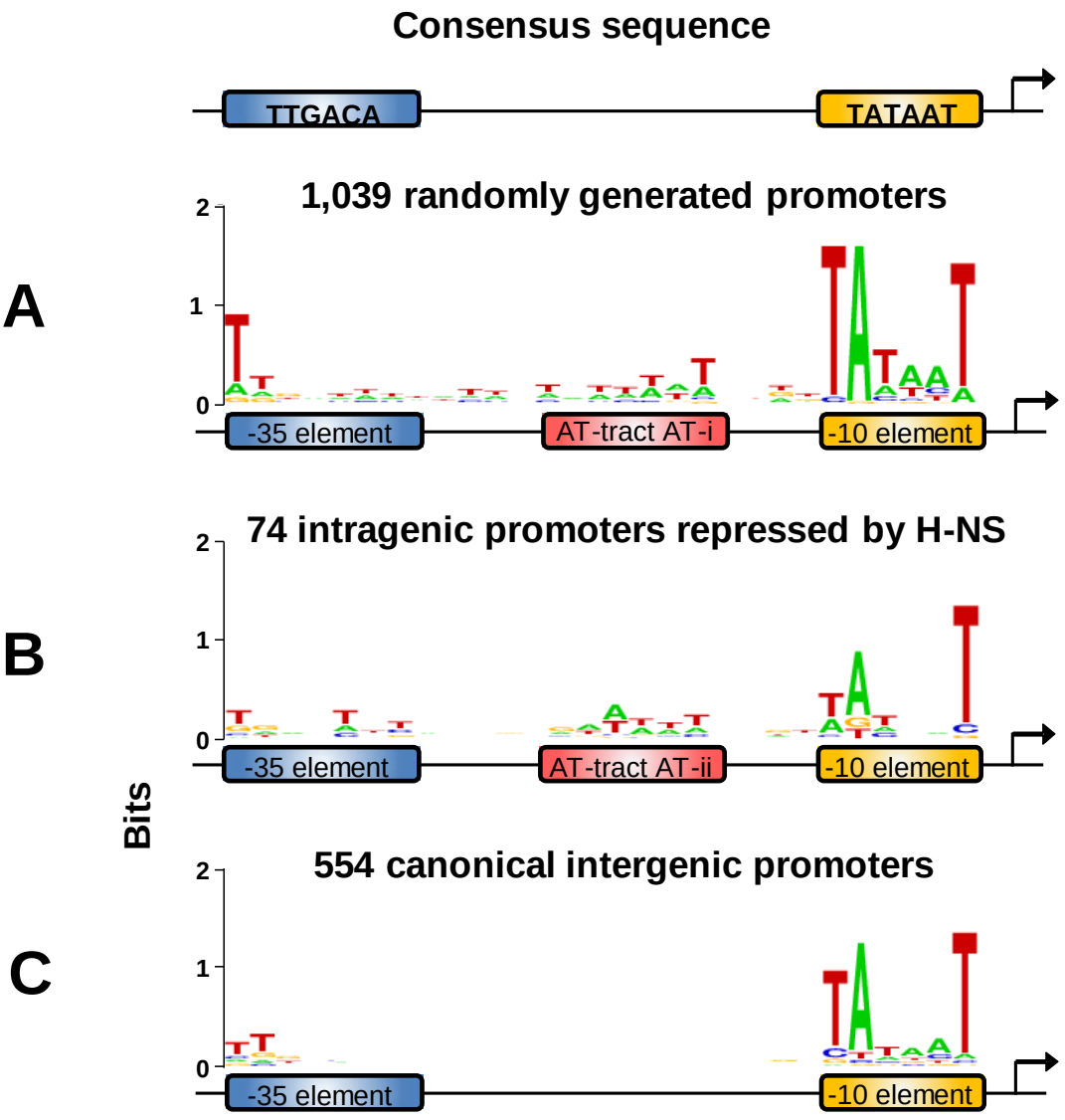


Figure 15: Conserved promoter elements

Sequence logos generated by aligning the DNA sequences from A) 1,039 randomly generated promoters isolated from *lac*+ colonies B) 74 intragenic promoters identified inside H-NS repressed genes and C) 554 canonical intergenic promoters. Top diagram shows the consensus sequence for σ^{70} -10 (yellow) and -35 (blue) promoter elements. Canonical promoter data from Mitchell et al., 2003; intragenic promoters identified by Singh et al., 2014.

logos, the best conserved feature was the promoter -10 element. Conversely, -35 elements were poorly conserved; only the first two bases of the hexamer were ever apparent. Notably, randomly generated and intragenic promoters had an AT-tract in the spacer region (Figure 15A-B). This AT-tract was situated between positions -17 and -23 in randomly generated promoters. In intragenic promoters the AT-tract was shorter, situated between positions -17 to -21. This element was not enriched at canonical promoters. This chapter explores the role of the AT-tract in σ^{70}-mediated transcription. Results from this chapter are published in Warman *et al*., 2020.

3.2. Results

3.2.1. AT-tracts can activate cryptic -10 elements

The promoter -10 element is important for DNA unwinding and transcription initiation, whilst the -35 element drives recruitment of RNAP. Hence, -10 hexamers alone are ineffective (Miroslavova and Busby, 2006). At most canonical promoters the absence of a -35 element is compensated for by transcription factors that help recruit RNAP to the DNA in response to environmental signals (Lee *et al.*, 2012).However, intragenic and randomly generated promoters are unlikely to contain transcription factor binding sites. This suggests that these promoters must recruit RNAP in some other way. To test whether the AT-tract contributed, a set of synthetic promoters was made. The promoters all had a consensus σ^{70} -10 element. The σ^{70} -10 hexamer was supplemented with an AT-tract (5'-TATTTAT-3'), derived from the randomly generated promoters, or a partial -35 element (5'-TT-3'). The different promoters are illustrated schematically in Figure 16. Each promoter was fused to *lacZ* in plasmid pRW50. Hence, β-galactosidase activity can be measured as a proxy for promoter activity. As

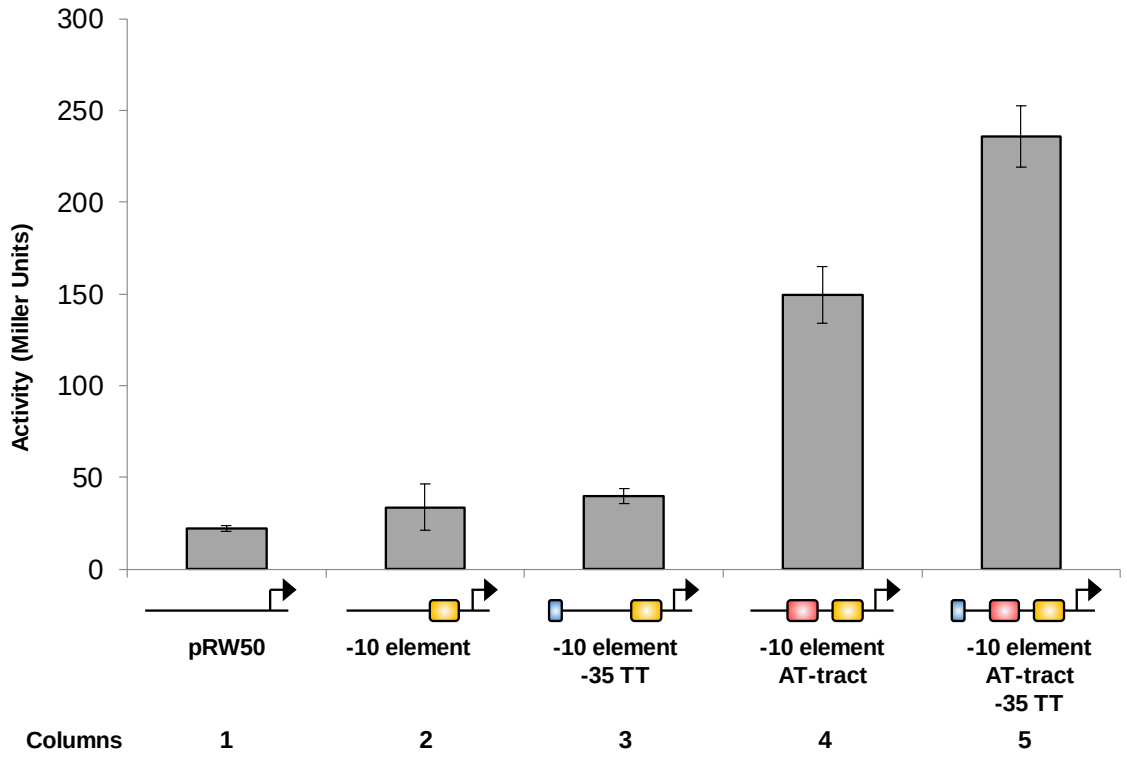


Figure 16: Promoter activity of sequences with an AT-tract

Promoter sequences were made with a consensus -10 element 'TATAAT' (yellow box), a partial -35 element 'TT' (blue box) and/or and an AT-tract (red box). These constructs were fused upstream of *lacZ* in plasmid pRW50. Empty pRW50 was used as a negative control. For each promoter, LacZ activity was measured in triplicate. Data is presented as the mean of the replicates in Miller Units. Error bars show +/- SD.

expected, the -10 element alone was unable to drive *lacZ* expression (Figure 16, column 2). Addition of the partial -35 element had no impact (column 3). Conversely, addition of an AT-tract resurrected promoter activity (column 4). This activation increased further when a partial -35 element was also present (column 5).

3.2.2. AT-tracts alter promoter DNA bending

AT-rich DNA sequences have different bending properties, which can influence promoter interactions with RNAP (Singh *et al.*, 2011). DNA curvature can be detected by altered electrophoretic mobility during native PAGE. Promoters with or without an AT-tract were excised from the plasmid pRW50 and tested for bending. Those with AT-tracts (lanes 1 and 2) had different mobility during electrophoresis (Figure 17, lanes 1 and 2).

3.2.3. Most AT-tract sequences stimulate transcription

The sequence of AT-tracts in randomly generated and intragenic promoters was different (Figure 15A-B). The AT-tracts are subsequently referred to as AT-i (from randomly generated promoters) and AT-ii (from intragenic promoters). To test which sequence was better at activating transcription, a promoter dependent on each AT-tract was made. Figure 18 shows the LacZ activity for promoters with AT-i (columns 1 and 2) or AT-ii (columns 3 and 4). In the absence of a partial -35 element, both AT-tracts stimulated similar amounts of *lacZ* expression. When in combination with a partial -35 element, AT-ii had a bigger effect. Importantly, both sequences were able to revive an otherwise transcriptionally inactive -10 element.

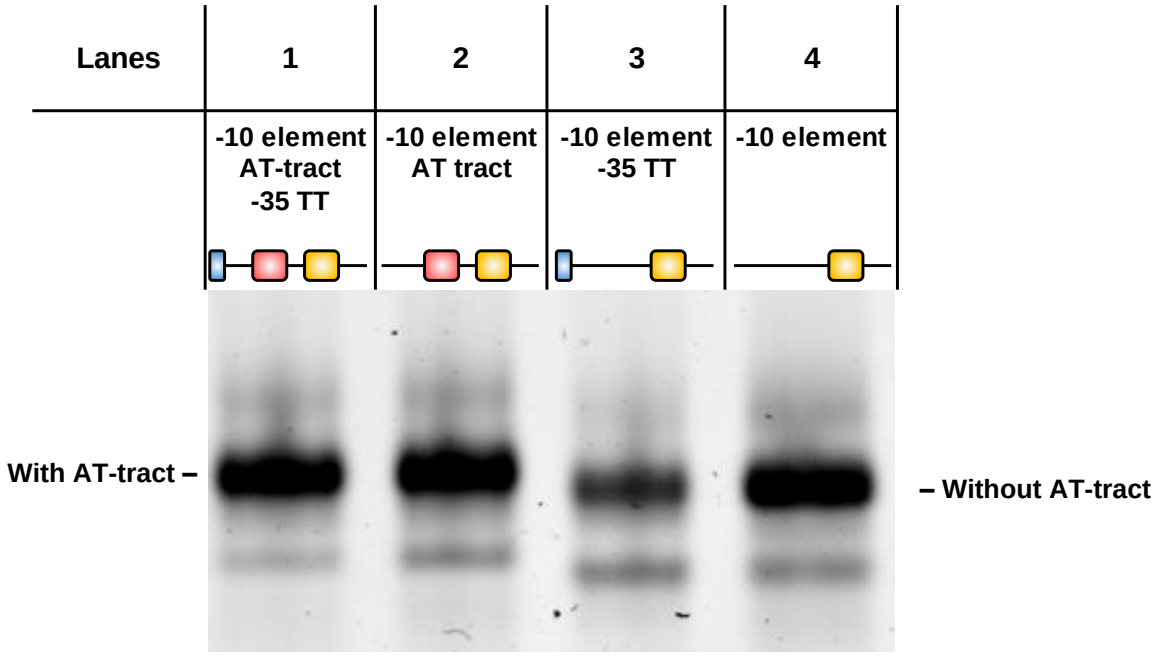


Figure 17: Effect of the AT-tract on DNA bending

Promoter sequences of equal length, with (lanes 1 and 2) or without (lanes 3 and 4) an AT-tract were separated by gel electrophoresis on a 7% native acrylamide gel. Bands indicate DNA dyed with ethidium bromide (Warman *et al.,* 2020).

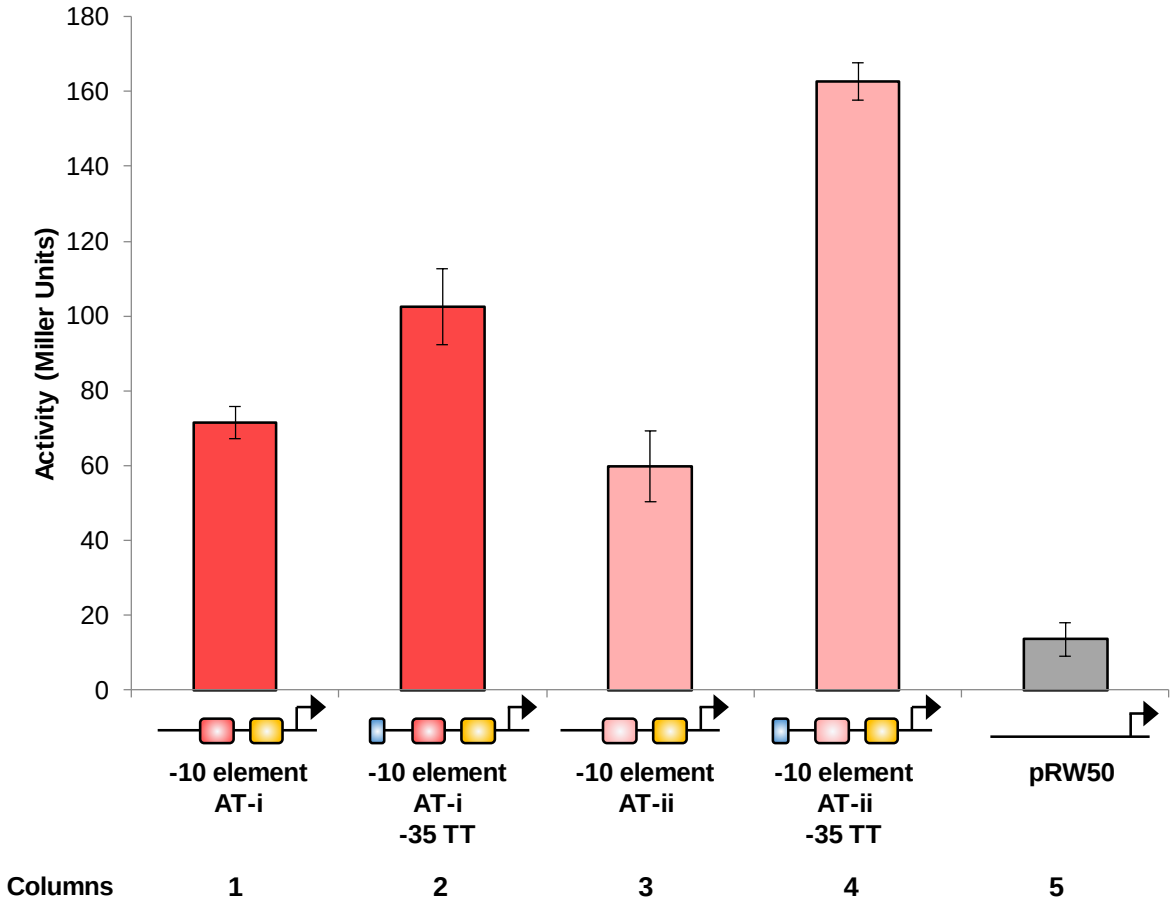


Figure 18: Promoter activity with two different AT-tract sequences

Promoter sequences containing a consensus -10 element 'TATAAT' (yellow box), an AT-tract with sequence 'TATTTAT' (red box, AT-i) or sequence 'AATTT' (pink box, AT-ii), and/or a partial -35 element 'TT' (blue box) were fused upstream of *lacZ* in pRW50. Empty pRW50 was used as a negative control. For each promoter, LacZ activity was measured in triplicate. Data is presented as the mean of the replicates in Miller Units. Error bars show +/- SD.

To further explore the role of the AT-tract sequence, promoters were made with a randomised AT-tract; bases between positions -17 and -23 were equally likely to be A or a T. These promoters were cloned into pRW50 and used to transform JCB387. Transformants were selected on MacConkey agar, then lysates of culture grown from each individual colony were used to measure *lacZ* expression. The data are presented as a box plot in Figure 19, with each data point taken from a single colony (i.e. expected to represent a different AT-tract). Again, promoters with both the partial -35 element and an AT-tract were more active, on average, than promoters with just an AT-tract. All AT-tract sequences were able to activate transcription. One AT-tract resulted in lower activity than the rest (indicated by a red arrow on the plot); this AT-tract had the sequence 5'-TTTTTAA-3'

3.2.4. Activation by AT-tracts requires σ^{70} residue R451

To confirm the effect of the AT-tract *in vitro*, selected promoters were cloned upstream of the *λoop* terminator in plasmid pSR. Transcripts terminating at *λoop* can be detected following electrophoresis. The pSR origin of replication produces the RNAI transcript which serves as an internal and loading control. Plasmids were mixed with RNAP and wild-type (WT) σ^{70} (Figure 20, lanes 1, 3, 5, and 7). Both AT-tracts stimulated transcription, and transcript amount increased when the -35 element was included. Reactions were also done with the R451A σ^{70} mutant (lanes 2, 4, 6, 8). This mutant was previously shown to be defective at initiating transcription at promoters reliant on a T at position -18 (Singh *et al*., 2011). These data show that R451A σ^{70} holoenzyme is also unable to generate transcripts from promoters dependent on an AT-tract. The -9A-10T promoter is a derivative of the *cbpA* promoter, modified to have a strong -10

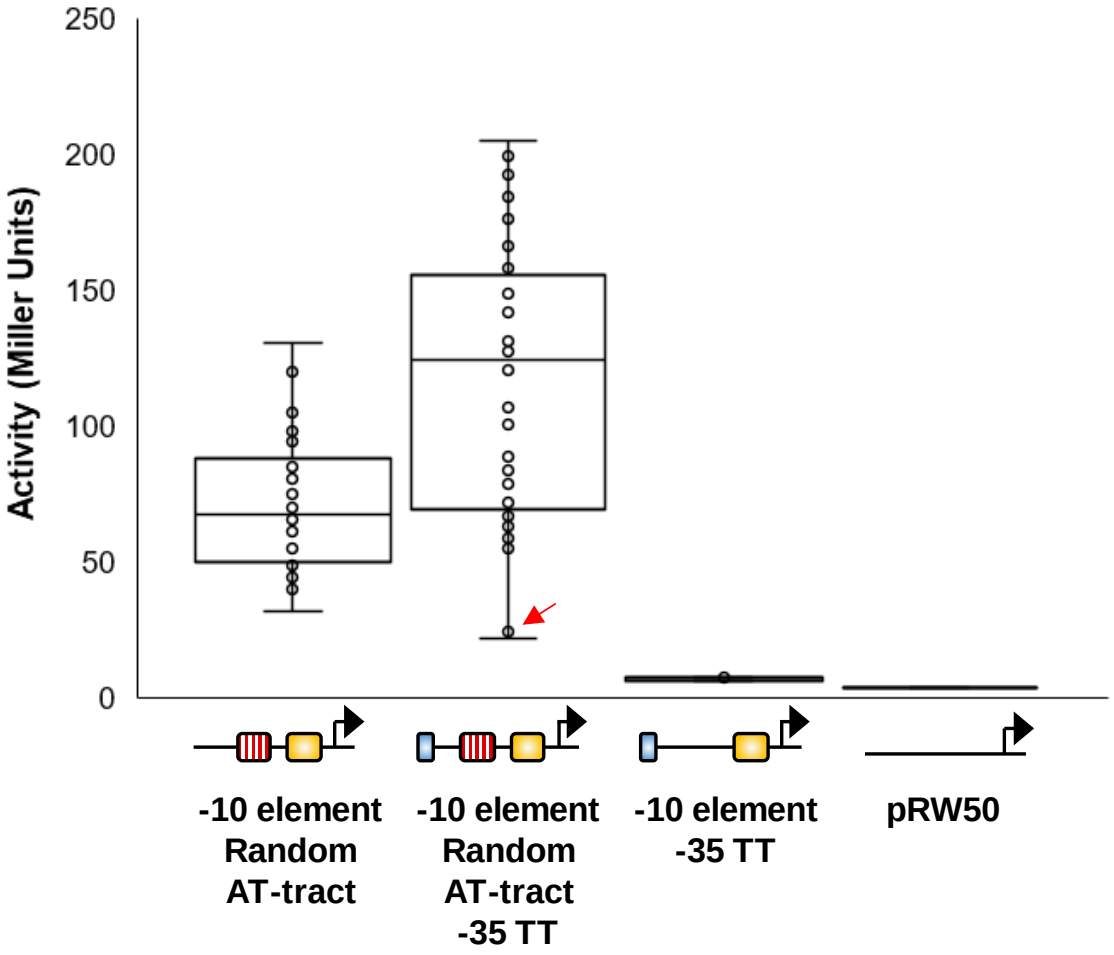


Figure 19: Promoter activity with a randomised AT-tract

Promoter sequences containing a -10 element (yellow box), with a partial -35 element (blue box) and/or a randomised AT-tract (pink striped box), were fused upstream of *lacZ* in pRW50 to create a fusion library. This library was used to transform JCB387 *E. coli* and LacZ activity of each colony was measured. Data is presented as a box plot with each individual point shown, compared to the activity of a promoter without an AT-tract (measured previously in Figure 3) and baseline activity (empty pRW50) (Warman *et al.* 2020).

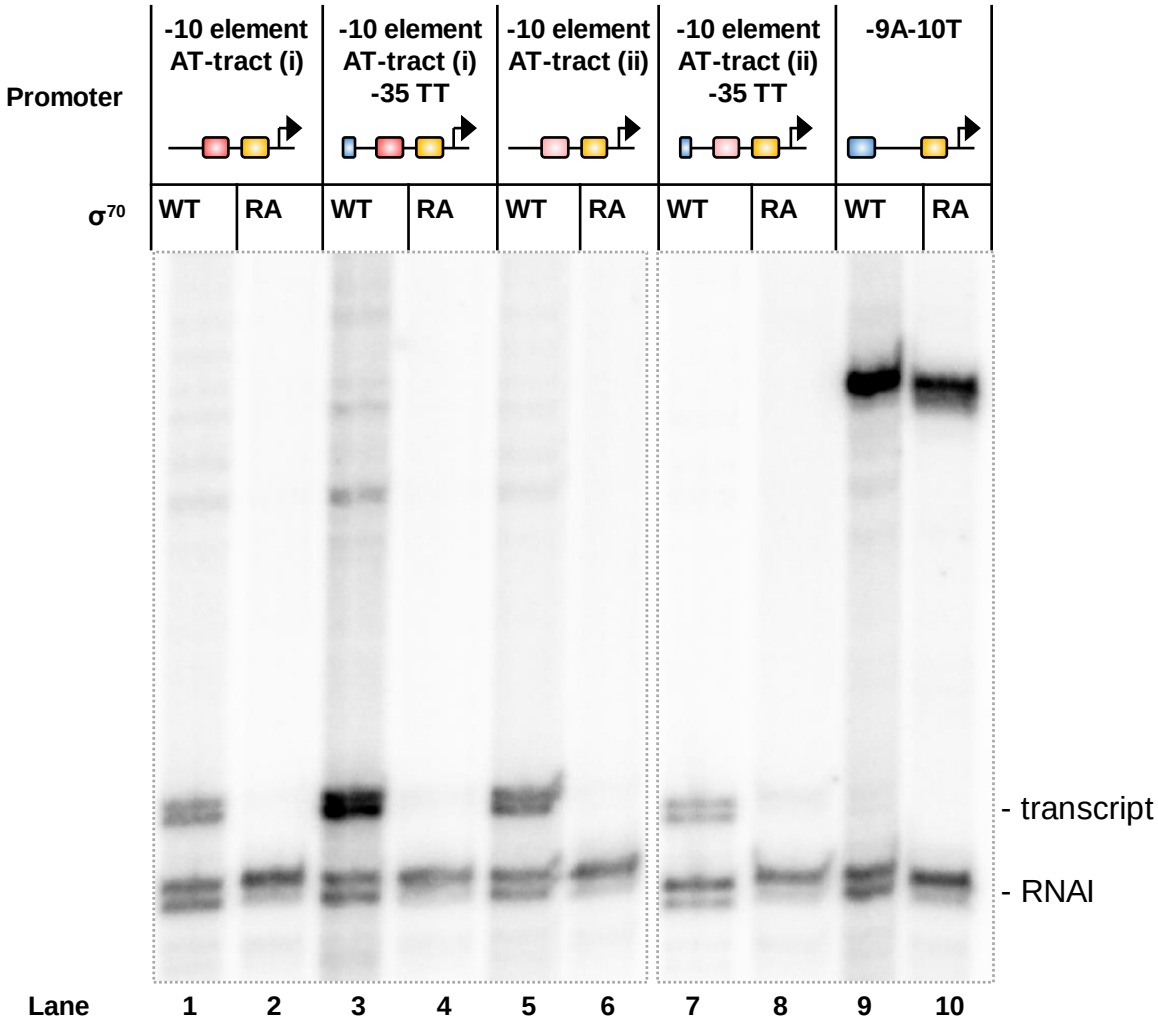


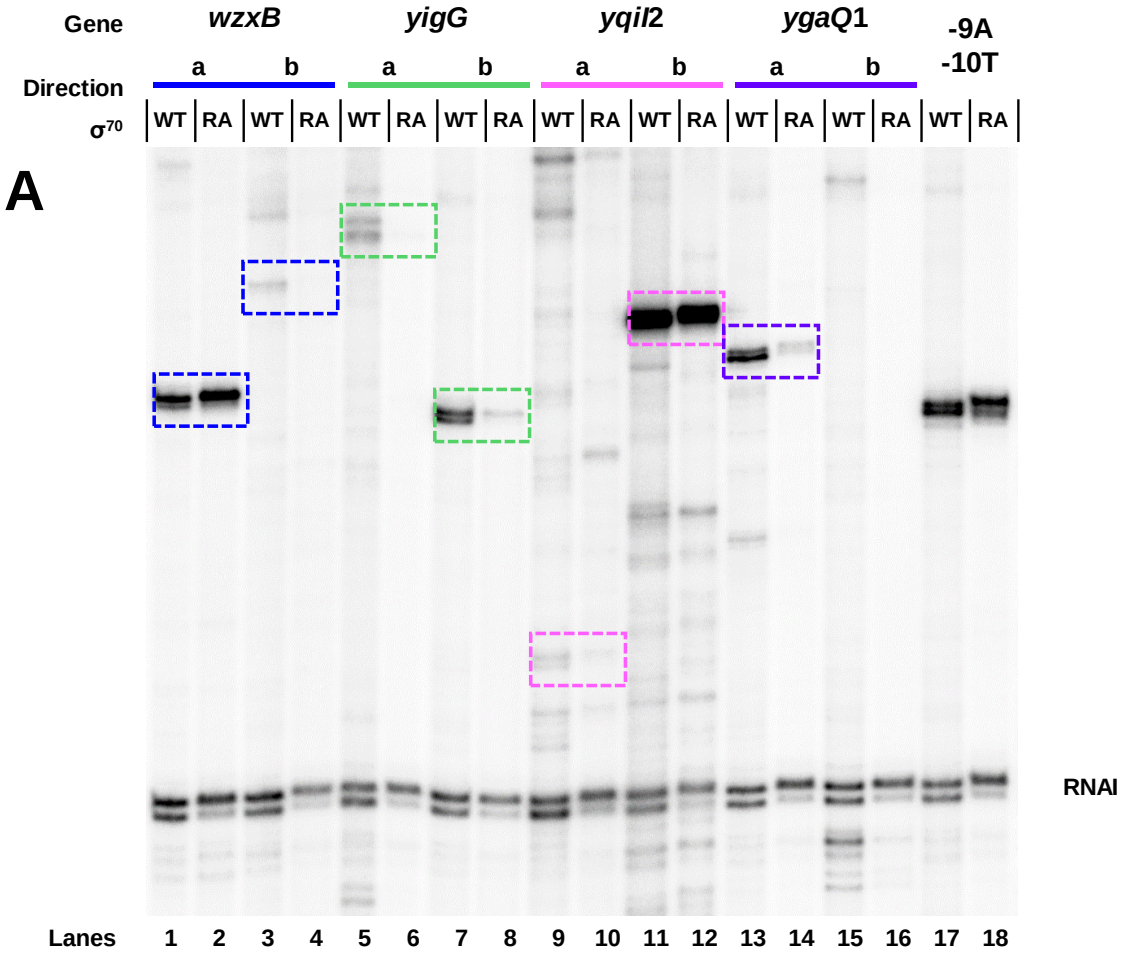
Figure 20: Effect of σ^{70} mutation RA451 on *in vitro* transcription initiation from AT-tract promoters

Promoter sequences containing a consensus -10 element 'TATAAT' (yellow box), an AT-tract with sequence 'TATTTAT' (red box) or sequence 'AATTT' (pink box), and/or a partial -35 element 'TT' (blue box) were fused into pSR. Transcription was initiated by RNA polymerase with either wild-type (WT) σ^{70} or the R451A (RA) σ^{70} mutant. The gel image shows transcripts derived from these promoters. Bands at 107/108 nt are RNA-I transcripts derived from the pSR plasmid and used as a loading control. -9A-10T is a constitutively active promoter which is not dependent on R451 (Warman *et al.,* 2020).

element and a good match to the -35 element. This promoter does not rely on R451 and produces similar amounts of transcript with the WT and mutant σ^{70}.

3.2.5. Many H-NS repressed intragenic promoters require R451

We identified numerous intragenic promoters in H-NS-bound regions, containing an AT-tract. We reasoned that these promoters would also be dependent on σ^{70} R451. To test this, twelve intragenic promoters identified by Singh *et al*. (2014) were selected and assayed for activity by *in vitro* transcription. The promoters were within the coding sequences of *wzxB, yigG, ygaQ*, *trmA*, *wcaD, ybcV, yhcD, lpxD ribB* and *stfP.* Two non-overlapping promoters within *yqiI (yqiI*1 and *yqiI*2) were also included. DNA sequence spanning promoter regions was amplified by PCR and cloned in pSR. Because Singh *et al*. (2014) identified some promoters described above on the basis of RNAP binding, rather than transcription initiation, TSSs were not known. This applied to *wzxB, yigG, ygaQ* and *yqiI* 2. Therefore, since transcription might initiate from these sites in either direction, DNA fragments were cloned in both orientations (a and b). Transcription initiating from each promoter *in vitro* was detected using RNAP with WT σ^{70} or the σ^{70} R451A derivative (Figure 21). Strong transcription was observed for promoters in *wzxB* a, *yigG* b, *yqiI*2 b and *ygaQ* a (Figure 21A; lanes 1, 7, 11 and 13); and for *wcaD, yqiI* 1 and *lpxD* (Figure 21B; lanes 3, 9 and 11). Transcription from promoters within *yigG*, *ygaQ, wcaD* and *lpxD* was greatly reduced by the R451A mutation (Figure 21A; lanes 8 and 14, Figure 21B; lanes 4 and 12). Conversely, transcription derived from the *wzxB, yqiI*2 and *yqiI*1 DNA fragments was unchanged (Figure 21A; lanes 2 and 12, Figure 21B; lane 10). Fainter bands indicated weaker promoters within *wzxB* b, *yigG* a and *yqiI*2 a (Figure 21A; lanes 3, 5, and 9). Faint


A
Gene
wzxB
yigG
yqiI2
ygaQ1
-9A
-10T
Direction
a
b
σ70
WT
RA
RNAI
Lanes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

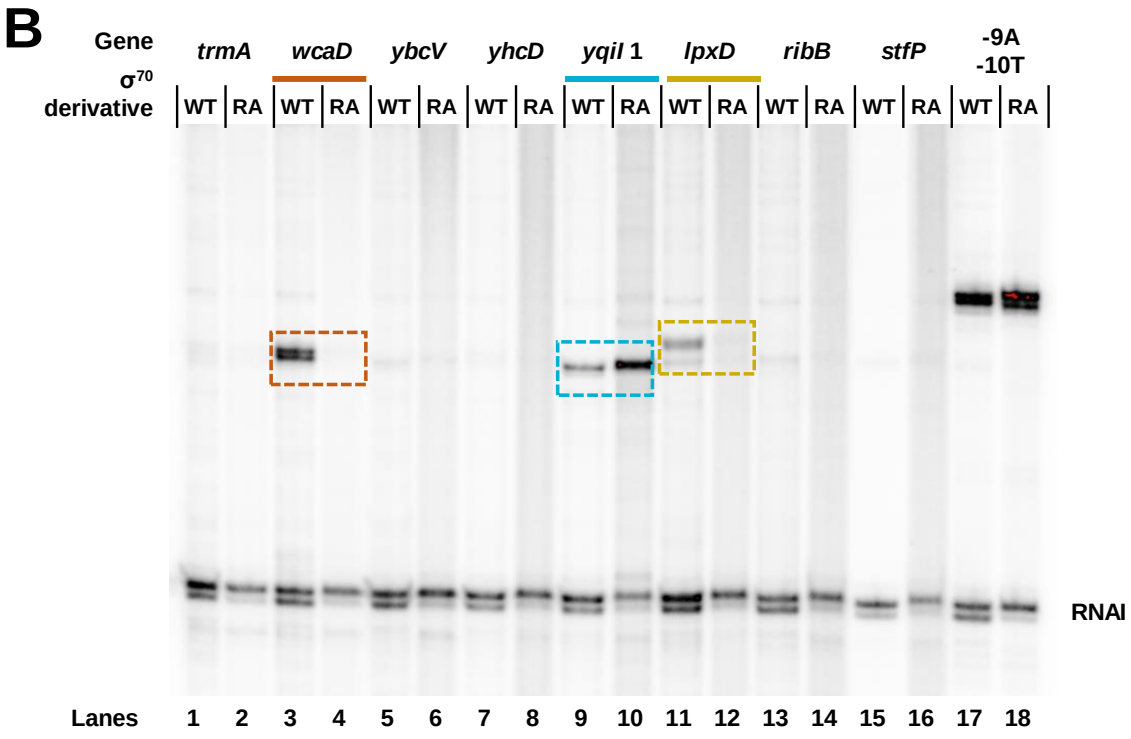

B
Gene
σ70
derivative
trmA
wcaD
ybcV
yhcD
yqiI 1
lpxD
ribB
stfP
-9A
-10T
WT
RA
RNAI
Lanes
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Figure 21: Effect of σ^{70} mutation RA451 on *in vitro* transcription initiation from intragenic promoters

A) Promoter regions isolated from within *wzxB, yigG, yqiI* and *ygaQ* were fused into the plasmid pSR in both 'a' and 'b' orientations. B) Promoter regions from within *trmA, wcaD, ybvC, yhoD, yqiI*, *lpxD, ribB* and *stfP* were fused into pSR in the 'sense' direction. For both sets of promoters, transcription was initiated by RNA polymerase with either wild-type (WT) σ^{70} or the R451A (RA) σ^{70} mutant. The gel image shows transcripts derived from these promoters. Coloured boxes indicate transcripts from strong promoters. Bands at 107/108 nt are RNAI transcripts. -9A-10T is a constitutively active promoter which is not dependent on R451 (Warman *et al.* 2020).

bands were also seen for reactions with the *trmA, ybcV, yhcD, ribB* and *stfP* DNA fragments (Figure 21B; lanes 1, 5, 7, 13 and 15). All of these transcripts were lost with the mutant σ^{70}.

3.2.1. R451-dependent intragenic promoters also require an AT-tract

These results show that not all promoters with AT-tracts require the interaction with σ^{70} R451. To identify differences that might explain the varying dependence, promoter elements were compared and ranked by similarity to the consensus σ^{70} promoter elements (Table 8). All promoters except *lpxD* had AT-tracts (at least 5 out of 7 A/T bases between positions -17 and -23). Only promoters with at least 4/6 positions matching the -35 element consensus (*wzxB, yqiI*2 and *yqiI*1) were able to initiate transcription independently of the σ^{70} R451 residue (Figure 21). Conversely, R451-dependent promoters had weaker -35 elements and were also more likely to have extended -10 elements. This suggests that, for a promoter to be active, it must have either an AT-tract (with or without an extended -10 element), or a -35 element. If this hypothesis is correct, removing the AT-tract from R451-dependent promoters would render them inactive. In contrast, promoters with a strong -35 element would not require the AT-tract. To test this, the entire spacer region was replaced with a 60 % GC sequence lacking an AT-tract. Extended -10 elements were preserved. These 'GC-spacer' promoters were assayed with WT or R451A σ^{70} alongside the native promoter sequences (Figure 22). Most of the promoters behaved as predicted; *lpxD, yqaQ, yigG* and *wcaD* were defective both in the absence of R451 and when the AT-tract was removed. The extended -10 elements in *ygaQ, yigG* and *wcaD* were not able to independently stimulate transcription. Also as expected, R451-independent promoters

Table 8: Predicted intragenic promoter elements.

Promoters are aligned at the -10 position. TSSs are in bold underlined, predicted -10 elements are highlighted in orange, AT-tracts in red and -35 elements in blue. The similarity of each promoter element to the σ^{70} consensus promoter is shown, and whether the promoter is dependent on R451 (Warman *et al*., 2020).

Gene	Promoter sequence	-35 consensus	-10 consensus	R451 dependent
lpxD	**cacata**acgtcg**tgattgg**cgac**aatacg**gcggtt**g**	2/6	2/6	Yes
ygaQ	**ttatga**attacat**ggaatatc**t**gg**t**aactt**gtcagtt**g**	3/6	3/6	Yes
yigG	**tatata**aaaact**tatgtttt**t**gg**t**acatt**agcagta**t**	3/6	4/6	Yes
wcaD	**atcaaa**acgcct**aaaacag**a**tg**c**tatgat**tctcgca**g**	3/6	5/6	Yes
yqil 1	**tttact**ttattg**gcatttt**tagc**tagctt**aatata**a**	4/6	3/6	No
yqil 2	**ttgacc**tgatgct**aacaaca**tcat**tatatt**tgcccat**g**	5/6	5/6	No
wzxB	**ttgaaa**ccatat**tcaacaa**acat**tatcat**atatagc**a**	5/6	5/6	No
-9A -10T	**ttggca**tatgaa**attttga**ggat**tatact**acactt**a**	5/6	5/6	No

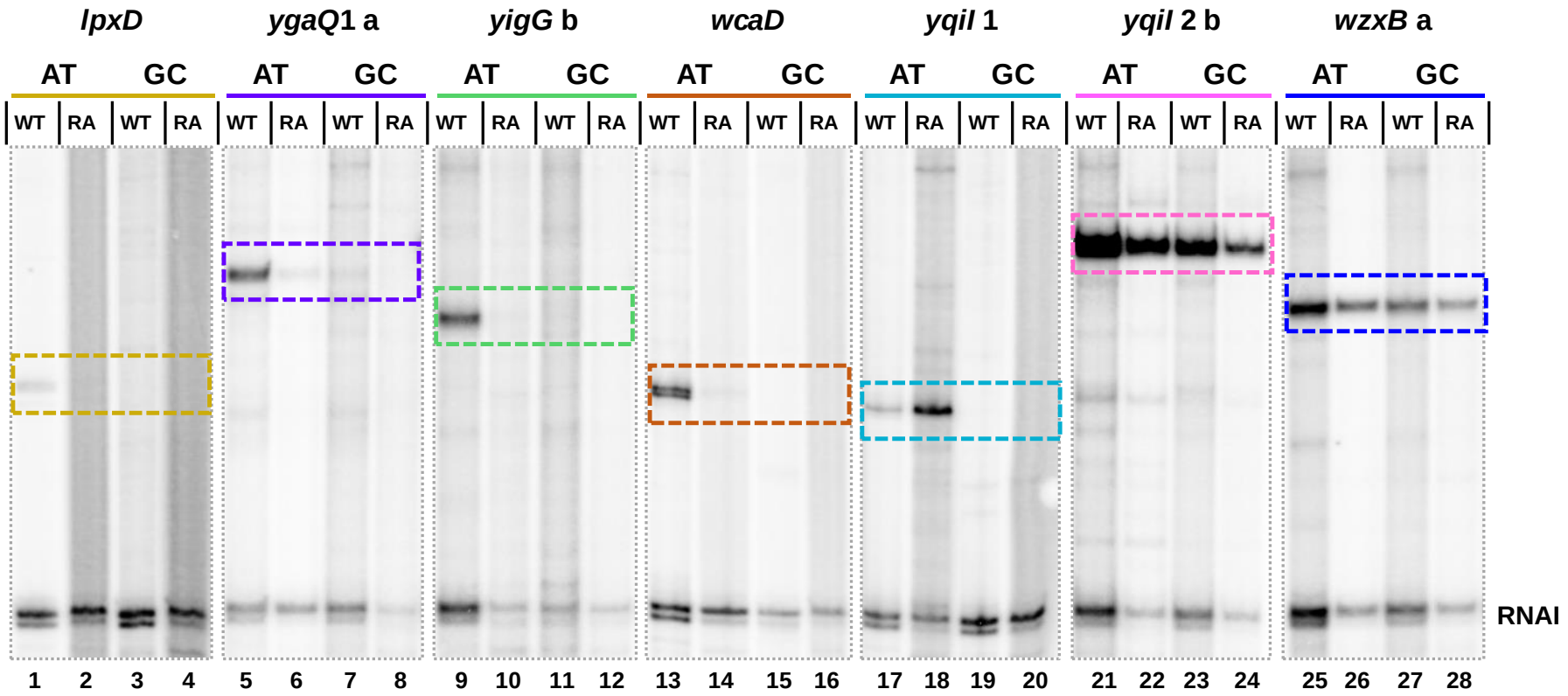


Figure 22: Effect of removing the AT-tract on *in vitro* transcription from intragenic promoters

AT-rich spacer sequences were replaced with GC-rich spacers. These 'GC' promoters were fused into the plasmid pSR. Transcription was initiated by RNA polymerase with either wild-type (WT) σ^{70} or the R451A (RA) σ^{70} mutant. The gel image shows transcripts derived from both native and GC promoters. Bands at 107/108 nt are RNA-I transcripts derived from the plasmid. -9A-10T is a constitutively active promoter which is not dependent on R451 (Warman *et al*., 2020).

*yqiI*2 and *wzxB* were not significantly affected by the loss of an AT-tract. The only promoter which did not fit the hypothesis was *yqiI*1, which was able to initiate transcription with the mutant σ factor, but only when the AT-tract was present (Figure 22, lane 18).

3.3. Discussion

These experiments show that a correctly positioned AT-tract is able to function as an accessory promoter element, stimulating transcription at promoters which would otherwise be silent. This provides an explanation for the abundance of promoters inside AT-rich genes; not only are AT-rich DNA sequences more likely to contain a -10 element, they are also more likely to have an AT-tract in the correct location. The mechanism of activation is likely driven by altered DNA bending rather than sequence; any combination of A and T in the AT-tract region is capable of increasing transcription. Previous studies have suggested that the base at position -18 can modulate transcription levels, with a T having the most stimulatory effect, followed by an A (Singh *et al.*, 2011). The AT-i and AT-ii sequences had an A or a T respectively at position -18 and this difference did not have a significant effect on the activity of the promoter. Both AT-tracts were more effective in combination with the partial -35 element.

The AT-tract facilitates an interaction between σ^{70} residue R451 and the DNA backbone. Mutating R451 results in a σ^{70} unable to participate in transcription initiation at promoters dependent on an AT-tract, but is sufficient for transcription at canonical promoters. However, it is important to note that the R451A σ^{70} mutant decreases transcription from all promoters; the data presented here has been adjusted to show

relative changes. This suggests that R451 has a general role in transcription initiation, but is essential only for promoters with an AT-tract.

Most of the intragenic promoters tested were rendered inactive by removal of either the AT-tract or R451. However, some intragenic promoters are not dependent on the AT-tract/R451 interaction; this is often because their promoter elements more closely match the –10 and -35 hexamers for a σ^{70} promoter. Hence, the promoters form tighter complexes with RNAP, via the contact between the -35 hexamer and σ^{70} region 4.2 (Paget, 2015). This negates the requirement for an AT-rich spacer. Two intragenic promoters do not fit this model. The TSS in *lpxD* does not have an upstream AT-tract, and yet promoter activity is dependent on both the spacer sequence and the R451 residue. The *yqiI*1 promoter does not require the R451 residue for transcription, but does depend on the AT-tract. Overall, these results suggest that AT-tracts are able to activate promoters with sub-optimal promoter elements, but that the spacer sequence/σ^{70} interaction is more complicated than previously described.

Chapter 4. **Widespread Divergent Transcription at Prokaryotic Promoters**

4.1. Introduction

Promoters can be identified by mapping RNAP binding sites, RNA 5' ends or matches to the consensus DNA sequence. However, it is difficult to identify intragenic promoters within horizontally acquired genes; H-NS binding blocks access of the transcription machinery and RNA will not be produced. To combat this, previous work combined σ^{70} ChIP-seq analysis and PPP-seq experiment, to map RNAP binding and RNA 5' ends in a Δ*hns* strain. This identified 668 promoters usually repressed by H-NS, many of which were intragenic. The study also found many σ^{70} binding sites where no nearby RNA 5' end was found. Such sites were excluded from the prior study. However, there are many reasons why intragenic promoters may not produce detectable RNA. For example, transcription of non-coding RNA from intragenic promoters is terminated by Rho (Cardinale *et al.*, 2008). Furthermore, in the previous chapter, we were able to detect transcription from some such RNAP binding sites. In this chapter, I have explored these sites further. I show that such RNAP binding sites are often associated with divergent transcription. Further inspection reveals that divergent transcription is a common property of bacterial promoters. Divergence results from the inherent symmetry of promoter -10 elements.

4.2. Results

4.2.1. Intragenic promoter sequences frequently drive divergent transcription

Using σ^{70} ChIP-seq data (Singh *et al.*, 2014), σ^{70} binding peaks in Δ*hns,* but not WT cells, were selected. All such peaks were inside genes, and four peaks (within *wzxB, yigG, yqiI* and *ygaQ*) corresponded to sites studied in Chapter 3. A 150 bp region of DNA under each peak was amplified using PCR and fused to the *lacZ* gene in plasmid pRW50. As the direction of transcription was unknown, and could not be inferred from the ChIP-seq data, DNA was cloned in both orientations, referred to as 'a' and 'b'. Promoter activity was measured by LacZ assay. Note that fragments tested were short enough that H-NS was not expected to bind and repress transcription (Haycocks *et al.*, 2015). Figure 23 shows LacZ activity for each of the 33 fragments in both the 'a' and 'b' orientations. 30 of the fragments had an active promoter in at least one direction. Hence, σ^{70} binding peaks were associated with transcription despite the absence of a transcript detected by RNA-seq. Intriguingly, the majority of the promoter regions drove divergent transcription; LacZ activity was observed for both a and b orientations.

4.2.2. Canonical promoters are more directional

There was a chance that divergent transcription was an artefact due to some inherent property of the plasmid. To rule this out, an equivalent number of canonical promoters (i.e those from intergenic regions upstream of genes) were cloned in sense and antisense orientations and assayed for their ability to drive LacZ expression. Antisense transcription is not expected from canonical promoters. Briefly, 30 canonical promoters were randomly selected from a list compiled by Mitchell *et al*. (2003). We excluded any promoters known to have an opposing canonical promoter within 75 bp, for example,

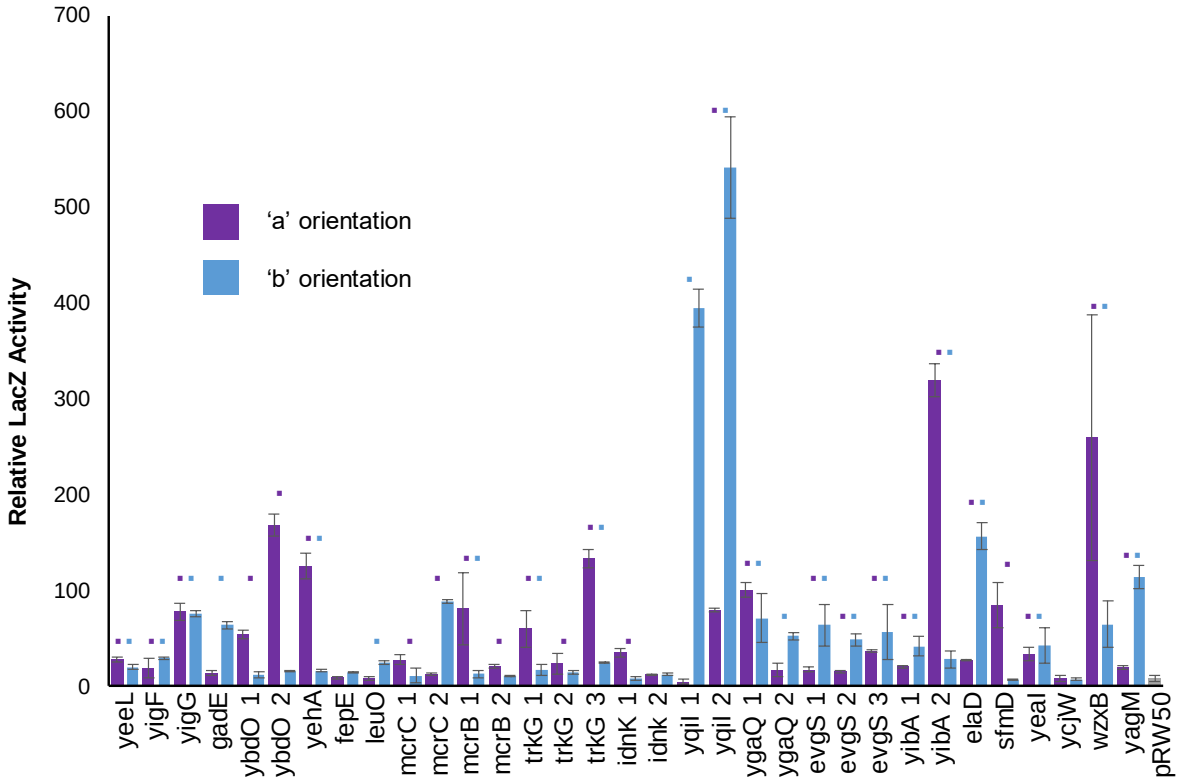


Figure 23: Intragenic promoter assays.

150 bp DNA fragments corresponding to intragenic σ^{70} binding sites were fused upstream of *lacZ* in pRW50**.** Each fragment was assayed in the forward (a, purple bars) and reverse (b, blue bars) orientation. Assays were done in triplicate. Activity values were normalised to the mean activity of *ygaQ* a, which was used as a standard in all assays. Empty pRW50 was used as a control. A dot (·) denotes where the activity of the 'a' orientation (purple ·) or 'b' orientation (blue ·) was at least two fold higher than background. Error bars: ± SD.

many promoters between divergent genes. 150 bp of DNA sequence surrounding each of the 30 promoters was fused to LacZ in the plasmid pRW50 and assayed for LacZ activity. The results from this assay are shown in Figure 24. Only the promoters for the genes *adiY*, *gdhA*, *pheL* and *ssuE* displayed antisense transcription 2-fold higher than background (empty pRW50 activity). Five of the promoters assayed were not active in the sense or antisense directions. This inactivity is most likely because the assay conditions lacked a key inducer or because important regulatory DNA was not present. The DNA fragments from intragenic regions (Figure 23), and the assayed canonical promoters (Figure 24), were classified as either directional (activity in one orientation), divergent (activity in both orientations) or inactive (no activity). Pie charts comparing the proportions of each class of promoter region are shown in Figure 25. More than half of intragenic promoters drove divergent transcription. In comparison, only 13.3 % of canonical promoters had antisense promoter activity, and antisense activity tended to be very low.

4.2.3. Intragenic divergent promoter elements frequently overlap

We next wanted to provide an explanation for the high occurrence of divergent transcription from intragenic DNA fragments. Five regions which had promoters in both orientations (*yibA, wzxB, yigG, yqil*2 and *ygaQ*1) were selected for further analysis. The exact transcription start site of each promoter was mapped by primer extension. Briefly, RNA was extracted from cells carrying the various *lacZ* fusions. Reverse transcription was done using the labelled primer D49724, which binds to the *lacZ* mRNA. After extension, the size of the resulting radiolabelled cDNA identifies the 5' end of the transcript. The different cDNA fragments were separated by electrophoresis

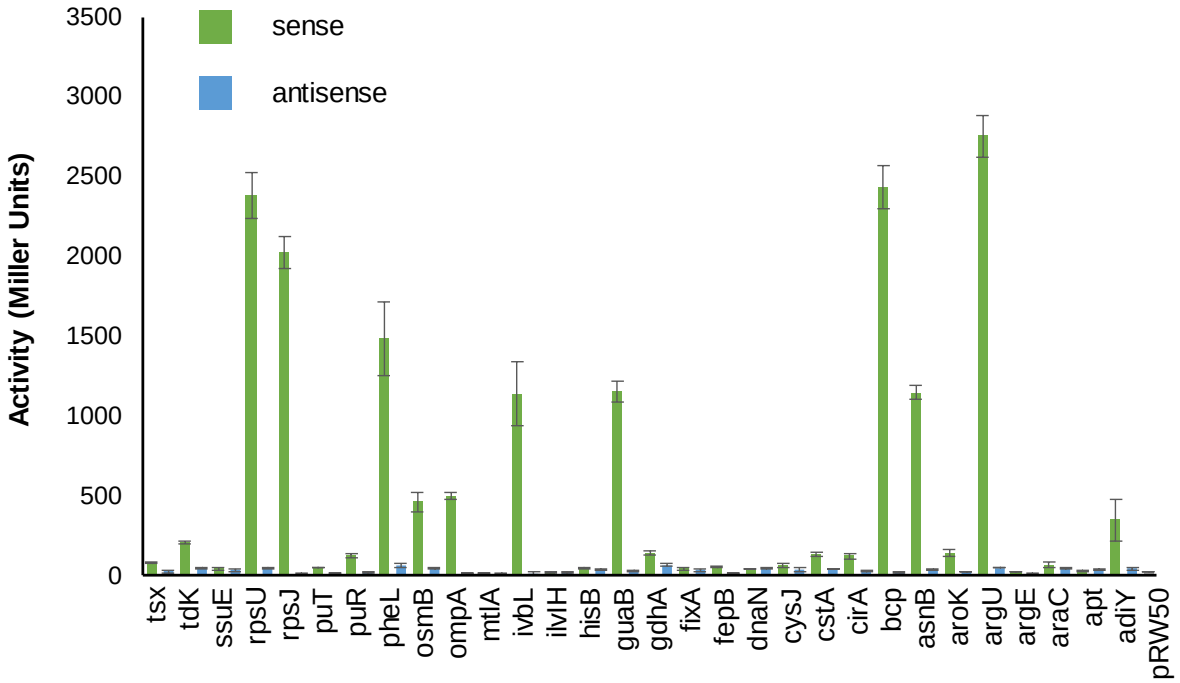


Figure 24: Canonical promoter assays.

150 bp DNA sequences containing canonical promoters were fused upstream of *lacZ* in pRW50. Promoter regions were assayed in both orientations, sense (green bars) and antisense (blue bars). Empty pRW50 was used as a negative control. For each promoter, LacZ activity was measured in triplicate. Data is presented as the mean of the replicates in Miller Units. Error bars show +/- SD.

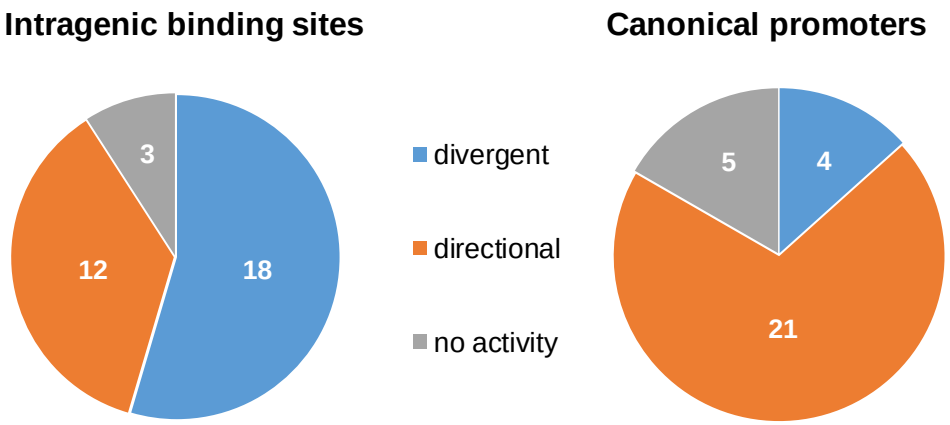


Figure 25: Directionality of intragenic and canonical promoter regions.

Pie charts representing the proportion of assayed promoters which were directional (activity in one direction), divergent (activity in both directions) or inactive (no activity in either direction).

on a polyacrylamide gel (Figure 26) alongside an M13 sequencing reaction for calibration. Transcript-derived cDNA was observed for all promoter regions. Start site mapping was confirmed by *in vitro* transcription; promoters were cloned into the plasmid pSR and transcribed by RNAP/σ^{70} holoenzyme. Transcripts were separated on a polyacrylamide gel alongside a GA ladder (Figure 27). TSS positions were calculated from bands that appeared in both Figure 26 and Figure 27. The sizes of the cDNA band and RNA band, derived from the same promoter, differ as they are generated in different ways. Both molecules can be used to determine the TSS, by considering the position of the promoter relative to the primer or the terminator utilised in the primer extension or *in vitro* transcription experiments respectively. Additional bands were seen in the *in vitro* transcription assay, but as these bands were not seen *in vivo*, they were not investigated further. The start sites are shown on annotated sequences in Figure 28. Predicted promoter elements were identified upstream of the TSSs and are shown in yellow (-10 elements) and blue (-35 elements). Most of the promoters on both strands had an AT-tract (highlighted in red).

Four of the intragenic regions, *wzxB, yigG, yqiI*2 and *ygaQ*1, had overlapping promoter elements on opposite DNA strands. These are shown in more detail in Figure 29. The promoters within *wzxB* had overlapping -10 elements; the TSS of the top promoter corresponded to the extended -10 element of the promoter below. In *yigG*, the transcription start sites aligned with the -10 elements of the opposite strand. The -35 elements of the promoters in *yqiI* 2 overlapped with the opposing AT-tracts, and in *ygaQ* the -35 elements almost completely overlapped. These promoter sequences were bidirectional; transcription on opposite strands initiated at the same shared

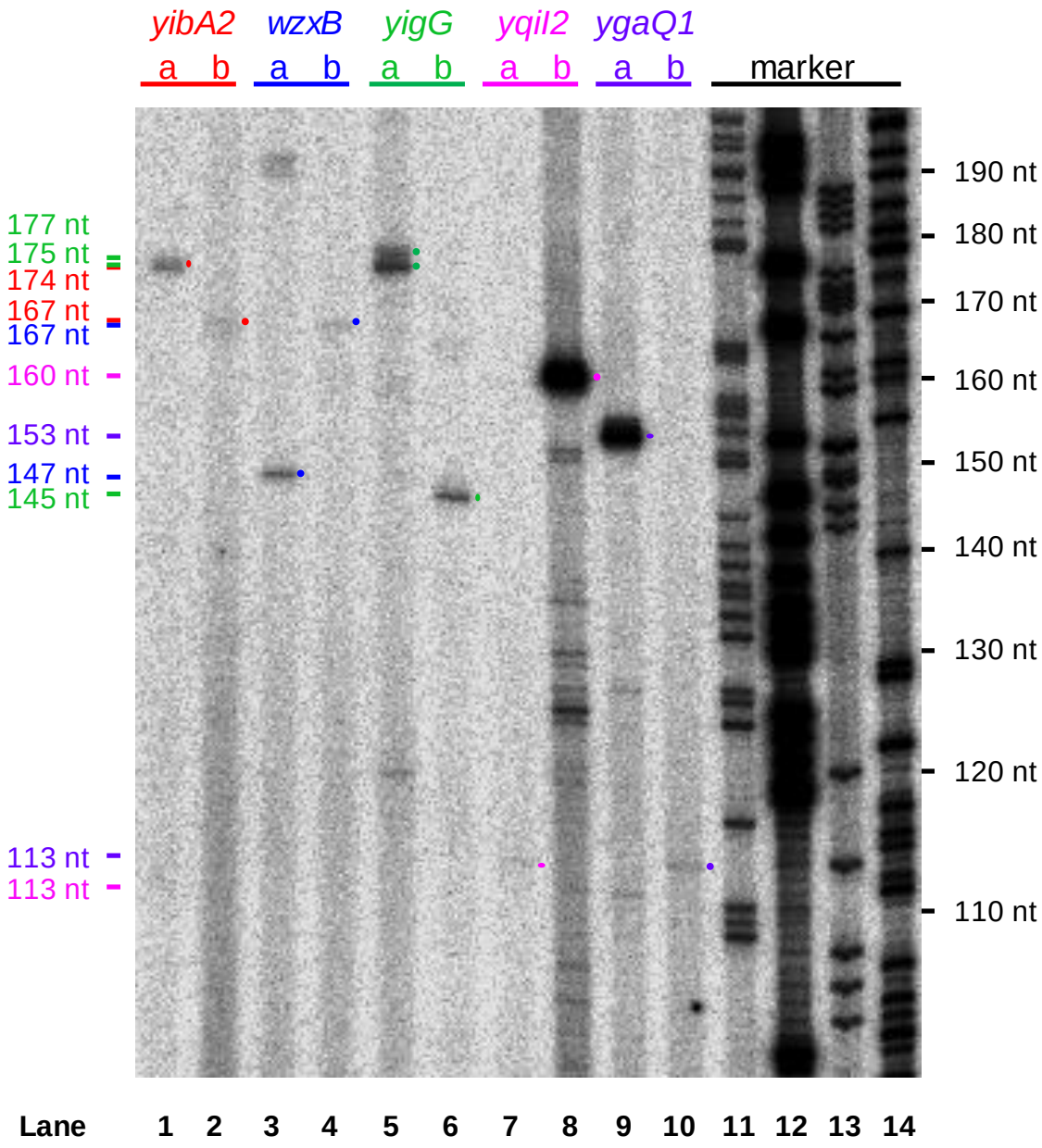


Figure 26: TSS mapping by primer extension

RNA transcripts derived from intragenic promoters cloned in pRW50 were reverse-transcribed using a labelled primer. The resulting labelled cDNA fragments were separated on a 6% polyacrylamide gel. A sequencing reaction of M13 phage DNA was used as a marker to enable mapping of RNA 5' ends and therefore transcription start sites. Band lengths are marked.

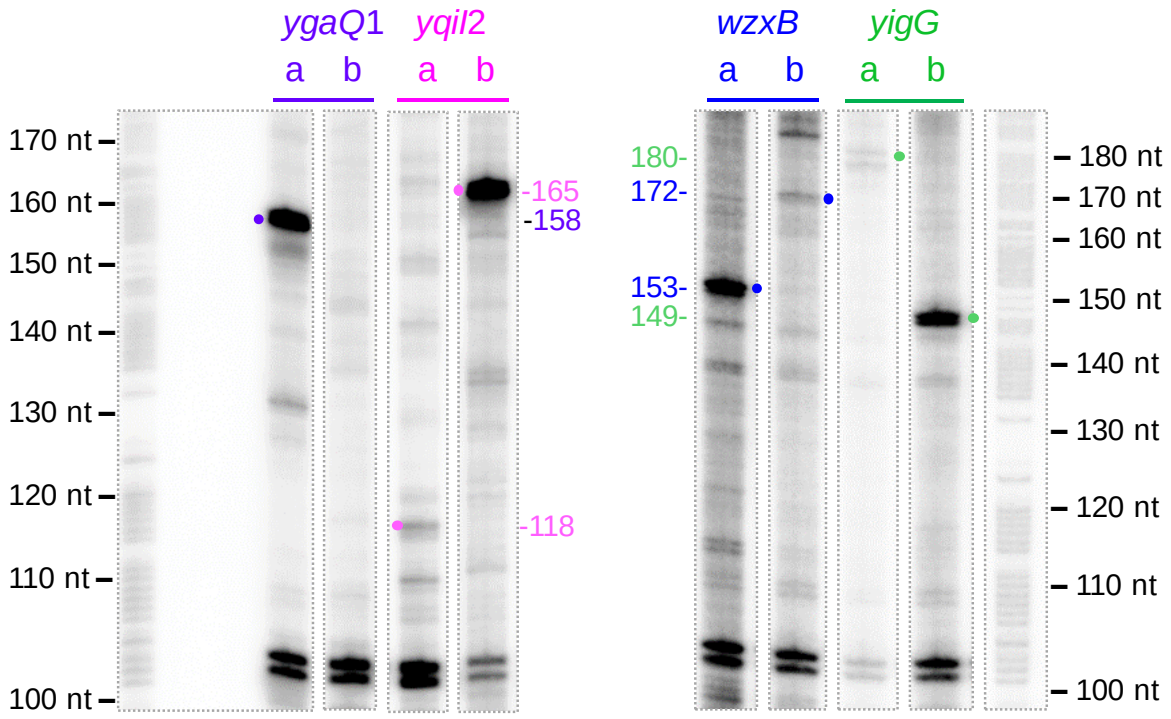


Figure 27: TSS mapping by in vitro transcription

RNA transcripts derived from intragenic promoters fused into pSR were generated with RNAP and WT σ^{70}. Transcripts were separated on a 6% polyacrylamide gel alongside a GA ladder of a known sequence. Band lengths are marked.

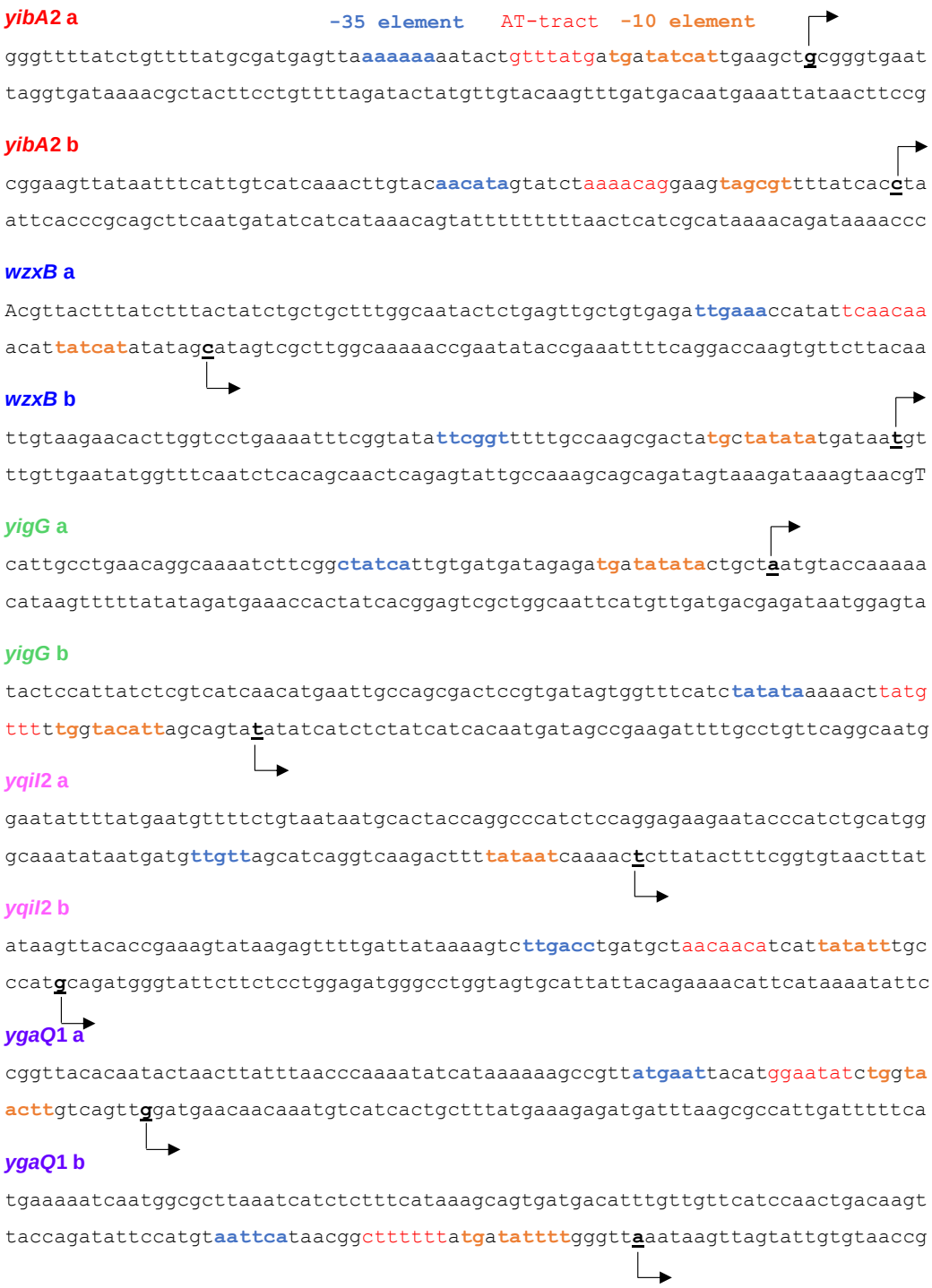


Figure 28: Predicted promoter elements in intragenic promoters.

Sequences of the 150 bp promoter regions are shown in the a and b orientations. Transcription start sites are bold and underlined with arrows indicating the direction of transcription. Predicted -10 elements are highlighted in orange, AT-tracts in red and -35 elements in blue.

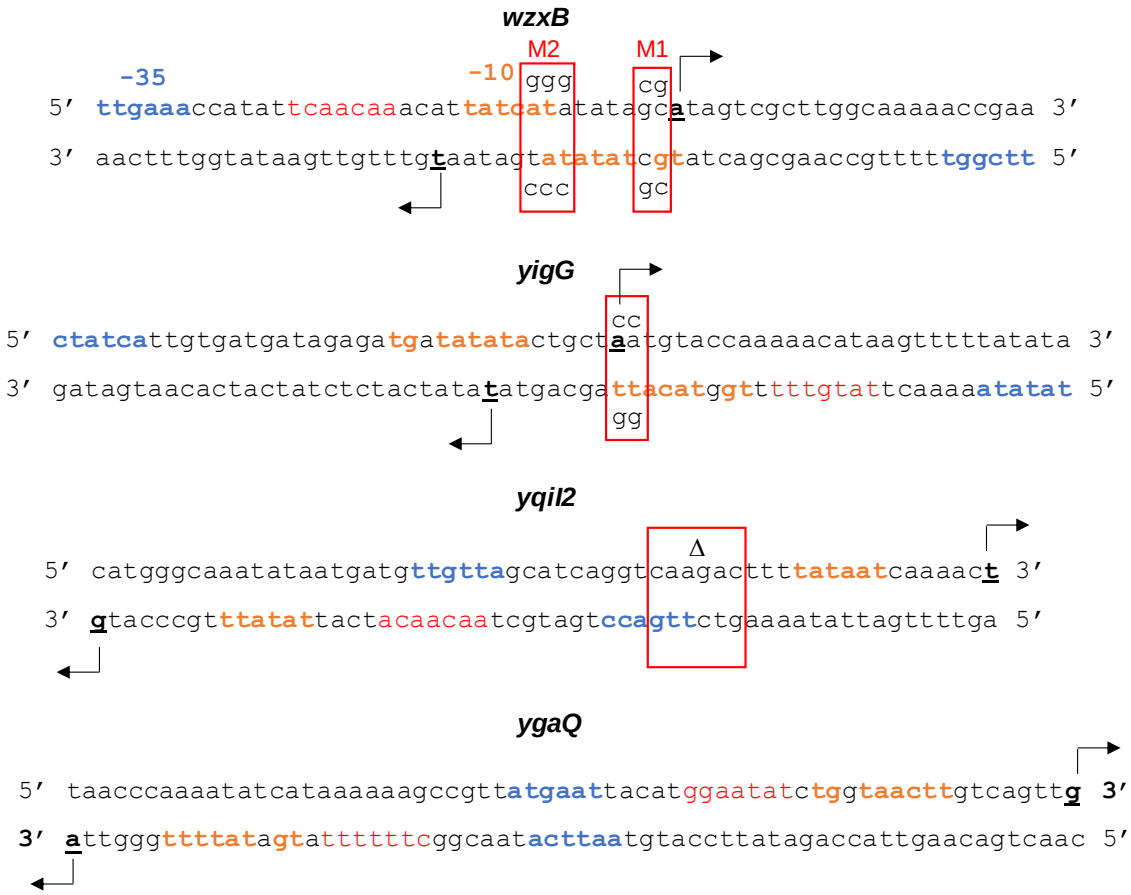


Figure 29: Sequences of bidirectional intragenic promoters

Sequences of the bidirectional part of promoter regions are shown in the ‘a’ orientation – for the ‘b’ orientation read the bottom strand from 5' to 3'. Transcription start sites are underlined. Predicted -10 elements are shown in orange, AT-tracts in red and -35 elements in blue. The mutated sequences are enclosed in a red box with the new bases/deleted bases indicated.

promoter region. The promoter elements in *yibA* did not overlap and so were not included in the following analysis.

To confirm that these bidirectional promoters were reliant on shared sequence elements, mutations were designed to simultaneously reduce transcription in both directions (red boxes in Figure 29). In *wzxB,* the overlapping section of the -10 element for each promoter was replaced with a GC rich sequence. Separately, a mutation was introduced to disrupt the extended -10 element on the bottom strand. The mutation in *yigG* replaced part of the -10 element on the bottom strand with 'GG', which consequently swapped the start site of the opposite promoter with a C. A short section of the *yqiI* promoter was deleted, removing part of the spacer sequence of the top strand promoter and part of the -35 element of the bottom strand promoter. These mutated promoters were fused to *lacZ* in the plasmid pRW50 and levels of LacZ activity were used to assess changes in promoter activity. As shown in Figure 30, mutations (dashed columns) decreased transcription in both orientations. The most effective mutation was the deletion in *yqiI*, particularly for the b orientation, which shows the importance of the -35 element for the activity of this promoter. Mutating the extended -10 element in *wzxB* b was also particularly inhibitory which suggests and important role for this extended motif.

4.2.4. Widespread divergent transcription from bidirectional promoters

The occurrence of divergent promoters in both intragenic promoter regions and canonical promoters was higher than expected (Figure 25). To determine global prevalence of closely spaced divergent TSSs in *E. coli*, 3 TSS maps, generated by

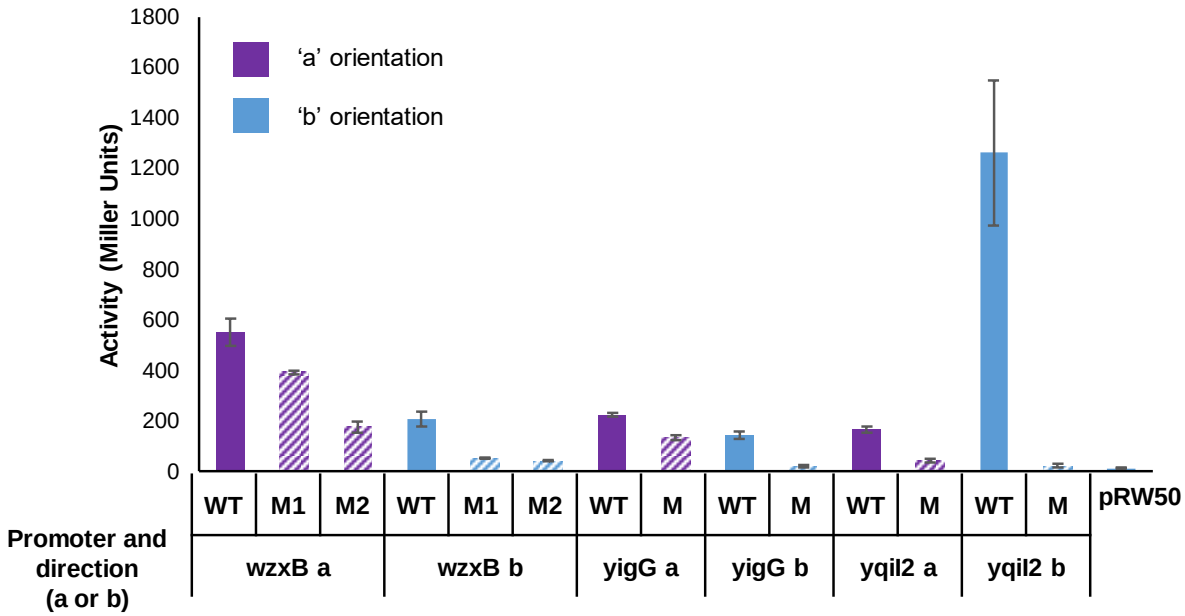


Figure 30: Effect of mutating shared promoter elements on promoter activity.

Wild-type (WT, solid bars) or mutated (M, dashed bars) intragenic promoters were fused upstream of *lacZ* in pRW50. Promoter regions were assayed in both directions, orientation 'a' (purple bars) and 'b' (blue bars). Empty pRW50 was used as a negative control. For each promoter, LacZ activity was measured in triplicate. Data is presented as the mean of the replicates in Miller Units. Error bars show +/- SD.

RNA 5′ polyphosphatase sequencing (PPP-seq), dRNA-seq and cappable-seq, were combined (data from Singh *et al*., 2014; Thomason *et al*., 2015; Ettwiller *et al*., 2016). The distance between each TSS on the top strand, and the closest TSS on the bottom strand, was calculated. Figure 31 presents this data as a heat map which shows that TSSs on opposite strands tend to co-locate. A total of 5,292 bidirectional promoters were identified, representing 19 % of all *E. coli* promoters. These bidirectional promoters are defined as regions with two opposing TSSs separated by between 29 and 7 base pairs. The frequency of promoter pairs separated by each distance in this range were plotted in Figure 32. There are 6 distances which appear more frequently; bottom strand TSSs preferentially appear at positions -29, -23, -18, -12, -10 and -7 with respect to top strand TSSs (Figure 32A). Bidirectional promoter sequences with TSSs separated by these distances were combined to generate sequence logos (Figure 32B). The configurations of overlapping promoter elements were determined and are shown below the sequence logos. The 29 bp spacing was the only configuration where the top strand -10 element did not align with either the -10 element or the TSS of the bottom strand. TSSs separated by 23 and 18 bp had directly overlapping -10 elements. The most important positions of the -10 hexamer (underlined bases) also acted as key bases in the -10 element on the opposite strand, demonstrating the symmetrical nature of the -10 element. For example, in the -18 spacing the last T of the -10 element served as the second A of the opposing -10 element. In the -12 and -7 bp configurations these important positions acted as the TSSs for the opposing promoters. There was a distinct decrease in the number of TSS pairs aligning at the same position (i.e. a spacing of 0); likely because a TSS tends to be an A or a G making the opposite TSS an uncharacteristic T or C.

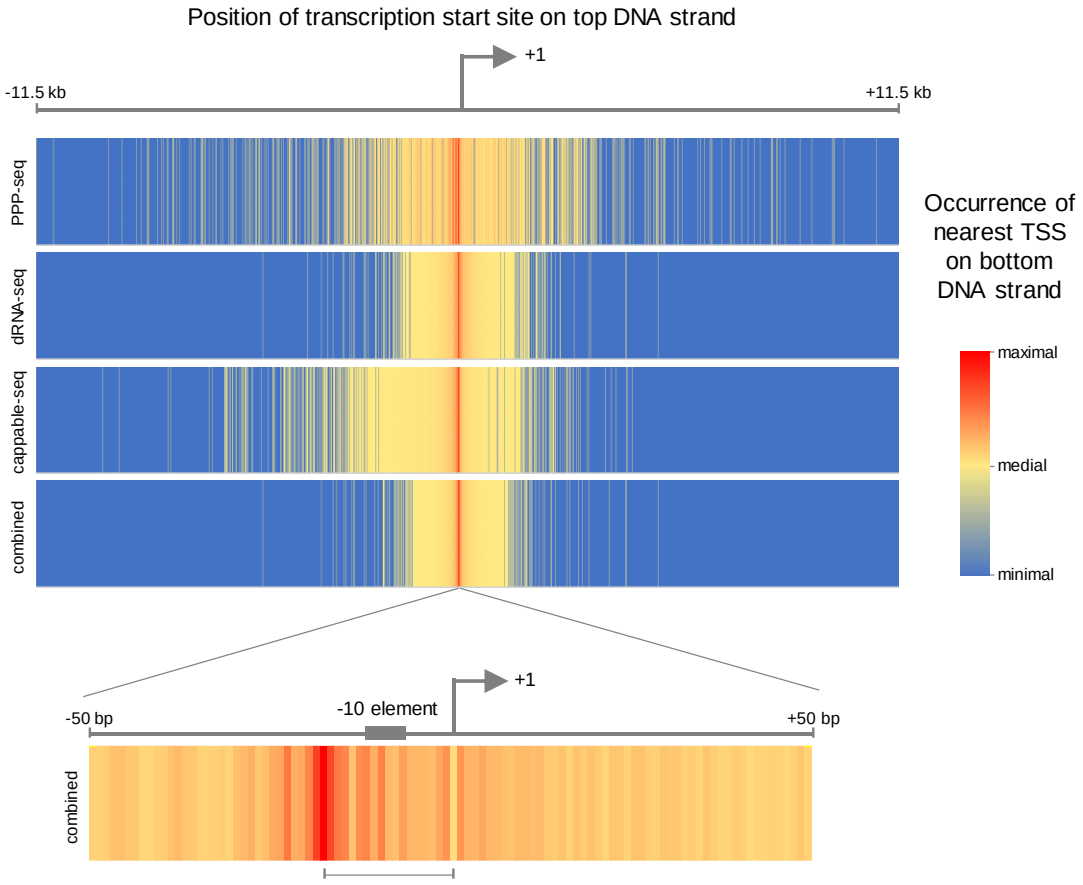


Figure 31: Distances between TSSs on opposite strands.

Heatmaps made using global transcription start site (TSS) data from *E. coli*. TSSs on the top chromosome strand are aligned at the centre of the heatmap (bent arrow, labelled +1). Heatmap colour indicates abundance of bottom strand TSSs at that position. The expansion shows the occurrence of bottom strand TSSs in a 50 bp window either side of all top strand promoters. TSS data from Singh *et al*., 2014; Thomason *et al*., 2015; Ettwiller *et al*., 2016.

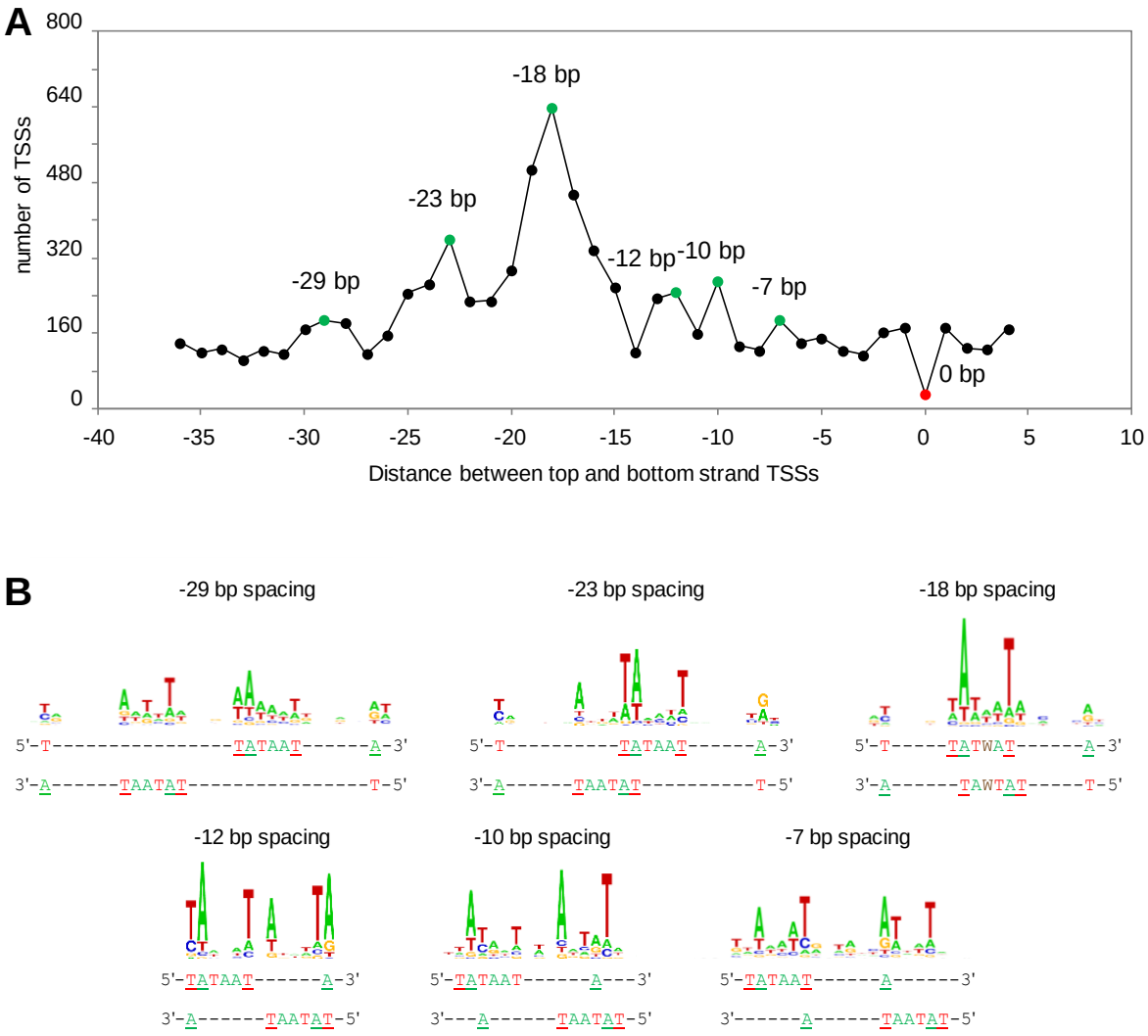


Figure 32: Distances between TSSs on opposite strands in bidirectional promoter regions.

A) The distance between opposing TSSs plotted against frequency. There is a preference for bottom strand TSSs appearing 18 bp upstream of the top strand TSS. B) Sequence conservation of bidirectional promoter regions. TSSs spaced between 29 and 7 bp apart were collected and aligned to generate sequence logos for each distance. The corresponding promoter elements on each strand are shown.

4.2.5. Reciprocal stimulation between divergent transcription start sites

We were particularly interested in the -18 bp spacing, which was most overrepresented. This configuration has a T at position -18, due to the opposite strand's TSS most often being A. The T at position -18 has been shown to be stimulatory (Singh *et al.*, 2011), and frequently occurs in the AT-tracts seen in spurious promotes. To investigate this 18 bp spacing further, a synthetic promoter sequence was made that matched the sequence logo in Figure 33A. Schematics of these promoters are shown in Figure 33B. Briefly, these promoters had -10 elements overlapping at 5 out of 6 positions, an A as a TSS at position +1 and a T at position -18. DNA upstream and downstream of the 18 bp promoter region was GC-rich to avoid AT-tracts or -35 elements (Figure 21B, promoters i and ii). Derivatives of the promoter sequence were made where the start site on the top strand was replaced with a C (iii) thus replacing the bottom strand -18 with a G (iv). Similarly, the bottom strand TSS was mutated to a C (vi) so the top strand -18 became a G (v). These bidirectional promoters were fused to *lacZ* in the plasmid pRW50 and promoter activity was measured by LacZ assay (Figure 34). The LacZ activity driven by the top strand promoter was 2.4-fold above background (empty pRW50) (column i). In contrast, the bottom strand promoter resulted in activity 15.6-fold above background (column ii). This is likely because the bottom strand promoter had a perfect consensus -10 element instead of a 5'-TATTAT-3'. Mutating the start site or the -18 position significantly decreased promoter activity in both directions. Activity for the mutated top strand promoters dropped to below the significance threshold (<2X background) (columns iii and v). Mutating the bottom strand -18 T (column iv) and TSS (column vi) decreased activity by 32 % and 40 % respectively. These results were confirmed by *in vitro*

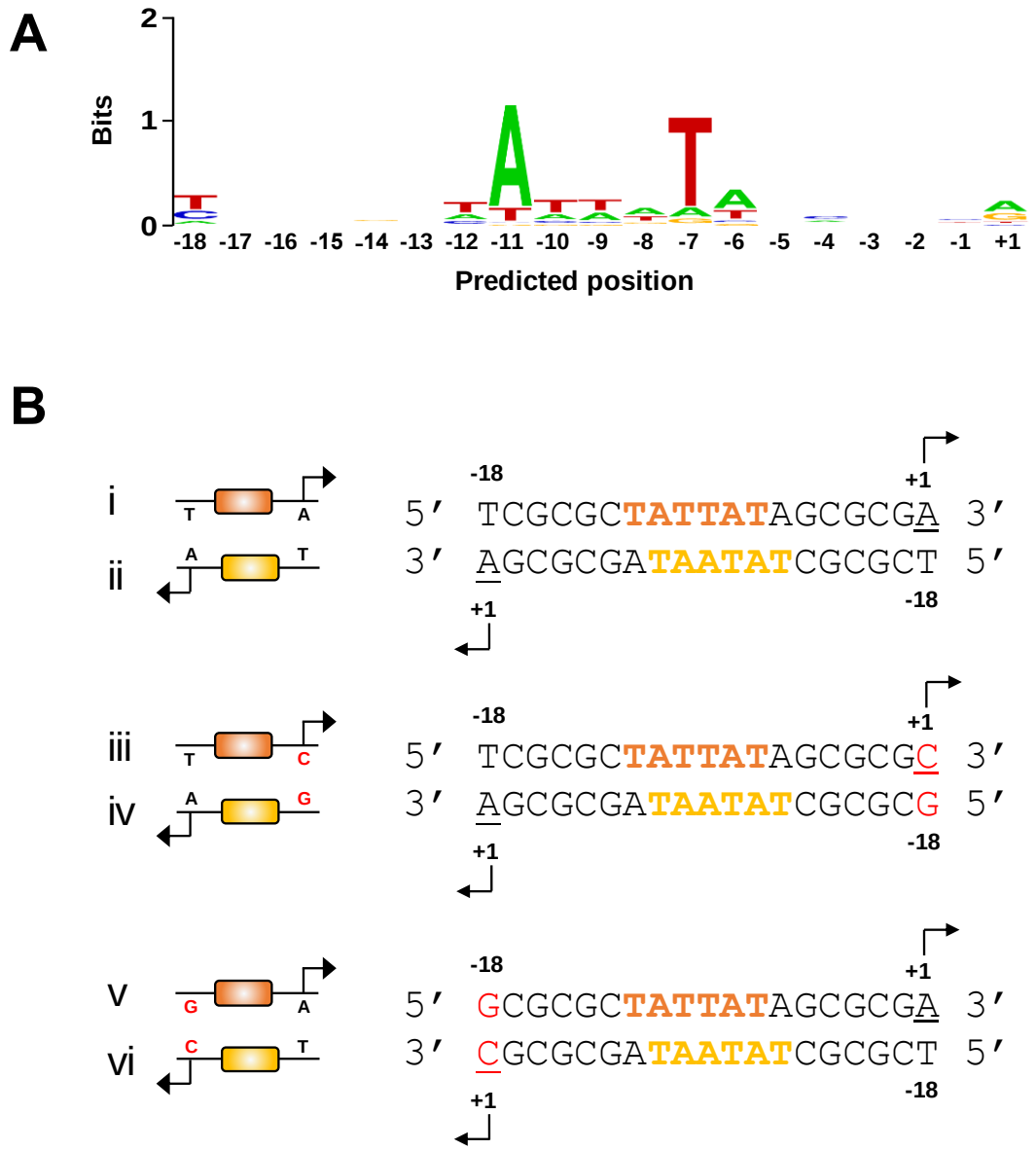


Figure 33: Sequence of an 18 bp bidirectional promoter

A) Sequence logo generated from 638 promoters where divergent TSSs are 18 bp apart. B) Schematics of the bidirectional promoters. The -10 elements are in bold and the TSSs are underlined. These promoters were tested with a consensus sequence (i and ii), with the top strand TSS mutated from A::C (iii and iv) and with the bottom strand TSS mutated from A::C (v and vi). Mutated bases are shown in red. The top strand promoters are indicated in orange and the bottom strand promoters in yellow.

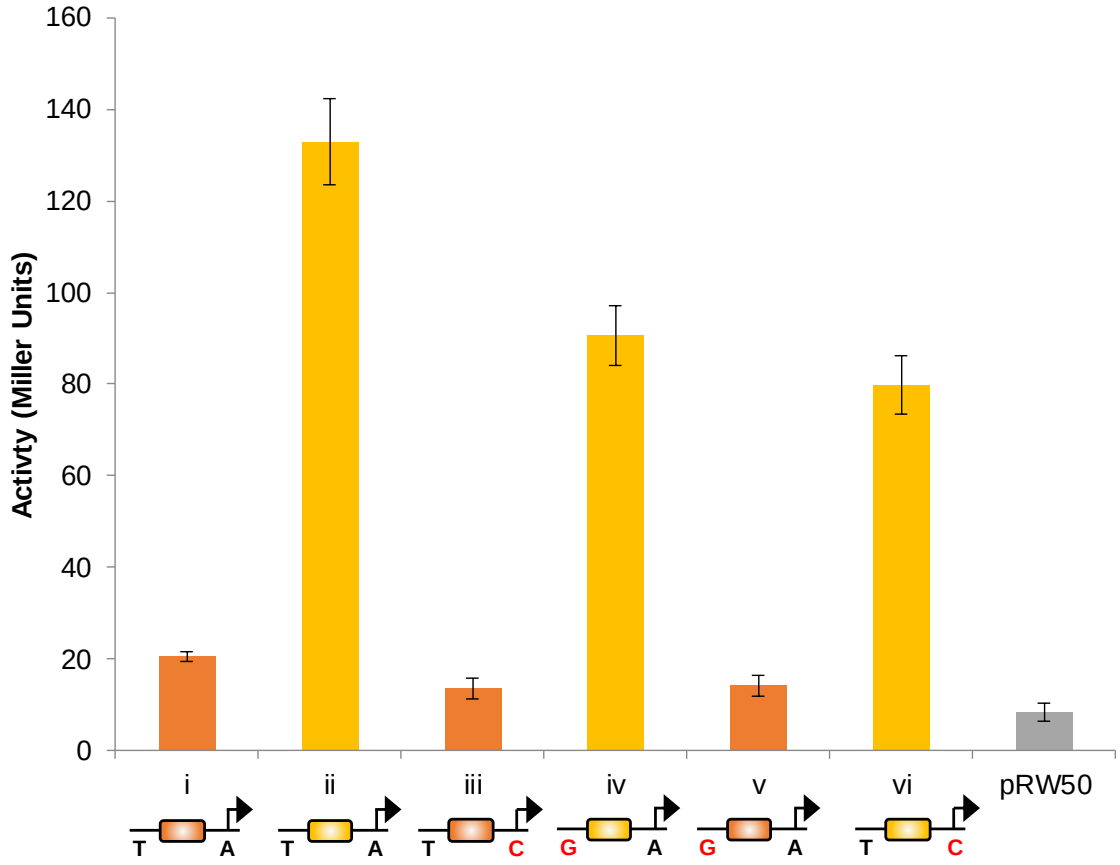


Figure 34: Activity of an 18 bp bidirectional promoter.

Promoter sequences were fused upstream of *lacZ* in pRW50. Orange bars show the activity of the promoter on the top strand whereas yellow bars show the activity of the promoter on the bottom strand. Bases which have been mutated are shown in red. Empty pRW50 was used as a negative control. For each promoter, LacZ activity was measured in triplicate. Data is presented as the mean of the replicates in Miller Units. Error bars show +/- SD

transcription; the amount of transcript produced from both promoters decreased when mutations were introduced (Figure 35, even numbered lanes). Additionally, none of the promoters were able to initiate transcription with the R451A mutated σ factor (Figure 35, odd numbered lanes).

4.3. Discussion

Our initial screen of σ^{70} binding sites identified 27 targets with no detectable transcript in RNA-seq data. Nevertheless, these were found to be genuine promoters. This emphasises the difficulty in finding intragenic promoter by conventional TSS mapping approaches; not only are intragenic promoters usually hidden by H-NS, their RNA transcripts are subject to rapid termination so are difficult to isolate by RNA-seq. There are methods which could be used to isolate these unstable transcripts. Treatment of the cells with bicyclomycin, an inhibitor of the Rho terminator, may increase the length of spurious transcripts by preventing early termination (Peters *et al.*, 2009). Another way to identify small unstable transcripts is by extracting them while they are still in complex with the RNAP. This can be done using native elongating transcript sequencing (NET-seq). RNAP complexes are crosslinked to the nascent RNA and immunoprecipitated. RNA libraries are then extracted and processed, then sequenced and mapped to the genome to identify the locations of the 3' transcript ends. However, this protocol is most frequently used in eukaryotes (Jin *et al.*, 2017); it has previously been adapted for use in prokaryotes (Larson *et al.*, 2014), but our attempts to use NET-seq with a Δ*hns E. coli* strain were unsuccessful.

Analysis of the intragenic promoters identified here revealed that most were able to drive divergent transcription, and that divergence was frequently due inherent

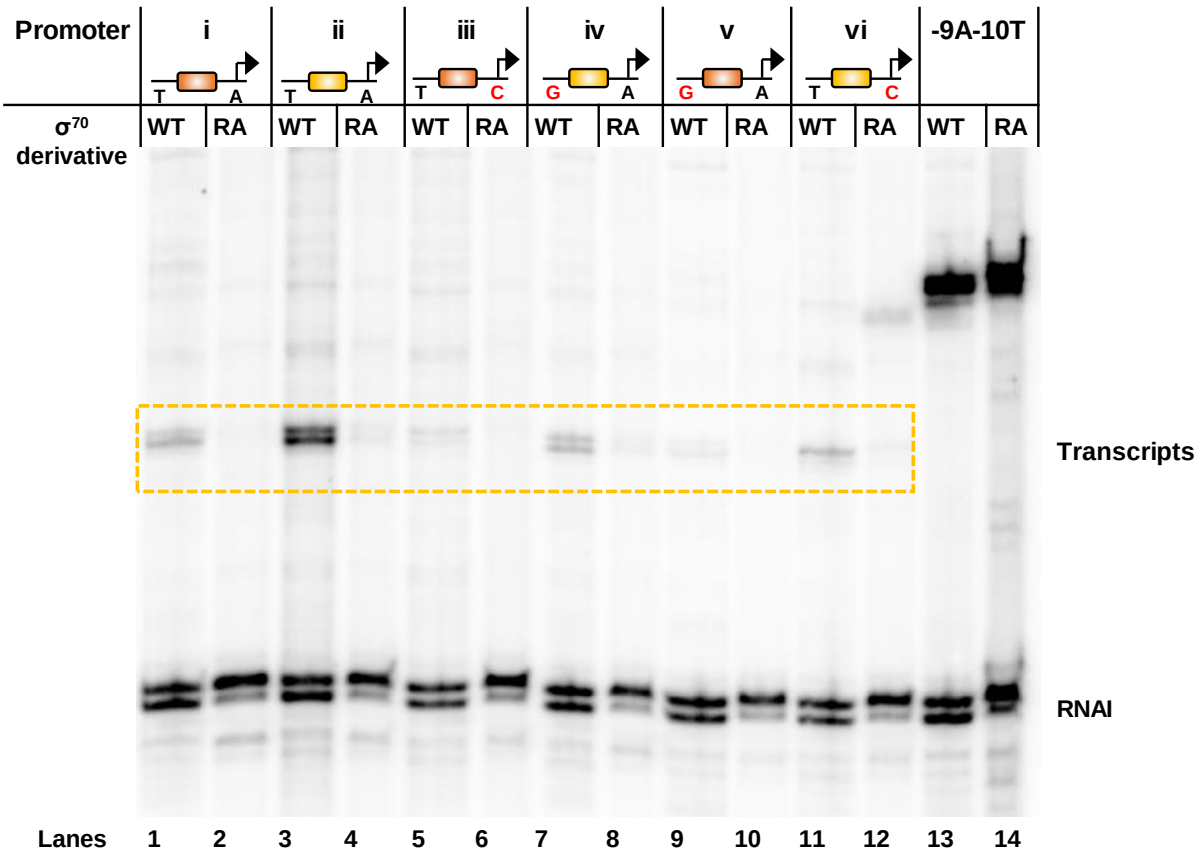


Figure 35: Effect of σ^{70} RA451 on *in vitro* transcription of bidirectional promoters.

Bidirectional promoter sequences were fused into the plasmid pSR in the forward (orange) and reverse (yellow) orientations. Bases which have been mutated are indicated in red. Transcription was initiated by RNA polymerase with either wild-type (WT) σ^{70} or the R451A (RA) σ^{70} mutant. The gel image shows transcripts derived from these promoters. Bands at 107/108 nt are RNA-I transcripts derived from the pSR plasmid and used as a loading control. -9A-10T is a constitutively active promoter which is not dependent on R451.

symmetry. Hence, promoters overlap on opposite DNA strands. In fact, our data show that bidirectional promoters are widespread in *E. coli.* TSSs also have a key role; a T at position -18 corresponds to a +1 A on the opposite strand. This bidirectional promoter configuration, with TSSs spaced 18 bp apart, is overrepresented (Figure 20). Another possible conformation of bidirectional promoters involves the AT-tract, which could act as a -10 element on the opposite strand. This would correspond to the -29 bp spacing seen in Figure 20, which also has a -18 T. However, as AT-tracts appear to be limited to intragenic promoters, they are not conserved in the whole genome data set.

While divergent transcription is seen frequently at canonical promoters, it is more common at intragenic promoters in AT-rich genes. Again, this was expected; we had shown previously that promoter occurrence is linked to AT-content. As both strands contain the same AT content, they have the same probability of containing a randomly generated promoter element. The difference between intragenic and canonical promoters also suggests a link to promoter evolution. In eukaryotes, recently acquired promoters are inherently bidirectional and evolve directionality over time (Jin *et al.*, 2017). This effect is comparable to how divergent transcription is seen more frequently in horizontally acquired DNA.

Chapter 5. **Distribution of the Transcriptional Machinery and Associated Factors in Cells Lacking H-NS**

5.1. Introduction

H-NS is a NAP which binds horizontally acquired AT-rich DNA to silence intragenic promoters (Singh *et al.*, 2014). Without H-NS, intragenic promoters sequester RNAP which reduces availability for canonical gene expression (Lamberte *et al.*, 2017). Lamberte *et al*. (2017) visualised this change in RNAP localisation by super resolution microscopy. In this chapter, protein distribution was measured by ChIP-seq in a WT and Δ*hns* background. We predicted that the binding pattern RNAP-associated proteins, such as alternative σ, elongation and termination factors, would be altered in cells lacking H-NS. Three proteins were assayed for changes in distribution; σ^{38}, NusA and Rho. The alternative σ factor σ^{38} accumulates in a Δ*hns* mutant background (Yamashino *et al.*, 1995). Additionally, as σ^{38} recognises σ^{70}-like promoter elements, we predicted it would also recognise spurious promoters. The terminator Rho and its cofactor NusA are important for termination of spurious transcription (Cardinale *et al.*, 2008; Qayyum *et al.*, 2016). For these reasons, we explored how the genomic distribution and intracellular protein levels of σ^{38}, Rho and NusA, are affected by the loss of H-NS. Results from this chapter are published in Warman *et al*., 2020.

5.2. Results

5.2.1. σ38 targets intragenic promoters

The stationary phase σ factor σ^{38} (RpoS) is structurally similar to σ^{70}, particularly in the σ2 and σ4 regions responsible for interacting with DNA (Paget, 2015). Consequently, σ^{38} recognises sequences similar to σ^{70}, with a preference for extended -10 elements (Typas, Becker and Hengge, 2007). Differences in the -10 hexamer are better tolerated by σ^{38}, which prefers the sequence 5'-TATACT-3' (Becker and Hengge-Aronis, 2004). The intragenic promoters aligned in Figure 15 frequently contained an extended -10 element with a conserved C at position -8. This suggests that σ^{38} promoters could be frequent in AT-rich DNA. To identify σ^{38}-bound intragenic promoters, a ChIP-seq experiment was done in strain RPB104Δ*hns.* The parental strain RPB104 was gifted by J. T. Wade (New York State Department of Health) and encodes an SPA-tagged σ^{38}. The *hns* gene was substituted with a kanamycin resistance cassette using the Gene Doctoring protocol (Lee *et al.*, 2009). Cultures were grown in duplicate to stationary phase (OD_{600} = 3.0), when expression of *rpoS* is at its peak. Proteins were crosslinked to DNA by formaldehyde treatment, then DNA was fragmented by sonication. The σ^{38}-DNA complexes were immobilised with sepharose A beads and anti-FLAG antibody, blunted and poly(A) tailed. An Illumina barcode was subsequently attached to the end of the DNA fragments by ligation, one barcode for each replicate. Finally σ^{38}-DNA complexes were de-crosslinked. This resulted in two libraries of barcoded DNA fragments corresponding to σ^{38} binding sites. The DNA libraries were combined and sequenced on an Illumina MiSeq. Sequences from the Δ*hns* libraries, and sequences from a σ^{38} ChIP-seq in the WT strain (Wong *et al.*, 2017), were mapped to the *E. coli* genome. Read counts for

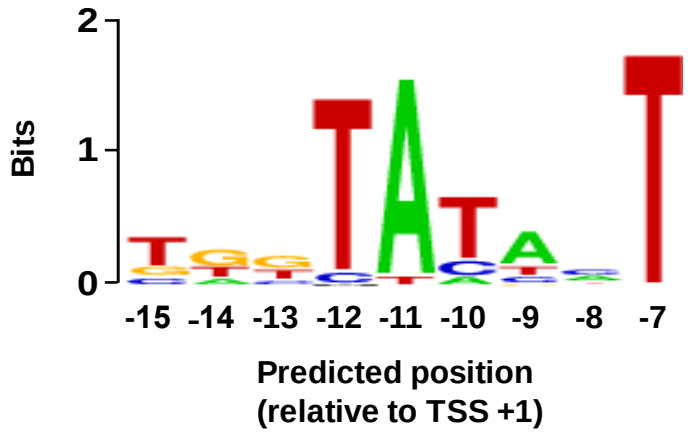


Figure 36: A conserved extended -10 element in intragenic promoters

A second sequence logo was generated by aligning 74 intragenic promoters identified inside H-NS repressed genes (Singh et. al, 2014). This sequence is similar to the -10 element in the first sequence logo (Figure 15B), but shows conservation of a T and a G at predicted positions -15 and -14, and a C at -9 (Warman *et al.,* 2020).

each position were calculated and normalised. In total, we identified 890 σ^{38} binding peaks in the starting strain and 905 peaks in the Δ*hns* derivative As expected, σ^{38} binds to intragenic regions in the absence of H-NS. Five examples of σ^{38} bound genes are shown in Figure 37. The genes are all bound by H-NS in WT cells and σ^{38} is largely excluded. Notably, these are genes that also had intragenic σ^{70} promoters. Hence, it is possible that the same promoters are targeted by both σ factors. All σ^{38} binding sites were classified based on their position relative to genes and whether they were bound by H-NS in WT. The pie charts in Figure 38 show that almost half of σ^{38} binding peaks in the Δ*hns* strain were intragenic, compared to 38 % in WT. Only 12 % of σ^{38} peaks in WT were in H-NS bound regions, but in the *hns* deletion 36 % of peaks corresponded to high H-NS occupancy. Hence, when *hns* is deleted, σ^{38} binds to intragenic promoters in H-NS bound regions.

The global redistribution of σ^{38} in cells lacking *hns* is visualised as a scatter plot (Figure 39). Regions of the genome were separated in to 500 bp bins and each data point represents one bin. The $\log_{10}$ of the σ^{38} binding signal in each region was calculated for WT and Δ*hns* cells. Regions were also ranked by H-NS binding, and the top 10 % (bins with the highest H-NS binding signal) are shown in purple. When *hns* is lost, the σ^{38} binding signal decreases at the majority of regions (grey points). However, σ^{38} binds preferentially to high-H-NS regions (purple points), suggesting a redistribution to AT-rich genes and possibly intragenic promoters.

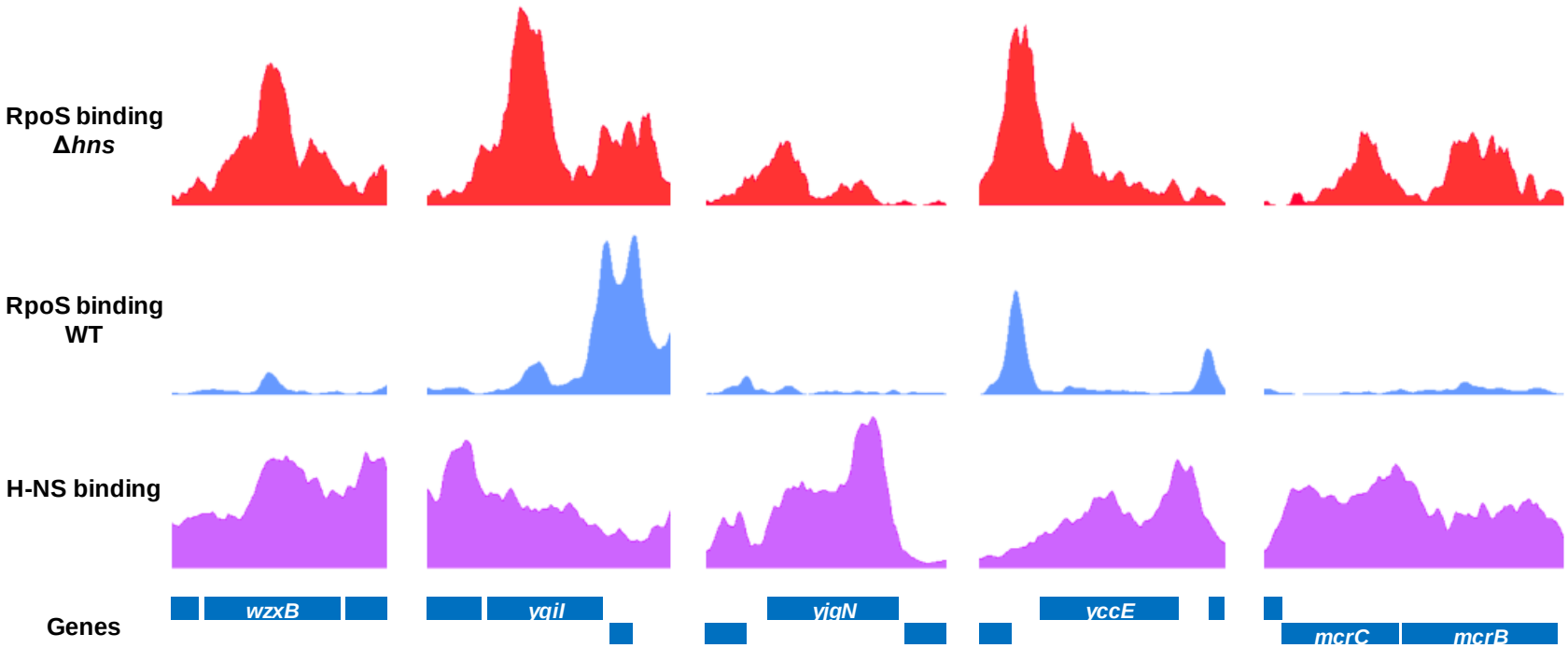


Figure 37: Examples of intragenic σ^{38} binding sites.

σ^{38} binding peaks generated from ChIP-seq data in a Δ*hns* strain (red peaks) and WT strain (blue peaks). H-NS coverage in these regions is shown in purple. Genes are shown in dark blue (Warman *et al.,* 2020).

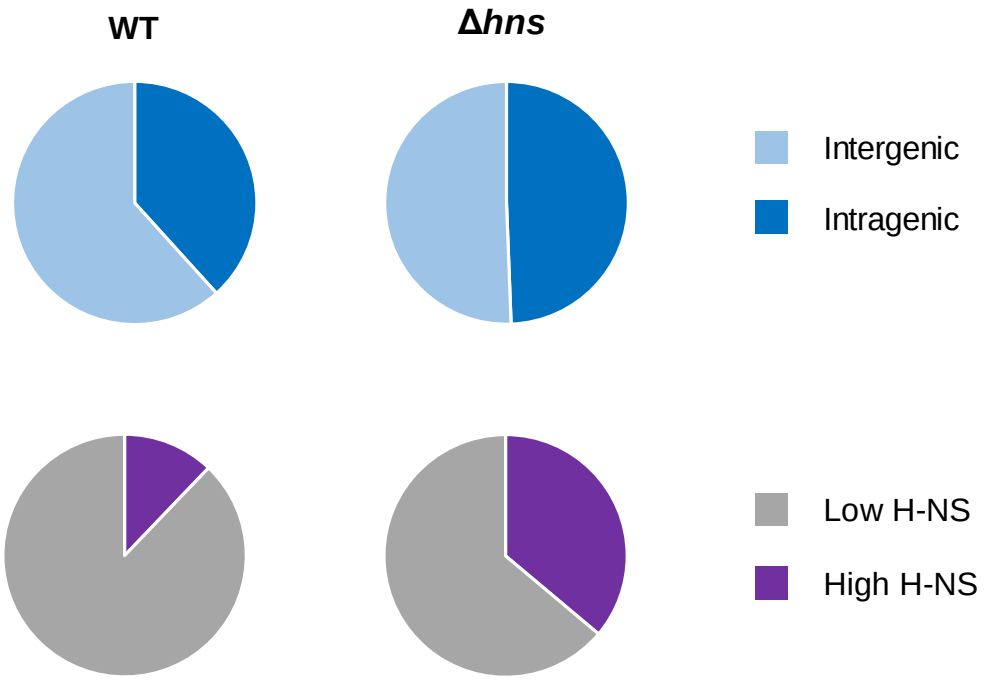


Figure 38: σ^{38} distribution in WT and Δ*hns* strains.

Pie charts indicating the proportion of σ^{38} binding sites identified inside genes (dark blue) or in intergenic regions (light blue), and in areas with a high H-NS binding signal (purple) or low H-NS (grey) (Warman *et al.*, 2020).

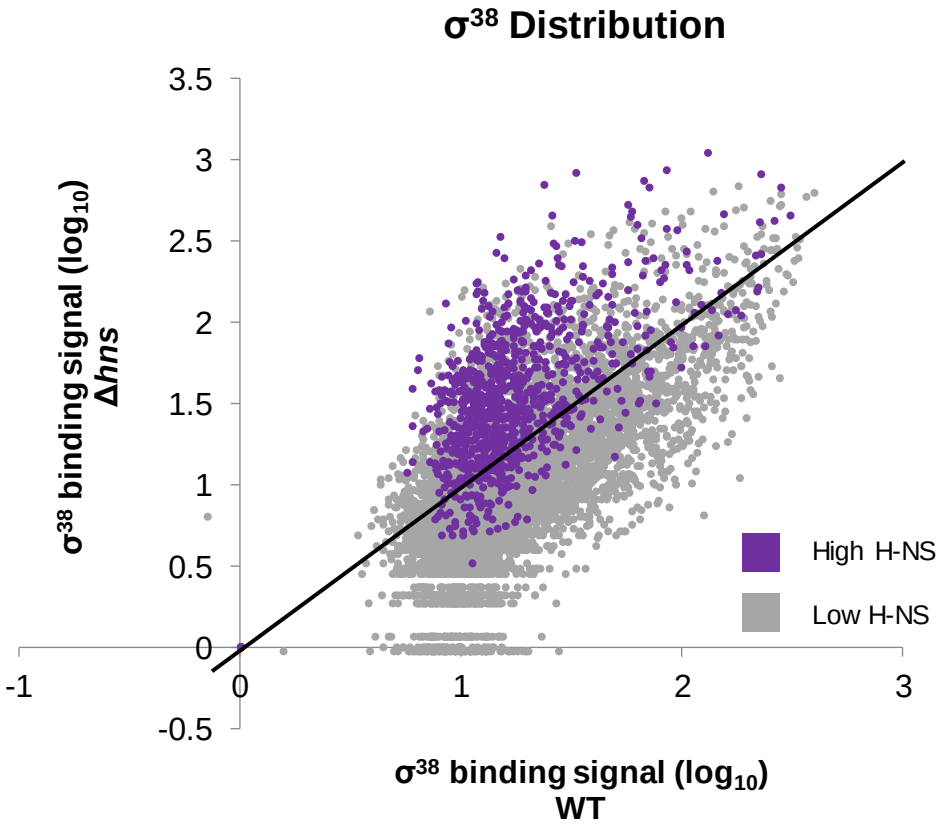


Figure 39: Changes in σ^{38} binding when H-NS is lost

Binding signal of σ^{38} in the WT strain is plotted against binding signal in the Δ*hns* strain. Each data point represents a 500 bp region of the genome. Points above the centre line show where σ^{38} binds preferentially in the absence of H-NS, and below the line are regions where σ^{38} binds less. Regions which are usually bound by high levels of H-NS (the top 10% when sorted from high to low H-NS binding signal) are shown in purple. (Warman *et al.,* 2020).

5.2.2. Some intragenic promoters can be recognised by both σ^{38} and σ^{70}

Many intragenic σ^{38} binding sites aligned with DNA bound by σ^{70} in the σ^{70} ChIP-seq assay (Singh *et al.*, 2014). To understand if σ^{38} and σ^{70} might be using the same promoters, five intragenic regions were assayed by *in vitro* transcription with σ^{38}. The five regions were within the AT-rich genes *yccE, yfdF, ykgH, yjgN* and *yjgL*, all targeted by H-NS. Previously, derivatives of these genes had been made where all intragenic σ^{70} -10 elements had been inactivated (Lamberte *et al.*, 2017). These mutated genes were also used as a template for *in vitro* transcription with σ^{38} to determine if the same promoters were recognised by both σ^{70} and σ^{38}. The transcripts derived from WT and mutated intragenic promoters are shown in Figure 40. Only faint transcripts were observed for promoters inside the gene *yccE* and there were no visible transcripts originating within *ykgH*. Transcripts from WT intragenic promoters in *yfdF, yjgN* and *yjgL* were detected, but not when the -10 elements were mutated. Hence, some of the -10 elements previously shown to be used by σ^{70} could also be recognised by σ^{38}. This is consistent with our σ^{38} ChIP-seq analysis that also identified σ^{38}-dependent transcription.

5.2.3. Intragenic promoters do not require the R451-equivalent arginine in σ38

The σ^{70} and σ^{38} σ factors share similar sequences, particularly in region σ2 which includes the σ^{70} R451 residue. This arginine is highly conserved between different σ factors (Barne *et al.*, 1997) and aligns to the σ^{38} arginine residue R166 (Figure 41). To explore whether σ^{38} R166 is important for intragenic promoter function, the arginine was mutated to an alanine. This R166A derivative was used for *in vitro* transcription.

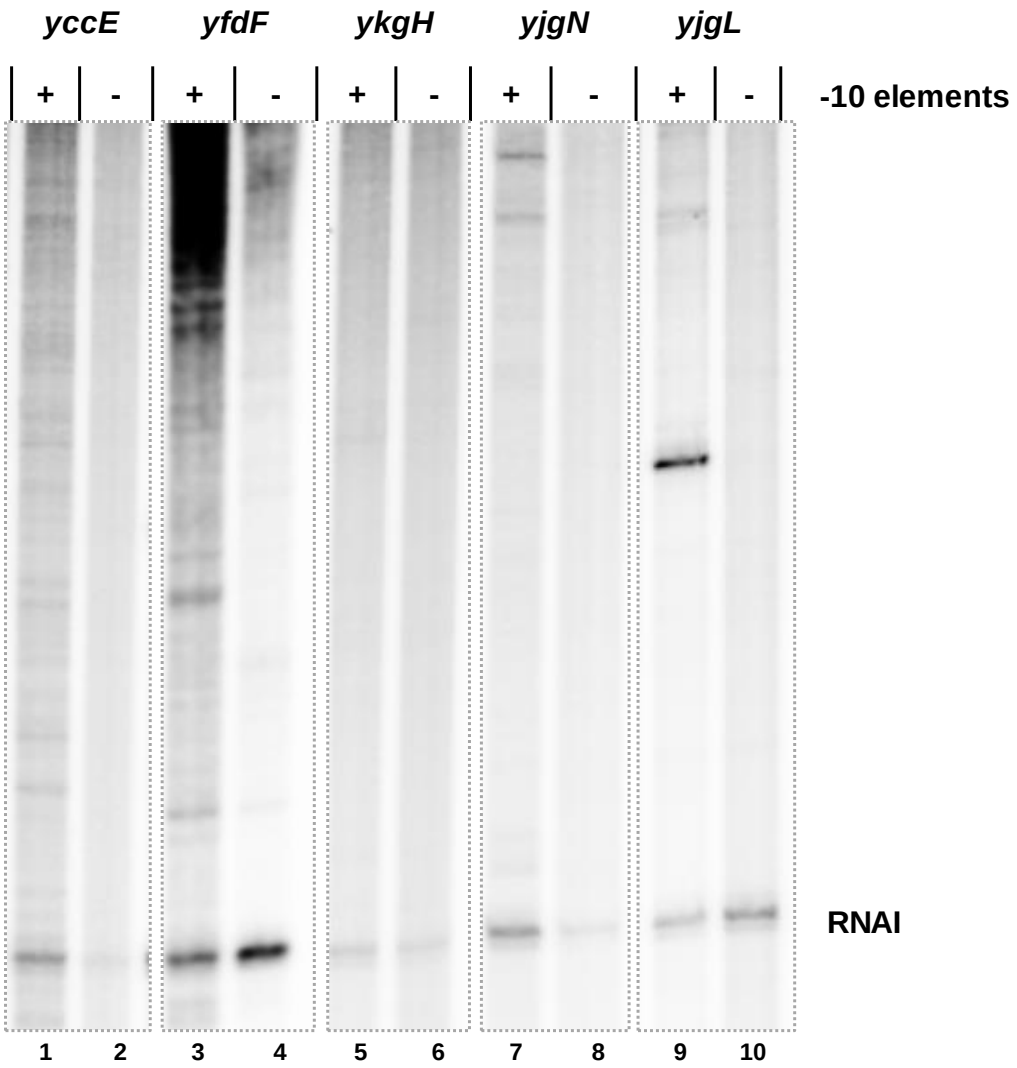


Figure 40: Effect of disrupting σ^{70} -10 elements on σ^{38}-dependent *in vitro* transcription.

Five genes *yccE, yfdF, ykgH, yjgN* and *yjGL*, which are bound by both σ^{70} and σ^{38} in ChIP assays, were cloned into pSR (+ lanes). The internal -10 elements were disrupted and these genes were also cloned into pSR (- lanes). Transcription was initiated by RNA polymerase core enzyme and σ^{38}. The gel image shows transcripts derived from these promoters. Bands at 107/108 nt are RNA-I transcripts derived from the pSR plasmid and used as a loading control.

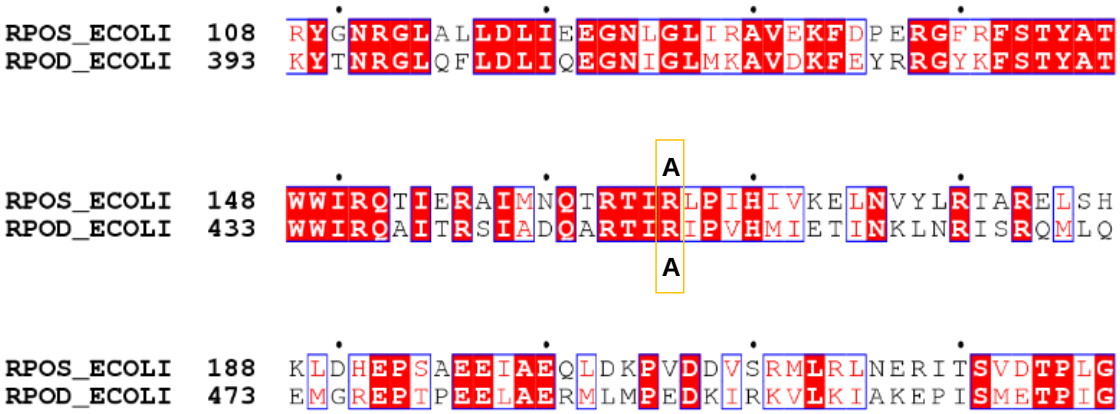


Figure 41: Comparison of the amino acid sequences of RpoD (σ^{70}) and RpoS (σ^{38}).

Sequences were aligned with Clustal Omega and visualised with ESPript 3. The arginine R451 in RpoD and the corresponding arginine in RpoS, R166, are boxed in yellow with the alanine mutation indicated. Red boxes highlight exact sequence matches.

Six intragenic regions from the genes *yfdF, yjgL, yqiI 2* a*, wzxB* a*, wzxB* b *and yigG* a were selected as templates for the assay. σ^{38} binding had been identified in these regions by ChIP-seq and confirmed *in vitro* (Figure 42, WT/odd numbered lanes). These regions had also previously been cloned in pSR to test for σ^{70} promoters (Figure 21), where transcription derived from *wzxB* b, *yigG* a and *yqiI*2 a was shown to be dependent on R451. The R166A mutant did not affect transcription initiating from any of the intragenic regions. (Figure 42, RA/even numbered lanes). Hence, the arginine at position 166 does not appear to be important for σ^{38}-dependent intragenic transcription.

5.2.4. Rho and NusA are sequestered at intragenic promoters

ChIP-seq experiments were also done for the proteins Rho and NusA in WT and Δ*hns* backgrounds. We predicted that these proteins would be recruited to sites associated with spurious transcription, as they are involved in termination of untranslated RNA (Cardinale *et al.*, 2008). Rho and NusA do not bind to the DNA but instead associate with elongation complexes by binding to nascent RNA. These elongation complexes are crosslinked to DNA during ChIP. Consequently, immunoprecipitation using antibodies for Rho and NusA isolated the DNA bound by the elongation complex. For these ChIP experiments, a spiked-in control was used for normalisation. Briefly, processing variation can affect the number of sequencing reads in each sample. To account for this, each *E. coli* culture contained a fixed proportion of *Salmonella* cells. Hence, final sequenced libraries therefore contained a mixture of reads derived from *Salmonella* and *E. coli*, which were separated by mapping to their respective genomes.

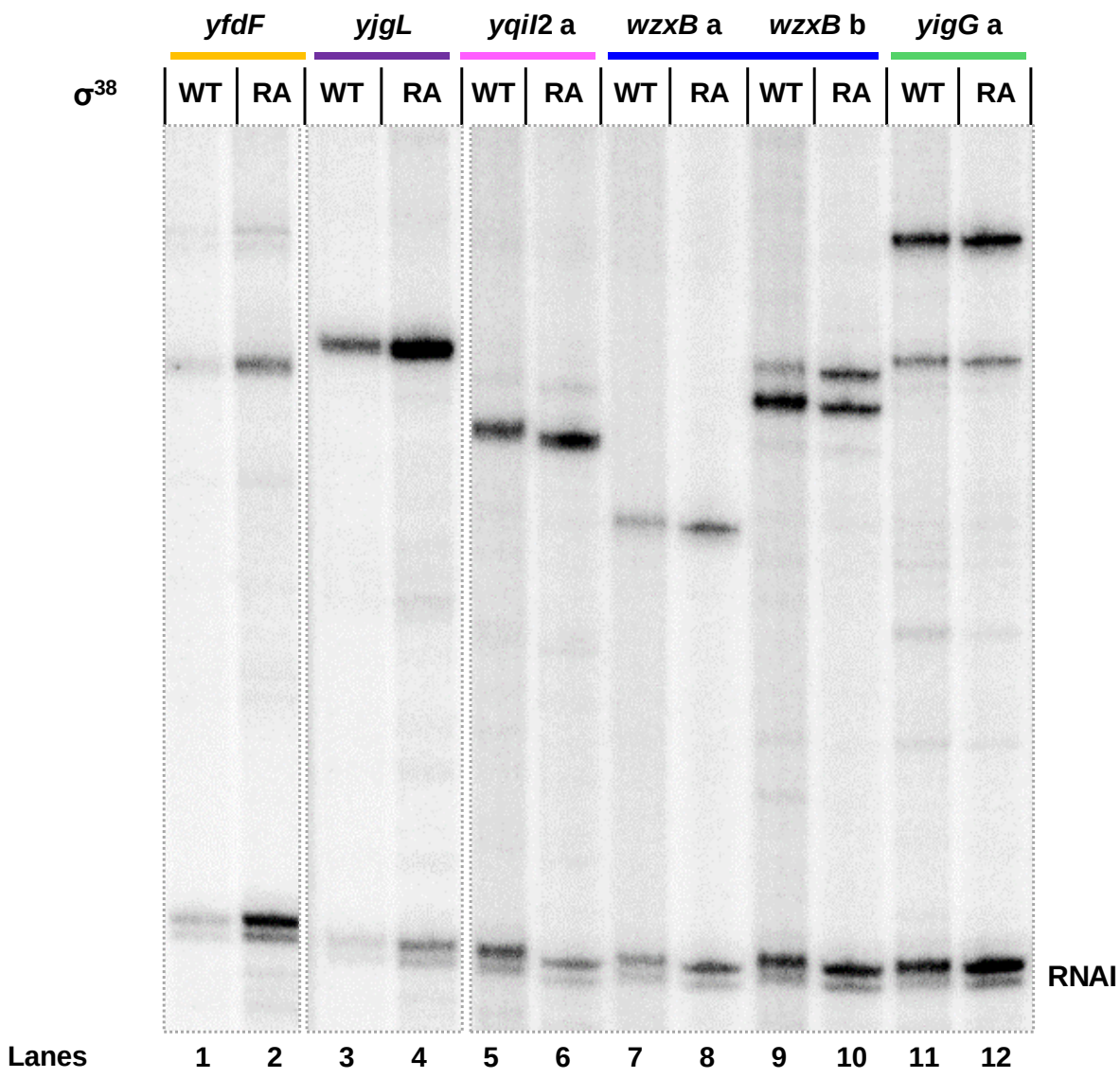


Figure 42: Effect of the σ^{38} R166A mutation on *in vitro* transcription at intragenic promoters.

Intragenic σ^{38} promoters cloned into pSR were assayed with RNAP core enzyme and WT or R166A σ^{38}. The gel image shows transcripts derived from these promoters. Bands at 107/108 nt are RNAI transcripts derived from the plasmid.

Since the *Salmonella* reads came from the same volume of cells, and the same culture, differences represented processing variation. This information was used to normalise the *E. coli* reads. Differences in distribution were again assessed by plotting WT binding signal against Δ*hns* binding signal (Figure 43). Both proteins were recruited to high H-NS regions (purple points) when *hns* was deleted. Binding signal accross the rest of genome (grey points) was decreased.

5.2.5. Rho and NusA protein levels are unchanged in Δ*hns*

The redistribution of RNAP and associated proteins, in cells lacking H-NS, would reduce availability of these proteins for canonical transcription. The cell responds to this by making more copies of σ^{38}, but RNAP core enzyme levels remained unchanged (Lamberte *et al.*, 2017; Yamashino *et al.*, 1995). To investigate whether Rho and NusA protein levels are increased when *hns* is deleted, a western blot was done with the anti-Rho and anti-NusA antibodies. Cultures of WT and Δ*hns E. coli* were grown and cells were isolated at different growth points; OD_{600} 0.4, 0.9 and 1.8. Following lysis, 5 µg of total protein was separated by gel electrophoresis and transferred to a nitrocellulose membrane. The membrane was incubated with primary antibody (anti-Rho and anti-NusA) then the appropriate secondary antibody conjugated to HRP. The correct protein bands were identified by comparing size to the ladder and are shown in Figure 44. There was no difference in protein amount either between each growth stage, or between the WT and Δ*hns* strain.

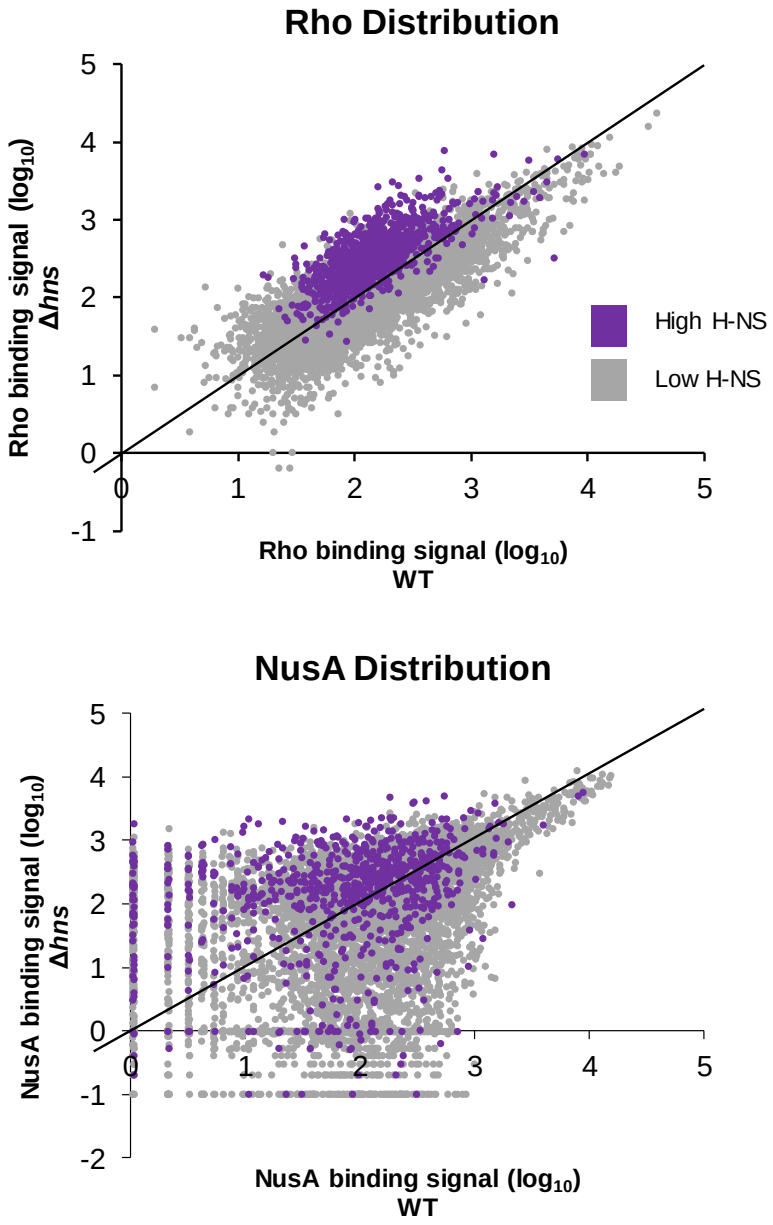


Figure 43: Changes in Rho and NusA binding when H-NS is lost.

ChIP binding signal of Rho/NusA in the WT strain is plotted against binding signal in the Δ*hns* strain. Each data point represents a 500 bp region of the genome. Points above the centre line show where Rho/NusA binds preferentially in the absence of H-NS, and below the line are regions where Rho/NusA binds less. Regions which are usually bound by high levels of H-NS (the top 10% when sorted from high to low H-NS binding signal) are shown in purple.

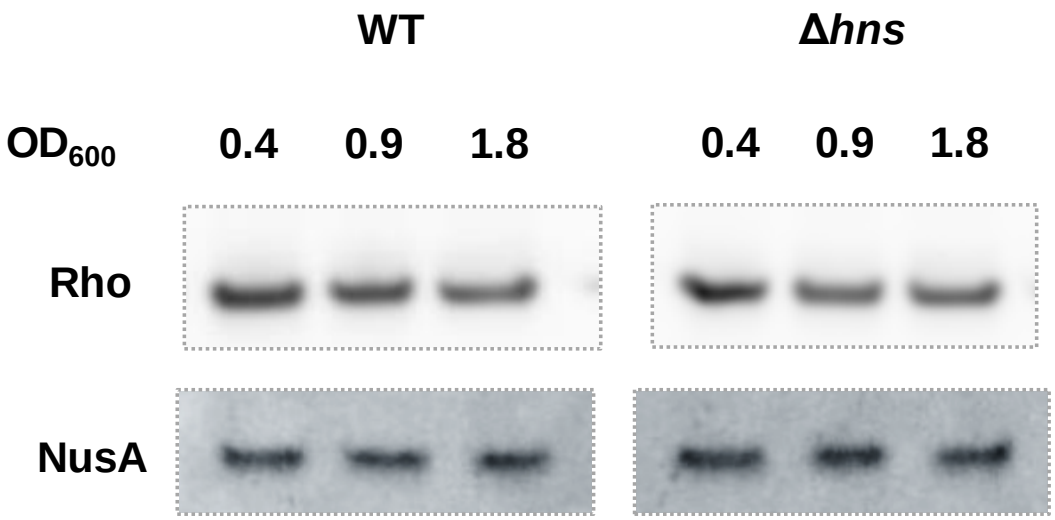


Figure 44: Changes in Rho and NusA protein levels when H-NS is lost.

Western blots of Rho protein and NusA protein in lysates from 2 strains, WT and Δ*hns* and from 3 growth points; at OD_{600} 0.4, 0.9 and 1.8.

5.3. Discussion

The data presented in this chapter suggest that RNAP-associated proteins are recruited to intragenic promoters in cells lacking H-NS. Intragenic promoters are frequently targeted by σ^{38}, which recognises similar promoter elements to σ^{70} (Figure 40). Previous experiments have shown that σ^{38} preferentially recognises -10 elements with a C at position -13 (Becker and Hengge-Aronis, 2001). This feature is seen in the annotated σ^{70} promoter element for *wzxB* b. However it is absent in the other predicted promoter elements (see Figure 28 for sequences), suggesting that σ^{38} is either able to use these sequences or is acting at different promoter elements in these regions. TSS mapping of σ^{38}-dependent promoters is required to accurately identify the promoter elements used. This would reveal whether intragenic σ^{38} promoters tend to contain an AT-tract, as seen in σ^{70} intragenic promoters. If σ^{38} promoters do contain an AT-tract, it does not interact with the conserved arginine R166. Hence, the AT-tract may not function to stimulate σ^{38}-dependent transcription.

The Rho and NusA ChIP experiments show that these proteins are also recruited to sites which would usually be bound by H-NS. We predict that these are sites of spurious transcription, which would require termination by Rho. Binding of σ^{38}, Rho and NusA to non-canonical promoters decreases the binding signal of these proteins in the rest of the genome. The same effect is seen with RNAP and results in a global downshift in housekeeping transcription. If σ^{38} is sequestered away from its canonical promoters then global σ^{38}-dependent transcription would decrease. However, σ^{38} levels increase in cells lacking H-NS, which may ameliorate the cost of sequestering σ^{38} at intragenic promoters. In contrast, Rho and NusA protein levels remain the same when H-NS is lost. Redistribution of these proteins could therefore cause a downshift

in global termination rates, and an increase in levels of non-coding transcripts associated with terminator readthrough. The resulting disruption of transcription regulation may contribute to the toxicity seen in cells lacking H-NS.

Final Conclusions

This work reveals why promoters occur so frequently within intragenic AT-rich DNA sequences. A relationship between genome AT-content and promoter occurrence has been suggested previously (Lamberte *et al.*, 2017; Singh *et al.*, 2014), and is supported by experiments that show promoters occurring at higher rates in more AT-rich DNA sequences (Figure 14) (Warman *et al.*, 2020). Here we show that promoters inside AT-rich genes depend on a unique AT-rich promoter element, the AT tract. Promoters reliant on this element tend to lack good -35 elements. This suggests that the AT-tract acts to stabilise the promoter/RNAP complex, making the -35 element redundant.

The interaction between the AT-tract and RNAP involves a specific arginine residue in σ^{70}, R451. Interestingly, an equivalent arginine in the alternative σ factor σ^{38} does not have an equivalent role. Consequently, whether the AT-tract is recognised by σ factors other than σ^{70} is unclear. Despite acting independently of the AT-tract σ^{38} does recognise many intragenic promoters in AT-rich regions. It is possible that σ^{38} binding is less specific than that of σ70 or that σ38 does recognise AT-tracts but in a way different to σ70.

Studying the intragenic promoters identified in this work led to an unexpected observation; many intragenic promoter regions drove transcription in both directions. This was often due to overlapping promoter elements permitting transcription on opposite strands. We have refer to such promoters as a “bidirectional”. This led us to investigate the occurrence of bidirectional promoters across the *E. coli* genome. We found that TSSs are frequently spaced 18 bp apart, corresponding to bidirectional promoters with overlapping -10 elements. This conformation is explained by the

inherent symmetry of the -10 element; the sequence complementary to 5'-TATAAT-3' is likely to contain the key bases (T_{-12}, A_{-11} and T_{-7}) required for an overlapping functional -10 element. Additionally, when separated by 18 bp, TSSs, which are most frequently adenine, pair with a T_{-18} in the opposite promoter. Promoters with a T at this position are more active (Singh *et al.*, 2011), and this effect is dependent on R451.

Further analysis of TSS distances in multiple species shows that bidirectional promoters are prevalent across prokaryotes and some archaea. Figure 45A (by D. Grainger, awaiting publication) presents heatmaps generated from multiple TSS datasets, as in Figure 31, and shows the same trend for bidirectional promoters. The preferred spacing between TSSs in each species reflects the properties of the promoter elements in each species; in *Bacillus subtilis*, for example, TSSs are most frequently spaced between 10 and 12 bp apart (Grainger and Forrest, awaiting publication). This is presumably because the promoter recognised by SigA, the primary σ factor in *B. subtilis*, often has AT-rich sequence surrounding the TSS (Jarmer *et al.*, 2001). Hence, this region acts as a -10 element on the opposite strand. This is another example of how the intrinsic symmetry of promoter sequences allow them to function in both directions.

Our observations challenge the typical model of directional transcription initiation, and is reminiscent of recent observations for eukaryotic transcription. Bidirectional transcription occurs very frequently in eukaryotes; however, transcription is derived from separate core promoters spaced >100 bp apart (Rhee and Pugh, 2012). Both promoters become accessible when activator proteins remodel the chromatin to create a large nucleosome free region. This is in contrast to our model of prokaryotic bidirectional transcription, which requires just a short section of DNA with shared

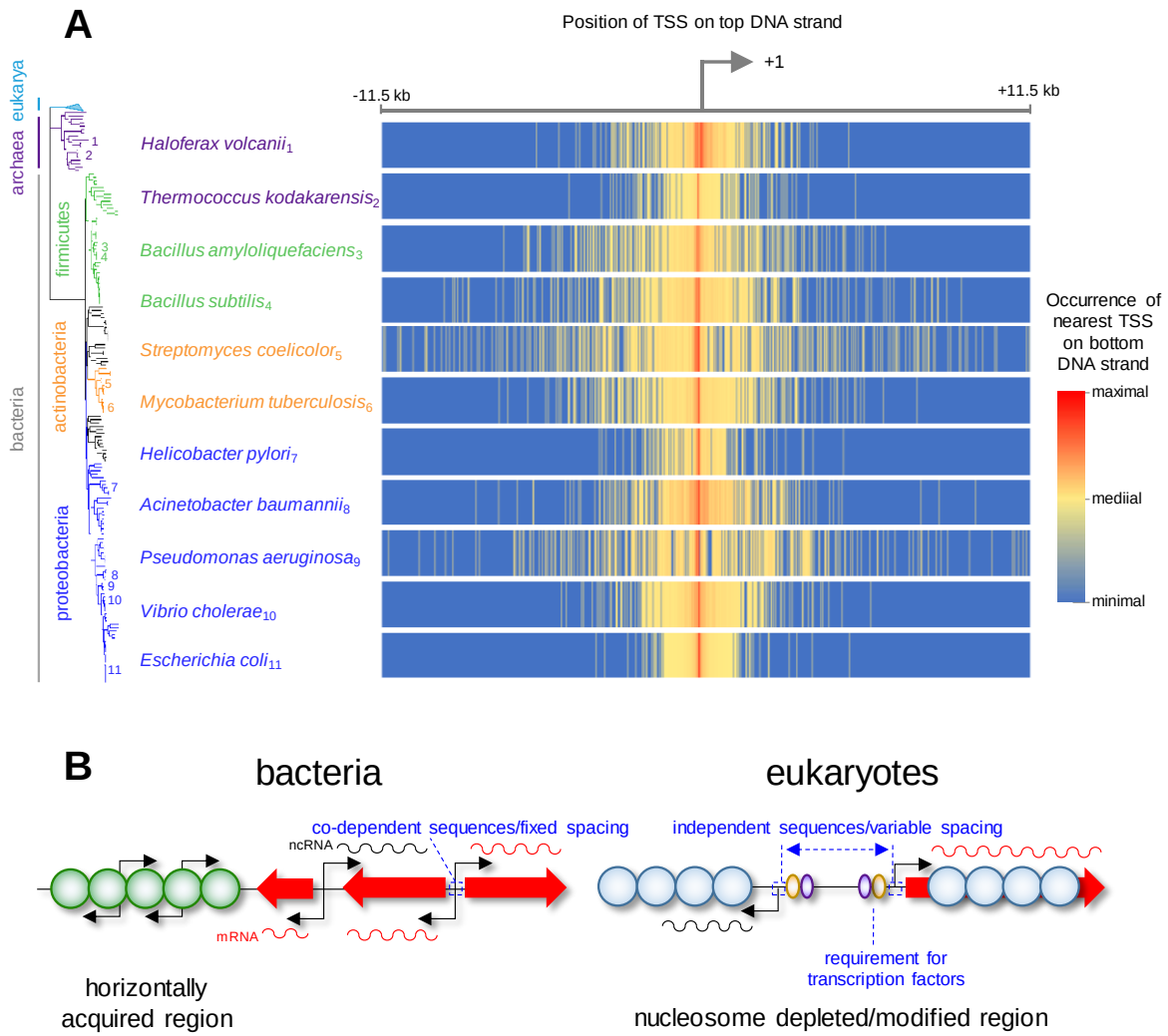


Figure 45: Bidirectional transcription in prokaryotes

A) Heatmaps indicate abundance and position of TSSs on the bottom DNA strand, relative to the nearest top strand promoter (bent arrow). Species and phylogenetic relationships are indicated to left of heatmaps. B) Bidirectional promoters have a different basis in bacteria and eukaryotes.

overlapping core promoter elements (Figure 45B). In eukaryotes, bidirectionality has been linked to the evolution of promoters; less evolved promoters are more bidirectional (Jin *et al.*, 2017). Perhaps the bidirectional promoters seen here are a consequence of horizontal gene transfer, as foreign DNA can be considered less evolved. This would explain why bidirectional transcription is seen more frequently in intragenic promoters compared to canonical promoters.

Our final observation relates to how transcriptional machinery is sequestered at intragenic promoters in cells lacking H-NS. Rho, NusA and σ^{38} all display this effect, presumably because they are recruited to promoters and transcripts associated with spurious transcription. We show that the cell does not make more copies of NusA or Rho in response to *hns* deletion. We therefore predict that redistribution of Rho and NusA reduces their activity at housekeeping genes. As previously observed with RNAP (Lamberte *et al.*, 2017), this redistribution would decrease the fitness of the cell. Future work will determine whether redistribution in an Δ*hns* background is also seen with other elongation or termination factors, or with other σ factors. Depending on the function of the protein, this would have varied and widespread effects on the cells transcriptional landscape.

List of References

Ali, S.S., Soo, J., Rao, C., *et al.* (2014) Silencing by H-NS potentiated the evolution of Salmonella. *PLoS pathogens*, 10 (11): e1004500.

Amit, R., Oppenheim, A.B. and Stavans, J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophysical Journal*, 84 (4): 2467–2473.

Aoyama, T., Takanami, M., Ohtsuka, E., *et al.* (1983) Essential structure of E. coli promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res*, 11 (17): 5855–64.

Arfin, S.M., Long, A.D., Ito, E.T., *et al.* (2000) Global Gene Expression Profiling in Escherichia coliK12 THE EFFECTS OF INTEGRATION HOST FACTOR. *Journal of Biological Chemistry*, 275 (38): 29672–29684.

Arold, S.T., Leonard, P.G., Parkinson, G.N., *et al.* (2010) H-NS forms a superhelical protein scaffold for DNA condensation. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (36): 15728–15732.

Balsalobre, C., Johansson, J., Uhlin, B.E., *et al.* (1999) Alterations in Protein Expression Caused by the hha Mutation in Escherichia coli: Influence of Growth Medium Osmolarity. *Journal of Bacteriology*, 181 (10): 3018–3024.

Baños, R.C., Vivero, A., Aznar, S., *et al.* (2009) Differential Regulation of Horizontally Acquired and Core Genome Genes by the Bacterial Modulator H-NS. *PLOS Genetics*, 5 (6): e1000513.

Barne, K.A., Bown, J.A., Busby, S.J.W., *et al.* (1997) Region 2.5 of the Escherichia coli RNA polymerase σ70 subunit is responsible for the recognition of the 'extended −10' motif at promoters. *The EMBO Journal*, 16 (13): 4034–4040.

Barth, M., Marschall, C., Muffler, A., *et al.* (1995) Role for the histone-like protein H-NS in growth phase-dependent and osmotic regulation of sigma S and many sigma S-dependent genes in Escherichia coli. *Journal of Bacteriology*, 177 (12): 3455–3464.

Battesti, A., Tsegaye, Y.M., Packer, D.G., *et al.* (2012) H-NS Regulation of IraD and IraM Antiadaptors for Control of RpoS Degradation. *Journal of Bacteriology*, 194 (10): 2470–2478.

Beatty, C.M., Browning, D.F., Busby, S.J.W., *et al.* (2003) Cyclic AMP Receptor Protein-Dependent Activation of the Escherichia coli acsP2 Promoter by a Synergistic Class III Mechanism. *Journal of Bacteriology*, 185 (17): 5148–5157.

Becker, G. and Hengge-Aronis, R. (2001) What makes an Escherichia coli promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S). *Mol Microbiol*, 39 (5): 1153–65.

Blot, N., Mavathur, R., Geertz, M., *et al.* (2006) Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome. *EMBO Reports*, 7 (7): 710–715.

Bougdour, A., Cunning, C., Baptiste, P.J., *et al.* (2008) Multiple pathways for regulation of σS (RpoS) stability in Escherichia coli via the action of multiple anti-adaptors. *Molecular Microbiology*, 68 (2): 298–313.

Bougdour, A., Wickner, S. and Gottesman, S. (2006) Modulating RssB activity: IraP, a novel regulator of σS stability in Escherichia coli. *Genes & Development*, 20 (7): 884–897.

Braun, V. and Mahren, S. (2005) Transmembrane transcriptional control (surface signalling) of the Escherichia coli Fec type. *FEMS Microbiology Reviews*, 29 (4): 673–684.

Brown, K.L. and Hughes, K.T. (1995) The role of anti-sigma factors in gene regulation. *Molecular Microbiology*, 16 (3): 397–404.

Brown, N.L., Stoyanov, J.V., Kidd, S.P., *et al.* (2003) The MerR family of transcriptional regulators. *FEMS Microbiology Reviews*, 27 (2–3): 145–163.

Browning, D.F. and Busby, S.J.W. (2004) The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2 (1): 57–65.

Burmann, B.M., Schweimer, K., Luo, X., *et al.* (2010) A NusE:NusG complex links transcription and translation. *Science (New York, N.Y.)*, 328 (5977): 501–504.

Busby, S. and Ebright, R.H. (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell*, 79 (5): 743–746.

Busby, S. and Ebright, R.H. (1999) Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology*, 293 (2): 199–213.

Campbell, E.A., Muzzin, O., Chlenov, M., *et al.* (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol Cell*, 9 (3): 527–39.

Carafa, Y. d'Aubenton, Brody, E. and Thermes, C. (1990) Prediction of rho-independent Escherichia coli transcription terminators: A statistical analysis of their RNA stem-loop structures. *Journal of Molecular Biology*, 216 (4): 835–858.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., *et al.* (2008) Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. *Science (New York, N.Y.)*, 320 (5878): 935–938.

Chattopadhyay, R. and Roy, S. (2002) DnaK-Sigma 32 Interaction Is Temperature-dependent IMPLICATION FOR THE MECHANISM OF HEAT SHOCK RESPONSE. *Journal of Biological Chemistry*, 277 (37): 33641–33647.

Chen, C.Y. and Richardson, J.P. (1987) Sequence elements essential for rho-dependent transcription termination at lambda tR1. *The Journal of Biological Chemistry*, 262 (23): 11292–11299.

Chen, J., Chiu, C., Gopalkrishnan, S., *et al.* (2020) Stepwise Promoter Melting by Bacterial RNA Polymerase. *Molecular Cell*.

Chen, J., Darst, S.E., Thirumalai, D., *et al.* (2010) Promoter melting triggered by bacterial RNA polymerase occurs in three steps. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (28): 12523–12528.

Cho, B.-K., Knight, E.M., Barrett, C.L., *et al.* (2008) Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. *Genome Research*, 18 (6): 900–910.

Cordeiro, T.N., Schmidt, H., Madrid, C., *et al.* (2011) Indirect DNA Readout by an H-NS Related Protein: Structure of the DNA Complex of the C-Terminal Domain of Ler. *PLOS Pathogens*, 7 (11): e1002380.

Dame, R.T. (2005) The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology*, 56 (4): 858–870.

Dame, R.T., Noom, M.C. and Wuite, G.J.L. (2006) Bacterial chromatin organization by H-NS protein unravelled using dual DNA manipulation. *Nature*, 444 (7117): 387–390.

Dame, R.T. and Wuite, G.J.L. (2003) On the Role of H-NS in the Organization of Bacterial Chromatin: From Bulk to Single Molecules and Back…. *Biophysical Journal*, 85 (6): 4146–4148.

Dame, R.T., Wyman, C. and Goosen, N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. *Nucleic Acids Research*, 28 (18): 3504–3510.

Dame, R.T., Wyman, C., Wurm, R., *et al.* (2002) Structural Basis for H-NS-mediated Trapping of RNA Polymerase in the Open Initiation Complex at the rrnB P1. *Journal of Biological Chemistry*, 277 (3): 2146–2150.

David, L., Huber, W., Granovskaia, M., *et al.* (2006) A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (14): 5320–5325.

Decker, K.B. and Hinton, D.M. (2013) Transcription Regulation at the Core: Similarities Among Bacterial, Archaeal, and Eukaryotic RNA Polymerases. *Annual Review of Microbiology*, 67 (1): 113–139.

deHaseth, P.L., Zupancic, M.L. and Record, M.T. (1998) *RNA Polymerase-Promoter Interactions: the Comings and Goings of RNA Polymerase*.

Dorman, C.J. (2004) H-NS: a universal regulator for a dynamic genome. *Nature Reviews Microbiology*, 2 (5): 391–400.

Dorman, C.J. and Dillon, S.C. (2010) Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8 (3): 185–195.

Dorman, C.J. and Dorman, M.J. (2017) Control of virulence gene transcription by indirect readout in Vibrio cholerae and Salmonella enterica serovar Typhimurium. *Environmental Microbiology*.

Dorman, C.J., Schumacher, M.A., Bush, M.J., *et al.* (2020) When is a transcription factor a NAP? *Current Opinion in Microbiology*, 55: 26–33.

Dornenburg, J.E., DeVita, A.M., Palumbo, M.J., *et al.* (2010) Widespread Antisense Transcription in Escherichia coli. *mBio*, 1 (1).

Durand, S., Gilet, L. and Condon, C. (2012) The Essential Function of B. subtilis RNase III Is to Silence Foreign Toxin Genes. *PLOS Genetics*, 8 (12): e1003181.

Ebright, R.H. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol*, 304 (5): 687–98.

Ettwiller, L., Buswell, J., Yigit, E., *et al.* (2016) A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics*, 17 (1): 199.

Feklistov, A., Barinova, N., Sevostyanova, A., *et al.* (2006) A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. *Mol Cell*, 23 (1): 97–107.

Feklistov, A. and Darst, S.A. (2011) Structural Basis for Promoter −10 Element Recognition by the Bacterial RNA Polymerase σ Subunit. *Cell*, 147 (6): 1257–1269.

Gao, R., Mack, T.R. and Stock, A.M. (2007) Bacterial Response Regulators: Versatile Regulatory Strategies from Common Domains. *Trends in biochemical sciences*, 32 (5): 225–234.

Gao, Y., Foo, Y.H., Winardhi, R.S., *et al.* (2017) *Charged residues in the H-NS linker drive DNA binding and gene silencing in single cells*.

Georg, J. and Hess, W.R. (2011) cis-Antisense RNA, Another Level of Gene Regulation in Bacteria. *Microbiology and Molecular Biology Reviews*, 75 (2): 286–300.

Ghosh, P., Ishihama, A. and Chatterji, D. (2001) Escherichia coli RNA polymerase subunit omega and its N-terminal domain bind full-length beta' to facilitate incorporation into the alpha2beta subassembly. *Eur J Biochem*, 268 (17): 4621–7.

Gingeras, T.R. (2007) Origin of phenotypes: Genes and transcripts. *Genome Research*, 17 (6): 682–690.

Goldman, S.R., Ebright, R.H. and Nickels, B.E. (2009) Direct Detection of Abortive RNA Transcripts in Vivo. *Science*, 324 (5929): 927–928.

Gordon, B.R.G., Li, Y., Cote, A., *et al.* (2011) Structural basis for recognition of AT-rich DNA by unrelated xenogeneic silencing proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 108 (26): 10690–10695.

Gourse, R.L., Ross, W. and Gaal, T. (2000) UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology*, 37 (4): 687–695.

Grainger, D.C. (2016a) Structure and function of bacterial H-NS protein. *Biochemical Society Transactions*, 44 (6): 1561–1569.

Grainger, D.C. (2016b) The unexpected complexity of bacterial genomes. *Microbiology (Reading, England)*, 162 (7): 1167–1172.

Grainger, D.C., Goldberg, M.D., Lee, D.J., *et al.* (2008) Selective repression by Fis and H-NS at the Escherichia coli dps promoter. *Molecular Microbiology*, 68 (6): 1366–1377.

Guo, X., Myasnikov, A.G., Chen, J., *et al.* (2018) Structural Basis for NusA Stabilized Transcriptional Pausing. *Molecular Cell*, 69 (5): 816-827.e4.

Gusarov, I. and Nudler, E. (1999) The Mechanism of Intrinsic Transcription Termination. *Molecular Cell*, 3 (4): 495–504.

Haycocks, J.R.J., Sharma, P., Stringer, A.M., *et al.* (2015) The Molecular Basis for Control of ETEC Enterotoxin Expression in Response to Environment and Host. *PLOS Pathogens*, 11 (1): e1004605.

Helmann, J.D. (1991) Alternative sigma factors and the regulation of flagellar gene expression. *Molecular Microbiology*.

Hengge-Aronis, R. (2002) Signal Transduction and Regulatory Mechanisms Involved in Control of the σS (RpoS) Subunit of RNA Polymerase. *Microbiol Mol Biol Rev*, 66 (3): 373–95.

Herbert, K.M., Zhou, J., Mooney, R.A., *et al.* (2010) E. coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase. *Journal of Molecular Biology*, 399 (1): 17–30.

Herring, C.D., Glasner, J.D. and Blattner, F.R. (2003) Gene replacement without selection: regulated suppression of amber mutations in Escherichia coli. *Gene*, 311: 153–63.

Hook-Barnard, I.G. and Hinton, D.M. (2009) The promoter spacer influences transcription initiation via σ70 region 1.1 of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A*, 106: 737–42.

Hsu, L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1577 (2): 191–207.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, 3 (3): 318–356.

Jain, S., Gupta, R. and Sen, R. (2019) Rho-dependent transcription termination in bacteria recycles RNA polymerases stalled at DNA lesions. *Nature Communications*, 10 (1): 1–12.

Jarmer, H., Larsen, T.S., Krogh, A., *et al.* (2001) Sigma A recognition sites in the Bacillus subtilis genome. *Microbiology,* 147 (9): 2417–2424.

Jensen, T.H., Jacquier, A. and Libri, D. (2013) Dealing with Pervasive Transcription. *Molecular Cell*, 52 (4): 473–484.

Jin, Y., Eser, U., Struhl, K., *et al.* (2017) The ground state and evolution of promoter region directionality. *Cell*, 170 (5): 889-898 e10.

Johansson, J., Eriksson, S., Sondén, B., *et al.* (2001) Heteromeric Interactions among Nucleoid-Associated Bacterial Proteins: Localization of StpA-Stabilizing Regions in H-NS of Escherichia coli. *Journal of Bacteriology*, 183 (7): 2343–2347.

Johansson, J. and Uhlin, B.E. (1999) Differential protease-mediated turnover of H-NS and StpA revealed by a mutation altering protein stability and stationary-phase survival of Escherichia coli. *Proceedings of the National Academy of Sciences*, 96 (19): 10776–10781.

Ju, X., Li, D. and Liu, S. (2019) Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nature microbiology*, 4 (11): 1907–1918.

Kapanidis, A.N., Margeat, E., Ho, S.O., *et al.* (2006) Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*, 314 (5802): 1144–1147.

Kapranov, P., Cheng, J., Dike, S., *et al.* (2007) RNA Maps Reveal New RNA Classes and a Possible Function for Pervasive Transcription. *Science*, 316 (5830): 1484–1488.

Keseler, I.M., Mackie, A., Peralta-Gil, M., *et al.* (2012) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res*, 41 (Database issue): D605-12.

Kolb, A., Kotlarz, D., Kusano, K., *et al.* (1995) Selectivity of the Escherichia coli RNA polymerase EG38 for overlapping promoters and ability to support CRP activation. *Nucleic Acids Res*, 23 (5): 819–826.

Kornberg, R.D. (1999) Eukaryotic transcriptional control. *Trends in Cell Biology*, 9 (12): M46–M49.

Koslover, D.J., Fazal, F.M., Mooney, R.A., *et al.* (2012) Binding and Translocation of Termination Factor Rho Studied at the Single-Molecule Level. *Journal of Molecular Biology*, 423 (5): 664–676.

Kotlajich, M.V., Hron, D.R., Boudreau, B.A., *et al.* (2015) Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria Proudfoot, N.J. (ed.). *eLife*, 4: e04970.

Kröger, C., Dillon, S.C., Cameron, A.D.S., *et al.* (2012) The transcriptional landscape and small RNAs of Salmonella enterica serovar Typhimurium. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (20): E1277-1286.

Lamberte, L.E., Baniulyte, G., Singh, S.S., *et al.* (2017) Horizontally acquired AT-rich genes in Escherichia coli cause toxicity by sequestering RNA polymerase. *Nature Microbiology*, 2: 16249.

Lang, B., Blot, N., Bouffartigues, E., *et al.* (2007) High-affinity DNA binding sites for H-NS provide a molecular basis for selective silencing within proteobacterial genomes. *Nucleic Acids Research*, 35 (18): 6330–6337.

Lange, R. and Hengge-Aronis, R. (1994) The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes & Development*, 8 (13): 1600–1612.

Larson, M.H., Greenleaf, W.J., Landick, R., *et al.* (2008) Applied force reveals mechanistic and energetic details of transcription termination. *Cell*, 132 (6): 971–982.

Larson, M.H., Mooney, R.A., Peters, J.M., *et al.* (2014) A Pause Sequence Enriched at Translation Start Sites Drives Transcription Dynamics In Vivo. *Science*, 344 (6187): 1042–7.

Lawson, M.R., Ma, W., Bellecourt, M.J., *et al.* (2018) Mechanism for the Regulated Control of Bacterial Transcription Termination by a Universal Adaptor Protein. *Molecular Cell*, 71 (6): 911-922.e4.

Le, T.T., Yang, Y., Tan, C., *et al.* (2018) Mfd Dynamically Regulates Transcription via a Release and Catch-Up Mechanism. *Cell*, 172 (1): 344-357.e15.

Lee, D.J., Bingle, L.E.H., Heurlier, K., *et al.* (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic Escherichia colistrains. *BMC Microbiology*, 9 (1 %@ 1471-2180): 252.

Lee, D.J., Minchin, S.D. and Busby, S.J.W. (2012) Activating Transcription in Bacteria. *Annual Review of Microbiology*, 66 (1): 125–152.

Levine, J.A., Hansen, A.-M., Michalski, J.M., *et al.* (2014) H-NST induces LEE expression and the formation of attaching and effacing lesions in enterohemorrhagic Escherichia coli. *PloS one*, 9 (1): e86618.

Liu, Y., Chen, H., Kenney, L.J., *et al.* (2010) A divalent switch drives H-NS/DNA-binding conformations between stiffening and bridging modes. *Genes & development*, 24 (4): 339–344.

Lloréns-Rico, V., Cano, J., Kamminga, T., *et al.* (2016) *Bacterial antisense RNAs are mainly the product of transcriptional noise*.

Lobell, R.B. and Schleif, R.F. (1990) DNA Looping and Unlooping by AraC Protein. *Science; Washington*, 250 (4980): 528.

Lodge, J., Fear, J., Busby, S., *et al.* (1992) Broad host range plasmids carrying the Escherichia coli lactose and galactose operons. *FEMS Microbiol Lett*, 74 (2–3): 271–6.

Lonetto, M.A., Brown, K.L., Rudd, K.E., *et al.* (1994) Analysis of the Streptomyces coelicolor sigE gene reveals the existence of a subfamily of eubacterial RNA polymerase sigma factors involved in the regulation of extracytoplasmic functions. *Proceedings of the National Academy of Sciences of the United States of America*, 91 (16): 7573–7577.

Lu, Z. and Lin, Z. (2019) Pervasive and dynamic transcription initiation in Saccharomyces cerevisiae. *Genome Research*, 29 (7): 1198–1210.

Lybecker, M., Zimmermann, B., Bilusic, I., *et al.* (2014) The double-stranded transcriptome of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 111 (8): 3134–3139.

Macchi, R., Montesissa, L., Murakami, K., *et al.* (2003) Recruitment of sigma54-RNA polymerase to the Pu promoter of Pseudomonas putida through integration host factor-mediated positioning switch of alpha subunit carboxyl-terminal domain on an UP-like element. *The Journal of Biological Chemistry*, 278 (30): 27695–27702.

Manak, J.R., Dike, S., Sementchenko, V., *et al.* (2006) Biological function of unannotated transcription during the early development of Drosophila melanogaster. *Nature Genetics*, 38 (10): 1151–1158.

Martin, R.G. and Rosner, J.L. (2001) The AraC transcriptional activators. *Current Opinion in Microbiology*, 4 (2): 132–137.

Martincorena, I., Seshasayee, A.S.N. and Luscombe, N.M. (2012) Evidence of non-random mutation rates suggests an evolutionary risk management strategy. *Nature*, 485 (7396): 95–98.

Martínez-Antonio, A. and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, 6 (5): 482–489.

McLeod, S.M., Aiyar, S.E., Gourse, R.L., *et al.* (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. *Journal of Molecular Biology*, 316 (3): 517–529.

Mecsas, J., Rouviere, P.E., Erickson, J.W., *et al.* (1993) The activity of sigma E, an Escherichia coli heat-inducible sigma-factor, is modulated by expression of outer membrane proteins. *Genes & Development*, 7 (12b): 2618–2628.

Merrikh, H., Ferrazzoli, A.E. and Lovett, S.T. (2009) Growth Phase and (p)ppGpp Control of IraD, a Regulator of RpoS Stability, in Escherichia coli. *Journal of Bacteriology*, 191 (24): 7436–7446.

Metzner, M., Germer, J. and Hengge, R. (2004) Multiple stress signal integration in the regulation of the complex sigma S-dependent csiD-ygaF-gabDTP operon in Escherichia coli. *Mol Microbiol*, 51 (3): 799–811.

Mika, F. and Hengge, R. (2005) A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of σS (RpoS) in E. coli. *Genes Dev*, 19: 2770–81.

Miroslavova, N.S. and Busby, S.J.W. (2006) Investigations of the modular structure of bacterial promoters. *Biochemical Society Symposia*, 73: 1–10.

Mitchell, J.E., Zheng, D., Busby, S.J.W., *et al.* (2003) Identification and analysis of 'extended –10' promoters in Escherichia coli. *Nucleic Acids Res*, 31 (16): 4689–95.

Mitra, P., Ghosh, G., Hafeezunnisa, Md., *et al.* (2017) Rho Protein: Roles and Mechanisms. *Annual Review of Microbiology*, 71 (1): 687–709.

Morgan, W.D., Bear, D.G., Litchman, B.L., *et al.* (1985) RNA sequence and secondary structure requirements for rho-dependent transcription termination. *Nucleic Acids Research*, 13 (10): 3739–3754.

Mulligan, M.E., Brosius, J. and McClure, W.R. (1985) Characterization in vitro of the effect of spacer length on the activity of Escherichia coli RNA polymerase at the TAC promoter. *J Biol Chem*, 260 (6): 3529–38.

Murakami, K.S. (2013) X-ray Crystal Structure of Escherichia coli RNA Polymerase σ70 Holoenzyme*. *J Biol Chem*, 288: 9126–34.

Murakami, K.S. (2015) Structural Biology of Bacterial RNA Polymerase. *Biomolecules*, 5: 848–64.

Murakami, K.S. and Darst, S.A. (2003) Bacterial RNA polymerases: the wholo story. *Curr Opin Struct Biol*, 13 (1): 31–9.

Murakami, K.S., Masuda, S. and Darst, S.A. (2002) *Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution*.

Nicolas, P., Mäder, U., Dervyn, E., *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in Bacillus subtilis. *Science (New York, N.Y.)*, 335 (6072): 1103–1106.

Nieto, J.M., Carmona, M., Bolland, S., *et al.* (1991) The hha gene modulates haemolysin expression in Escherichia coli. *Molecular Microbiology*, 5 (5): 1285–1293.

Nieto, J.M., Madrid, C., Prenafeta, A., *et al.* (2000) Expression of the hemolysin operon in Escherichia coli is modulated by a nucleoid-protein complex that includes the proteins Hha and H-NS. *Molecular & general genetics: MGG*, 263 (2): 349–358.

Noom, M.C., Navarre, W.W., Oshima, T., *et al.* (2007) H-NS promotes looped domain formation in the bacterial chromosome. *Current Biology*, 17 (21): R913–R914.

Oberto, J., Nabti, S., Jooste, V., *et al.* (2009) The HU Regulon Is Composed of Genes Responding to Anaerobiosis, Acid Stress, High Osmolarity and SOS Induction. *PLOS ONE*, 4 (2): e4367.

Okazaki, Y., Furuno, M., Kasukawa, T., *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, 420 (6915): 563–573.

Page, L., Griffiths, L. and Cole, J.A. (1990) Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria. *Arch Microbiol*, 154 (4): 349–54.

Paget, M.S. (2015) *Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution*., 5: 1245–1265.

Panyukov, V.V. and Ozoline, O.N. (2013) Promoters of Escherichia coli versus Promoter Islands: Function and Structure Comparison. *PLOS ONE*, 8 (5): e62601.

Park, H.-S., Östberg, Y., Johansson, J., *et al.* (2010) Novel role for a bacterial nucleoid protein in translation of mRNAs with suboptimal ribosome-binding sites. *Genes & Development*, 24 (13): 1345–1350.

Pérez-Rueda, E. and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Research*, 28 (8): 1838–1847.

Peters, J.M., Mooney, R.A., Grass, J.A., *et al.* (2012) Rho and NusG suppress pervasive antisense transcription in Escherichia coli. *Genes & Development*, 26 (23): 2621–2633.

Peters, J.M., Mooney, R.A., Kuan, P.F., *et al.* (2009) Rho directs widespread termination of intragenic and stable RNA transcription. *Proceedings of the National Academy of Sciences*, 106 (36): 15406–15411.

Peters, J.M., Vangeloff, A.D. and Landick, R. (2011) Bacterial Transcription Terminators: The RNA 3′-End Chronicles. *Journal of molecular biology*, 412 (5): 793–813.

Peterson, C.N., Ruiz, N. and Silhavy, T.J. (2004) RpoS Proteolysis Is Regulated by a Mechanism That Does Not Require the SprE (RssB) Response Regulator Phosphorylation Site. *Journal of Bacteriology*, 186 (21): 7403–7410.

Piper, S.E., Mitchell, J.E., Lee, D.J., *et al.* (2009) A global view of Escherichia coli Rsd protein and its interactions. *Molecular bioSystems*, 5 (12): 1943–1947.

Qayyum, M.Z., Dey, D. and Sen, R. (2016) Transcription Elongation Factor NusA Is a General Antagonist of Rho-dependent Termination in Escherichia coli. *The Journal of Biological Chemistry*, 291 (15): 8090–8108.

Raghunathan, N., Kapshikar, R.M., Leela, J.K., *et al.* (2018) Genome-wide relationship between R-loop formation and antisense transcription in Escherichia coli. *Nucleic Acids Research*, 46 (7): 3400–3411.

Rappas, M., Bose, D. and Zhang, X. (2006) Bacterial enhancer-binding proteins: unlocking sigma54-dependent gene transcription. *Curr Opin Struct Biol*, 17 (1): 110–6.

Ray-Soni, A., Bellecourt, M.J. and Landick, R. (2016) Mechanisms of Bacterial Transcription Termination: All Good Things Must End. *Annual Review of Biochemistry*, 85 (1): 319–347.

Rhee, H.S. and Pugh, B.F. (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature*, 483 (7389): 295–301.

Richardson, J.P. (2003) Loading Rho to Terminate Transcription. *Cell*, 114 (2): 157–159.

Roberts, J.W. (2019) Mechanisms of Bacterial Transcription Termination. *Journal of Molecular Biology*, 431 (20): 4030–4039.

Rojo, F. (1999) Repression of Transcription Initiation in Bacteria. *Journal of Bacteriology*, 181 (10): 2987–2991.

Rouvière-Yaniv, J., Yaniv, M. and Germond, J.-E. (1979) E. coli DNA binding protein HU forms nucleosome-like structure with circular double-stranded DNA. *Cell*, 17 (2): 265–274.

Sabrina, S.A., Emily, B., Sandy Jeehoon, B., *et al.* (2011) The 5.5 Protein of Phage T7 Inhibits H-NS through Interactions with the Central Oligomerization Domain. *Journal of Bacteriology*, 193 (18): 4881–4892.

Salgado, H., Peralta-Gil, M., Gama-Castro, S., *et al.* (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Research*, 41 (Database issue): D203-213.

Santero, E., Hoover, T.R., North, A.K., *et al.* (1992) Role of integration host factor in stimulating transcription from the sigma 54-dependent nifH promoter. *Journal of Molecular Biology*, 227 (3): 602–620.

Savery, N.J. (2007) The molecular mechanism of transcription-coupled DNA repair. *Trends in Microbiology*, 15 (7): 326–333.

Schlax, P.J., Capp, M.W. and Record Jr, T.M. (1995) Inhibition of Transcription Initiation buIacRepressor. *Journal of Molecular Biology*, 245 (4): 331–350.

Schwartz, E.C., Shekhtman, A., Dutta, K., *et al.* (2008) A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem Biol*, 15 (10): 1091–103.

Seila, A.C., Calabrese, J.M., Levine, S.S., *et al.* (2008) Divergent transcription from active promoters. *Science (New York, N.Y.)*, 322 (5909): 1849–1851.

Selinger, D.W., Cheung, K.J., Mei, R., *et al.* (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. *Nature Biotechnology*, 18 (12): 1262–1268.

Sharma, C.M. and Vogel, J. (2014) Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology*, 19: 97–105.

Shin, M., Lagda, A.C., Lee, J.W., *et al.* (2012) Gene silencing by H-NS from distal DNA site. *Molecular Microbiology*, 86 (3): 707–719.

Shin, M., Song, M., Rhee, J.H., *et al.* (2005) DNA looping-mediated repression by histone-like protein H-NS: specific requirement of Eσ70 as a cofactor for looping. *Genes & Development*, 19 (19): 2388–2398.

Singh, K., Milstein, J.N. and Navarre, W.W. (2016) Xenogeneic Silencing and Its Impact on Bacterial Genomes. *Annual Review of Microbiology*, 70 (1): 199–213.

Singh, S.S. and Grainger, D.C. (2013) H-NS Can Facilitate Specific DNA-binding by RNA Polymerase in AT-rich Gene Regulatory Regions. *PLoS Genetics*, 9 (6).

Singh, S.S., Singh, N., Bonocora, R.P., *et al.* (2014) Widespread suppression of intragenic transcription initiation by H-NS. *Genes & development*, 28 (3): 214–9.

Singh, S.S., Typas, A., Hengge, R., *et al.* (2011) *Escherichia coli p70 senses sequence and conformation of the promoter spacer region*., 39: 5109–5118.

Sobetzko, P., Travers, A. and Muskhelishvili, G. (2012) Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proceedings of the National Academy of Sciences*, 109 (2): E42–E50.

Soper, T., Mandin, P., Majdalani, N., *et al.* (2010) *Positive regulation by small RNAs and the role of Hfq*.

Soper, T.J. and Woodson, S.A. (2008) The rpoS mRNA leader recruits Hfq to facilitate annealing with DsrA sRNA. *RNA*, 14 (9): 1907–17.

Spurio, R., Dürrenberger, M., Falconi, M., *et al.* (1992) Lethal overproduction of the Escherichia coli nucleoid protein H-NS: ultramicroscopic and molecular autopsy. *Molecular & general genetics: MGG*, 231 (2): 201–211.

Srinivasan, R., Scolari, V.F., Lagomarsino, M.C., *et al.* (2015) The genome-scale interplay amongst xenogene silencing, stress response and chromosome architecture in Escherichia coli. *Nucleic Acids Research*, 43 (1): 295–308.

Staroń, A., Sofia, H.J., Dietrich, S., *et al.* (2009) The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Molecular Microbiology*, 74 (3): 557–581.

Stoebel, D.M., Free, A. and Dorman, C.J. (2008) Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology,* 154 (9): 2533–2545.

Storz, G. and Hengge, R. (2010) *Bacterial Stress Responses*. 2nd ed. ASM Press.

Storz, G., Vogel, J. and Wassarman, K.M. (2011) Regulation by Small RNAs in Bacteria: Expanding Frontiers. *Molecular Cell*, 43 (6): 880–891.

Swinger, K.K., Lemberg, K.M., Zhang, Y., *et al.* (2003) Flexible DNA bending in HU–DNA cocrystal structures. *The EMBO Journal*, 22 (14): 3749–3760.

Swinger, K.K. and Rice, P.A. (2004) IHF and HU: flexible architects of bent DNA. *Current Opinion in Structural Biology*, 14 (1): 28–35.

Thomason, M.K., Bischler, T., Eisenbart, S.K., *et al.* (2015) Global Transcriptional Start Site Mapping Using Differential RNA Sequencing Reveals Novel Antisense RNAs in Escherichia coli. *Journal of Bacteriology*, 197 (1): 18–28.

Thomsen, N.D., Lawson, M.R., Witkowsky, L.B., *et al.* (2016) Molecular mechanisms of substrate-controlled ring dynamics and substepping in a nucleic acid-dependent hexameric motor. *Proceedings of the National Academy of Sciences of the United States of America*, 113 (48): E7691–E7700.

Tobe, T., Yoshikawa, M., Mizuno, T., *et al.* (1993) *Transcriptional Control of the Invasion Regulatory Gene virB of Shigella flexneri: Activation by VirF and Repression by H-NS*., 175: 6142–6149.

Tomar, S.K. and Artsimovitch, I. (2013) NusG-Spt5 Proteins—Universal Tools for Transcription Modification and Communication. *Chemical Reviews*, 113 (11): 8604–8619.

Treviño-Quintanilla, L.G., Freyre-González, J.A. and Martínez-Flores, I. (2013) Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. *Current Genomics*, 14 (6): 378–387.

Typas, A., Becker, G. and Hengge, R. (2007) The molecular basis of selective promoter activation by the σS subunit of RNA polymerase. *Molecular Microbiology*, 63 (5): 1296–1306.

Typas, A. and Hengge, R. (2006) Role of the spacer between the −35 and −10 regions in σs promoter selectivity in Escherichia coli. *Molecular Microbiology*, 59 (3): 1037–1051.

Uyar, E., Kurokawa, K., Yoshimura, M., *et al.* (2009) Differential binding profiles of StpA in wild-type and h-ns mutant cells: a comparative analysis of cooperative partners by chromatin immunoprecipitation-microarray analysis. *Journal of Bacteriology*, 191 (7): 2388–2391.

Valentin-Hansen, P., Søgaard-Andersen, L. and Pedersen, H. (1996) A flexible partnership: the CytR anti-activator and the cAMP–CRP activator protein, comrades in transcription control. *Molecular Microbiology*, 20 (3): 461–466.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., *et al.* (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature*, 448 (7150): 157–162.

Wade, J.T. and Grainger, D.C. (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. *Nat Rev Microbiol*, 12 (9): 647–53.

Wang, H., Yehoshua, S., Ali, S.S., *et al.* (2014) A biomechanical mechanism for initiating DNA packaging. *Nucleic Acids Research*, 42 (19): 11921–11927.

Warman, E.A., Singh, S.S., Gubieda, A.G., *et al.* (2020) A non-canonical promoter element drives spurious transcription of horizontally acquired bacterial genes. *Nucleic Acids Research*.

Weber, H., Polen, T., Heuveling, J., *et al.* (2005) Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol*, 187 (5): 1591–603.

Will, W.R., Whitham, P.J., Reid, P.J., *et al.* (2018) Modulation of H-NS transcriptional silencing by magnesium. *Nucleic Acids Research*, 46 (11): 5717–5725.

Williamson, H.S. and Free, A. (2005) A truncated H-NS-like protein from enteropathogenic Escherichia coli acts as an H-NS antagonist. *Molecular Microbiology*, 55 (3): 808–827.

Winardhi, R.S., Gulvady, R., Mellies, J.L., *et al.* (2014) Locus of enterocyte effacement-encoded regulator (Ler) of pathogenic Escherichia coli competes off histone-like nucleoid-structuring protein (H-NS) through noncooperative DNA binding. *The Journal of biological chemistry*, 289 (20): 13739–13750.

Wong, G.T., Bonocora, R.P., Schep, A.N., *et al.* (2017) Genome-Wide Transcriptional Response to Varying RpoS Levels in Escherichia coli K-12. *J Bacteriol*, 199 (7).

Yamamoto, K., Yamanaka, Y., Shimada, T., *et al.* (2018) Altered Distribution of RNA Polymerase Lacking the Omega Subunit within the Prophages along the Escherichia coli K-12 Genome. *mSystems*, 3.

Yamashino, T., Ueguchi, C. and Mizuno, T. (1995) Quantitative control of the stationary phase-specific sigma factor, sigma S, in Escherichia coli: involvement of the nucleoid protein H-NS. *The EMBO Journal*, 14 (3): 594–602.

Yura, T., Nagai, H. and Mori, H. (1993) Regulation of the Heat-Shock Response in Bacteria. *Annual Review of Microbiology*, 47 (1): 321–350.

Zhang, A., Derbyshire, V., Salvo, J.L., *et al.* (1995) Escherichia coli protein StpA stimulates self-splicing by promoting RNA assembly in vitro. *RNA (New York, N.Y.)*, 1 (8): 783–793.

Zhang, G. and Darst, S.A. (1998) Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. *Science*, 281 (5374): 262–6.

Zhang, Y., Feng, Y., Chatterjee, S., *et al.* (2012) Structural Basis of Transcription Initiation. *Science*, 338 (6110): 1076–80.

Zhou, Y. and Gottesman, S. (2006) Modes of Regulation of RpoS by H-NS. *Journal of Bacteriology*, 188 (19): 7022–7025.

Zhou, Y., Gottesman, S., Hoskins, J.R., *et al.* (2001) The RssB response regulator directly targets sigma(S) for degradation by ClpXP. *Genes & Development*, 15 (5): 627–637.