Imperial College London

The sigma^{54} activator bypass problem *in vivo* and *in vitro*

Jorrit Schäfer

Division of Cell and Molecular Biology,
Department of Life Sciences,
Faculty of Natural Sciences.

SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY DECEMBER 2014

Abstract

Tight regulation of gene expression is crucial for the survival of an organism, and allows certain genes to be switched on or off depending on the growth conditions and the state of differentiation. Transcription initiation is the most highly regulated step of gene expression, preventing wastage and being subject to the action of sophisticated signalling pathways. The RNA polymerases are further regulated through multiple different activators (stimulating transcription) and repressors (inhibiting transcription). Notably, promoter recognition specificity in bacteria is regulated by several dissociable sigma factors which can bind to the RNA polymerase core enzyme.

Sigma^{54} (σ^{54}) is the major alternative sigma factor in *E. coli* and historically known for its role in nitrogen metabolism. One unique property of σ^{54}-dependent transcription is that it recognises -12 and -24 promoter elements rather than the traditional -10 and -35 sequences. Another distinguishing feature of σ^{54}-dependent transcription initiation is that it absolutely requires the activity of a cognate activator and ATP hydrolysis, potentially giving tight control over gene expression and a wider dynamic range than with σ^{70}-dependent transcription.

Interestingly, this activator requirement has been shown to be bypassed *in vitro* with σ^{54} mutants where the interaction with the -12 promoter DNA element is disrupted. However, σ^{54}-dependent transcription is not readily observed *in vivo* for the same mutants, suggesting further barriers exist *in vivo* inhibiting activator independence from occurring. The -12 promoter element has been shown to be somewhat expendable for σ^{54} promoter binding, and activator-independent σ^{54} transcription would obviate the need for additional proteins and ATP. This raises the question as to why σ^{54}-dependent transcription has not evolved towards activator independence over time, by bypassing the interaction with the -12 promoter DNA.

In this study, I show that σ^{54}-dependent inhibition of transcription was detected for several genes with σ^{54} binding sites in their promoter region (*ytfJ, chaC, patA, argT, mdfA ybhK,* and *acrD*). Additionally, local transcriptional repression in the RNASeq data directly correlated with proximal σ^{54} binding sites, suggesting a novel role for σ^{54} as a transcriptional repressor. Strikingly, the -12 consensus motifs were more conserved in promoters linked to σ^{54}-dependent transcriptional inhibition, underlining the importance of this promoter element in this newly discovered putative repressive function.

However, extensive screens for genes inhibiting bypass transcription *in vivo* failed to identify any major repressive genes that keep activator independence in check *in vivo*. The

mechanisms maintaining activator bypass transcription at low levels *in vivo* still remain to be characterised.

Table of Contents

Declaration of originality

I, Jorrit Schäfer declare that I am the sole author of the written work in this thesis and that all the results are original. Results in Chapter 5 were obtained in collaboration with Vertis Biotechnologie AG (Freising, Germany), all other experiments were performed by myself. Any work not my own is referenced accordingly.

Copyright declaration

Acknowledgements

I am very grateful to Prof. Martin Buck for the opportunity to do this PhD project, and for his invaluable knowledge and friendly guidance throughout the course of the project.

I would also like to thank all the members of the Buck lab (past and present) for their assistance and providing an enjoyable working environment. In particular, I would like to express my gratitude to Dr. Christoph Engl for his supervision, expertise and help.

A big thank you also goes to the lab of Dr. Joseph Wade (University of Albany, NY) for sharing their data on genomewide σ^{54} ChIP binding.

Finally, I would also like to thank my parents for their continuous support over the course of my PhD.

List of Figures

List of Tables

Abbreviations

^{32}P	Radio-labelled Phosphate
AAA+	ATPases Associated with Various Cellular Activities
ADP	Adenosine 5'-Diphosphate
Amp	Ampicillin
Amp^R	Ampicillin resistant
Ara	Arabinose
ATP	Adenosine 5'-Triphosphate
bEBP	Bacterial Enhancer Binding Protein
BGAL	β-Galactosidase
BLAST	Basic local alignment search tool
Bp	Base pairs
Cam	Chloramphenicol
Cam^R	Chloramphenicol resistant
cDNA	complementary DNA
CR	Conserved Region in RNAP ω Subunit
CTD	C-terminal Domain
dNTP	Deoxyribose nucleotide triphosphate
DRRS	Δ*rpoN* reporter strain
E	Core RNAP
$E\sigma^{54}$	σ^{54}RNAP holoenzyme
$E\sigma^{70}$	σ^{70}RNAP holoenzyme
FlgR	Flagellar Regulator

FRT	Flippase recognition target
GFP	Green fluorescent protein
HTH	Helix-Turn-Helix
IHF	Integration Host Factor
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
Kan	Kanamycin
Kan^R	Kanamycin resistant
kDa	kilodalton
Lac^-	Unable to grow on lactose
Lac^+	Able to grow on lactose
LB	Luria Broth
M-MLV RT	Moloney Murine Leukemia Virus Reverse Transcriptase
mRNA	Messenger RNA
NifA	Nitrogen Fixation Protein A
NMR	Nuclear Magnetic Resonance
nt	Nucleotide
NTD	N-terminal Domain
NtrC	Nitrogen Regulatory Protein C
o/n	Overnight
ONPG	Ortho-Nitrophenyl-β-Galactoside
P1*vir*	Virulent P1 bacteriophage
PCR	Polymerase Chain Reaction
P_i	Phosphate
pIV	Phage fi Protein IV
ppGpp	Guanosine 5'-Diphosphate 3'-Diphosphate
Psp	Phage shock protein
PspF	Phage Shock Protein
$PspF_{1-275}$	Truncated phage shock protein F with residues 1-275
Rho	Rho factor protein
RNAP	RNA Polymerase
RNAPII	RNA Polymerase II
RNASeq	RNA Sequencing
RPc	Closed Complex
RP_I	Intermediate Complex

RPo	Open Complex
rRNA	ribosomal RNA
SABRS	Sigma^{54} activator bypass reporter strain
SABT	σ^{54} activator bypass transcription
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
sRNA	Small regulatory RNA
T7RNAP	RNA polymerase from T7 bacteriophage
TEC	Transcription Elongation complex
TEMED	N, N, N', N'-Tetramethylethylenediamine
Tet	Tetracycline
Tet^{R}	Tetrycacline resistant
TFIIH	Transcription factor II H
Tn*10*	Tn10 transposon
trans	On the other side (From a different protomer)
tRNA	Transfer RNA
UAS	Upstream Activation Sequence
V4R	Vinyl 4 Reductase
w/v	weight/volume
WT	Wild Type
WTRS	Wild type reporter strain
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
α	RNAP Alpha Subunit (RpoA)
β	RNAP Beta Subunit (RpoB)
β'	RNAP Beta Prime Subunit (RpoC)
σ	Sigma factor
σ^{18}(FecI)	Sigma18 (FecI)
σ^{24}(E)	Sigma 24 (RpoE
σ^{28}(F)	Sigma28 (RpoF)
σ^{32}(H)	Sigma 32 (RpoS)
σ^{38}(S)	Sigma 38 (RpoS)
σ^{54}	Sigma 54 (RpoN)
σ^{54}_{RI}	σ^{54} Region I
σ^{54}_{RII}	σ^{54} Region II
σ^{54}_{RIII}	σ^{54} Region III

$\sigma^{54}_{\Delta R1}$	σ^{54} with Region I deletion
σ^{70}	Sigma70 (RpoD)
$\sigma^{70}_{1.1}$	σ^{70} Region 1.1
σ^{70}_{2}	σ^{70} Region 2
$\sigma^{70}_{2.3}$	σ^{70} Region 2.3
$\sigma^{70}_{2.4}$	σ^{70} Region 2.4
σ^{70}_{3}	σ^{70} Region 3
σ^{70}_{4}	σ^{70} Region 4
$\sigma^{70}_{4.2}$	σ^{70} Region 4.2
$\Phi P_{gln\mathrm{Ap2}}$-*gfp*	*gfp* reporter gene fused to a gln_{Ap2} promoter
ΦP_{pspA}-*lacZ*	*lacZ* reporter gene fused to a *pspA* promoter
ω	RNAP Omega Subunit (RpoZ)

Chapter 1: Introduction

1.1 Gene expression and transcription in bacteria

Throughout the three domains of life, organisms must adapt to rapidly changing environments in order to survive. Such changes are often driven by the regulated expression of genes encoded in their genome. Gene expression typically consists of two steps. The first step is transcription, where genomic DNA is transcribed into messenger RNA (mRNA), structural RNA, and regulatory non coding RNA. Where the gene product is a protein, translation is the second step, and mRNA is decoded by ribosomes to produce proteins.

Tight regulation of gene expression allows certain genes to be switched on or off depending on the growth conditions and the state of differentiation, and is therefore crucial for the development and survival of an organism. Transcription initiation is the most highly regulated step of gene expression and is subject to the action of sophisticated signalling pathways to prevent wastage.

RNA synthesis is catalysed by the multi-subunit RNA polymerase (RNAP) in all living organisms and even some viruses. In bacteria, the RNAPs are further regulated through multiple different activators (stimulating transcription) and repressors (inhibiting transcription), as well as by the intrinsic properties of the DNA sequences with which they interact. In *E. coli* over 100 transcription factors modifying the activity of the RNAP have been identified (Ishihama, 2000). Notably, promoter recognition specificity in bacteria is further regulated by a dissociable sigma (σ) factor.

Sigma factors are required for bacterial transcription initiation, recruiting core RNAP to target gene promoters. The σ factors bind to the core RNAP ($\alpha_2\beta\beta'\omega$) to form the RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma$). The holoenzyme is capable of recognising and binding consensus DNA sequences upstream of the transcription start site, forming a closed complex (transcriptionally silent). Here the DNA to be transcribed sits outside of the holoenzyme in a double stranded form.

The closed promoter complex structure undergoes conformational changes involving several intermediate complexes, during which the DNA duplex is melted and the template DNA is loaded into the active site of RNAP (Kontur et al, 2006; Feklistov and Darst, 2011). These conformational changes result in formation of the open complex (transcriptionally active).

Short abortive transcripts up to 12 nucleotides long are generated prior to entering elongation (Carpousis, 1980; Goldman, 2009). The σ factor is usually released at initiation or during early elongation, and the RNAP complex produces a nascent RNA chain with high accuracy and a processivity of up to 100 nucleotides per second (McLure, 1985; Mooney, 1998). Elongation continues until transcription termination, which is Rho-independent when the RNA forms a hairpin loop. Alternatively, transcription termination can be Rho-dependent where the protein factor Rho destabilises the template/RNA interaction and releases the synthesised RNA from the elongation complex. This process is summarised below (Fig. 1).

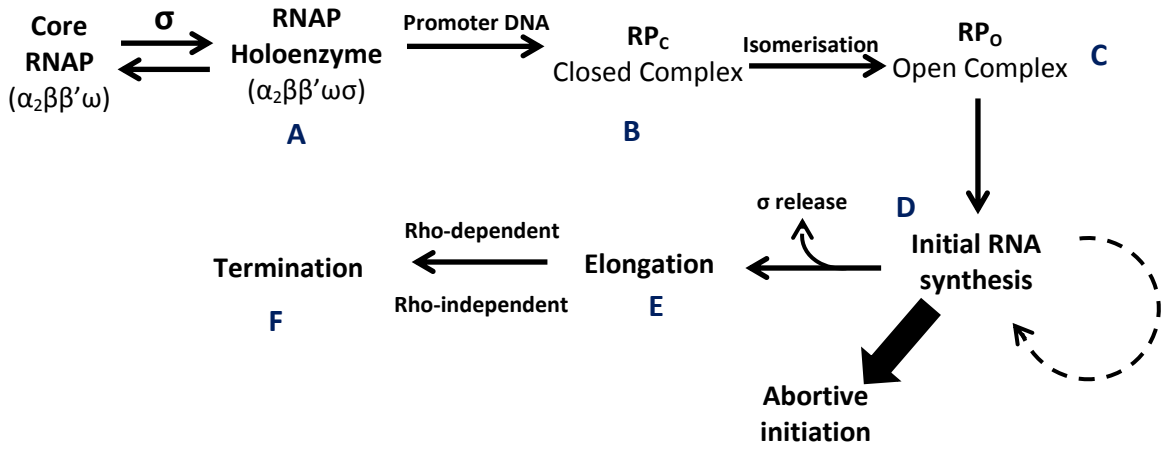


Fig. 1.1: **Schematic outlining bacterial transcription initiation.**
Holoenzyme formation (**A**) is followed by promoter recognition and closed complex formation (**B**). The closed complex isomerises into an open active complex (**C**), after which RNA synthesis can commence (**D**). Next, the σ factor is released and transcription proceeds into the elongation phase (**E**) until a transcriptional terminator is reached which can be either Rho-dependent or Rho-independent (**F**).

Following transcriptional termination the nascent RNA acts as a template for protein synthesis, as a structural element, or as a regulatory molecule in its own right. In the bacterial cytoplasm the mRNA produced is bound by ribosomes, and decoded 3 nucleotides at a time. As the mRNA is decoded it is paired with transfer RNAs (tRNA) that contain 3 nucleotide anti-codons complementary to the mRNA. These tRNAs carry certain amino acids (depending on the tRNA) that can be chained together to form polypeptides from the N-terminus to C-terminus, as the mRNA is read by the ribosomes. This continues until a stop codon (UAG/UAA/UGA) is reached, and the polypeptide is released from the ribosome. In doing so different polypeptides and proteins are formed depending on the mRNA sequence, a process known as translation.

1.2 Bacterial core RNAP

1.2.1 Multi-subunit RNAPs

The multi-subunit RNAP is the main component of transcription in living organisms and generally highly conserved. Bacteria have a single 400 kDa core RNAP catalysing the synthesis of RNA, which is structurally and mechanistically similar to the multi subunit RNAP found in eukaryotes and archaea (Ebright, 2000). Eukaryotes contain several types of RNAP responsible for the transcription of different types of RNA, as well as some organelle-specific RNAPs such as those found in mitochondria and chloroplasts. RNAP II is the most studied eukaryotic RNAP as it is responsible for the transcription of mRNA, and is also the most similar to RNAPs in bacteria and archaea (Ebright, 2000). Despite differences between multi-subunit RNAPs throughout the 3 domains of life, they are believed to have a common ancestor (Werner and Grohmann, 2011) and therefore share a great deal of homology (Table 1.2.1).

However, these multi-subunit RNAPs are not structurally or mechanistically related to the single-subunit RNAPs found in many viruses (eg. T7 RNAP), chloroplasts, and mitochondria. These RNAPs are capable of recognising promoter sequences without the need for additional regulatory factors (Cheetham, 2000; Steitz, 2009).

	Core Enzyme Homologues					No. of Subunits
Bacterial	β′	β	α_i	α_{ii}	ω	5
Archaeal	RpoA	RpoB	RpoD	RpoL	RpoK	12
Eukaryotic RNAP I	A190	A135	AC40	AC19	RPB6	14
Eukaryotic RNAP II	RPB1	RPB2	RPB3	RPB11	RPB6	12
Eukaryotic RNAP III	C160	C128	AC40	AC19	RPB6	17

Fig 1.2.1: Conserved RNAP subunits in the three domains of life (Werner and Grohmann, 2011).

The 5 main subunits of bacterial core RNAP ($\alpha_2\beta\beta'\omega$) are highly conserved, with homologues throughout the 3 domains of life (Bacteria, Archaea, Eukaryotes).

1.2.2 Bacterial core RNAP structure

The bacterial core RNAP resembles a crab claw containing the β/β' 'jaw' domains interacting with the downstream DNA-binding site. The subunit composition of the core bacterial RNAP being: two alpha subunits, one beta subunit, one beta prime subunit and an omega subunit (Fig 1.2.2). The central channel is defined by the β and β' subunits and is involved in the binding of the duplex DNA template (Vassylyev, 2002). This central channel is divided into two parts by the conserved β' bridge helix (Haugen, 2008), thought to be involved in the nucleotide addition cycle (Jovanovic, 2011). At the back of central channel is a catalytic Mg^{2+} ion, which is chelated by a sequence conserved in all bacterial and eukaryotic cellular RNAPs (Zhang, 1999).

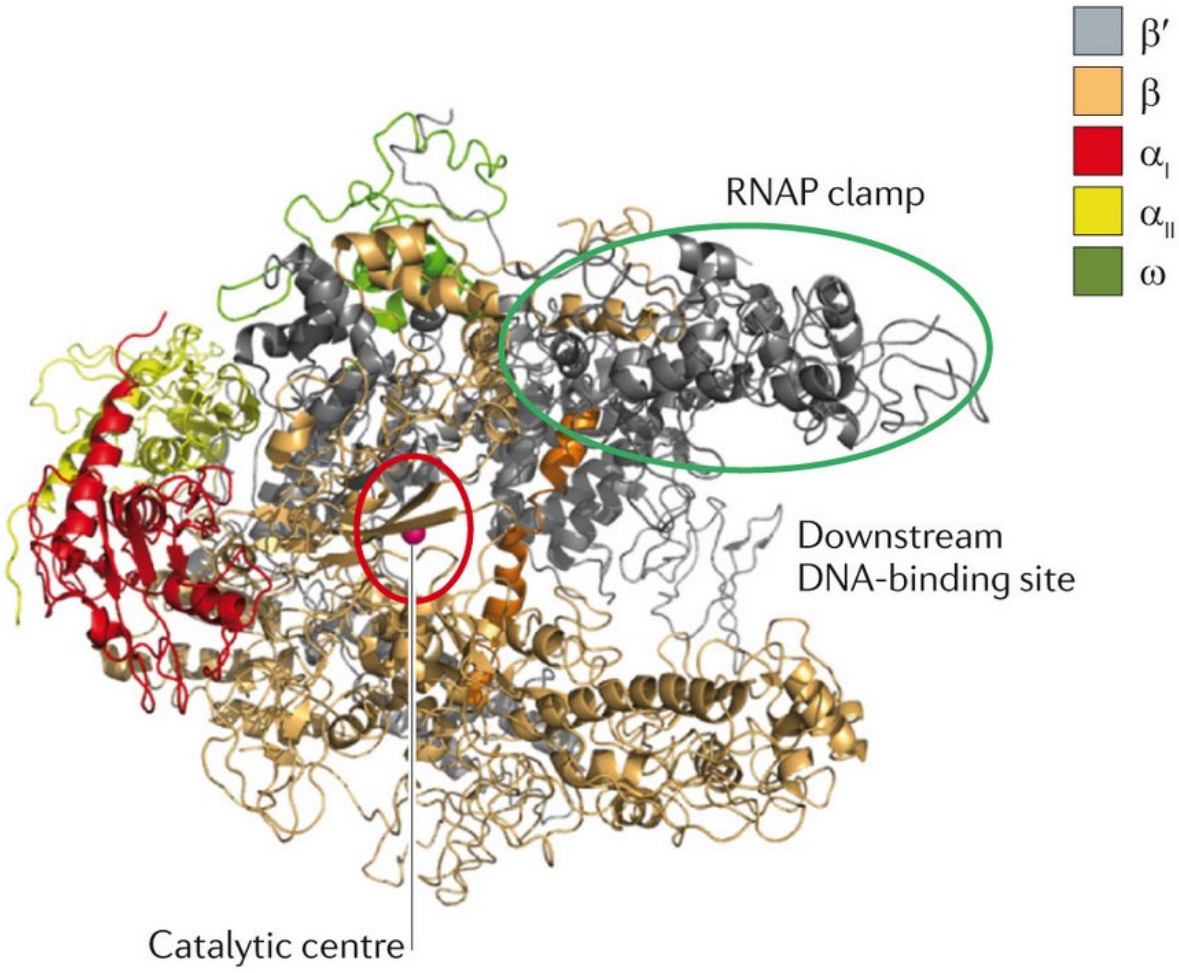


Fig 1.2.2: Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution.
The bacterial core RNAP consists of 5 subunits: two alpha subunits (red and yellow), one beta subunit (beige), one beta prime subunit (grey), and an omega subunit (green). The structure resembles a crab claw with a Mg^{2+} ion at the catalytic centre of the molecular (magenta). The bridge helix (orange) links the two beta subunits (Werner and Grohmann, 2011).

1.2.3 The alpha subunit

rpoA encodes the alpha (α) subunit of RNAP consisting of 329 amino acids (37 kDa). The multi-subunit RNAP contains two α subunits which are involved in 3 main functions: initiating RNAP assembly upon homodimerisation (Haugen, 2008), promoter recognition through sequence-specific protein-DNA interactions, and being the target of positive transcriptional regulators such as Catabolite Activator Protein (Ebright and Busby, 1995).

The two α subunits are identical in sequence but interact with different regions of the RNAP complex. The α_i and α_{ii} subunits are responsible for binding to the beta and beta prime

subunits respectively (Ebright, 2000; Arthur, 2000; Severinov, 2000). Each α subunit is consists of a C-terminal and an N-terminal domain, joined by a flexible linker (Fig 1.2.3).

The N-terminal domain of the α subunit (residues 1-235) is involved in the homodimerisation during RNAP assembly ($2\alpha \rightarrow \alpha_2 \rightarrow \alpha_2\beta \rightarrow \alpha_2\beta\beta' \rightarrow \alpha_2\beta\beta'\omega$), as well as contacts with the beta and beta prime subunits (Ebright and Busby, 1995).

The smaller C-terminal domain of the α subunit (residues 249-329) serves as a flexibly tethered DNA-binding, activator-binding adaptor (Ebright and Busby, 1995). The C-terminal domain interacts directly with the A/T-rich upstream promoter elements, present upstream of the -35 element in certain strong promoters (Rao, 1994). The C-terminal domain has also been shown to be subject to positive transcriptional regulation of activators such as the catabolite activator protein (Zou, 1992).

Fig 1.2.3: Domain organisation of the α subunit of bacterial core RNAP.
The bacterial α subunit consists of an N-terminal domain (1-235) involved in homodimerisation, RNAP assembly and interactions with the beta and beta prime subunits of RNAP. This domain is connected to the C-terminal domain via a flexible linker region. The C-terminal domain (249-329) contains regions involved in the interaction with upstream promoter DNA, the UP element.

1.2.4 The beta subunit

rpoB encodes the beta (β) subunit of the bacterial RNAP consisting of 1342 amino acids (151 kDa), which combined with the beta prime subunit forms the main catalytic channel of RNAP (Fig 1.2.2). The β subunit contains 3 domains important for RNAP function: the upstream lobe, the downstream lobe, and the β flap (Fig 1.2.4).

The upstream lobe interacts with the σ subunit in order to maintain strand separation at the leading edge of the transcription bubble (Severinov, 2000).

The downstream lobe (residues 186-433) is part of an evolutionary variable region and involved in the formation of stable heparin-resistant open complexes (Severinov, 2000).

The β flap (residues 835-935) is involved in transcription initiation and elongation. During transcription initiation the β flap interacts with the Region 4 of the σ^{70} subunit using a hydrophobic patch at the tip of the flap domain, positioning it for the recognition of the -35 promoter element (Geszvain, 2004). Unlike for the -35 promoter element of σ^{70}, a β flap deletion does not disrupt the recognition of the σ^{54} promoters (Wigneshweraraj, 2003). During transcription elongation the β flap covers the RNA exit channel, and contacts a nascent RNA hairpin stem-loop structure that obstructs transcription (Toulokhonov, 2001).

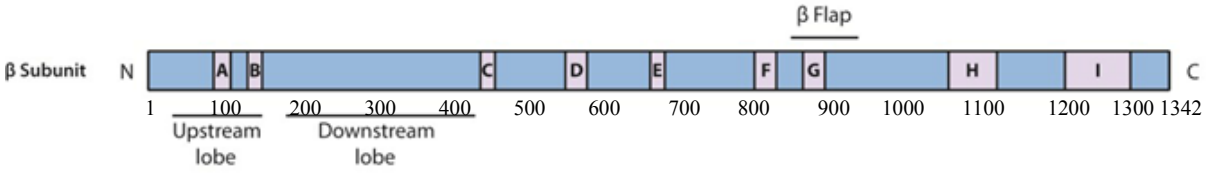


Fig 1.2.4: Domain organisation of the β subunit of bacterial core RNAP.
The bacterial 1342 amino acid β subunit of RNAP contains 9 highly conserved regions in *E. coli* and yeast (A-I), separated by variable (blue) regions (Sweetser, 1987). 3 key domains of the β subunit include the upstream lobe, the downstream lobe, and the β flap.

1.2.5 The beta prime subunit

rpoC encodes the beta prime (β')subunit of the bacterial core RNAP, and is made up of 1407 amino acids (155 kDa). The β'subunit, along with the beta subunit, is crucial for the enzymatic activity of the RNAP enzyme, particularly at the promoter melting stage (Young, 2004). The β'subunit include several important domains for RNAP function including:

- The **zipper** removes the RNA product from the DNA/RNA hybrid, and is also involved in binding to the σ subunit (Nudler, 2009).

- The **lid** interacts with the RNA exit channel and is involved in strand separation, displacing the RNA transcript from the DNA/RNA hybrid, and termination (Naryshkina, 2006).
- The highly conserved **zinc finger** is involved in binding zinc atoms, and thought to be involved in RNAP assembly in eubacteria and chloroplasts (Markov, 1999).
- The **coiled-coil** binds to the σ subunit for non-template strand recognition (Arthur, 2000).
- The **rudder** is also located in the coil-coil region and is involved in the stabilisation of the DNA–RNA hybrid as it translocates through RNAP (Korzheva, 2000).
- The **NAD-FDGD motif** is involved in chelation of the active center Mg^{2+} using 3 aspartate residues (Zaychikov, 1996).
- The **bridge helix** divides the main active site from the secondary channel, and is also thought to be involved in translocation and catalysis (Nedialkov, 2013).
- The **upstream clamp** is thought to restrict the movement of the trigger loop and jaw domain, which would otherwise interfere with loading of DNA (Saecker, 2011).
- The **trigger loop** is important for initial RNA synthesis until a complete DNA–RNA hybrid is formed, transcription fidelity during nucleotide incorporation, and preventing aberrant transcription termination at non-terminator sites (Foqueau, 2013).
- The **jaw** domain stabilises the open complex and binds the duplex DNA (Wigneshweraraj, 2005).

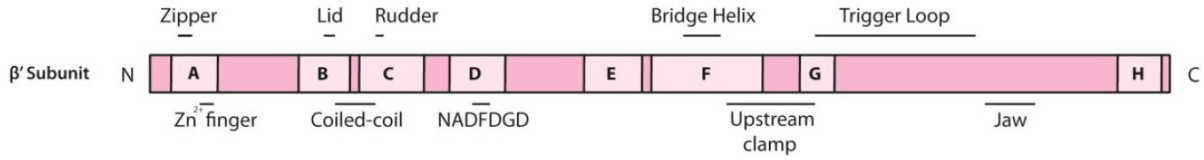


Fig 1.2.5: Domain organisation of the β' subunit of bacterial core RNAP.

The bacterial 1407 amino acid beta prime subunit of RNAP contains 8 conserved regions in *E. coli* and yeast (A-H), separated by variable (pink) regions (Sweetser, 1987). Key domains of the beta prime subunit include the zipper, Zn^{2+} finger, lid, coiled-coil, rudder, NAD-FDGD motif, bridge helix, upstream clamp, trigger loop, and jaw domain.

1.2.6 The omega subunit

rpoZ encodes the 91 amino acid omega (ω) subunit of the bacterial core RNAP (10 kDa), making it the smallest RNAP subunit. While the ω subunit was not considered part of the RNAP complex until the late 1990s, it has now been shown to have several structural and functional roles for RNAP.

The ω subunit is involved in the folding of the β' subunit, physical protection of the β' subunit, and recruiting β' subunit during RNAP assembly (Mathew, 2006). As such the ω subunit plays a key part in transcriptional regulation (Gunnelius, 2014).

The ω subunit was thought to modulate RNAP activity in the stringent response by influencing its sensitivity to ppGpp (Vrentas, 2005), however this finding was later questioned after obtaining a high-resolution X-ray structure (Vrentas, 2008). Moreover, the ω subunit has also been implicated in heat stress adaptation in *Bacillus cereus* (Periago, 2003). Finally, the ω has also been suggested to play a role in stationary phase survival (Mathew, 2006).

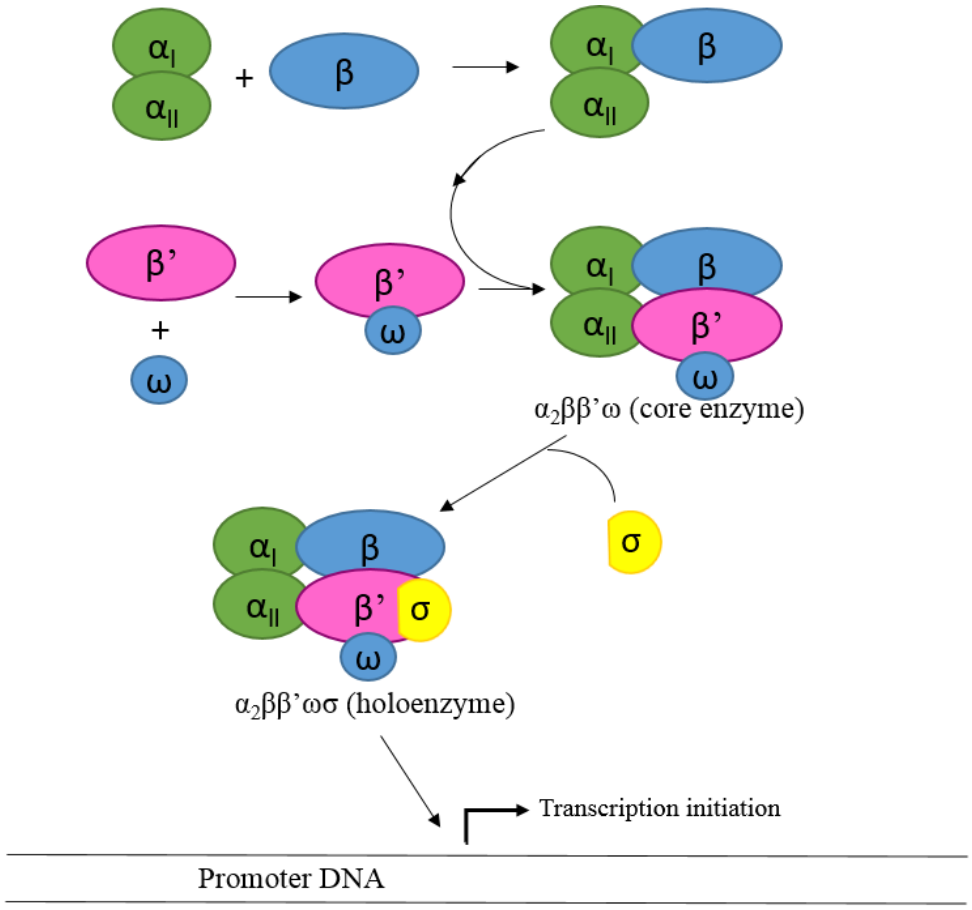


Fig 1.2.6: The bacterial ω subunit is involved in RNAP assembly.

The ω subunit of the bacterial core RNAP interacts with the β' subunit, aiding it in RNAP assembly by recruiting it to the $\alpha_2\beta$ subassembly.

1.3 The sigma (σ) factors

The bacterial core RNA polymerase complex ($\alpha_2\beta\beta'\omega$) is sufficient for transcription elongation and termination, but is unable to initiate transcription. Bacterial transcription

initiation also requires a dissociable σ factor. Sigma factors are essential intermediaries in promoter recognition, RNAP recruitment, DNA melting, and transcriptional regulation (Gross, 1998).

Bacteria need to respond to rapidly changing environments in order to survive. Tight regulation of gene expression is therefore crucial for the survival of an organism and allows certain genes to be switched on or off depending on the growth conditions and the state of differentiation. Bacteria achieve this through the regulated expression of several different sigma factors which operate under various cellular conditions, and drive the expression of distinct promoter classes under different conditions (Guo, 1999). Bacteria like *Streptomyces coelicolour* have over 60 different σ factors regulating their complex life cycle of physiological and morphological differentiation (Kim, 2008).

Although many bacteria contain multiple sigma factors, one single sigma factor is usually responsible for the bulk of gene transcription and crucial for cell viability. In *E. coli* this primary sigma factor is σ^{70}. It is involved in growth and metabolism by regulating the expression of "housekeeping genes". In *E. coli*, 7 types of sigma factor that have been characterised (σ^{70}, σ^{54}, σ^{38}, σ^{32}, σ^{28}, σ^{24} and σ^{18}), each regulating the transcription of specific groups of genes (Table 1.3).

Organism	σ	Gene	Function
E. coli	σ^{70} (σ^{D})	*rpoD*	Housekeeping genes
	σ^{H} (σ^{32})	*rpoH*	Heat shock
	σ^{E} (σ^{24})	*rpoE*	Extreme heat shock, periplasmic stress (ECF)
	σ^{F} (σ^{28})	*fliA*	Flagellar-based motility
	σ^{S} (σ^{38})	*rpoS*	Stationary phase adaptations
	σ^{N} (σ^{54})	*rpoN*, *glnF*	Nitrogen-regulated genes
	σ^{fecI}	*fecI*	Ferric citrate uptake (ECF)

Table 1.3: The *E. coli* σ factors.

There are 7 σ factors in *E. coli*, which can be divided into two categories: σ^{54}- and σ^{70}-like (σ^{70}, σ^{32}, σ^{24}, σ^{28}, σ^{38}, σ^{18}) σ factors.

Sigma factors achieve specificity for the transcription of certain genes by recognising specific bacterial promoters. These bacterial promoters usually contain two short sequence elements upstream of the transcription start site, and typically vary between genes allowing different genes to be regulated independently by different sigma factors. Typically bacterial promoters

contain a -35 (TTGACA) element and a -10 (TATAAT) element, found 35 and 10 bases upstream of the transcription start site respectively. While conserved on average, these elements are generally only 50-70% intact.

The *E. coli* σ factors are classified into two structurally unrelated families that share no sequence homology: σ^{70} and σ^{54}. The σ^{54} family consists solely of σ^{54}, whereas all the other σ factors fall into the σ^{70} class of σ factors. The σ^{54} is different from σ^{70}-like σ factors in that it has a different amino acid sequence, promoter element, and mechanism of action (Buck, 2000). In particular, the σ^{54} family recognises -24 and -12 promoter elements (as opposed to -35/-10 elements), and requires both an activator and ATP hydrolysis to initiate RNAP transcription.

While sigma factors are important intermediaries in transcriptional regulation in their own right, they are also sometimes subject to negative regulation themselves by proteins known as anti-sigma factors. Examples include Rsd, which has been shown to inhibit the activity of σ70-dependent transcription (Hofmann, 2011). FlgM is another anti-sigma factor responsible for inhibiting σ28, which is involved in flagellar biosynthesis (Daughdrill, 1997; Hughes, 1998). DnaK is a general regulator of the heat shock response, which in E. coli inhibits the heat shock sigma factor σ32 (Hughes, 1998).

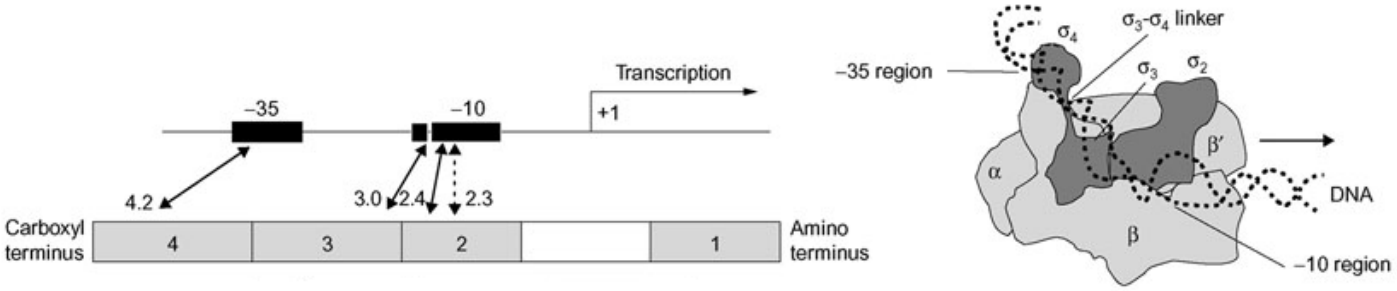
1.3.1 The sigma^{70} (σ^{70}) family

The σ^{70} family of σ factors recruit the bacterial core RNA polymerase to specific promoter elements that are situated 10 and 35 bases upstream of the transcription start site. Members of the σ^{70} family are involved in DNA melting, act as targets for activator proteins (eg. PhoB, CRP, and λcl), and play a role in transcription initiation itself (Paget, 2003).

Sequence conservation studies among the σ^{70} family members revealed 4 conserved domains (Fig 1.3.1). Regions 2 and 4 are highly conserved in all members of the σ^{70} family, and are involved in binding to RNA polymerase, promoter recognition, and DNA melting.

The C-terminal part of region 4 includes a helix-turn-helix motif involved in binding the -35 element, while subdomain 2.4 is involved in recognizing the -10 bacterial promoter element. Residues in region 2.3 are implicated in melting of the duplex DNA (Fig 1.3.1).

The N-terminal part of region 3 (region 3.0) is involved in recognising the extended −10 promoter element (^{15}TG^{14}), which can be present in promoters lacking a -35 promoter element. Region 1.1 functions as an autoinhibitory domain to mask DNA binding determinants in free σ^{70} (Gruber, 2003). Hence promoter recognition only occurs in the RNAP holoenzyme.



σ^{70} subdomain	Function
$\sigma^{70}_{1.1}$	**Autoinhibitory domain**
$\sigma^{70}_{2.3}$	**DNA melting**
$\sigma^{70}_{2.4}$	**Binding to -10 promoter element**
$\sigma^{70}_{3.0}$	**Binding to extended -10 promoter element**
$\sigma^{70}_{4.2}$	**Binding to -35 promoter element and core RNAP**

Fig 1.3.1: Domain organisation of σ^{70} and domain function.

The σ^{70} protein can be divided into 4 domains based on its sequence conservation with other members of the σ^{70} family (1-4). A model for the interaction of the RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma$) with promoter DNA is also shown (Paget, 2003).

1.3.2 The sigma^{70} (σ^{70}) family subgroups

While all bacteria typically have an essential σ factor, the number of genes encoding σ factors can vary greatly from 1 in *Mycoplasma genitalium* to 64 in *Streptomyces coelicolour* (Gruber, 2003). Therefore, the σ^{70} family is further subdivided into 4 groups based on σ factor structure and function (Lonetto, 1992).

Group 1 of the σ^{70} family includes essential primary σ factors closely related to the *E. coli* σ^{70}, which are typically involved in the growth of exponentially growing cells. Region 1 (involved in autoinhibition) is generally only conserved in Group 1 of the σ^{70} family (Fig 1.3.2).

Group 2 consists of σ factors closely related to the essential sigma factors but are not required for cell growth. In *E. coli* one such Group 2 σ factor is σ^{S}, the general stress regulator of *E. coli*. It is involved transcribing over 70 genes that mitigate the effects of oxidative stress, UV-radiation, heat shock, hyperosmolarity, acidic pH, and ethanol (Hennge-Aronis, 2002).

Group 3 includes alternative σ factors more distantly related to the primary σ factor, and are involved in the transciption of genes in response to certain conditions like heat shock, sporulation, and flagellar biosynthesis (Paget, 2003).

Group 4 includes the extracytoplasmic function sigma factors that diverge in sequence relative to most other sigma factors. The σ factors in Group 4 typically lack the amino-terminal part of region 3 (Fig 1.3.2) involved in binding extended -10 promoter elements (Helmann, 2002).

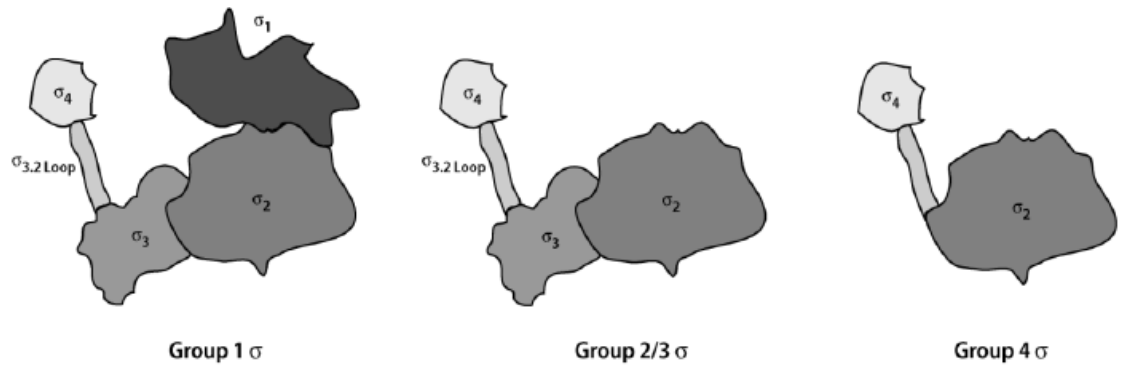


Fig 1.3.2: Domain conservation in the σ^{70} family groups.

All 4 domains of σ^{70} are conserved in group 1 of the σ^{70} family (primary σ factors). In groups 2 and 3 (alternative σ factors) region 1 is generally not conserved, while in group 4 (extracytoplasmic function σ factors) only regions 2 and 4 are conserved (adapted from Gruber, 2003).

1.3.3 Sigma^{70}-dependent (σ^{70}) transcription initiation

Binding of the dissociable σ^{70} factor to the RNAP core enzyme ($\alpha_2\beta\beta'\omega$) forms the RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma$), which is capable of recognizing -10 and -35 promoter elements. The interaction of the RNAP holoenzyme with promoter DNA initiates a series of conformational changes, starting with the formation of an inactive close promoter complex (Fig 1.3.3, A).

In the **closed promoter complex** the $\sigma^{70}{}_2$ and $\sigma^{70}{}_4$ bind to the -10 and -35 promoter elements respectively (Mecsas, 1991). The holoenzyme covers the promoter region from -55 to +1, protecting the double stranded DNA from degradation (Haugen, 2008). The $\sigma^{70}{}_{3.2}$ loop protrudes into the active channel (Fig 1.3.3, A) where it is thought to play a role in initiating nucleotide substrate binding (Murakami, 2002). Also located in the active channel is $\sigma^{70}{}_{1.1}$, which inhibits free σ^{70} from binding to promoter DNA by masking $\sigma^{70}{}_4$, which is involved in recognising the -35 promoter element (Dombroski, 1992). It has been suggested that the negatively charged $\sigma^{70}{}_{1.1}$ domain is important for opening up the active channel of RNAP, allowing duplex DNA to enter prior to DNA melting (Darst, 2002).

During an **intermediate step** (Fig 1.3.3, B), aromatic residues (W, Y and F) of $\sigma^{70}{}_{2.3}$ are thought to be exposed to the nontemplate bases of the -10 promoter element (Spassky, 1988). The A_{-11} and T_{-7} are flipped out of the ssDNA base-stack and buried in complementary protein pockets in $\sigma^{70}{}_2$ (Feklistov, 2014). The interactions between the protein and DNA are believed to stabilise a short region of transiently melted DNA, extending from the A_{-11} to the +1 transcription start site (Murakami, 2003). This melting is hypothesised to allow the DNA to unwind, and downstream DNA to bend across the active site of the RNAP (Vassylyev, 2002). The entry of the duplex DNA into the active channel simultaneously displaces the $\sigma^{70}{}_{1.1}$ region through an unknown mechanism (Murakami, 2002).

As the **open complex** is formed (Fig 1.3.3, C), DNA melting extends beyond the transcription start site (+1) creating the transcription bubble. Subsequently the single-stranded DNA is guided into the active site through a positively charged channel as the downstream

duplex DNA is clamped into position by the β and β' subunits (Murakami, 2002). As nucleotides enter the secondary channel a short RNA product is formed.

Due to a steric clash between the nascent RNA chain and the $\sigma^{70}_{3.2}$ loop, only an abortive transcript is formed until the $\sigma^{70}_{3.2}$ loop is removed. During abortive initiation the RNAP remains fixed by binding the -35 element DNA while pulling downstream DNA towards it in a process known as 'DNA scrunching' (Kapanidis, 2006). Once a product of about 12 nucleotides is formed the emerging RNA chain displaces the $\sigma^{70}_{3.2}$ loop, marking the **end of abortive transcription** (Fig 1.3.3, D).

Displacement of the $\sigma^{70}_{3.2}$ from the active site channel by the abortive transcript destabilises the interaction between the β flap and the σ^{70}_{4} domain. This destabilisation of the σ^{70}_{4} domain disrupts the binding with the -35 element, and releases core RNAP from the -35 promoter element. The energy stored in the scrunched complexes is also thought to fuel promoter escape (Revyakin, 2006). This step is known as **promoter clearance** (Fig 1.3.3, E) and allows the RNAP complex to slide downstream whilst elongating the nascent RNA chain.

This forms the **transcription elongation complex** (Fig 1.3.3, F). Formation of the transcription elongation complex does not require complete dissociation of σ^{70}, since binding of RNAP to σ^{70}_{2} and σ^{70}_{3} does not obstruct the duplex DNA or nascent RNA chain (Bar-Nahum, 2001; Mukhopadhyay, 2001). About 70-90% of early elongation complexes retain σ^{70} and a further 50-60% of mature elongation complexes also include a σ^{70} factor (Kapanidis, 2005).

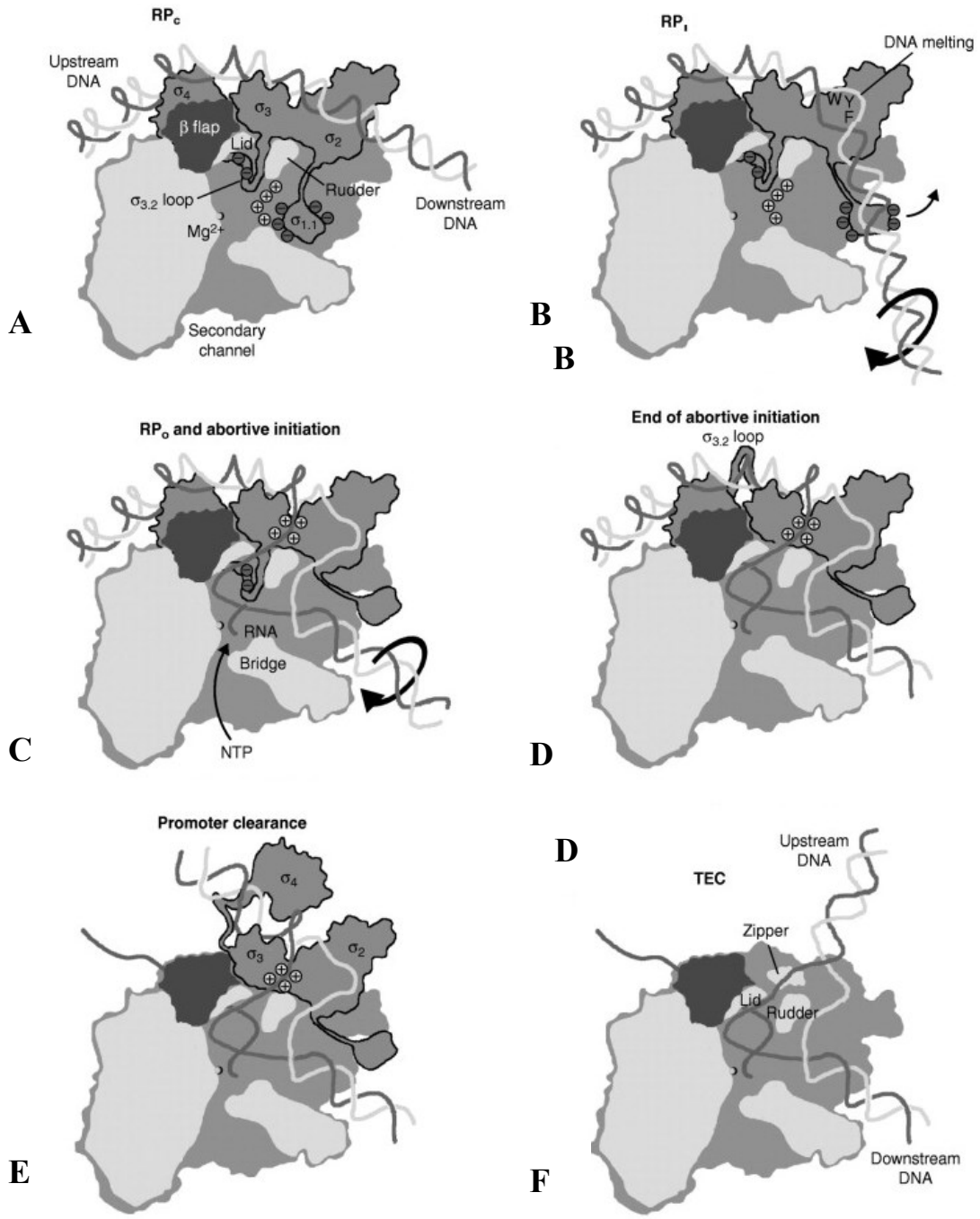


Fig 1.3.3: The 6 stages of σ^{70}-dependent transcription.

First a transcriptionally inactive **closed promoter complex** (**A**) is formed, where the holoenzyme covers the -6 to -56 promoter DNA and $\sigma^{70}{}_2$ and $\sigma^{70}{}_4$ interact with the -10 and -35 promoter elements respectively. During an **intermediate step** (**B**), $\sigma^{70}{}_{2.3}$ is exposed to single-stranded DNA of the -10 element, stabilising transiently melted DNA and formation of the transcription bubble. During **open complex formation and abortive initiation** (**C**), DNA melting extends past +1 to form the final transcription bubble. β and β' pincers clamp on the downstream duplex DNA, and NTPs enter through the secondary channel to the active site for abortive RNA synthesis. At the **end of abortive initiation**

(**D**), the $\sigma^{70}_{3.2}$ loop is displaced by elongating RNA that passes under the β flap. During **promoter clearance (E)** the displacement of $\sigma^{70}_{3.2}$ helps destabilise σ^{70}_{4} and the -35 element, and RNAP is freed of σ^{70} to form the **transcription elongation complex (F).** Adapted from Murakami, 2003.

1.4 The sigma^{54} (σ^{54}) family

Historically, σ^{54} was first described as a positive regulator required for the expression of the gene encoding glutamine synthase (*glnA*), in enteric bacteria (Garcia, 1977). Subsequently this function as a positive regulator was expanded to include other genes required for nitrogen assimilation, such as those related to amino acid transport, degradation, and biological nitrogen fixation (Kustu, 1989).

Sometime after that it became apparent that σ^{54} was also involved in the transcription of many other bacterial pathways including: flagellar biosynthesis in *E. coli* (Zhao, 2010) and *Geobacter sulfurreducens* (Leang, 2009); phosphotransferase system-mediated regulation in Gram-positive and Gram-negative bacteria (Deutscher, 2006); Type III secretion system mediated pathogenicity in *Pseudomonas syringae* (Jovanovic, 2011); adaptation to cold shock in *B. subtilis* (Wiegeshoff, 2006); regulation of virulence and lipoprotein biosynthesis in *Borrelia burgdorferi* (Fischer, 2005); and motility, biofilm formation, luminescence, and colonisation in *Vibrio fischeri* (Wolfe, 2004).

The σ^{54} family has been detected in 60% of sequenced bacterial genomes (Wigneshweraraj, 2008), yet only about 40% of σ^{54}-dependent promoters in *E. coli* were found to be involved nitrogen metabolism regulation (Reitzer, 2001). Such wide ranging regulon involvement is unusual compared the other alternative σ factors in the σ^{70} family, which are each typically associated with a specific programmed response to a certain stress like heat shock or adaptation to starvation (Gruber, 2003).

The sigma^{54} (σ^{54}) family constitutes an evolutionary distinct family and shares little sequence homology with the σ^{70} family (Buck, 2000; Kill, 2005; Merrick, 1993). The σ^{54} family also has a different mechanism of open complex formation, and binds to highly conserved -12/-24 promoter sequences as opposed to -10/-35 promoter sequences. Additionally, σ^{54}-dependent open complex formation also requires both an enhancer binding protein and ATP (Fig 1.4), analogous to "activation" of RNA polymerase II in eukaryotes by the ATP-hydrolysing TFIIH (Jones, 2009). This is different from σ^{70}-dependent open complex formation, which is spontaneous.

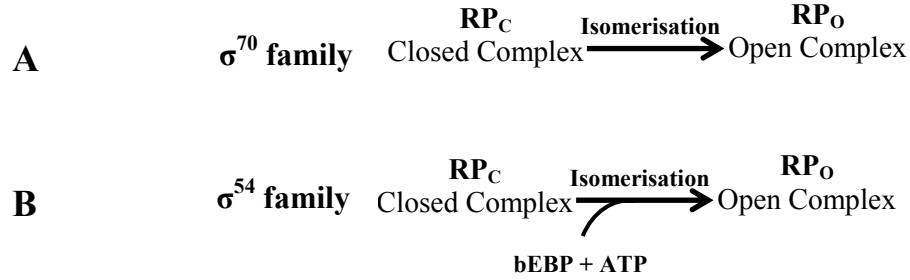


Fig 1.4: σ^{70}-dependent and σ^{54}-dependent open complex formation is distinct.

Formation of a transcriptionally active complex is spontaneous for σ^{70}-dependent transcription initiation (A), but requires a bacterial enhancer binding protein and ATP hydrolysis for σ^{54}-dependent transcription initiation (B).

1.4.1 The *E. coli* sigma^{54} (σ^{54})

The 477 amino acid σ^{54} protein is a major σ factor in *E. coli*, and encoded by the *rpoN* gene. Transcriptional activation by σ^{54} in *E.coli* targets multiple pathways including nitrogen assimilation, membrane stress, biofilm formation, heat shock, and carbon metabolism (Francke, 2011). While no high resolution structure of σ^{54} is available thus far, the structures of some regions have been studied in the Wemmer lab using NMR (see Fig 1.4.1). The *E. coli* σ^{54} protein contains two highly conserved domains (regions I and III), separated by a flexible linker domain (Region II).

Region I (residues 1-56) of σ^{54} is a glutamine and leucine rich domain involved in the interaction with bacterial enhancer binding proteins (bEBPs), core RNAP and the -12 "GC element". Analysis of σ^{54}_{R1} has shown this region is involved in 4 key functions: i) Preventing RNAP isomerisation and open complex formation (Wang, 1995; 1997b; Wang and Gralla, 1996; Cannon, 1999), ii) Inhibiting the interaction with melted DNA, iii) Binding to activators, and iv) Open complex formation and maintaining a stable open complex (Cannon, 1999; Gallegos. 1999). A leucine patch in σ^{54}_{R1} (residues 25-31) was found to be important in this repressive function of σ^{54} during transcription initiation (Syed and Gralla, 1998).

Region II (residues 57-119) is a less conserved linker domain of variable length (and sometimes absent), consisting of predominantly acidic amino acids. Although relatively little is known about the function of this domain, deletions in σ^{54}_{RII} of *Klebsiella pneumonia* σ^{54} reduce the activity of the holoenzyme in open complex formation (Southern and Merrick, 2000). While it is thought that σ^{54}_{RII} is involved in DNA binding (Cannon, 1999) and DNA melting (Wong, 1992), σ^{54}_{RII} was found to be expendable for DNA binding (Cannon, 1999). Moreover, the region was found to be dispensable entirely in *Aquifex aeolicus* and *Rhodobacter capsulatus* (Buck, 2000).

Region III (residues 120-463) is a 7-helix domain and contains the core RNAP binding domains (120-215), a -12 DNA binding domain (329-386) and a C-terminal *RpoN* box (454-463) which is characteristic of σ^{54} factors.

The NMR structure of σ^{54}_{RIII} in *Aquifex aeolicus* (Fig 1.4.1) showed that the N-terminal domain (residues 120-215 in *E. coli*) forms a four-helical bundle that forms the main core RNAP binding interactions (Hong, 2009).The putative helix-turn-helix domain of σ^{54}_{RIII} (residues 366-386) is involved in binding to the -12 promoter element (Merrick, 1992). The C-terminal domain of σ^{54}_{RIII} contains the *rpoN* box (454-463), which is a -24 "GG" promoter element DNA binding domain (Burrows, 2003; Doucleff, 2007; Sasse-Dwight, 1988a). The NMR structure of the C-terminal σ^{54}_{RIII} in *Aquifex aeolicus* (Fig 1.4.1) showed that the *rpoN* box forms a winged helix-turn-helix structure and inserted the recognition helix into the major groove of the -24 promoter DNA element (Doucleff, 2007). The *rpoN* box domain was also suggested to interact with the β flap of RNAP, suggesting a new role in transcription initiation: properly orienting the σ^{54}-RNAP holoenzyme at the site of melting via its interaction with the β-flap region of RNAP (Doucleff, 2007).

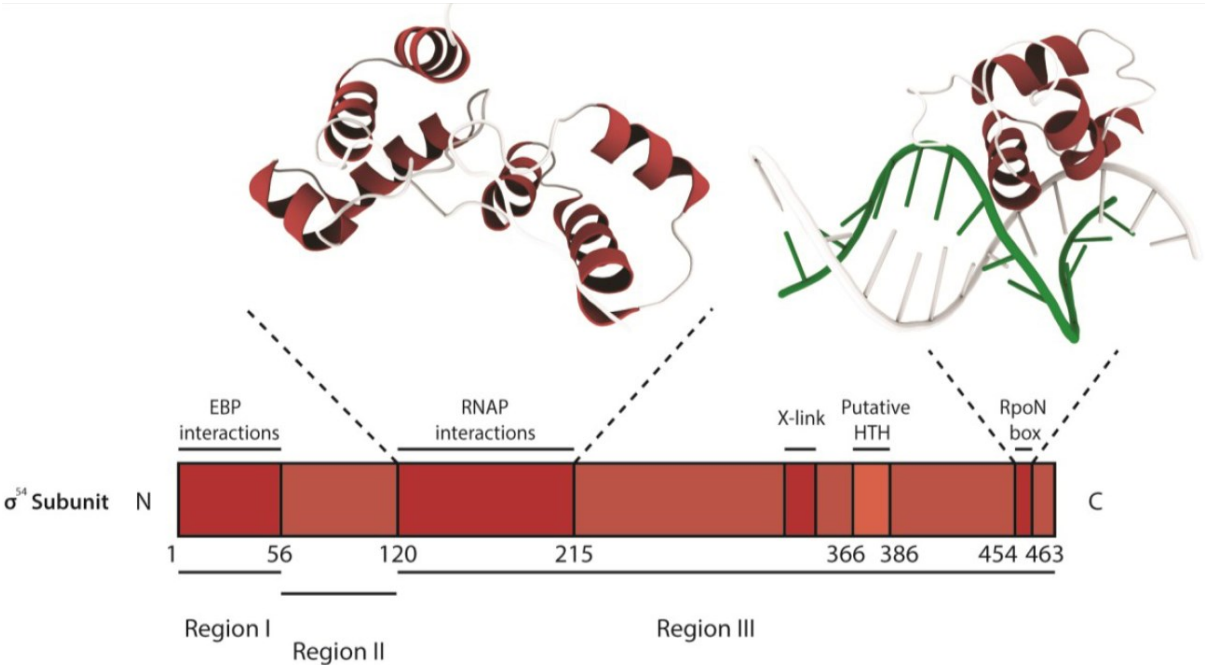


Fig 1.4.1: Domain organisation of the *E. coli* σ^{54} protein.

The *E. coli* σ^{54} has 2 domains important for function (regions I and III), joined by a flexible acidic linker region (region II). The N-terminal region I (1-56) is involved in binding the -12 promoter element, bacterial enhancer binding protein interactions and forming contacts with RNAP. Region III

(120-463) contains the major core RNAP binding domains (120-215) and a -12 promoter element binding domain (329-386). Region III also contains the C-terminal *RpoN* box (454-463), which serves as a -24 promoter element binding domain. NMR fragments of region III in *Aquifex aeolicus* are also shown (Doucleff, 2007; Hong, 2009).

1.4.2 Bacterial enhancer binding proteins (bEBPs)

A distinguishing feature of σ^{54}-dependent transcription initiation is that in all described examples it absolutely requires the activity of a cognate activator (Kustu, 1989), possibly giving it a tighter control over gene expression and a wider dynamic range than σ^{70}-dependent transcription. These activators (also known as a bacterial enhancer binding proteins, bEBPs) are part of the AAA+ superfamily (ATPases associated with diverse cellular activities), which are involved in a range of cellular functions including: membrane transport, proteolysis, protein refolding, transcription, DNA replication, and apoptosis (Ogure, 2001).

bEBPs often have 3 overall domains required for function (Fig 1.4.2): an **N-terminal regulatory domain**, a central **AAA+ catalytic domain** involved in σ^{54} binding and ATP hydrolysis, and a C-terminal helix-turn-helix **DNA binding domain** (Studholme, 2003; Salai, 2005; Bush, 2012). Based on these three domains, bEBPs can be subdivided into 5 main groups (Fig 1.4.2). This classification is largely based on the N-terminal regulatory domain present in most bEBPs, which responds to various environmental stimuli and subsequently regulates the catalytic AAA+ domain (Schumacher, 2006). This regulation is typically achieved in one of 3 ways: i) phosphorylation, ii) protein-protein interactions, and iii) ligand binding.

Group I contains the majority of bEBPs (NtrC, NtrC1, NtrC4, DctD, ZraR and FlgR), and includes those which are activated by phosphorylation of a conserved histidine or aspartate residue in the N-terminal regulatory domain as a part of the two component regulatory systems of bacteria (Studholme, 2003).

Group II bEBPs are regulated by the ligand-ligand interactions through the binding of small effector molecules such as toluene or aromatic compounds (Shingler, 1995; Delgado, 1995).

Group III bEBPS are regulated by protein-protein interactions via a GAF motif (cGMP-specific and –stimulated phosphodiesterases, *Anabaena* adenylate cyclases and *E. coli* FhlA) (Ho, 2000).

Group IV bEBPs are characterised by the lack of an N-terminal regulatory domain all together, and are controlled by *trans*-acting factors such as PspA (the negative regulator of the bEBP, PspF) or HrpV (the negative regulator of the co-dependent bEBP complex, HrpRS) (Brissette, 1991; Elderkin, 2002).

Group V bEBPs on the other hand lack a C-terminal DNA binding domain (Brahmachary, 2004). This group includes FlgR from *Helicobacter pylori*, which is phosphorylated by the sensor histidine kinase FlgS in order to drive the transcription of flagellar biosynthesis. The Group V bEBPs likely act from solution and not from DNA.

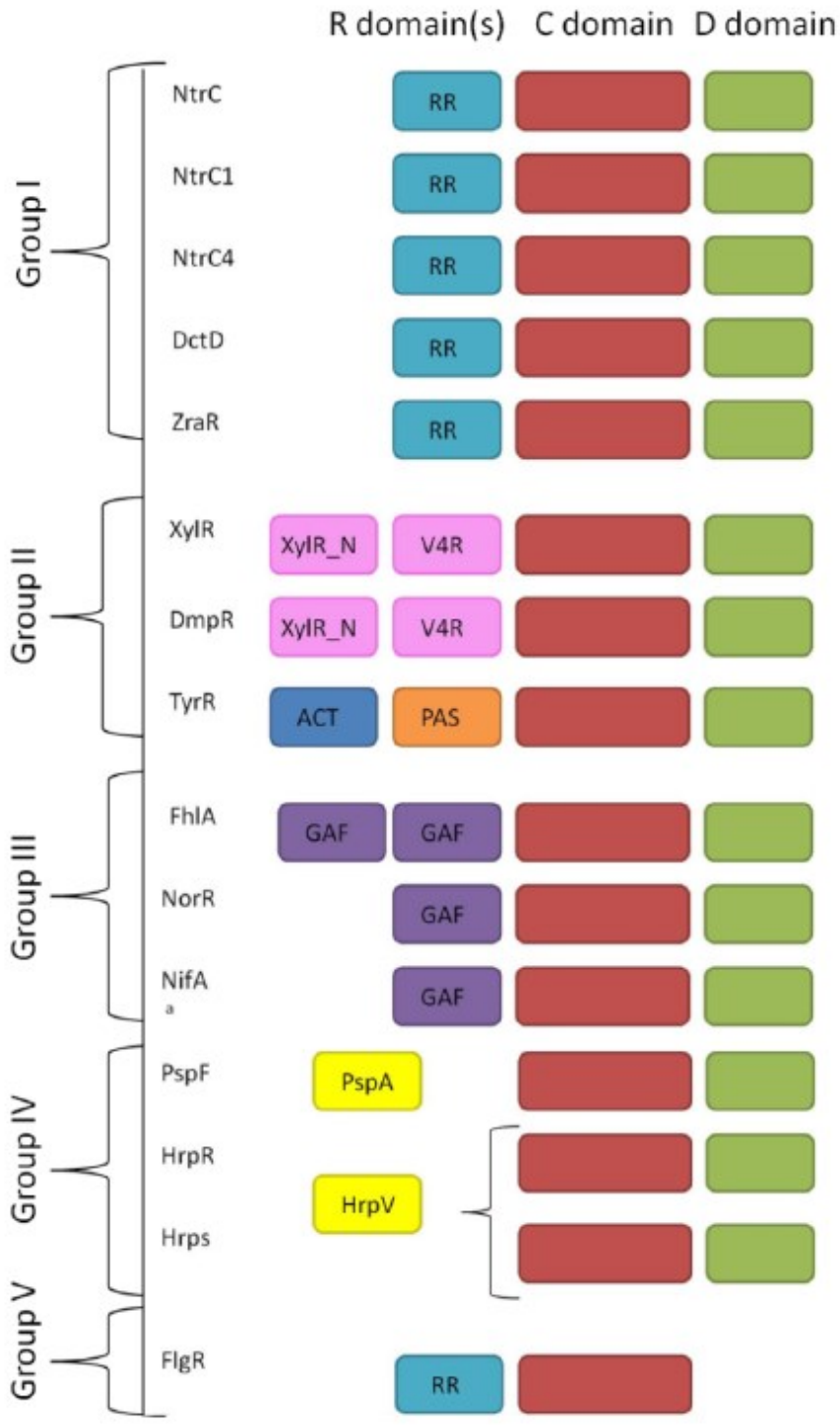


Fig 1.4.2: Domain architecture of bEBPs divided into 5 Groups (I-V).

The central catalytic AAA+ domain (red) is highly conserved in all groups and essential for σ^{54}-dependent transcription. The C-terminal is a DNA binding motif (green) that contain a helix-turn-helix motif responsible for recruiting the bEBPs to upstream activation sequences (UAS) or enhancer binding sites. This motif is absent in Group V bEBPs. The N-terminal regulatory domain is less conserved and lacking in some bEBPs (eg Group IV). Group I-III are subdivided based on their different N-terminal regulatory domains (adapted from Bush, 2012).

1.4.2.1 Bacterial enhancer binding proteins in *E. coli*

Multiple bEBPs have been characterised in *E. coli* (Table 1.4.2.1) influencing a range of pathways including: regulatory RNAs (GlrR), nitrogen metabolism (NtrC, NorR), phage shock (PspF), and zinc tolerance (ZraR). NtrC is the main bEBP in *E. coli* and is involved in activating nitrogen assimilation genes, and thought to affect the expression of up to 2% of the genome (Zimmer, 2000).

bEBP	Description	Reference
AtoC	Acetoacetate response regulator	Grigoroudis, 2007
FhlA	Formate hydrogen lyase activator	Hopper, 1995
GlrR	Regulatory RNA response regulator	Yamamoto, 2005
HyfR	Hydrogenase four regulator	Andrews, 1997
NorR	Nitric oxide reductase regulator	Studholme, 2003
NtrC	Nitrogen regulatory protein C	Morett, 1993
PrpR	Propionate regulator	Palacios, 2004
PspF	Phage shock protein F	Jovanovic, 1996
RtcR	RNA terminal phosphate cyclase regulator	Genschik, 1998
TyrR*	Tyrosine regulator	Verger, 2007
ZraR	Zinc resistance-associated regulator	Sallai, 2005

Table 1.4.2.1: *E. coli* bEBPs and the pathways they regulate.

**TyrR is not an activator of σ^{54}-dependent transcription, and lacks the GAFTGA motif that binds σ^{54}.*

1.4.3 Sigma^{54}-dependent (σ^{54}) transcription initiation

A distinguishing characteristic of σ^{54}-dependent transcription initiation is that it absolutely requires the activity of cognate activators: bEBPs (Kustu, 1989). This activator requirement allows σ^{54}-dependent transcription initially to be regulated at the open complex formation stage, and is akin to eukaryotic RNAP II transcriptional activation. This is in contrast to σ^{70}-dependent transcription initiation, where the closed complex is short-lived, as it spontaneously isomerises into an open complex or dissociates upon binding to the promoter (Browning, 2004). This activator requirement allows σ^{54}-dependent transcription to create swift and precise responses to environmental change.

Typically, σ^{54}-dependent transcription initiation can be divided into 3 steps (see Fig 1.4.3): i) **Closed complex formation**, ii) **Intermediate complexes**, and iii) **Open complex formation** (Friedman, 2012a).

1.4.3.1 Closed complex formation

Initially, σ^{54}_{RI} and σ^{54}_{RIII} of the σ^{54}RNAP holoenzyme bind to the -12 and -24 promoter elements via 3D diffusion to form a stable and inactive closed promoter complex (Fig. 1.4.2, A) (Ninfa, 1987; Friedman, 2012b). The -12 'GC' and -24 'GG' promoter elements are spaced one helical turn apart and face the same direction on the promoter DNA, a requirement for σ^{54}-dependent transcriptional activation (Buck, 1986). This binding to promoter DNA has been shown to be mediated primarily by the -24 promoter binding domain, as deletions in the -24 element (but not -12 element) eliminate promoter recognition (Wong, 1994; Hsieh, 1994).

Early melting of the DNA duplex at bases -11/-12 occurs to form a fork junction-like structure. This DNA fork junction structure is present within closed complexes and binds very strongly to σ^{54}, and most of the downstream DNA remains base-paired and the RNAP complex remains in a thermodynamically favourable closed state (Popham, 1989; Guo, 2000). Tight interactions between the fork junction and σ^{54} prevent access of single-stranded DNA to the active site of core RNAP, which is a key step in the process of DNA melting. Basal transcription from the complex is inherently very low, consistent with the lack of repressors observed for σ^{54}-dependent promoters (Buck, 2000). The fork junction also serves

as a nucleation point for the DNA melting, which is required for full open complex formation.

1.4.3.2 Intermediate and open complex formation

Bacterial enhancer binding proteins and ATP hydrolysis are required to overcome the energy barrier imposed by the promoter complex, and induce a conformational change that removes the inhibitory interactions present at the early melted -11/-12 fork junction and the transcription start site (Fig. 1.4.3, B).

These activators typically bind unusually far (for bacteria) from the transcription start site, 100-150 bases upstream (Buck, 2000). The bEBPs bind upstream activation sequences (UAS) as inactive dimers, and are then brought into proximity of the σ^{54}RNAP closed complex by a DNA looping event. Here the DNA bends up to 180°, which is typically mediated by the integration host factor (IHF), a DNA bending protein (Hoover, 1990). This allows the bEBP to interact directly with $\sigma^{54}{}_{RI}$ at the -12 fork junction forming the RNA polymerase intermediate complex.

ATP hydrolysis is linked to open complex formation via conformational changes in the catalytic AAA+ domain of the bEBP, which occur upon binding and sensing of ATP. The GAFTGA-containing L1 and L2 loops in the catalytic AAA+ domain adopt different conformations depending on the state of hydrolysis (Rappas, 2005; Rappas. 2006; Chen, 2010; Joly, 2010). Upon ATP binding, the loops assume an extended conformation unable to bind stably to $\sigma^{54}{}_{RI}$. ATP hydrolysis then increases the strength of this interaction, allowing for remodelling of the holoenzyme and formation of an open complex (Bush, 2012). In the post-nucleotide hydrolysis state, release of the γ-phosphate from ADP allows the L1 and L2 loops to disengage from σ^{54} (Chen, 2007).

In doing so, activator-driven ATP hydrolysis provides the necessary energy to drive the conformational changes required for stable association of the holoenzyme with melted DNA and producing an open and stable active complex (Fig. 1.4.3, C) (Wedel and Kustu, 1995). In this remodelled complex, the DNA may be melted outside the active channel and then loaded into the active site channel of RNAP so that transcription can begin (Wigneshweraraj, 2008).

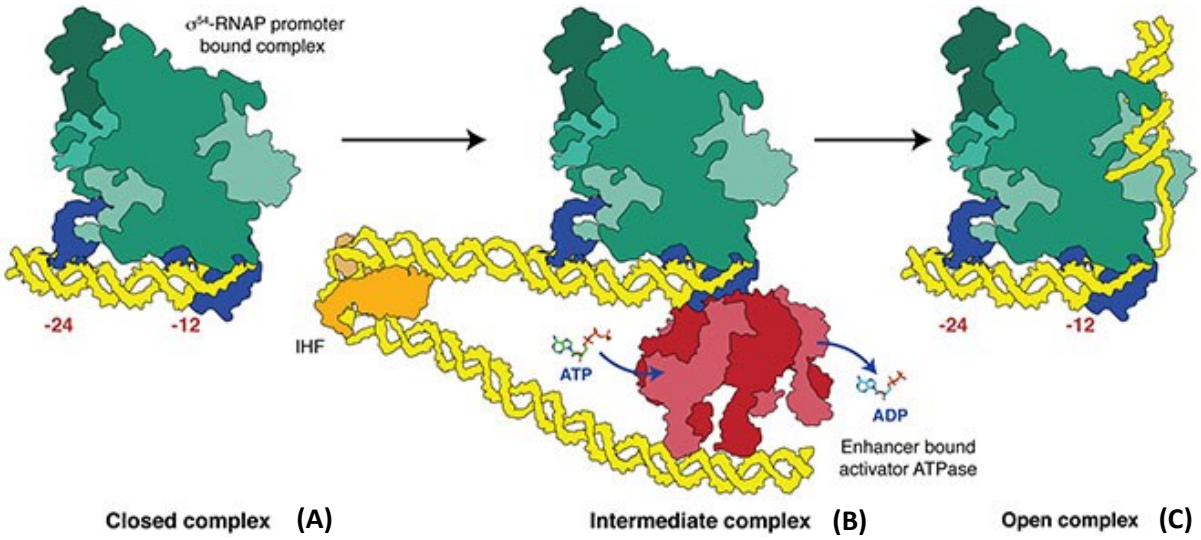


Fig. 1.4.3: 3 stages of σ^{54}-dependent transcription initiation.

Initially, RNAP-σ^{54} complex binds to -12 and -24 promoter consensus sequences forming a closed and transcriptionally silent complex (A). Distal bEBPs bound to upstream activation sequences are brought into proximity of the closed complex via DNA looping, aided by Integration host factor (IHF). Activator-mediated ATP hydrolysis results in the release of ADP+P_i (B). DNA melting follows ADP+P_i formation (and/or release by the activator) to form an open and transcriptionally active complex (C). (adapted from Wigneshweraraj, 2005).

1.4.4 Mutants of sigma^{54} (σ^{54}) that bypass the activator requirement

The defining characteristic of σ^{54}-dependent transcription is its similarity to eukaryotic RNAPII transcriptional activation through the use of cognate activators and their ATP hydrolysis (see 1.4.3). However, a number of σ^{54} mutants have been described as being able to bypass their normal activator requirement *in vitro*, and are therefore known as σ^{54} activator bypass mutants.

1.4.4.1 Disruption of a leucine patch in σ^{54}_{RI}

These activator bypass mutants were first described by Jonathan Wang in 1995, after screening σ^{54} proteins with alanine substitutions in the leucine patch of σ^{54}_{RI} (Wang, 1995). Three leucine patch mutants were able to form large colonies in nitrogen-limiting conditions in the absence of NtrC, a bEBP required for glutamine synthetase regulation, glutamine transport, and general nitrogen metabolism (Ninfa, 1987). This finding was confirmed by showing that these leucine mutants were able to transcribe the σ^{54}-dependent gln_{AP2} promoter *in vitro* in the absence of NtrC, while wild type σ^{54} could not (Wang, 1995).

A key property of the σ^{54}RNAP holoenzyme is the ability to form stable closed promoter complexes, and it was proposed that this leucine patch plays a role in maintaining this closed conformation. NtrC and other bEBPs are thought to disrupt the interactions with the leucine patch in the closed complex and allow open complex formation, explaining why substitutions in the leucine patch mimic the effect of the adding activator and ATP (Wang, 1995)

1.4.4.2 A two-step model for activation

Later, $\underline{\sigma}^{54}$ activator bypass mutant transcription (SABT) at the gln_{Hp2} and gln_{Ap2} promoters *in vitro* was found to be sensitive to heparin (Wang, 1996). Heparin is a highly negatively charged molecule that is known to destabilise unstable open promoter complexes. This finding was unusual because wild type σ^{54}-dependent transcription typically forms stable open complexes that are heparin-resistant. This led to the hypothesis that σ^{54}-dependent transcription involves a two-step model of activation (Fig 1.4.4). Initially, the enhancer complex overcomes the repressive effect of the σ^{54}_{RI} leucine patch and unlocks the melting

activity of σ^{54}RNAP. Second, different σ^{54} determinants are used to drive stabilisation of the open complexes, allowing the full transcription potential to be realised (Wang, 1996).

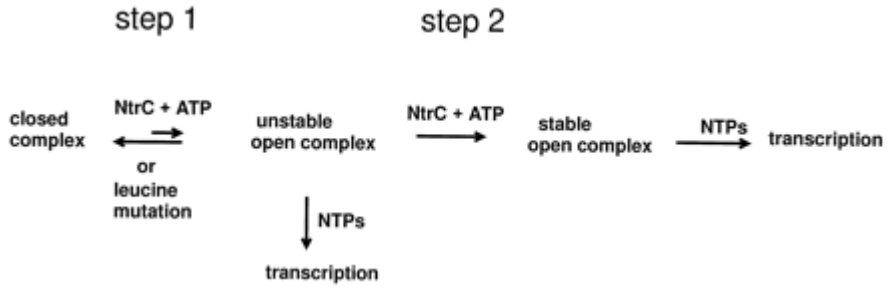


Fig 1.4.4: A two-step model of activation for σ^{54}-dependent transcription

Initially, NtrC and ATP unlock the ability to melt the promoter start site, overcoming the repressive effects of the leucine patch in σ^{54}_{R1} and forming an unstable promoter complex (step 1). Next the enhancer complex drives the formation of a stable open complex (step 2), capable of the full transcription potential *in vivo*. During one-round transcription assays *in vitro* these unstable open complexes can be driven to transcribe by addition of the rNTPS which "drag" the RNAP into elongation (Wang, 1996).

1.4.4.3 The activator bypass phenotype and the -12 DNA interaction

In addition to disruption in the N-terminal leucine patch of σ^{54}, a C-terminal R336A substitution in σ^{54}_{RIII} later also yielded the activator bypass phenotype (Chaney, 1999). This region is part of the σ^{54} helix-turn-helix motif involved in binding to the -12 element promoter DNA (see 1.4.1). Incidentally, σ^{54}_{RI} (where the leucine patch is located) also interacts with the -12 element promoter DNA.

Disruptions in σ^{54}_{RI}, the -12 promoter DNA, or conditions favouring transient DNA melting (high temperature, low salt concentration) all yielded heparin-sensitive pre-initiation complexes capable of transcription *in vitro* in the absence of activator or ATP. This led to the hypothesis that the tight interactions with the -12 DNA limit open complex formation in the absence of activator and ATP hydrolysis (Wang, 1997).

Combined, these suggested that disrupting the interaction between σ^{54} and the -12 promoter DNA in the closed promoter complex is key for generating an activator bypass mutant phenotype.

1.4.4.3 The σ^{54} activator bypass mutant phenotype is not readily observed *in vivo*

Despite having been first reported *in vivo* (Wang, 1995), SABT detected *in vitro* is not readily observed *in vivo* despite several unpublished attempts to do so (Martin Buck, Imperial College; Sydney Kustu, Berkeley). This implies that there are inhibitory mechanisms that prevent activator independence from being manifested *in vivo*.

1.4.5 Objectives

No clear unifying biological theme for σ^{54} in *E. coli* has been characterised, other than regulating and potentially integrating nitrogen metabolism and N assimilation (Reitzer, 2001). Yet a large proportion of functions associated with σ^{54} are seemingly unrelated to nitrogen metabolism, and include membrane stress, biofilm formation, heat shock, and carbon metabolism (Francke, 2011).

A distinguishing feature of σ^{54}-dependent transcription initiation is that it absolutely requires the activity of a cognate activator (Kustu, 1989), giving it a tighter control over gene expression and a wider dynamic range. Perhaps it is this unique property of σ^{54}-dependent gene expression that links many of the otherwise unrelated pathways σ^{54} regulates, by allowing for swift and precise responses to environmental change. Yet surprisingly, only a very small proportion of genes are transcribed by σ^{54} overall, less than 1% of promoters are known to be σ^{54}-dependent. This suggests that there are perhaps additional levels of complexity governing this enigmatic sigma factor.

Some bacteria like the pathogenic *Neisseria gonorrhoeae* have abandoned σ^{54}-dependent transcription altogether quite recently, their genomes still containing *rpoN* pseudogenes which have undergone deletion of the -12 promoter DNA-binding region (Laskos, 1998). Such mutants can result in activator-independent σ^{54}-dependent transcription in *E. coli* (Wang, 1997), raising the question as to why *N. gonorrhoeae* has not evolved towards activator independence instead of abandoning σ^{54}-dependent transcription entirely? Presumably barriers inhibiting activator independence exist *in vivo,* which are both biologically important and advantageous.

Minimal conditions allow activator-independent transcription to be observed *in vitro* for several σ^{54} mutants, yet cannot be readily detected *in vivo* in *E. coli* (see 1.4.4.3). This further supports the hypothesis that barriers are present *in vivo* to prevent σ^{54}-dependent transcription from evolving towards activator independence. Determining what these potential barriers could be and why they are in place, may help in answering why σ^{54} has not evolved towards

activator independence. Moreover, it could also shed light on some of the many other questions that still remain regarding σ^{54} after more than 3 decades of research.

Therefore, the overarching objective of this thesis is to **investigate the putative barriers that prevent bypass transcription from being readily observed *in vivo*, in σ^{54} activator bypass mutants**.

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Antibiotics

The following antibiotics were sterilised and used at the final concentrations below:

Antibiotic	Concentration
Ampicillin	100 µg/ml (50 µg/ml for chromosomal copies)
Chloramphenicol	30 µg/ml
Kanamycin	50 µg/ml (25 µg/ml for chromosomal copies)
Tetracycline	20 µg/ml (10 µg/ml for chromosomal copies)

Table 2.1 Antibiotics used in this project.

2.1.2 Media

The following types of media were used for cell growth:

Medium	Contents
Luria Broth (LB)	10 g/l peptone, 10 g/l NaCl, 5 g/l yeast extract
LB agar	10 g/l peptone, 10 g/l NaCl, 5 g/l yeast extract, 15 g/l agar
Minimal Gutnick Media	4.7 g/l KH_2PO_4, 13.5 g/l K_2HPO_4, 1 g/l K_2SO_4, 0.1 g/l $MgSO_4$-$7H_2O$, 5mM L-Glutamine (suppl. 0.4% glucose)
Top agar	10 g/l tryptone, 10 g/l NaCl, 5 g/l yeast extract, 8 g/l agar
Trace elements	0.5 g/l $FeCl_2$, 0.05 g/l $ZnCl_3$, 0.01 g/l $CuCl_2$, 0.01 g/l $CoCl_2$•$6H_2O$, 0.01 g/l H_3BO_3, 0.5 g/l, EDTA pH 7.0

Table 2.2 Types of media used in project.

2.1.3 Bacterial Strains

Bacterial strains used in this study are listed below:

Strain	Genetic background
Obtained	
KEIO pool *(from CGSC, Yale)*	In-frame single-gene substitution of 3985 genes (Baba, 2006). Parental BW25113: *rrnB3 ΔlacZ4787 hsdR514 Δ(araBAD)567 Δ(rhaBAD)568 rph-1.*
K1471 *(from CGSC, Yale)*	F-, *Δ(araA-leu)7697*, *[araD139]$_{B/r}$*, *Δ(codB-lacI)3*, *galK16*, *galE15*(GalS), *λ*$^-$, *e14-*, *mcrA0*, *relA1*, ***rpoN208::Tn10***, *rpsL150*(strR), *spoT1*, *mcrB1*, *hsdR2*
MVA4 *(G. Jovanovic)*	MC1061 **ΦP$_{pspA}$-*lacZ*** (Amp^R)
MG1655 *(Lab collection)*	F- λ- *ilvG- rfb-50 rph-1*
XL1-Blue *(Lab collection)*	*rec*A1 *end*A1 *gyr*A96 *thi*-1 *hsd*R17 *sup*E44 *rel*A1 *lac* F′ *pro*AB l*ac*IqZΔM15 Tn10 (Tet^R)
Constructed	
gln$_{Ap2}$-WTRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, pSB4A3- ΦP$_{glnAp2}$-*GFP,* pBAD18-*rpoN*
gln$_{AP2}$-SABRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, pSB4A3- ΦP$_{glnAp2}$-*GFP,* pBAD18-*rpoN$_{ΔR1}$*
gln$_{Ap2}$-DRRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, pSB4A3- ΦP$_{glnAp2}$-*GFP,* pBAD18-*empty*
pspA-WTRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, ΦP$_{pspA}$-*lacZ,* pBAD18-*rpoN*
pspA-SABRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, ΦP$_{pspA}$-*lacZ,* pBAD18-*rpoN$_{ΔR1}$*

pspA-DRRS	MG1655, LacZ$^-$, *rpoN208::Tn10*, ΦP_{pspA}-*lacZ,* pBAD18-*empty*

Table 2.3 Bacterial strains used in project.

2.1.4 Kits

The following commercially available kits were used:

Kit	Company	Use
DNeasy® Blood & Tissue Kit	Qiagen	Purification of chromosomal DNA
GoTaq® Green Master Mix	Promega	Analytical PCR
LigaFast™ Rapid DNA Ligation Kit	Promega	DNA ligation
pGEM®-T Easy Vector Kit	Promega	DNA cloning
QIAprep® Spin Miniprep Kit	Qiagen	Plasmid DNA purification
QIAquick® PCR Purification Kit	Qiagen	DNA gel extraction

Table 2.4 Kits used in project.

2.2 DNA Methods

2.2.1 Chromosomal DNA purification

Cells were grown in a 5 ml overnight culture, in LB with relevant antibiotics. Genomic DNA was purified from 1-2 x 10^9 cells using a DNeasy® Blood & Tissue Kit (Qiagen), according to the manufacturer's guidelines, and eluted in a final volume of 200 µl sterile water.

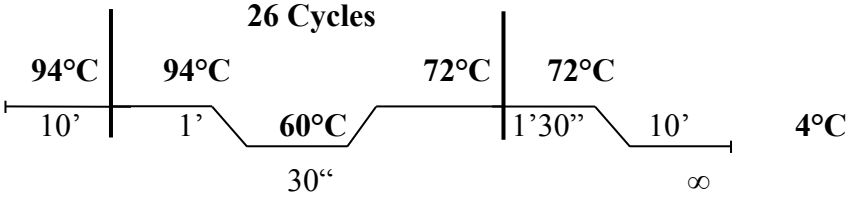
2.2.2 Plasmid DNA purification

Cells were grown in a 5 ml overnight culture, in LB with relevant antibiotics. Plasmid DNA was purified using a QIAprep® Spin Miniprep Kit (Qiagen), according to the manufacturer's guidelines, and eluted in a final volume of 30-50 µl sterile water.

2.2.3 PCR

GoTaq® Green Master Mix is a premixed ready-to-use solution containing bacterially derived Taq DNA polymerase, dNTPs, $MgCl_2$ and reaction buffers at optimal concentrations for efficient amplification of DNA templates by PCR.

For colony PCR, single colonies were resuspended in 50 µl ddH_2O. PCR reactions used a total volume of 25 µl containing: 0.5 µM forward primer, 0.5 µM reverse primer, 1 µL DNA (or 5µl colony suspension), 25 µl GoTaq® Green Master Mix, and 19µl water. The reactions were run in a Harlow Scientific Eppendorf Mastercycler under the following conditions:



5 µl of the reaction was subsequently analysed via agarose gel electrophoresis (see 2.2.4).

2.2.4 Agarose gel electrophoresis

50 ml gels were made of 1% (w/v) Hi-Res Agarose (Geneflow) and 5 µl Safe View (NBS Biologicals, UK) in TBE (89mM Tris Borate pH 8.3 and 2mM Na_2EDTA) buffer (Geneflow). DNA samples were loaded with 1x DNA loading dye (Fermentas) and run at 90 V for 45 min before being visualised with a Bio-Rad Gel Doc 2000 imaging system.

2.2.5 Gel Extraction

DNA fragments were excised from agarose gels and purified using a QIAquick® PCR Purification Kit (Qiagen). DNA was eluted in of 30-40 µl sterile water.

2.2.6 DNA Sequencing

For sequencing, 10µl DNA (isolated as in 2.2.2 above) and 10 µl 10µM sequencing primer were prepared separately and submitted to Beckman Coulter Genomics for sequencing.

2.2.7 Restriction Digest

Plasmid DNA was digested in a final reaction volume of 20 µl, containing 1-2 µl of the relevant Fermentas FastDigest enzyme and 2µl FastDigest buffer (1x). Samples were incubated at 37°C for 2 hours, and the digests were run on an agarose gel. When required, DNA bands of interest were extracted and purified.

2.2.8 DNA Ligation

Vector and insert (molar ratio = 1:3) were ligated overnight at 4°C for 1 hour in a final reaction volume of 20µl using the LigaFast™ Rapid DNA Ligation system (Promega). The reaction mix contained 10µl of 2x Reaction buffer, 1 µl of T4 DNA ligase , 1-3µl vector, 2-6 µl insert and was made up to a final volume of 20 µl with distilled water.

2.2.9 $CaCl_2$ competent cells

$CaCl_2$ competent cells were made using a modified $CaCl_2$ competent cell protocol. 1 ml overnight culture was used to inoculate 100 ml LB and left to grow to an OD_{600} of 0.4 to 0.7. Cells were centrifuged at 5000 rpm for 5 mins at 4°C and resuspended in 50 ml cold 0.1M $MgCl_2$. The cells were incubated on ice for 20 mins and again centrifuged at 5000 rpm for 5 mins at 4°C. They were resuspended in 50 ml cold 0.1M $CaCl_2$ and incubated on ice for 20 mins. The cells were then centrifuged at 5000 rpm for 5 mins at 4°C once more. They were resuspended in 1 ml cold 0.1M $CaCl_2$/ 18% glycerol, aliquoted and stored at -80°C.

2.2.10 Transformation (heat-shock)

50 µl $CaCl_2$ competent *E. coli* cells were thawed on ice and mixed with 1 µl of plasmid DNA. Cells were incubated on ice for 20 mins, transferred to a 42°C water bath for 1 min, then transferred back to ice for 5 mins to minimise damage to the cells. 500 µL LB medium was added to the cells and they were placed in a shaking incubator for 1 hour at 37°C, allowing them to recover. 200 µl cells were spread out on an LB plate with the relevant antibiotic and grown overnight at 37°C.

2.3 In Vitro Methods

2.3.1 Beta-galactosidase assay

β-Galactosidase is a large 120 kDa protein encoded by the *E. coli lacZ* gene and cleaves lactose into glucose and galactose. The synthetic compound o-nitrophenyl-β-D-galactoside (ONPG) is also recognised as a substrate, producing galactose and o-nitrophenol (yellow colour). Thereby beta-galactosidase activity can be used as a readout of LacZ activity.

$$\text{Miller Units} = \frac{1000 * OD420}{Time * Volume * \ OD600}$$

OD_{420} is the absorbance of yellow o-nitrophenol
OD_{600} is the cell density
Time is the reaction time in minutes
Volume is the volume of culture assayed in ml

5ml overnight cultures were grown in LB liquid medium and relevant antibiotics on a shaker at 37°C. These overnight cultured were subcultured 1:100 under the same conditions and to an OD_{600} of ≈0.5. At this point the OD_{600} was measured and 0.5ml of culture was added to 0.5ml Z buffer (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgSO_4$) + 1.25 µl β-mercaptoethanol (BME). Cells were lysed by adding 50 µL chloroform and incubated at room temperature for 2 minutes. 0.2ml of ONPG dissolved in Z buffer (4mg/ ml) was added to start the reaction. When the solution started turning yellow, the reaction was then stopped by adding 0.5 ml 1M $NaCO_3$. The reaction was centrifuged at 13000 rpm for 2 minutes, to remove the cells and cell debris from the supernatant. The OD_{420} was taken and the results expressed in Miller Units.

2.3.2 SDS-PAGE

Protein samples in 2x Laemmli Buffer were loaded on 10% SDS gels, using a BioRad Protean Tetra Cell system. Samples were boiled at 95°c for 5 minutes in 1x Laemmli Buffer. Denatured proteins were run at 110V for 45 minutes in 1x SDS buffer (25 mM Tris, 192 mM Glycine, 1% SDS). MagicMark^{TM} XP Western Protein Standard (20-220kDa) was used as a protein marker.

2.3.3 Abortive transcription assay

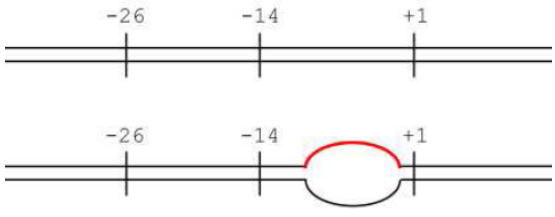
Abortive assays were run in a 10 µl final reaction volume. 8 µl reactions were made containing: 1x STA buffer (2.5 mM Tris-Acetate pH8.0, 8 mM Mg-acetate, 10 mM KCl, 1 mM DTT, 3.5% (w/v) PEG 8000), 5mM ATP, 1mM promoter template and 200nM of holoenzyme (with a core to sigma factor ratio of 1:5). Reactions were incubated for 10 mins at 37°C, followed by addition of 0.5 µl $PspF_{1-275}$ (final concentration of 5 µM). After 10 mins incubation at 37°C, 1.5 µl elongation mix was added (1 µl of 20 mg/ml heparin, 0.3 µl 10mM dinucleotide and 0.2 µl radiolabelled α-^{32}P nucleotide (3000 Ci/ mmol). Reactions were run for 30 mins. Samples were quenched with 2.5 µl 5x formamide stop dye and boiled at 95°C for 5 mins. Reactions were run at 300V on 20% urea gels containing: 20ml UreaGel Concentrate (from National Diagnostics), 2.5 ml SequGel Diluent (from National Diagnostics), 2.5 ml 10x TBE, 200 µl 10x APS and 20 µl TEMED. Gels were exposed for 30 mins to an IP plate and visualised on a phosphoimager and analysed on AIDA Image Analyzer software. The complete list of duplex DNA used in this study can be found in Figure 2.1 below.

WT/WT -- *HypA*

5’ -60…CACTGGCACAATTATT GCTTGTAGCTGGCAAT…+28 3’
-24 -12 +1
3’ -60…GTGACCGTGTTAATAACGAACATCGACCGTTA…+28 5’

-10-1/WT-- *HypA*

5’ -60…CACTGGCACAATTATTGCTTAGCATC ATTAAT…+28 3’
-24 -12 +1
3’ -60…GTGACCGTGTTAATAACGAACATCGACCGTTA …+28 5’

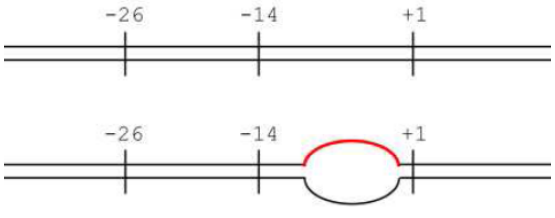


WT/WT -- *PrpB*

5’ -60…TTGTGGCACACCCCTT GC TTTG TC TTT ATCAA…+28 3’
-24 -12 +1
3’ -60…AACACCGTGTGGGGAACGAAACAGAAATAGTT…+28 5’

-10-1/WT -- *PrpB*

5’ -60…TTGTGGCACACCCC TTGCTTGACAGCG CGCAA…+28 3’
-24 -12 +1
3’ -60…AACACCGTGTGGGGAACGAAACAGAAATAGTT…+28 5’

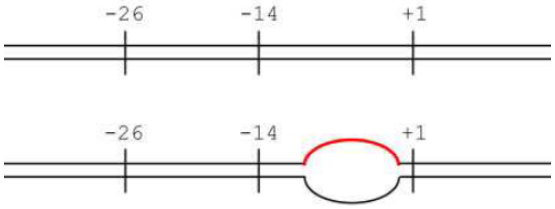


WT/WT --*YahE*

5’ -60…TACTGGAAGCGATTGTGCTTAATGATGGTGGTATT…+28 3’
-24 -12 +1
3’ -60…ATGACCTTCGCTAACACGAATTACTACCACCATAA…+28 5’

-10-1/WT --*YahE*

5’ -60…TACTGGAAGCGATTGTGCTTGCCAGGAACTTGATT…+28 3’
-24 -12 +1
3’ -60…ATGACCTTCGCTAACACGAATTACTACCACCATAA…+28 5’

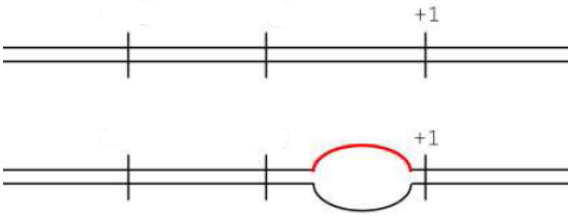


WT/WT --*CspA*

5’ -60…AGTTATCG TTGATACCCC TCGTAGTGCACATTCCT…+28 3’
+1
3’ -60…TCAATAGCAACTATGGGGAGCATCACGTGTAAGGA…+28 5’

-10-1/WT --*CspA*

5’ -60…AGTTATCG TTGATACCCC TCGTAACTAGTCGGCCT…+28 3’
+1
3’ -60…TCAATAGCAACTATGGGGAGCATCACGTGTAAGGA…+28 5’

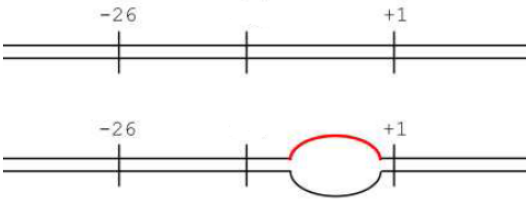


WT/WT --*CspH*

5’ -60…CGC TGGCT TTATTTT ATTAACCCCG TATAGTGCAGCGCA…+28 3’
-24 +1
3’ -60…GCGACCGAAATAAAATAATTGGGGCATATCACG TCGCGT…+28 5’

-10-1/WT --*CspH*

5’ -60…CGCTGGCT TTATTTT ATTAACCCCG TAGGTCATGATGCA…+28 3’
-24 +1
3’ -60…GCGACCGAAATAAAATAATTGGGGCATATCACGTCGCGT…+28 5’



WT/WT --*PspG*

5’ -60…TTTT GGC ATGATTCTTGTAATGCCAGCA…+28 3’
-24 -12 +1
3’ -60…AAAACCGTACTAAGAACATTACGGTCGT…+28 5’

-10-1/WT -- *PspG*

5’ -60…TTTT GGC ATGATTCTTGTAACTTACGCA…+28 3’
-24 -12 +1
3’ -60…AAAACCGTACTAAGAACATTACGGTCGT…+28 5’

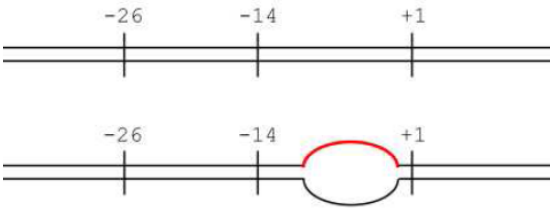


Fig. 2.1: Linear probes used during abortive initiation assays.

All the linear DNA probes were designed based their respective gene promoters, with the top strand representing the non-template strand and the bottom strand the template strand. The conserved -24 (GG) and -12 (GC) elements are annotated (blue), along with the transcription start site (green). The mismatched sequences are highlighted (red) and their positions are indicated in the name of the duplex.

2.3.4 Full length transcription assay

Full length transcription assays were run as 10 µl final reaction volume. 8 µl reactions were made containing: 1x STA buffer, 4mM dATP, 20 U RNase Inhibitor (Promega), 300 ng/µl DNA and 200nM holoenzyme (with a core to sigma factor ratio of 1:5). Reactions were incubated for 10 mins at 37°C, followed by addition of 0.5 µl $PspF_{1-275}$ (5 µM). After 30 mins incubation at 37°C, 1.5 µl elongation mix was added (1 µl 1mg/ml heparin, 0.3 µL 10µM nucleotide mix (ATP, UTP, GTP, CTP) and 0.2 µl radiolabelled α-^{32}P nucleotide (3000 Ci/ mmol). Reactions were quenched with 2.5 µl 5x Formamide stop dye solution and boiled at 95°C for 5 mins. Samples were run at 50W for 2hrs on a 4% denaturing sequencing gel containing: 8 ml UreaGel Concentrate (from National Diagnostics), 37 ml SequGel Diluent (from National Diagnostics), 5 ml 10x TBE, 500 µl 10x APS and 40 µl TEMED. The gels were dried, exposed overnight and imaged on the phosphoimager.

2.3.5 Inverse PCR

Inverse PCR was done for the identification of KEIO deletion mutants. Chromosomal DNA of colonies of interest was purified (2.2.1). Subsequently, the chromosomal DNA was digested with NcoI, a 6-base cutter (± every 4096 bases). These 4kbp fragments were religated (2.2.8) in a larger volume of 100 µl (instead of 20 µl) to favour relegation into circular DNA. This DNA was used as a template for PCR (2.2.3), using Kan-specific forward and reverse primers. Finally, products were run on a gel (2.2.4), extracted (2.2.5), and sequenced (2.2.6). Blast was used (http://blast.ncbi.nlm.nih.gov/Blast.cgi) to align sequenced regions to the *E. coli* chromosome, and determine the nature of the gene deletion.

2.4 In Vivo Methods

2.4.1 P1*vir* phage transduction

0.2 ml overnight culture of the donor strain was incubated with 0.2 ml of a P1*vir* bacteriophage lysate (obtained as given below) for 15 mins at room temperature. 18.5 ml LB, 0.5 ml 40% Glucose and 1ml 0.1 M $CaCl_2$ was added and the cells were incubated in a shaker at 37°C for 4-6 hrs until all bacterial cells have been lysed and the suspension became clear. 1-2 ml chloroform was added to kill any remaining bacteria and the mixture was then incubated at room temperature for 30 mins. The lysate was centrifuged at 16 000 g for 5 mins and the supernatant containing the P1*vir* (with a sub-population containing DNA of interest) was used to infect recipient strains.

0.2 ml of the new P1*vir* lysate was added to 0.2 ml overnight culture of the recipient strain and incubated for 20 mins at 37°C without shaking. The transduction was then stopped by adding 0.4 ml 1M Na-citrate. The suspension was mixed with 3 ml Top Agar (10 g/l tryptone, 5 g/l yeast, 5 g/l NaCl, 7 g/l agar) and spread on to plates containing kanamycin (50 µg/ml).

2.4.2 GFP fluorescence and cell growth

Cells were grown overnight in Mimimal Gutnick media complemented with 0.4% (w/v) glucose, 5mM L-glutamine and relevant antibiotics (in Greiner 96-Well Multiwell Plate). Growth was done in a FLUOstar Omega plate reader for 20 hours, with a starting OD_{600} of 0.01. GFP measurements were done with at a 400nm excitation and 510nm emission, and results were analysed with MARS Data Analysis Software.

2.4.3 RNA-Sequencing

5ml cell cultures were grown overnight at 42°C in minimal Gutnick media supplemented with 5mM L-Glutamine and 0.4% glucose (Gutnick et al, 1969). These were then sub cultured (1:100 dilution) in 100ml day cultures and grown exponential phase (OD_{600} 0.3-0.5). Cell growth was stopped by adding 10ml stop solution (95% ethanol, 5% phenol), and were then centrifuged at 5000 rpm for 5 mins (at 4°C). Supernatants were discarded and cell pellets were left to dry before being weighed and stored at -80°C.

Total RNA was isolated (at www.vertis-biotech.com) from the cell pellets using a bead mill and the mirVana RNA isolation kit (Ambion) including DNase treatment. Ribosomal RNA molecules were then depleted from the total RNA preparations using the RiboZero rRNA Removal Kit (Bacteria) (Epicentre).

The rRNA depleted RNAs were fragmented with RNase III. Then, the RNA fragments were poly(A)-tailed using poly(A) polymerase and the 5'PPP structures were removed using RNA 5' polyphosphatase (Epicentre). Afterwards, the RNA adapter was ligated to the 5'-phosphate of the RNA fragments. First-strand cDNA synthesis was performed using an oligo(dT)-adapter primer and M-MLV reverse transcriptase. The resulting cDNA was PCR-amplified using a high fidelity DNA polymerase. Finally, the cDNA was purified using the Agencourt AMPure XP kit (Beckman Coulter Genomics).

The cDNA pool was size fractionated in the size range of 200 – 450 bp using a differential clean-up with the Agencourt AMPure kit. The primers used for PCR amplification were designed for TruSeq sequencing according to the instructions of Illumina. The cDNA pools were sequenced on a Illumina HiSeq 2000 system using 50 bp read length. Finally. cDNA reads were mapped to the *E. coli* K-12 reference genome (NC_000913),

Chapter 3: High temperature increases bypass transcription *in vivo*

3.1 Introduction and objectives

Transcription by σ^{54}RNAP is typically activated by cognate enhancer binding proteins such as PspF and NtrC. Some variants of σ^{54} are known to bypass these activator requirements *in vitro*, and give rise to σ^{54} activator-independent transcription (see Chapter 1). However, σ^{54} activator bypass transcription (SABT) has not been observed *in vivo* (Sasse-Dwight, 1990; Hsieh, 1994; Buck, unpublished; Kustu, unpublished) unless promoter sequences were mutated (Wang, 1998). If SABT can be observed in a minimal setting *in vitro* but not *in vivo*, are there potential barriers that prevent SABT from occurring *in vivo*?

Unknown factors could stop σ^{54} transcription *in vivo*, when it occurs as SABT *in vitro*. It is also a distinct possibility that SABT does occur *in vivo* at low levels, yet the signal as yielded by a reporter construct is too weak to be readily observed (or the mRNA is degraded too quickly). Discovering conditions where SABT is readily observed *in vivo* may thus shed some light on the type of molecular mechanisms influencing and potentially repressing SABT *in vivo*.

While SABT has been observed *in vivo* at gln_{Ap2} mutant promoters, all of them lack the canonical C in the -12 promoter element (Wang, 1998). Some abiotic factors are known to increase SABT at the gln_{Ap2} promoter *in vitro*, such as high temperature and low salt concentration (Wang, 1997). This Chapter focuses on creating a reporter system for SABT *in vivo*, and to determine whether altering any abiotic factors allows SABT to be observed *in vivo* without mutating the promoter sequences.

3.2 σ^{54} Activator Bypass Reporter Strain Construction

In a standard lab strain, directly studying mRNA from SABT at σ^{54}-dependent promoters *in vivo* can be impractical, costly and time-consuming. Moreover, native chromosomal expression of wild type σ^{54} (from *rpoN* gene) will likely interfere with any observed SABT signal. Therefore, a reporter strain was constructed to circumvent these limitations and help study SABT *in vivo* (Fig 3.2.1).

For this reporter strain a *lacZ*⁻ MG1655 parental strain was used, with a triple translational terminator at the SacI site in the *lacZ* gene (Fig 3.2.1 A). To quantify σ^{54}-dependent SABT, a fusion between the σ^{54}-dependent *pspA* promoter and *lacZ* was transduced into the λ phage attachment site in *E. coli.* (Fig 3.2.1 B). Transcription of the *pspA* promoter is dependent on a functional σ^{54}RNAP holoenzyme (Model, 1997). A Tn*10* transposable element was inserted as *rpoN*208::Tn*10* into the *rpoN* locus via $P1_{vir}$ phage transduction, in order to disrupt native chromosomal σ^{54} production (Fig 3.2.1 C). Finally, a plasmid-borne copy of *Klebsiella pneumonia rpoN* with an N-terminal deletion on pBAD18 was transformed into this strain (Fig 3.2.1 D). An N-terminal deletion of σ^{54} ($rpoN_{\Delta R1}$) is a known strong σ^{54} activator bypass mutant *in vitro* that cannot be activated by a bEBP , since σ^{54}_{R1} is the bEBP binding site (Wang, 1997). Therefore unexpected bEBP activation will not be scored as a SABT false positive with this particular bypass mutant.

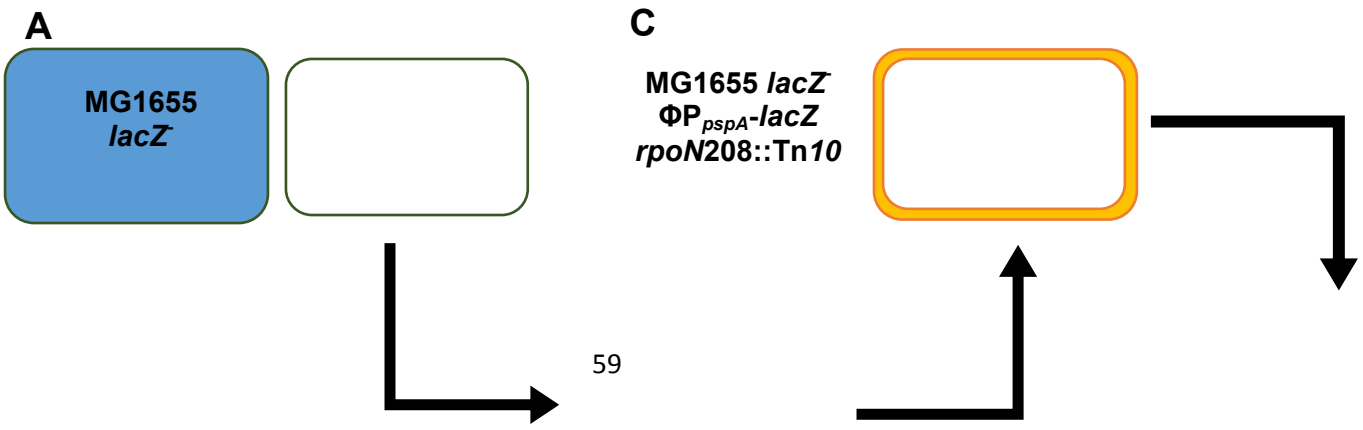


Fig 3.2.1: σ^{54} activator bypass reporter strain (SABRS) construction
A *lacZ*- MG1655 derivative strain (**A**) was transduced with an Amp^R fusion between the σ^{54}-dependent *pspA* promoter and *lacZ*, at the λ attachment site (**B**). Subsequently, an *rpoN*::Tn*10* insertion (Tet^R) was introduced into the *rpoN* locus, disrupting native chromosomal *rpoN* expression (**C**). Finally, an activator bypass mutant variant of *rpoN* with an N-terminal deletion of Region I was introduced into the strain (**D**), on pBAD18 (Cam^R).

3.2.1 *rpoN*208::Tn*10* abolishes σ^{54}-dependent transcription at the *pspA* promoter

In cells with the *ΦP$_{pspA}$-lacZ,* the *pspA* promoter controls LacZ production, which in turn can be monitored with β–galactosidase assays. The *ΦPpspA-lacZ* therefore provides a direct way of measuring σ^{54} activity *in vivo*. To verify that the Tn*10* insertion at the *rpoN* locus disrupts σ^{54} activity, ΦP$_{pspA}$-*lacZ* activity was measured with and without the Tn*10* insertion. ΦP$_{pspA}$-*lacZ* activity increased ≈100-fold in the presence of native chromosomal *rpoN* (Fig 3.2.2 A, grey) compared to the Tn*10* insertion strain (Fig 3.2.2 B, grey). This is in line with the chromosomal *rpoN* gene no longer being functional due to insertion of the Tn*10* transposon.

The strains were also transformed with the F3K3 plasmid encoding the phage protein pIV. F3K3 is a medium copy Kan^R plasmid induced by 400nm N-(beta-Ketocaproyl)-L-homoserine lactone. Secretin pIV from filamentous phage f1 is a potent inducer of the *pspA* promoter (Model, 1997). Expression of the secretin pIV further increases ΦP$_{pspA}$-*lacZ* activity 3-fold in the presence of wild type *rpoN* (Fig 3.2.2 A). The positive correlation between the addition of pIV (inducer of the *pspA* promoter) and a rise in σ^{54}-dependent transcription supports the notion that ΦP$_{pspA}$-*lacZ* expression is σ^{54}-dependent.

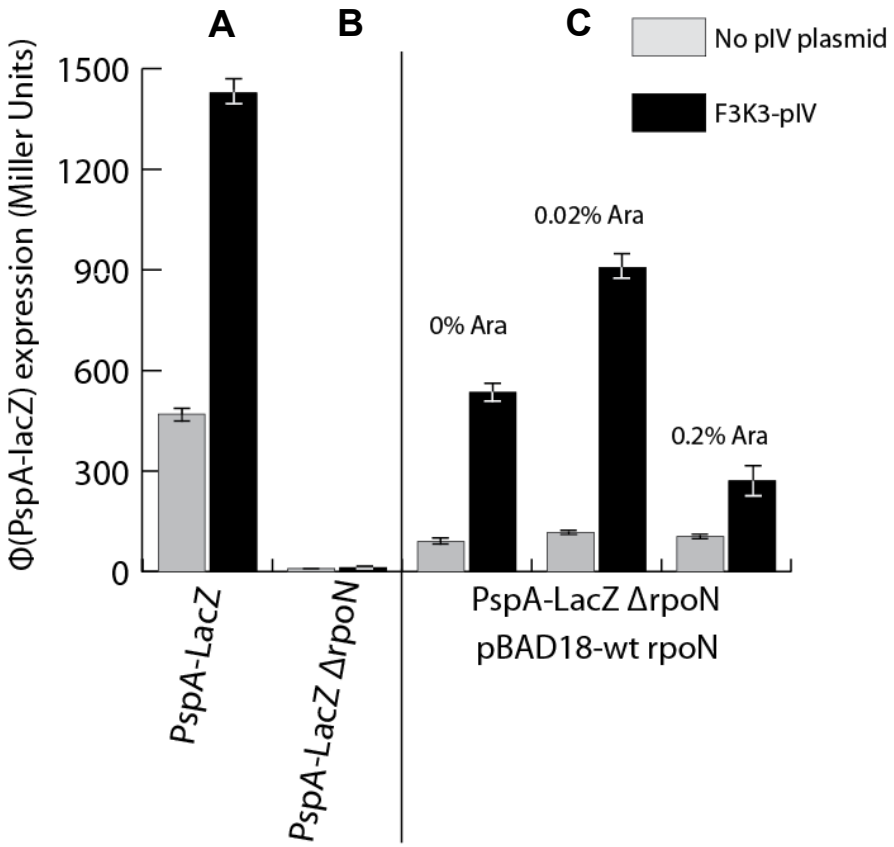


Figure 3.2.2: *ΦP$_{pspA}$-lacZ* expression induced by pIV and abolished by a Tn*10* insertion.

Strains from Fig 3.2.1 B and Fig 3.2.1 C were transformed with F3K3-pIV. Cells were grown to exponential phase (OD_{600} of 0.4) at 37°C and samples were assayed after inducing pIV expression for 1 hour (with 400nm N-(beta-Ketocaproyl)-L-homoserine lactone). β-galactosidase assays were used to determine *ΦPpspA-lacZ* expression *in vivo* in the presence of native chromosomal *rpoN* (**A**) and the *rpoN::*Tn*10* insertion (**B**). Later the strain with the *rpoN*::Tn*10* was complemented with arabinose-inducible pBAD18-*rpoN* (**C**) and the experiment was repeated under the same conditions after 2 hours of growth at different levels of arabinose (% w/v).

3.2.2 Plasmid-borne wild type σ^{54} expression restores σ^{54} activity in *rpoN*208::Tn*10* cells

In 3.2.1 it was shown that σ^{54}-dependent transcription was hardly detectable (<15 Miller Units) in *rpoN*208::Tn*10* cells, regardless of pIV expression (Fig 3.2.2 B). Here these *rpoN*208::Tn*10* cells were complemented with pBAD18 encoding wild type *Klebsiella pneumonia rpoN*. This was done to show that the loss of the σ^{54} activity is due to rpoN*208*::Tn*10* and can be (at least partially) restored with plasmid-borne wild type σ^{54}. If plasmid-borne *rpoN* does complement *rpoN*208::Tn*10* it would act as a proof of principle for introducing $rpoN_{\Delta R1}$ into the σ^{54} activator bypass reporter strain (SABRS) to study SABT *in vivo*, which is not thought to be active under standard growth conditions.

Plasmid-borne expression of a gene on certain plasmids (eg. pBAD) allows for a tighter control of expression and provides a wider dynamic range, which is preferred since σ^{54} availability is likely to affect σ^{54}-dependent transcription. pBAD18 is a plasmid that uses the pBAD promoter of the araBAD (arabinose) operon. pBAD18 also has a tight regulation, modulation and high levels of expression which can be inhibited with glucose (Guzman, 1995). Therefore pBAD18 was used as the plasmid of choice for complementing *rpoN*208::Tn*10* cells with wild type *rpoN* (and later $rpoN_{\Delta R1}$).

Klebsiella pneumonia rpoN was used since it is very similar in amino acid sequence to *E. coli rpoN* (80% sequence identity). It is commonly used in our lab to study σ^{54} function in *E. coli* and thus readily available (Oguiza, 1999; Cannon, 2001; Wigneshweraraj, 2003).

Upon complementation with *Klebsiella pneumonia rpoN*, ≈100 Miller Units of activity was observed in *rpoN*208::Tn*10* cells in the absence of secretin pIV (Fig. 3.2.2 C, grey). About 20% of native chromosomal σ^{54}-dependent transcription was restored (Fig 3.2.2 C and A). In the presence of pIV, ≈300-900 Miller Units were detected depending on the amount of σ^{54} being expressed, which is regulated by arabinose (Fig. 3.2.2 C, black). At the optimal arabinose concentration (0.02%), 65% of native chromosomal σ^{54}-dependent transcription was restored (Fig 3.2.2 C and A).

In the presence of pIV, complementation with *Klebsiella pneumonia rpoN* at 0% arabinose yielded 450 Miller Units of activity, around 40% of native chromosomal *rpoN* (Fig 3.2.2 C and A). Arabinose is an inducer of pBAD18, suggesting that even leaky levels of σ^{54} expression by pBAD18 are sufficient to observe σ^{54}-dependent transcription. This recovery increased to 65% at 0.02% arabinose, conditions which would increase the expression of plasmid borne-*rpoN*.

Interestingly a further increase of plasmid-borne expression (0.2% Ara) resulted in a decrease in ΦP_{pspA}-*lacZ* activity (Fig 3.2.2 C). The rate of induction/repression can be 1200-fold for pBAD based vectors (Guzman, 1995), therefore 0.2% arabinose induction may have detrimental side effects on both the cell itself and σ^{54}dependent transcription. Overexpression of some proteins (including membrane proteins and cytoplasmic proteins) are known to be toxic or create inclusion bodies (Doherty, 1993), which could account for the drop in σ^{54}-dependent transcription at 0.2% arabinose.

Without pIV, this recovery of function (≈20%) seemed to be relatively constant and largely independent of arabinose concentration (Fig 3.2.2 C). This indicates that σ^{54} availability may not be the rate-limiting factor in ΦP_{pspA}-*lacZ* transcription under these conditions.

Overall at least 20% recovery of native chromosomal *rpoN* function was observed when expressing plasmid-borne *Klebsiella pneumonia rpoN* in *rpoN*208::Tn*10* cells. Using a *Klebsiella* variant of the *rpoN* gene may explain why a complete functional recovery is not observed, however this data does show that the *Klebsiella pneumonia rpoN* is (at least partially) functional in *E. coli in vivo*. More importantly, it acts as a proof of principle that this setup with plasmid-borne *rpoN* (or $rpoN_{\Delta R1}$) and $\Phi P_{pspA}LacZ$ can be used to study σ^{54}-dependent transcription *in vivo*.

3.2.3 Introducing Klebsiella *pneumonia* $\sigma^{54}{}_{\Delta R1}$ to construct a functioning SABRS *in vivo*

In 3.2.1 *rpoN::*Tn*10* cells were shown to disrupt σ^{54}-dependent transcription of ΦP_{pspA}-*lacZ* due to the Tn*10* insertion, with virtually no background σ^{54} activity at the *pspA* promoter (Fig. 3.2.2 B). In 3.2.3 this loss of σ^{54}-dependent transcription was restored with a plasmid-borne *Klebsiella rpoN* (Fig 3.2.2 C). Therefore this parental strain is suitable for the study of SABT *in vivo* because it has known σ^{54}-dependent transcription of the ΦP_{pspA}-*lacZ* alongside very low background signal.

However, to create a working σ^{54} activator bypass reporter strain (SABRS) *in vivo*, a σ^{54} activator bypass mutant still needs to be introduced into this strain (Fig 3.2.1). Doing so will perform a dual function. First it will confirm (or possibly disprove) the initial hypothesis that SABT cannot be observed *in vivo* under standard lab growth conditions, presumably due to unknown inhibitory factors present *in vivo* but not *in vitro*. Secondly, this strain can act as a reporter strain that can be manipulated later in attempts to observe SABT *in vivo*.

Plasmid-borne $rpoN_{\Delta R1}$ was selected as a bypass mutant to study σ^{54} bypass transcription *in vivo*. All characterised σ^{54} activator bypass mutants typically disrupt the tight binding between σ^{54} and the DNA just downstream of the GC promoter element at -12 (Wang, 1997). This interaction normally prevents the binding of the holoenzyme to the non-template strand, a key step for isomerisation into the open complex and transcription initiation (Buck, 2000; Guo, 2000).

This tight σ^{54}/DNA interaction can be removed by changing a leucine patch in N-terminal Region I of σ^{54} (Syed, 1998). More importantly an N-terminal deletion of σ^{54} (amino acids 1-56) is well characterised and a known σ^{54} bypass mutant. It was also used previously to study how abiotic factors influence SABT *in vitro* (Wang, 1997). Since $\sigma^{54}{}_{R1}$ is the bEBP binding site, unexpected bEBP activation will also not be scored as a SABT false positive with this particular bypass mutant. Combined, these factors make $\sigma^{54}{}_{\Delta R1}$ the most suitable bypass mutant candidate for this study.

The *Klebsiella pneumonia* $rpoN_{\Delta R1}$ (σ^{54} bypass mutant) sequence was cloned into pBAD18 and transformed into MG1655 LacZ$^-$ ΦP_{pspA}-*lacZ rpoN*208::Tn*10*. This SABRS provided a framework for further study of σ^{54} bypass transcription *in vivo*.

3.2.4 Confirmation of reporter strain genetic backgrounds of using colony PCR

Before continuing to use the SABRS for further experiments, the genetic background of the strains used were confirmed via PCR, amplifying loci of interest. Ultimately, 3 strains were constructed for the study of SABT *in vivo*:

• Strain of interest: reporter strain with $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant (**SABRS**)

• Positive control: reporter strain with wild type σ^{54} (**WTRS**)

• Negative control: reporter strain with empty pBAD, lacking any σ^{54} (**DRRS**)

Amplification of the *rpoN* locus produced 1.5 kbp fragments in the MG1655 and MG1655 *LacZ*⁻ strains (lacking the *rpoN*208::Tn*10* insertion), characteristic of the wild type *rpoN* gene (Fig 3.2.3, 1-2). In the *rpoN*208::Tn*10* strains (Fig 3.2.3, 3-7) no PCR product was observed, most likely because the Tn*10* insertion is too large to amplify under these PCR conditions (13 kbp). This indicates that the Tn*10* transposon is present, disrupting *rpoN* expression.

Colony PCR of the pBAD18 plasmid confirmed the presence of the 1.5 kbp $rpoN_{\Delta R1}$ in pBAD18-$rpoN_{\Delta R1}$ (Fig 3.2.3, 9) and slightly larger 2kbp wild type *rpoN* in pBAD18-*rpoN* (Fig 3.2.3, 10). The empty pBAD18 plasmid only produces a very small product of several bases, which covers part of the pBAD18 multiple cloning site (Fig 3.2.3, 8). Presence of the *ΦPpspA-lacZ* was confirmed in the SABRS using *pspA*-UAS-specific forward and *lacZ* reverse primers (Fig 3.2.3, 11-12), all yielding the fragments expected.

MG1655

MG1655 *LacZ*⁻

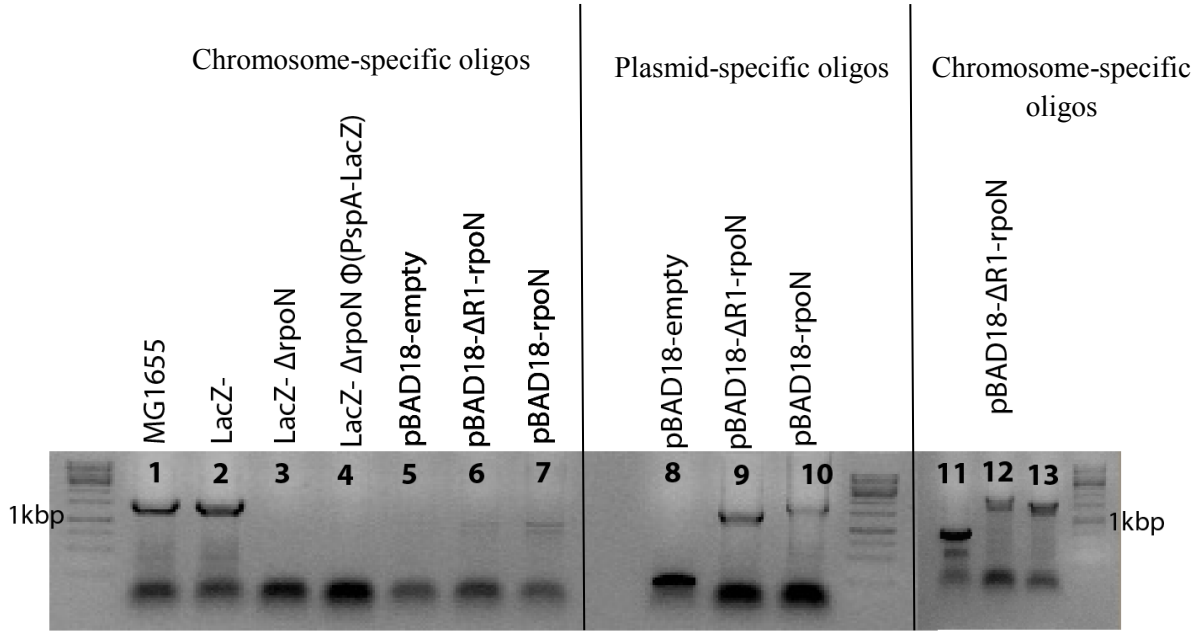


Fig 3.2.3: Colony PCR of *rpoN* locus, ΦP_{pspA}-*lacZ* and pBAD18 inserts.

Regions of interest were amplified by colony PCR and run on a 1% agarose gel for 45 mins. *rpoN*-specific primers were used to amplify the native *rpoN* locus (**1-7**), generating a 1.5 kbp fragment when it is intact (**1** and **2**). pBAD18-specific primers were used to amplify plasmid inserts in the DRRS, SABRS and WTRS (**8-10**). The *ΦPpspA-lacZ* construct was amplified using forward primers for the PspA-UAS and *LacZ* (near start site) reverse primers (**11**), PspA-UAS forward and *LacZ* (internal) reverse primers (**12**), and *LacZ* forward and reverse primers (**13**).

3.2.5 The SABRS (with rpoN$_{\Delta R1}$) does not cause SABT *in vivo* under standard conditions

In 3.2.1 it was shown that inserting a Tn*10* transposon into the *rpoN* gene reduces σ^{54}-dependent activity by 99% (Fig 3.2.2). This loss of function could be restored by up to 65% by introducing wild type *Klebsiella pneumonia rpoN* (3.2.2).

However, complementing that same strain with *Klebsiella pneumonia rpoN*$_{\Delta R1}$ (activator bypass mutant), does not result in σ^{54}-dependent transcription of ΦP_{pspA}-*lacZ* (Fig 3.2.4). Only ≈10 Miller Units of activity are observed, virtually identical to the levels observed in the uncomplemented *rpoN*208::Tn*10* strain (Fig 3.2.2 B). Therefore no SABT is observed under these conditions, in line the hypothesis that unknown factors exist *in vivo* that prevent bypass transcription from occurring.

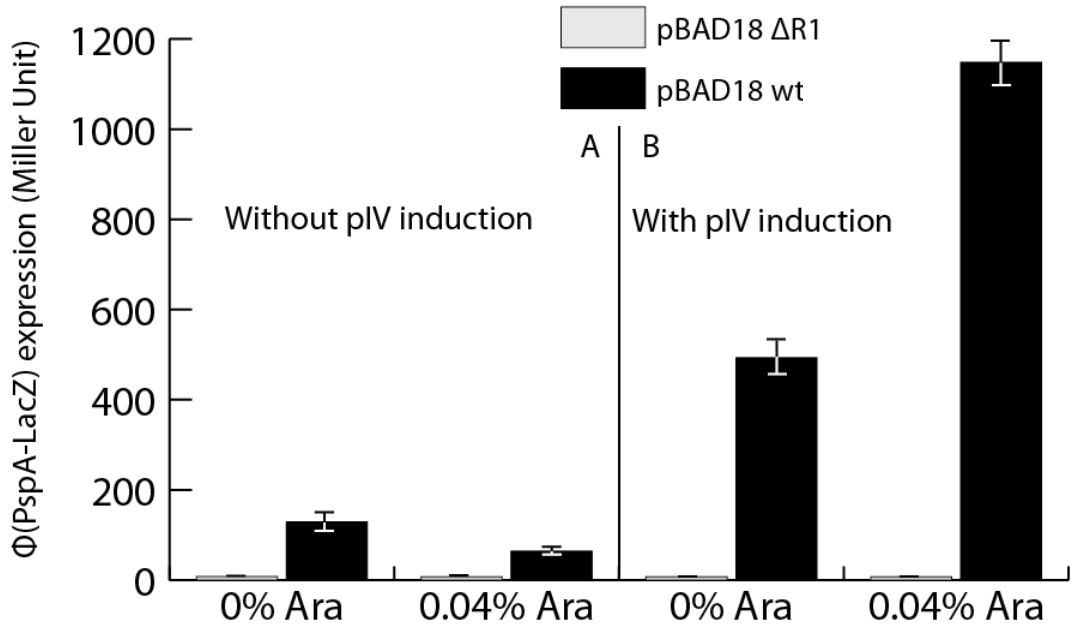


Fig 3.2.4: Complementing rpoN208::Tn*10* cells with arabinose-inducible pBAD18-*rpoN*$_{\Delta R1}$.
*rpoN*208::Tn*10* cells with F3K3-pIV and pBAD18-ΔR1-*rpoN* or pBAD18-*rpoN* were grown to exponential phase (OD_{600} of 0.4) at 37°C in pBAD18-inducing (0.04% w/v Ara) and non-inducing (0% w/v Ara) conditions. Samples were assayed before (**A**) and after (**B**) 1 hour F3K3-pIV induction (with 400nm N-(beta-Ketocaproyl)-L-homoserine lactone). β-galactosidase assays were used to determine *ΦPpspA-lacZ* expression *in vivo.*

3.2.6 pBAD18-rpoN$_{R1}$ is expressed in the SABRS

Lack of ΦP$_{pspA}$-*lacZ* activity suggests there are barriers to SABT *in vivo*, but does not exclude the possibility that the plasmid-borne bypass mutant (*rpoN*$_{\Delta R1}$) was not transcribed or translated within the *E. coli* cell.

To ensure that *rpoN*$_{\Delta R1}$ is expressed, SDS-PAGE and Coomassie staining was performed on cell extracts with 3 different plasmids: pBAD18-empty (Fig 3.2.5, 1), pBAD18-*rpoN* (Fig. 3.2.5, 2) and pBAD18-*rpoN*$_{\Delta R1}$ (Fig. 3.2.5, 3).

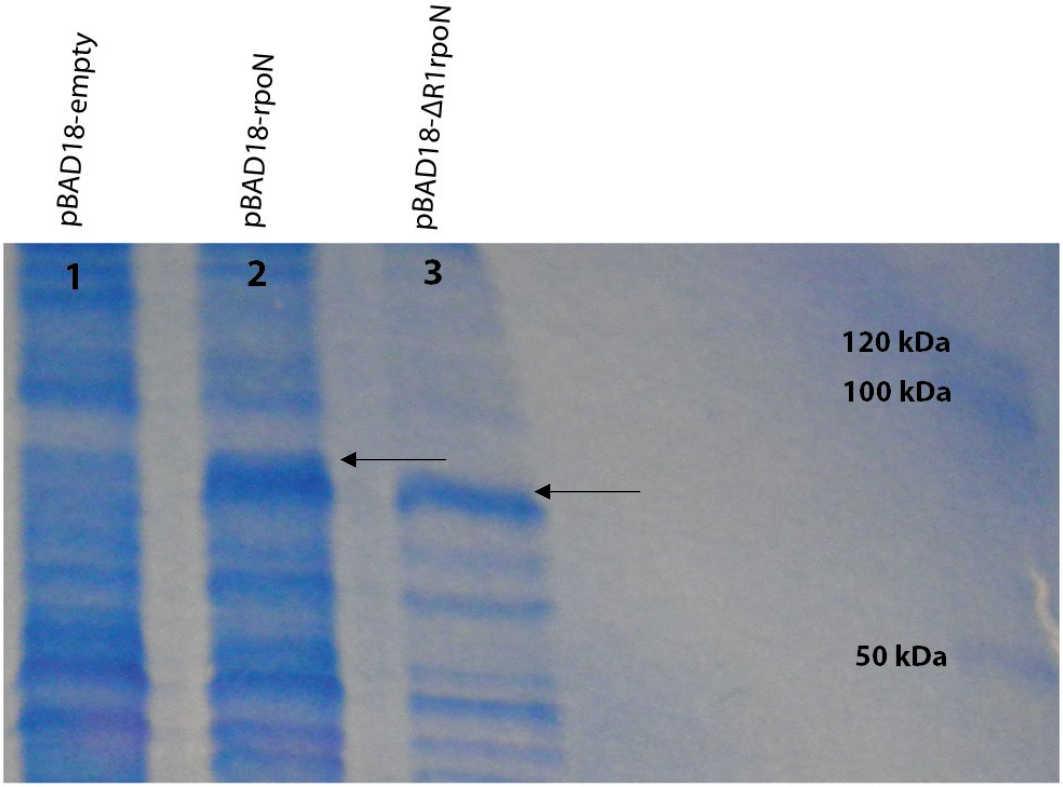


Fig 3.2.5: SDS-PAGE of cells with plasmid-borne *rpoN* and *rpoN*$_{\Delta R1}$ after overexpression.
5ml of cells with 3 different pBAD18 inserts were grown to exponential phase (OD_{600} of 0.4) at 37°C and induced with 0.02% arabinose for 1 hour to increase expression of the pBAD18 plasmid. Bacterial pellets were resuspended in Laemmli buffer, boiled for 5 mins at 95°C and run on a 10% SDS-PAGE gel at 200V for 45 minutes and visualised after Coomassie staining. The 3 plasmid inserts were: pBAD18-empty (**1**), pBAD18-*rpoN* (**2**), pBAD18- -*rpoN*$_{\Delta R1}$ (**3**).

Data from Fig. 3.2.2 C indicates that plasmid-borne wild type *Klebsiella pneumonia rpoN* is expressed because it functionally complements the chromosomal *rpoN* deletion in the *rpoN*208::Tn*10* strain. pBAD18 can express proteins at fairly high levels (Khlebnikov, 2002), hence σ^{54} should be readily visible on SDS-PAGE upon arabinose induction of pBAD18. Therefore pBAD18-*rpoN* expression can be used as a positive control for band intensity (and size approximation) when examining pBAD18-$rpoN_{\Delta R1}$ expression on the SDS-PAGE gel.

Despite being a 54 kDa protein, σ^{54} is known to run at ≈70 kDa on an SDS-PAGE gel. A protein at this size was readily visible in the lane with pBAD18-*rpoN* (Fig. 3.2.5, 2 arrowed), where *rpoN* is known to be expressed (Fig 3.2.2 C). A similarly sized protein was also easily detectable in the lane with pBAD18-$rpoN_{\Delta R1}$ (Fig. 3.2.5, 3 arrowed), but not in the negative control with an empty plasmid (Fig. 3.2.5, 1). This suggests that the proteins detected are a direct result of the plasmid insert (ie. σ^{54} and $\sigma^{54}{}_{\Delta R1}$). Finally, the protein expressed in lane 3 ($\sigma^{54}{}_{\Delta R1}$) appears to be slightly smaller than the one expressed in lane 2 (σ^{54}), which would be expected given the N-terminal deletion of region I. Additionally, RNASeq data confirmed that the bypass mutant was being transcribed from the pBAD construct (Chapter 5).

Combined, these outcomes suggest that plasmid-borne $rpoN_{\Delta R1}$ is indeed being expressed *in vivo* but that it simply does not cause SABT under standard conditions. This also shows that while σ^{54} bypass mutants allow SABT *in vitro* in the absence of bEBPs, they also lose the ability to do so in the presence of bEBPs *in vivo* (at least for the $\sigma^{54}{}_{\Delta R1}$ mutant). Consequently, any increase in SABT is presumed to be independent of bEBPs and attributable to other changes such as abiotic or biotic factors. The interaction between $\sigma^{54}{}_{\Delta R1}$ and bEBPs is discussed further in Chapter 6.

3.3 Results

SABT has been observed *in vitro* at the gln_{Ap2} promoter by increasing temperature or lowering salt concentration to favour transient DNA melting (Wang, 1997). These conditions allow for the formation of heparin sensitive pre-initiation complexes that can transcribe in the absence of ATP and bEBPs.

In order to determine whether these *in vitro* findings can be replicated *in vivo*, a SABRS reporter strain variant was constructed with a $\Phi P_{gln Ap2}$-*gfp* instead of ΦP_{pspA}-*lacZ* (gln_{Ap2}-SABRS). This gln_{Ap2}-SABRS was created to replicate the *in vivo* experiment as closely as possible to the *in vitro* conditions. The $\Phi P_{gln Ap2}$-*gfp* was introduced into the strain on a low copy Amp^R plasmid, pSB4A3.

The gln_{Ap2} promoter is dependent on σ^{54}RNAP for activation (Ninfa, 1987). This nitrogen-regulated promoter is responsible for the expression of glutamine synthase, and is upregulated under nitrogen-limiting conditions in *E. coli* (Kumar and Shimizu, 2010).

3.3.1 Wild type σ^{54} induces ΦP_{glnAp2}-*gfp* expression, SABT mutant does not

The ΦP_{pspA}-*lacZ* fusion had been examined for SABT previously (Fig 3.2.3), but ΦP_{pspA}-*lacZ* activity was only observed in the presence of wild type σ^{54} and not $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant (Fig 3.2.2 and 3.2.3). While the gln_{Ap2}-SABRS is likely to behave similarly with a ΦP_{glnAp2}-*gfp* reporter, SABT *in vivo* may vary between different σ^{54}-dependent promoters. Therefore examining the effect of σ^{54} and $\sigma^{54}{}_{\Delta R1}$ on ΦP_{glnAp2}-*gfp in vivo* will do two things: 1) confirm that the fusion construct is functional (with wild type σ^{54}), and 2) determine whether SABT is readily observed *in vivo* at the gln_{Ap2} promoter.

In order to maximise chances of observing SABT at the gln_{Ap2} promoter, cells were grown in mimimal Gutnick media (Gutnick, 1969) supplemented with 0.4% glucose (w/v) and 5mM L-Glutamine. These are nitrogen-limiting conditions and should activate transcription at the gln_{Ap2}-promoter (Kumar and Shimizu, 2010).

In Fig. 3.3.1, a signal output of ≈5000-11000 Units (arbitrary units) was observed in the gln_{Ap2}-WTRS (wild type σ^{54}). Similarly to the expression of ΦP_{pspA}-*lacZ* (Fig. 3.2.2 C), ΦP_{glnAp2}-*gfp* activity was highest (11000 Units) at 0.02% arabinose induction and lowest (5500 Units) at 0.2% arabinose induction at (Fig. 3.3.1). This supports the notion that ΦP_{glnAp2}-*gfp* activity is σ^{54}-dependent. Interestingly, arabinose induction only results in a small 20% increase in σ^{54}-dependent transcription (Fig. 3.3.1), perhaps suggesting that minimal Gutnick media does not activate ΦP_{glnAp2}-*gfp* as strongly as pIV induces ΦP_{pspA}-*lacZ* (200% increase, Fig. 3.2.2 C).

In the gln_{Ap2}-SABRS ($\sigma^{54}{}_{\Delta R1}$), only a constant and basal level (1000 Units) of σ^{54}-dependent transcription was observed (Fig. 3.3.1). This indicates that the $\sigma^{54}{}_{\Delta R1}$ bypass mutant does not cause any significant σ^{54}-dependent transcription at the gln_{Ap2} promoter *in vivo* under these conditions.

The same pattern of activity was observed in the presence of σ^{54} and $\sigma^{54}{}_{\Delta R1}$ for two different σ^{54}-dependent promoters (*pspA* and gln_{Ap2}), in line with the idea that the promoter fusions are a reliable output for fully functional σ^{54} activity *in vivo*.

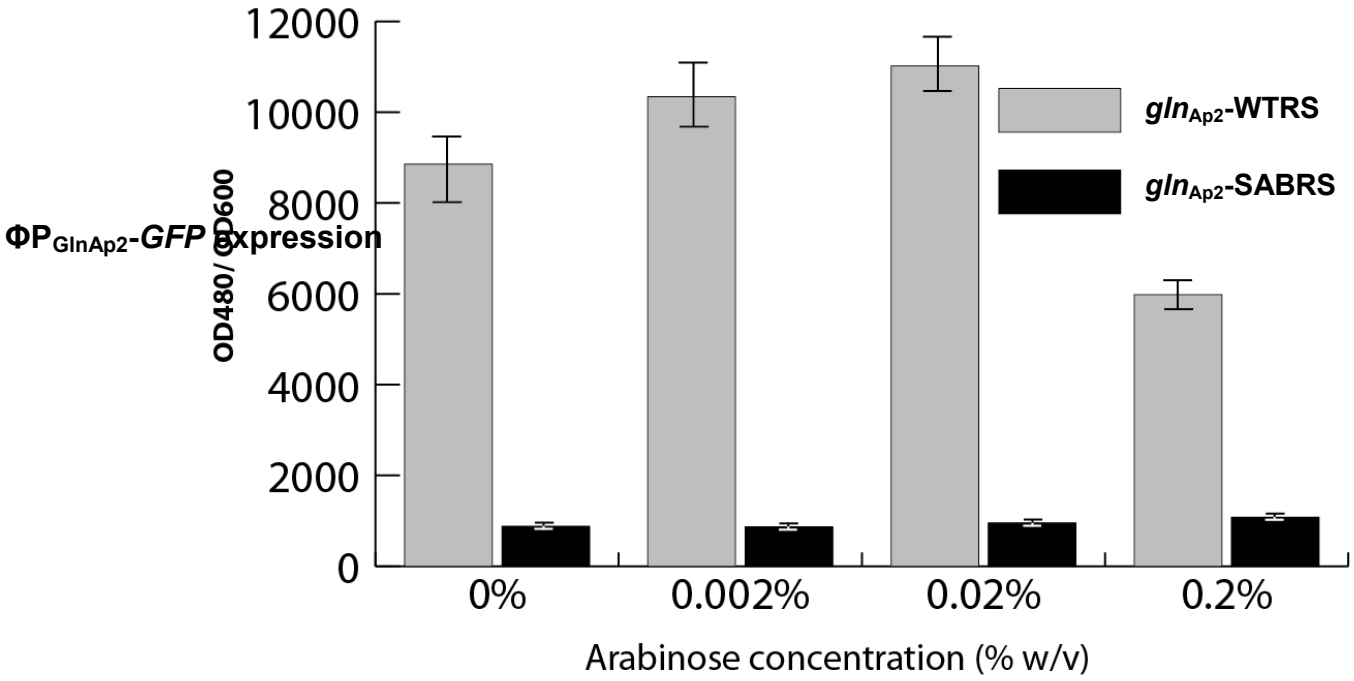


Fig 3.3.1: wild type σ^{54} induces ΦP_{GlnAp2}-*GFP*, $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant does not

gln$_{Ap2}$-WTRS (wild type σ^{54}) and *gln*$_{Ap2}$-SABRS ($\sigma^{54}{}_{\Delta R1}$) cells were grown for 20 hours at 37°C in minimal Gutnick media (supplemented with 0.4% glucose (w/v) and 5mM L-Glu) at different arabinose concentrations. GFP fluorescence (OD_{480}) was taken and normalised for cell density (OD_{600}).

3.3.2 Lower salt concentration does not induce ΦP_{glnAp2}-*gfp in vivo*, with σ^{54} or $\sigma^{54}{}_{\Delta R1}$

Low salt concentration is thought to favour transient DNA melting during transcription initiation *in vitro*, allowing σ^{54} to direct transcription in the absence of an activator. Lowering salt concentration *in vitro* from 200 mM to 50 mM increased σ^{54} bypass transcription from 0% to 22% of activator-dependent transcription. This is thought to occur by weakening the interaction of σ^{54}RNAP with the -12 element on the DNA (Wang, 1997).

However, lowering the salt conditions for the growth experiments *in vivo* did not result in an increase in σ^{54}-dependent transcription *in vivo*, it even reduced it in the gln_{Ap2}-WTRS (Fig. 3.3.2). Very low salt concentration (15mM) also reduced cell growth in both strains (not shown). The fluorescence levels observed at low salt concentration are thought to be background fluorescence from the inoculation cultures, which have not grown much since.

It should be noted that these salt concentrations refer to their respective levels in the extracellular environment. Unlike the *in vitro* experiments, it is harder to manipulate the salt concentration where transcription itself is taking place (ie. within the cell). It is likely that homeostatic cell processes try to maintain a relatively constant intracellular salt concentration despite large changes in extracellular salt concentrations.

Therefore this experiment was unable to confirm that low salt concentrations increase SABT *in vivo*. This was in part due to complications in altering the intracellular salt concentrations, but also because low salt concentrations were prohibitive to cell growth and expression of the fusion protein (also in the positive control, gln_{Ap2}-WTRS).

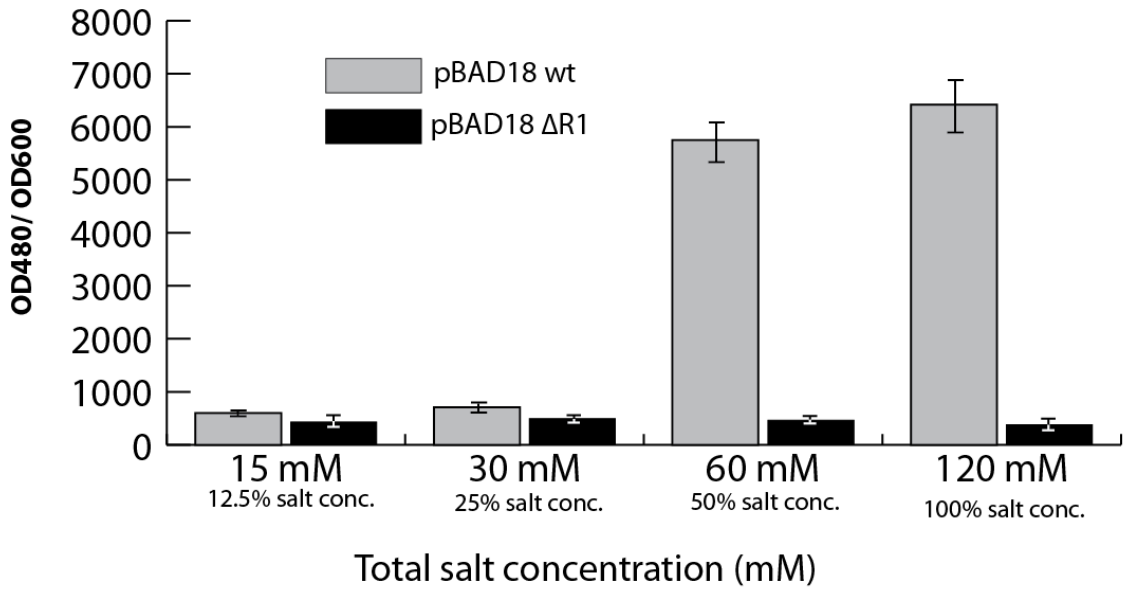


Fig 3.3.2: Effect of salt concentration on bypass transcription.

gln_{Ap2}-WTRS (wild type σ^{54}) and gln_{A} ***gln*$_{Ap2}$-WTRS** ells were grown for 20 hours at 37°C in minimal Gutnick media (supplemented ***gln*$_{Ap2}$-SABRS** (w/v) and 5mM L-Glutamine) at varying salt concentrations. GFP fluorescence (OD_{480}) was measured and normalised for cell density (OD_{600}).

ΦP$_{GlnAp2}$-*GFP* expression

3.3.3 L-glutamine concentration does not induce ΦP$_{GlnAp2}$-*gfp in vivo*, with σ^{54} or $\sigma^{54}{}_{\Delta R1}$

Next the effect L-glutamine concentration on SABT *in vivo* was examined. While L-glutamine concentration (a nitrogen source) was not one of the abiotic factors studied *in vitro*, nitrogen availability is an important factor *rpoN* regulation and expression (Kumar and Shimizu, 2010). Therefore any putative inhibitory proteins influencing SABT under normal conditions may be subject to a nitrogen dependent regulation themselves. Low levels of L-glutamine in minimal media (nitrogen-limiting conditions) are inducing conditions for σ^{54}-dependent transcription of the gln_{Ap2} promoter. These conditions may therefore alleviate the repressive effects of the putative repressive SABT proteins *in vivo*, and so allow SABT to be observed.

In the gln_{Ap2}-WTRS, L-Glutamine concentration did not cause major differences in ΦP_{glnAp2}-*gfp* expression. However, a 75% decrease in activity was observed comparing 5x L-Glutamine to 2x L-Glutamine (Fig. 3.3.3). It is likely that 5x L-Glutamine is no longer a nitrogen-limiting condition, therefore σ^{54}-dependent transcription of the gln_{Ap2} promoter may no longer be activated. At 0% L-Glutamine cell growth was severely stunted, the fluorescence observed is likely to be background fluorescence.

Changes in L-Glutamine concentration over the 0 mM to 7.5 mM range did not result in SABT being readily observed *in vivo* (Fig. 3.3.3). Any putative inhibitory proteins do not appear to be downregulated under these conditions. It is possible that SABT is subject to the action global of repressors that are not influenced by σ^{54}-specific regulation or nitrogen-limiting conditions. Potential candidates may include HNS, a global negative regulator of transcription thought to affect σ^{54} (Belik, 2008), or dksA which reduces the half-life of unstable promoter complexes (Perederina, 2004).

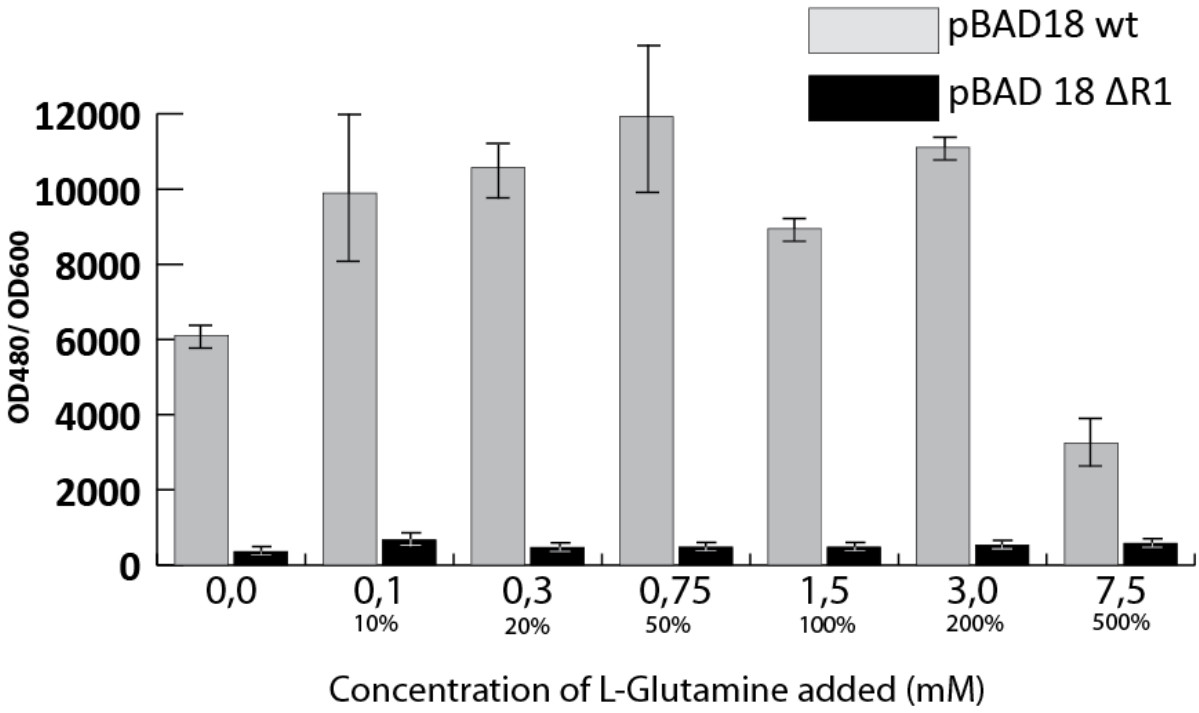


Fig 3.3.3: Effect of L-Glutamine on bypass transcription *in vivo*.

gln_{Ap2}-WTRS (wild type σ^{54}) and gln_{Ap2}-SABRS ($\sigma^{54}{}_{\Delta R1}$) cells were grown for in minimal Gutnick media supplemented with 0.4% glucose at different concentra e. GFP fluorescence (OD_{480}) was measured and normalised for cell density (OD_{600}).

gln_{Ap2}-WTRS

gln_{Ap2}-SABRS

GlnAp2-*GFP* expression

3.3.4 High temperature increases SABT at gln_{Ap2} *in vivo,* reduces wild type σ^{54} transcription

High temperature is thought to destabilise the σ^{54}/-12 DNA interactions in the σ^{54}RNAP closed complex to favour transient DNA melting *in vitro* (Wang, 1997). Increasing temperature from 37°C to 42°C showed a 2-fold increase in σ^{54} bypass transcription *in vitro* (Wang, 1997).

Here we see that a similar trend is seen *in vivo*, a 4-fold increase in SABT was observed upon increasing the temperature from 37°C to 42°C (Fig. 3.3.4). This accounts for almost 50% of wild type σ^{54}-dependent transcription at 42°C (Fig. 3.3.4).

Interestingly, a 50% decrease in wild type σ^{54}-dependent transcription is observed between 30°C and 37°C, but also 37°C and 42°C (Fig. 3.3.4). GFP is stable at temperatures much higher than 42°C, suggesting that GFP denaturation is not the cause for the inverse correlation between temperature and ΦP_{glnAp2}-*gfp* expression in the *gln*$_{Ap2}$-WTRS (Sokalingam, 2012). Perhaps some protein components of σ^{54}-dependent transcription (such as bEBPs) may be impaired at higher temperatures (especially 42°C), which could adversely affect ΦP_{glnAp2}-*gfp* expression. This has also been suggested in other studies (Wang, 1997).

This hypothesis could explain the sudden increase in SABT at 42°C: potential proteinaceous barriers inhibiting bypass transcription may be impaired at high temperatures, in addition to the aforementioned destabilisation of the σ^{54}/-12 DNA interactions in the σ^{54}RNAP closed complex. It is unclear whether the factors increasing SABT at high temperature in the *gln*$_{Ap2}$-SABRS and decreasing expression in the *gln*$_{Ap2}$-WTRS are the same.

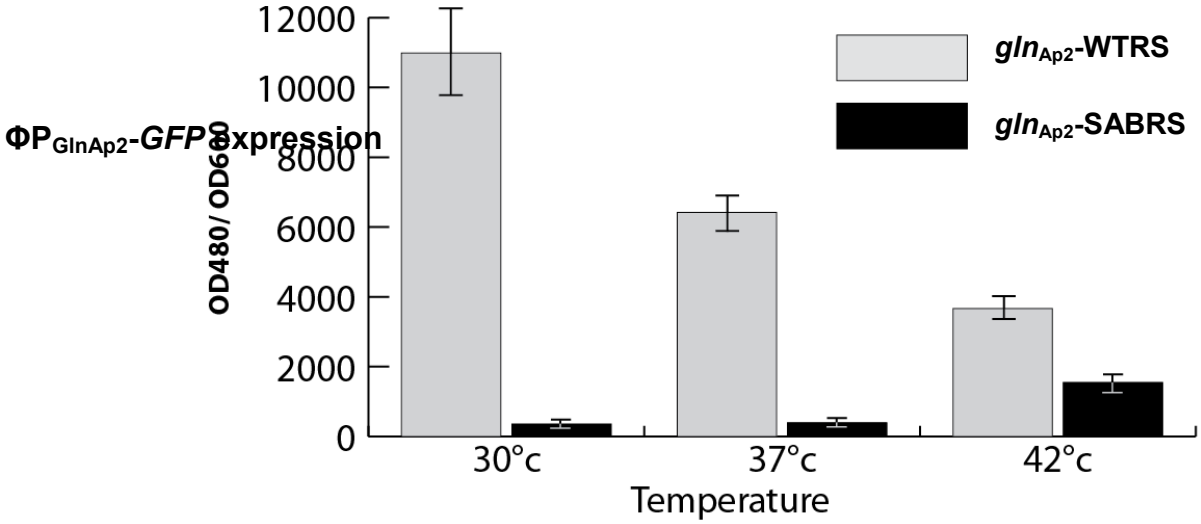


Fig 3.3.4: Effect of temperature on SABT at *gln*$_{Ap2}$ promoter *in vivo*.

gln$_{Ap2}$-WTRS (wild type σ^{54}) and *gln*$_{Ap2}$-SABRS ($\sigma^{54}{}_{\Delta R1}$) cells were grown for 20 hours at different temperatures in minimal Gutnick media (supplemented with 0.4% glucose and 5 mM L-Glutamine). GFP fluorescence (OD_{480}) was measured and normalised for cell density (OD_{600}).

3.3.5 High temperature does not increases SABT at the *pspA* promoter *in vivo*

SABT at the gln_{Ap2} promoter was shown to increase significantly from 37°C to 42°C, whereas wild type σ^{54}-dependent transcription of the gln_{Ap2} promoter decreased over the same temperature change. This was in line with *in vitro* observations for SABT over those temperature ranges (Wang, 1997), and raised the question as to whether this effect to the unique to the gln_{Ap2} promoter or whether SABT could be observed *in vivo* at other σ^{54}-dependent promoters.

To address this question the ΦP_{pspA}-*lacZ* variant of the SABRS was used (*pspA*-SABRS). SABT has not been observed at the *pspA* promoter *in vivo*, but it has *in vitro* (Zhang, unpublished). The *pspA* promoter is known to be induced during heat shock response, and should therefore be active at 42°C (Brissette, 1991). In line with this, ΦP_{pspA}-*lacZ* activity increased 5-fold in the *pspA*-WTRS (wild type σ^{54}), when increasing the temperature from 37°C to 42°C (Fig 3.3.5). However, in the *pspA*-SABRS ($\sigma^{54}{}_{\Delta R1}$) SABT was not observed at 42°C (Fig 3.3.5).

These results show that SABT can be observed *in vivo* at 42°C at the gln_{Ap2} promoter (Fig 3.3.4), but this is not the case for the *pspA* promoter (Fig 3.3.5). These findings indicate that SABT at all σ^{54}-dependent promoters may not be affected equally, perhaps suggesting their inhibition is a byproduct of more general repression as opposed to inhibition of σ^{54}-dependent promoters specifically. This supports the hypothesis in 3.3.3, that inhibition of SABT *in vivo* may be a byproduct of global repressors.

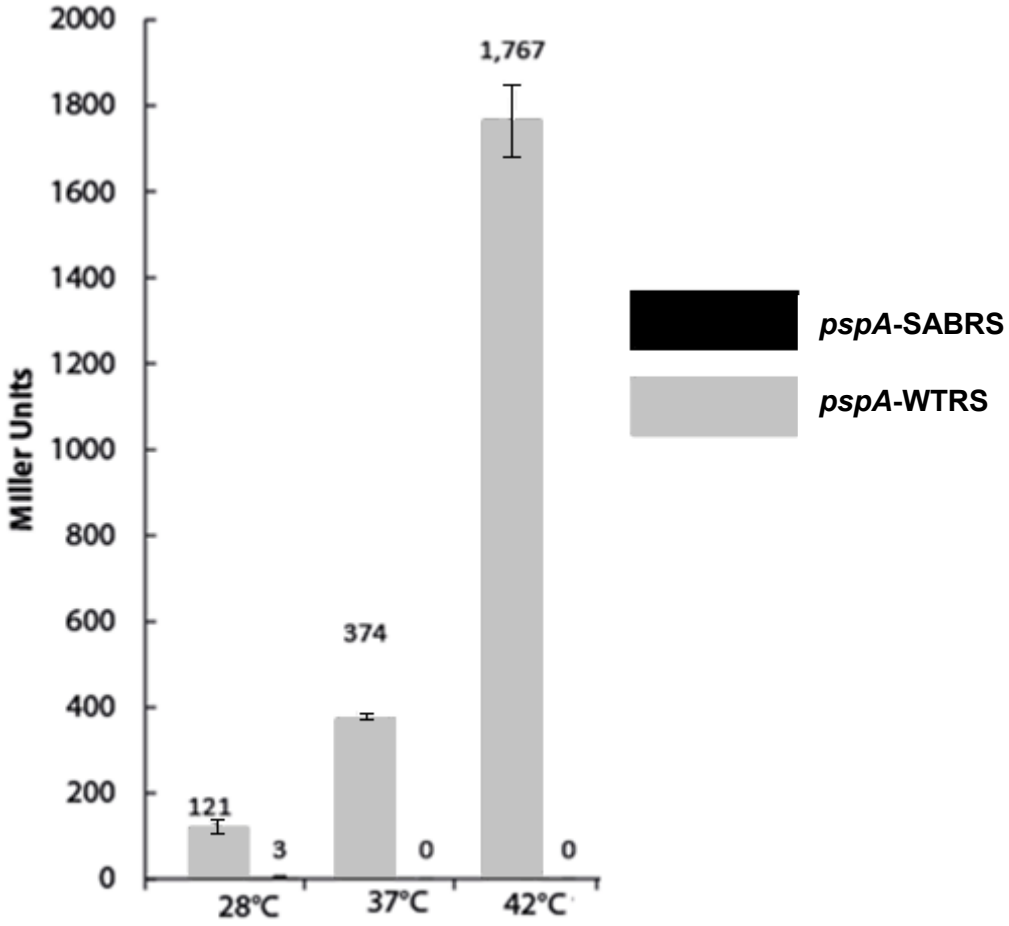


Fig 3.3.5 Effect of temperature on ΦP$_{pspA}$-*lacZ* activity in *pspA*-WTRS and *pspA*-SABRS.

pspA-WTRS (wild type σ^{54}) and *pspA*-SABRS ($\sigma^{54}{}_{\Delta R1}$) cells were grown for 20 hours at different temperatures minimal media with 5mM L-Glutamine and 0.4% glucose. ΦP$_{pspA}$-*lacZ* activity was measured using β-galactosidase assay. In the *pspA*-WTRS, ΦP$_{pspA}$-*lacZ* activity was low at 28°C, but 3x higher at 37°C and 15x higher at 42°C (black). The ΦP$_{pspA}$-*lacZ* fusion was not strongly induced at 42°C in the *pspA*-SABRS, with only background levels of activity at all 3 temperatures (grey).

3.3.6 Examining global bypass transcription *in vitro* using chromosomal DNA

In Fig 3.2.3 and 3.3.1 it was shown that SABT is not observed *in vivo* under standard conditions at the *pspA* or gln_{Ap2} promoters. In 3.3.3 and 3.3.5 it was suggested that this could possibly be due to proteins repressing transcription globally. This inhibition could be achieved in several ways including modification of the DNA topology, which is known to affect gene expression (Drolet, 2006). If there are barriers to SABT *in vivo* through changes in DNA structure, perhaps they can be studied through their loss or conservation during a chromosomal DNA extraction. This chromosomal DNA could then be used as a template for SABT *in vitro*. If we do observe inhibition of SABT *in vitro* using a chromosomal DNA template, we know that the barriers to SABT *in vivo* are preserved during a chromosomal DNA extraction. The chromosomal DNA could then be modified further in an attempt to allow SABT to occur again *in vitro*, and possibly identifying the barriers involved.

Chromosomal DNA was purified from the *pspA*-SABRS (DNeasy Blood & Tissue Kit, Qiagen). The protocol was modified to reduce DNA shearing and to maximise the yield of supercoiled DNA. Modifications include inverting tubes instead of vortexing, gravity flow as opposed to centrifugation, 12°C where possible (Carbone, 2012). Experiments were run with and without $pspF_{1\text{-}275}$, a bEBP variant lacking the DNA binding domain. $PspF_{1\text{-}275}$, without the DNA binding domain, allows it to act as an activator of σ^{54}RNAP from solution (Schumacher, 2007). About 70-100 units of background SABT was observed with RNAP only (Fig 3.3.6, A). 60-70% more SABT is observed with $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant (Fig 3.3.6, B). With wild type σ^{54} we detect over 2-fold σ^{54}-dependent transcription, compared to RNAP only, (Fig 3.3.6, C).

The levels of transcription with $\sigma^{54}{}_{\Delta R1}$RNAP and σ^{54}RNAP are higher than RNAP only, therefore σ^{54}-dependent transcription is detected (Fig 3.3.6 B and C). Since SABT was detected *in vitro* with chromosomal DNA and $\sigma^{54}{}_{\Delta R1}$, any putative changes to DNA thought to inhibit SABT *in vivo* were not seen *in vitro* (or not preserved in the chromosomal DNA extraction). As such, the hypothesis that DNA topology may play a role in inhibiting SABT *in vivo* could not be confirmed or disproven at this point.

Interestingly there appears to be little correlation in the levels of σ^{54}-dependent transcription and $pspF_{1\text{-}275}$, even for wild type σ^{54} which typically requires an activator. *In vitro* assays measuring SABT typically involve DNA that is almost exclusively σ^{54} promoter DNA. However using chromosomal DNA σ^{54} promoters are far less common, with only about 250 expected promoters over 4.5Mbp (Wade, 2015; Appendix C). Therefore the effective concentration of σ^{54}-promoter DNA is very low even at high DNA concentrations. $PspF_{1\text{-}275}$ acts from solution *in trans*, so only stable closed promoter complexes would be strongly activated. Therefore a strong signal would not be expected, especially when the existing signal is already low. Perhaps if the initial signal was higher (without $PspF_{1\text{-}275}$), the effects of adding $PspF_{1\text{-}275}$ would be more noticeable.

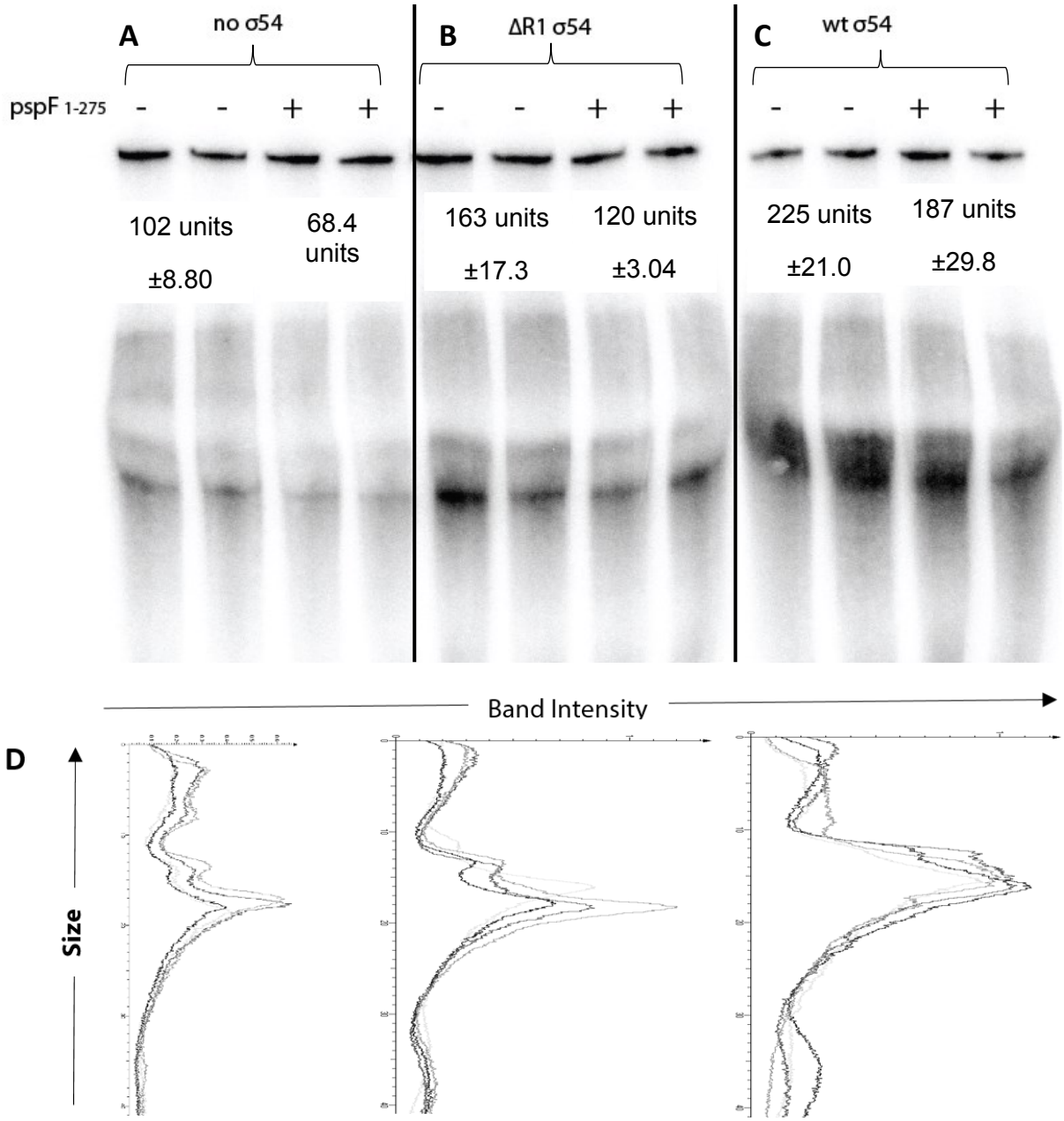


Fig 3.3.6 Global bypass transcription *in vitro* using chromosomal DNA.

Total *E. coli* genomic chromosomal DNA was prepared (DNeasy Blood & Tissue Kit, Qiagen), and used as a template for full length *in vitro* transcription assays. The experiment was done under 3 different σ^{54} conditions: core RNAP only (A), bypass mutant σ^{54} with RNAP (B) and wild type σ^{54} with RNAP (C). This was done both with and without σ^{54} activator $PspF_{1\text{-}275}$ (a variant lacking the DNA binding domain). Total transcription for each of the 3 classes was converted into arbitrary units, and the standard deviations within samples determined. Transcription results were converted into band intensity profiles (D).

3.4 Summary

In this Chapter it was shown that SABT does not occur *in vivo* under standard growth conditions (LB media at 37°C) at the *pspA* or gln_{Ap2} promoters (3.2.3 and 3.3.1), confirming the initial hypothesis that SABT is not readily observed *in vivo*.

Low salt concentration is known to increase SABT at the gln_{Ap2} promoter *in vitro* (Wang, 1997). However, results were unable to show that low salt or L-glutamine concentration increase SABT *in vivo* at the gln_{Ap2} promoter (3.3.2 and 3.3.3). The data was inconclusive however, in part due to adverse effects of these abiotic conditions on cell growth.

Next it was shown that increased temperature (42°C) and minimal Gutnick media allow SABT to be observed at the gln_{Ap2} promoter, but not at the *pspA* promoter (3.3.4 and 3.3.5). This is in line with *in vitro* studies showing that SABT at the gln_{Ap2} promoter has a positive correlation with temperature (Wang, 1997).

This raised the question as to whether potential global regulators could be influencing SABT *in vivo* by altering DNA topology. However, full length *in vitro* transcription assays using purified chromosomal DNA were unable to confirm that this is the case. *In vitro* assays did show that SABT can be observed *in vitro* using purified chromosomal DNA as a template for transcription, in line with the hypothesis that barriers to SABT are present *in vivo* but are alleviated in the minimal *in vitro* setting.

Chapter 4: Genome-wide screens for genes regulating SABT *in vivo*

4.1 Introduction and objectives

4.1.1 Recap and objectives

In Chapter 3 it was examined whether σ^{54} activator bypass transcription (SABT) could be observed *in vivo,* under several abiotic conditions known to favour SABT *in vitro* (Wang, 1997). In Fig 3.3.4, increased temperature (42°C) gave rise to SABT *in vivo* at the gln_{Ap2} promoter.

This is consistent with *in vitro* studies, where elevated temperature is thought to partially inactivate some repressive components for SABT (Wang, 1997). These repressive components are believed to include the inhibitory features of σ^{54} and the DNA sequence it binds to. Unlike *in vitro* experiments, SABT is not readily observed *in vivo* under standard growth conditions. Therefore it is possible further inhibitory proteins may prevent SABT from being readily observed *in vivo*. If so, which proteins could they be and why do they regulate SABT so tightly at 37C° *in vivo*?

Anti-sigma factors have been described, such as Rsd which inhibits the activity of σ^{70}-dependent transcription (Hofmann, 2011). FlgM is another anti-sigma factor responsible for inhibiting σ^{28}, which is involved in flagellar biosynthesis (Daughdrill, 1997; Hughes, 1998). DnaK is a general regulator of the heat shock response, which in *E. coli* inhibits the heat shock sigma factor σ^{32} (Hughes, 1998). An unknown anti-sigma factor may stop σ^{54} transcription *in vivo*, when it occurs as SABT *in vitro*. So far however, no such anti-sigma factor has been described for σ^{54}.

Reducing the expression of such putative inhibitory proteins may allow SABT to be readily observed *in vivo*. Alternatively, a gene product may be limiting for SABT *in vivo*. Examples may include the RNA chaperone Hfq, which protects against cleavage from RNAse E (Večerek, 2008). In this case overexpression may be required in order to observe SABT *in vivo*.

Identifying such putative regulatory proteins would provide key insights into the mechanisms regulating σ^{54}-dependent transcription, and potentially why σ^{54}-dependent transcription has not evolved towards activator independence. Chapter 4 focuses on searching and identifying such putative regulatory proteins, by screening large libraries of mutants for SABT *in vivo*.

4.1.2 Introduction to KEIO collection

The KEIO collection is a library of in-frame single-gene deletions of 3985 non-essential genes in *E. coli* (Baba, 2006). The library was constructed by excising genes using targeted recombination, via the λ red recombinase system. Open reading frames were substituted with selectable kanamycin cassettes flanked by flippase recognition target (FRT) sites. The FRT sites allow for subsequent removal of the Kan cassette by inserting a plasmid containing the FRT recombinase (eg. pCP20), leaving a 102bp in-frame scar (Fig 4.1).

Out of 4288 genes targeted only 303 were unsuccessfully disrupted. Most of these are essential, encoding a variety of products including: tRNAs, ribosomal RNAs, metabolic enzymes, and DNA/RNA polymerase subunits (Baba, 2006). Transferring this deletion library into the σ^{54} activator bypass reporter strain (SABRS), could indicate whether any of these single non-essential genes in *E. coli* repress σ^{54} activator bypass transcription (SABT) *in vivo*.

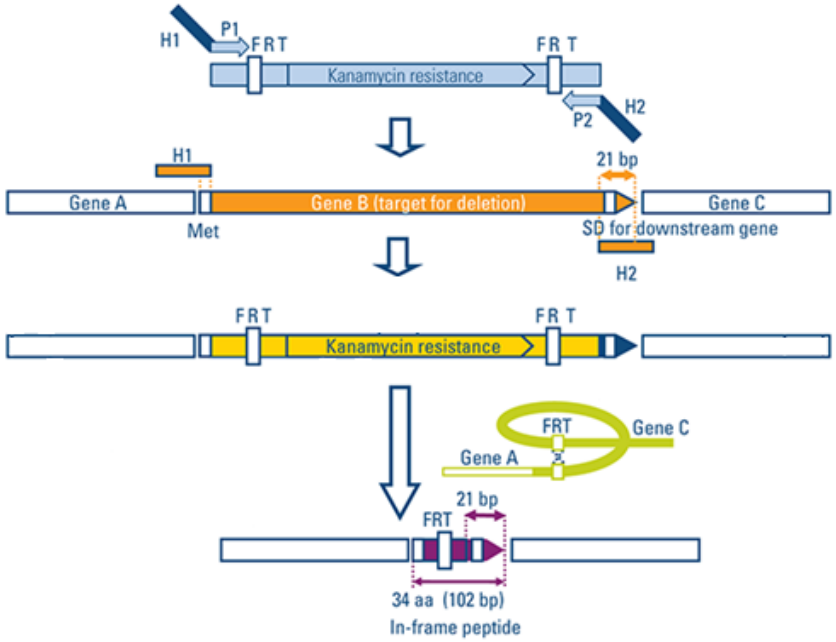


Fig 4.1.1: Construction of KEIO single gene deletion mutants.

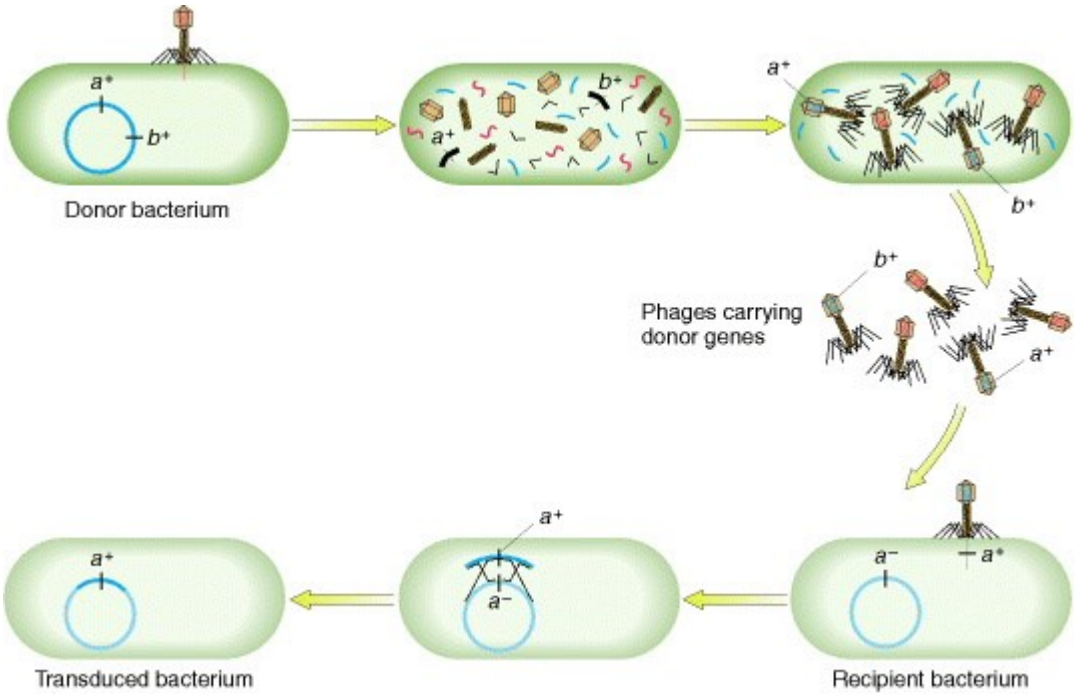
Gene knockout primers have 20-nt 3′ ends for priming upstream (P1) and downstream (P2) of the FRT sites flanking the kanamycin resistance gene in pKD13, and 50-nt 5′ ends homologous to upstream (H1) and downstream (H2) chromosomal sequences for targeting the gene deletion. H1 includes the gene B (target) initiation codon. H2 includes codons for the six C-terminal residues, the stop codon, and 29-nt downstream (adapted from Baba, 2006).

4.1.3 Introduction to generalised P1$_{vir}$ phage transduction

Transduction is the movement of genes from a bacterial donor strain to a bacterial recipient through a phage or vector. In 1965 it was discovered that when P1 phage lysed a donor cell, the bacterial chromosome is fragmented. The newly formed phage particles could accidentally incorporate a sequence of bacterial DNA into the phage head instead of phage DNA (Ikeda, 1965). This discovery is the origin of phage transduction and its use as a molecular biology technique to transfer genetic material from one bacterial strain to another (Fig 4.2).

There are two types of phage transduction: generalised (for example with P1 and P22) and specialized (for example with phage λ). During specialised transduction phages can only contain restricted parts of the chromosome, during generalised transduction any part of the chromosome can be incorporated.

Fig



4.1.2 Schematic of phage transduction

Upon entering a donor cell, phages replicate and produces new bacteriophages. 1 in 10 000 phages will contain chromosomal DNA. After cell lysis, phages carrying this chromosomal DNA can enter new recipient cells, and transfer DNA from the donor cell into the recipient cell chromosome (adapted from An Introduction to Genetic Analysis, 2000).

P1 phage is commonly used for generalised transduction of chromosomal DNA up to 90kb, and uses the Cre/Lox recombinase system. $P1_{vir}$ is a modified P1 phage, containing a mutation in the repressor binding site (which represses lytic genes in the lysogenic state). Therefore $P1_{vir}$ phage transductions ensures replication and cell lysis.

However, using $P1_{vir}$ phage to transfer gene deletions does impose some limitations. The $P1_{vir}$ receptor (on host cell) is a terminal glucose of an outer membrane lipopolysaccharide. Mutants lacking this receptor (eg. GalU mutants) are resistant to $P1_{vir}$ infection and cannot be transduced via P1 (Frankin, 1969). Additionally, there is no way simple way to prevent the transfer of genetic material directly adjacent to the gene deletion of interest (up to 90kbp) from the donor strain.

4.1.4 Identifying KEIO mutants using inverse PCR

P1$_{vir}$ phage transduction allows for the transfer of KEIO library mutants into a SABT reporter strain, which can subsequently be screened for SABT. However, positive hits include thousands of possible mutants without any discernable marker other than the Kan cassette. Below the inverse PCR method is described (Fig 4.1.3), used to determine gene deletions of interest.

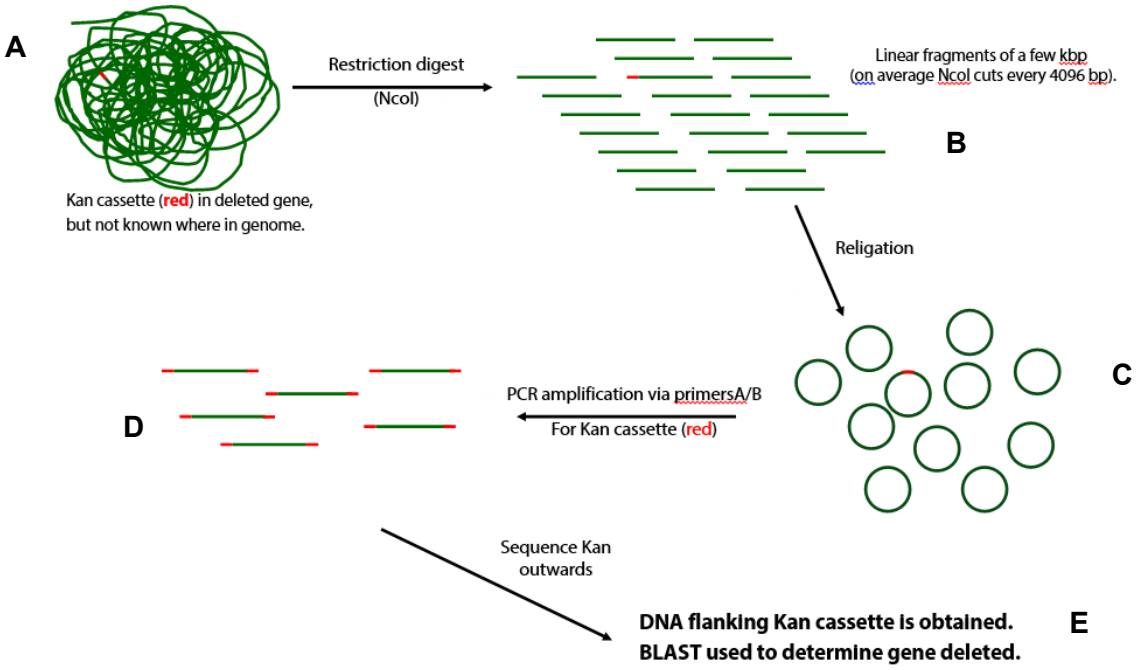


Fig 4.1.3 Schematic describing inverse PCR to identify KEIO mutants

The target chromosome (**A**) is digested into ±4kbp fragments using NcoI (**B**). These fragments are religated to form circular DNA (**C**), which is amplified using Kan-specific primers, to amplify only the DNA containing the target deletion (and Kan cassette). The PCR product is sequenced and located using BLAST.

4.2 Results and Discussion

The aim of this Chapter was to search for putative regulatory proteins affecting SABT *in vivo*. Here single gene deletion libraries such as the KEIO collection (or overexpression libraries) were transferred into the *pspA*-SABRS from Chapter 3 using $P1_{vir}$ phage transduction. The *pspA*-SABRS contains the $\sigma^{54}_{\Delta R1}$ activator bypass mutant that gives rise to SABT *in vitro,* however SABT is not readily observed *in vivo* at 37°C in minimal media or LB.

The resulting deletion/overexpression mutants in the *pspA*-SABRS were subsequently screened for SABT *in vivo* by growing them on X-Gal and MacConkey plates (sensing ΦP_{pspA}-*lacZ* reporter activity). During the screening process mutants were also grown under various conditions thought to facilitate SABT, including: 0.001% arabinose, 42°C and minimal media with lactose.

For mutants scoring positive with X-Gal and MacConkey, LacZ activity was quantified using β-galactosidase assays. Where relevant, the mutants of interest were subsequently also identified. This screening process is summarised below in (Fig 4.2.1).

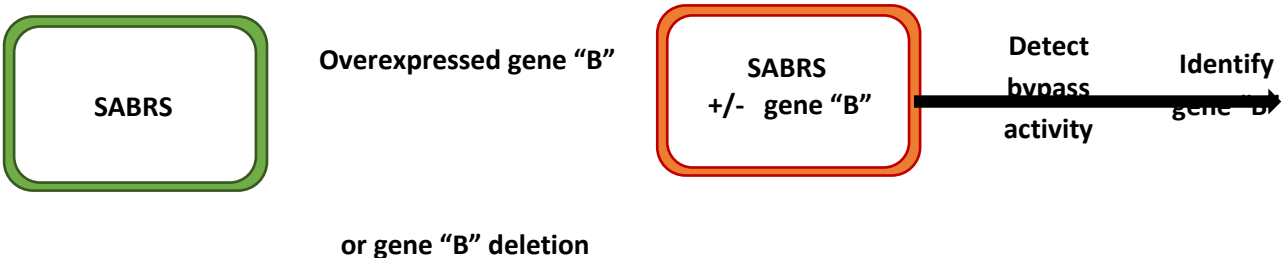


Fig 4.2.1: Screening process to identify putative regulatory proteins affecting SABT *in vivo*
The *pspA*-SABRS is transduced with phage lysate from single gene deletion libraries (eg. The pooled KEIO collection) or overexpression libraries (eg. *Salmonella* Sau3A library). Resulting transductants colonies are screened for SABT (through ΦP_{pspA}-*lacZ*) on X-Gal or MacConkey plates with kanamycin. Mutants with the highest apparent activity on the screening plates are quantified for SABT using β-galactosidase assays. Where β-galactosidase activity mutants is thought to be SABT, the mutants are identified by the inverse PCR method (described in 4.1.3).

4.2.1 KEIO deletions and ΦPpspA-LacZ activity in the *pspA*-SABRS

The KEIO collection contains 3985 single gene deletions, including known *E. coli* anti-sigma factors like Rsd, FlgM and DnK (Hughes, 1998). To ensure adequate coverage of the entire library, 5-fold coverage was used (ie. ≈20 000 colonies were screened). Transductants were screened for on LB agar with X-Gal or MacConkey. The mutants with the highest apparent ΦP_{pspA}-*lacZ* activity were selected, and their β-galactosidase activity was quantified below (Fig. 4.2.2).

Several of those mutants show significantly higher activity than the *pspA*-SABRS negative control strain (Fig. 4.2.2). Compared to *pspA*-SABRS, mutant activity ranged from 150% (Δ3) to 1250% (Δ*hldE*), the latter mutant being 60% of the *pspA*-WTRS positive control (labelled wt). A number of deletion mutants resulted in elevated β-galactosidase activity *in vivo*, potentially through SABT.

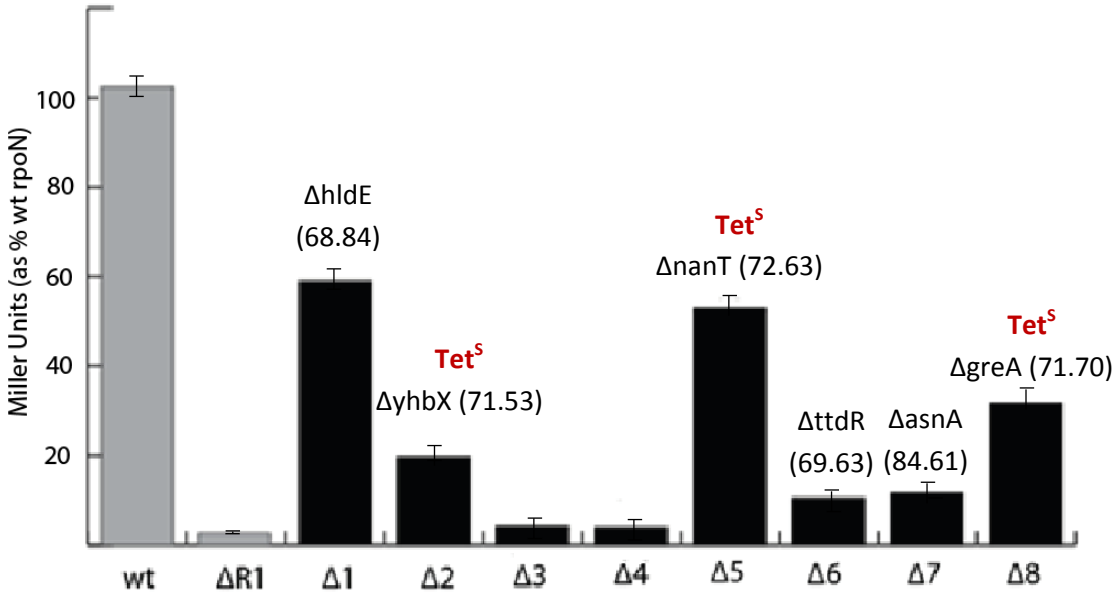


Fig 4.2.2: β-galactosidase activity of 8 KEIO mutants displaying putative SABT *in vivo*.

pspA-SABRS was transduced with $P1_{vir}$ lysate from the KEIO collection library. Mutants of interest were assayed for Φ*PpspA-lacZ* activity using liquid culture β-galactosidase assays. Four mutants (Δ1, Δ2, Δ5 and Δ8) display higher activity than the background *pspA*-SABRS (labelled ΔR1). 3 of those mutants (Δ2, Δ5 and Δ8) lost their chromosomal resistance to tetracycline (*rpoN208*::Tn*10*) during

the transductions process. They had gained a wild type *rpoN* locus (section 4.2.3, due to close linkage of the KO and the *rpoN* locus). Inverse PCR was also used to identify the Δ1, Δ2, Δ5, Δ6, Δ7 and Δ8 mutants as Δ*hldE*, Δ*yhbX*, Δ*nanT*, Δ*ttdR*, Δ*asnA* and Δ*greA* respectively. Their chromosomal locations are shown in brackets (% centisome).

4.2.2 Co-transduction of wild type chromosomal *rpoN*

3 of the selected mutants had lost their resistance to tetracycline during the transduction process, so one concern was that the *rpoN*::Tn*10* had been lost. It was possible that the native chromosomal *rpoN* from the donor strain (KEIO library) had been co-transduced back into the reporter strain (in addition to the Kan^R marked single gene deletion). If this was the case it could result in a false positive result as ΦP_{pspA}-*lacZ* activity might be generated by wild type σ^{54} rather than the $\sigma^{54}{}_{\Delta R1}$.

$P1_{vir}$ can package about 90kbp of DNA, therefore any co-transduced genes would have to be within 90kbp of *rpoN* on the *E. coli* chromosome. The locus of *rpoN* on the chromosome is 72.06 and 90kbp around that area is about 2 minutes. All 3 of the mutants with high activity and tetracycline sensitivity (Δ*yhbX*, Δ*nanT* and Δ*greA*) map back to within less than a third of that distance (Fig. 4.2.2). Moreover, colony PCR of the *rpoN* locus showed that the native chromosomal *rpoN* had been co-transduced back into the reporter strain (Fig. 4.2.3). Therefore Δ*yhbX*, Δ*nanT* and Δ*greA* are possibly false positive in the screen, and did not arise through $\sigma^{54}{}_{\Delta R1}$-dependent transcription of ΦP_{pspA}-*lacZ*.

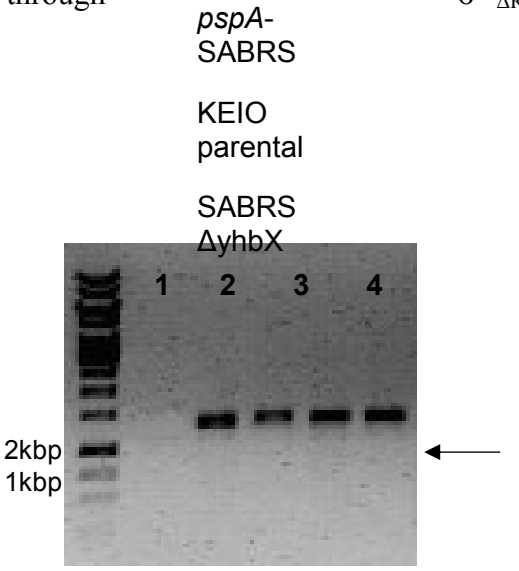


Fig 4.2.3: Colony PCR of *rpoN* locus in deletion mutants.

rpoN-specific primers were used to amplify the native *rpoN* locus, generating a 1.5 kbp fragment where it was intact (arrowed). A wild type *rpoN* locus was not observed in the *pspA*-SABRS (**1**), suggesting the *rpoN*208::Tn*10* was present. Native chromosomal *rpoN* loci were present in the Δ*yhbX*, Δ*nanT* and Δ*greA* mutants (**3, 4,** and **5**); and the KEIO background strain (**2**). This indicates those 3 mutants had lost the *rpoN*208::Tn*10*.

4.2.3 Screening KEIO library mutants for SABT at increased arabinose concentration

The $\sigma^{54}_{\Delta R1}$ bypass mutant is expressed from the pBAD plasmid, which is inducible with arabinose. Earlier experiments showed that ΦP_{pspA}-*lacZ* activity was highest in the WTRS between 0% and 0.02% arabinose (Fig 3.2.2). 5000 more KEIO library mutants were screened at 0.001% arabinose, in an attempt to increase a potential SABT signal and identify more putative inhibitory genes.

At 0.001% arabinose, two deletion mutants were found to have activity significantly higher than the background (Fig 4.2.4, BS3 and BS5). The BS5 mutant had lost its *rpoN208*::Tn*10* marker and gained the wild type *rpoN* locus, as in section 4.2.2 (not shown). In the BS3 mutant all chromosomal markers were retained (ΦP_{pspA}-*lacZ*, *rpoN*::Tn*10* and pBAD18-σ^{54}_{DR1}).

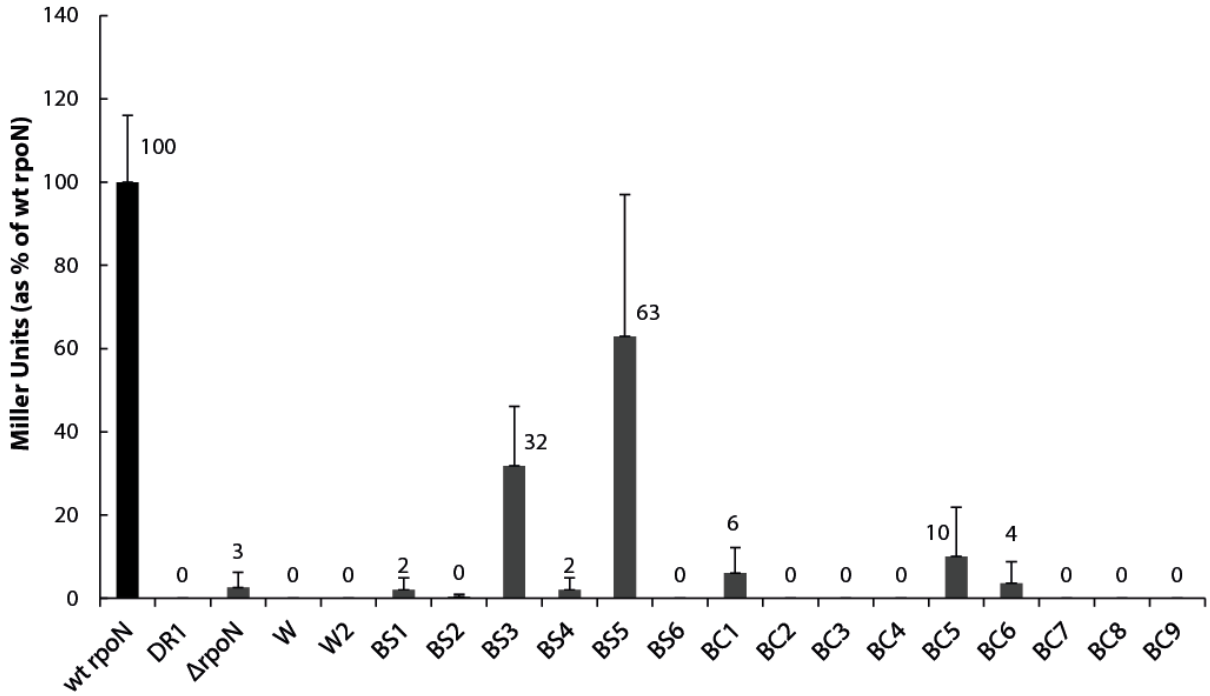


Fig 4.2.4: β-galactosidase activity of KEIO mutants screened at 0.001% arabinose.
pspA-SABRS was transduced with KEIO library lysate, and transductants screened for ΦPpspA-*lacZ* activity on plates with 0.001% arabinose and X-Gal/MacConkey. Mutant activity was assayed from 3 groups: white colonies (**W**, **W2**), blue small colonies (**BS1-BS6**), and colonies with a blue center (**BC1-BC6**).

If the β-galactosidase activity observed in Fig 4.2.4 was due to $\sigma^{54}{}_{\Delta R1}$-dependent transcription of Φ*PspA-lacZ*, this activity should be responsive to arabinose concentration (ie. $rpoN_{\Delta R1}$ expression). To verify whether the BS3 mutant did indeed display SABT *in vivo*, an arabinose induction curve was done and β-galactosidase was activity assayed (Fig. 4.2.5).

As seen in Fig 4.2.5., the BS3 mutant activity at 0.001% arabinose was only 3% compared to wt σ^{54}, but 32% previously (Fig 4.2.4). The high levels of activity observed earlier were not reproducible (Fig 4.2.5). Since the error margin initially was quite substantial (Fig 4.2.5), this former result was considered to be false positive and within the error range of the experiment.

Additionally, over the 0% to 0.001% arabinose range β-galactosidase activity for the BS3 mutant did not increase much, whereas the wt *rpoN* control strain showed a 2.4x increase over the same range (Fig. 4.2.5). Therefore the BS3 deletion mutant was not presumed to result in SABT *in vivo*, and the gene deletion was not characterised.

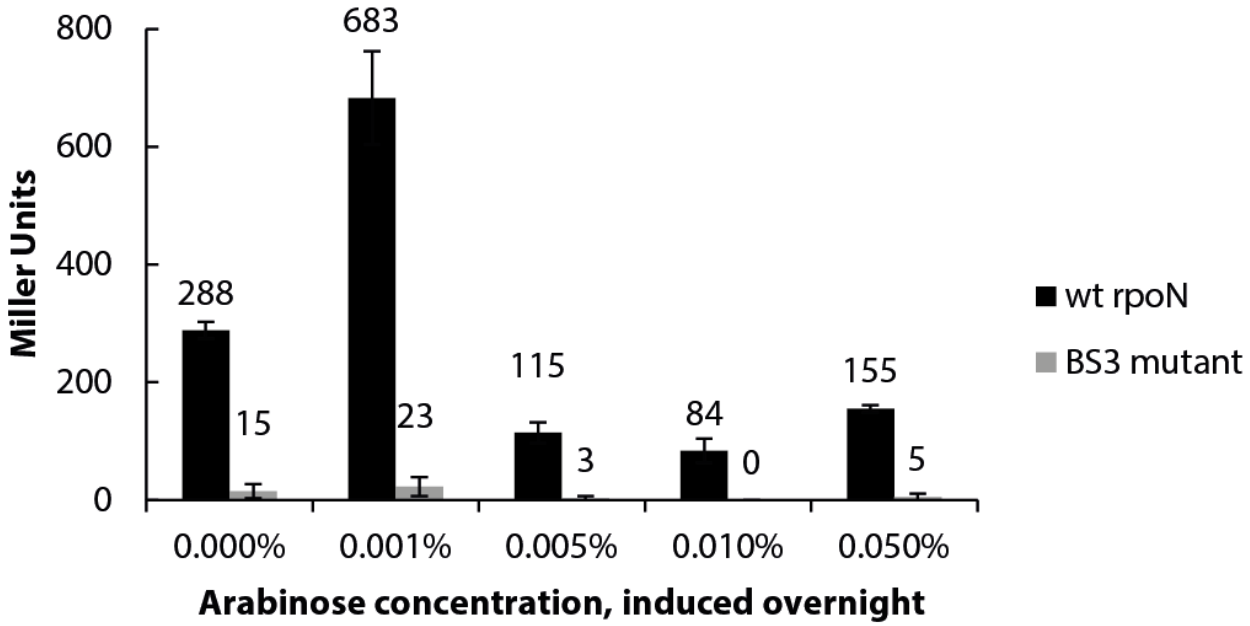


Fig 4.2.5 β-galactosidase activity of BS3 mutant an arabinose concentration.

Reporter strain cells with wt *rpoN* (black) and $rpoN_{\Delta R1}$ (grey) were grown overnight an induced with arabinose at various concentrations. The *pspA*-WTRS (black) increases in ΦP*pspA-lacZ* activity with increased arabinose concentration, peaking at 0.001%. The BS3 mutant (grey) has significantly lower activity and does not appear to be arabinose inducible.

4.2.4 Screening KEIO mutants for ΦPpspA-*lacZ* activity at increased temperature

KEIO mutants were screened for SABT earlier under standard conditions (4.2.1) and at 0.001% arabinose (4.2.3). Bypass transcription had been observed previously at the gln_{Ap2} promoter at 42°C *in vivo* (3.3.4). In an attempt to screen for new putative bypass mutants under conditions potentially favouring SABT, 5000 new KEIO library mutants were also screened at 42°C.

8 mutants scored positive in the screen on XGal/MacConkey, they were selected and their ΦP*pspA-lacZ* expression quantified. However, as seen in Fig 4.2.6**.**, no single gene deletions resulted in activity considerably higher than the *pspA*-SABRS (labelled ΔR1) negative control.

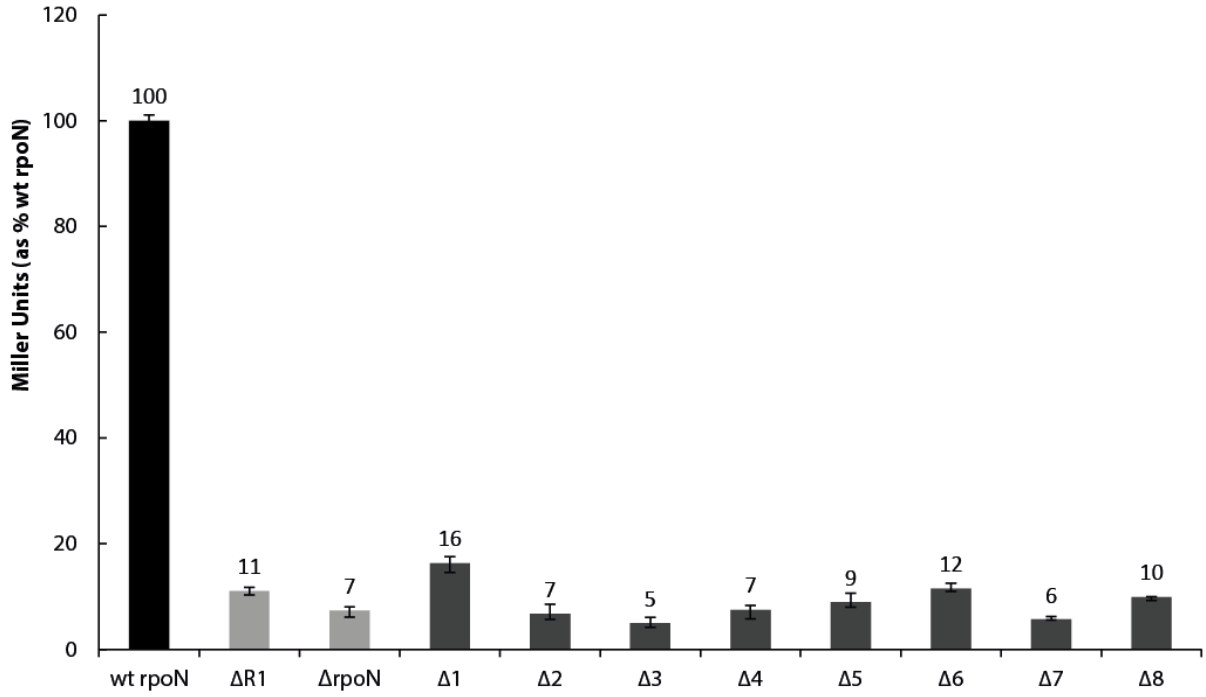


Fig 4.2.6: β-galactosidase activity of KEIO mutants screened at 42°C.

pspA-SABRS was transduced with the KEIO library lysate, and transductants screened for ΦP_{pspA}-*lacZ* activity on plates with X-Gal/MacConkey at 42°C. Mutant activity quantified with β-galactosidase assays for 8 colonies with the highest apparent signal in the screen. None of the mutants show activity significantly higher than the parental reporter strain (labelled ΔR1) and none of the deletions result in significant increase in Φ*pspA-lacZ* signal under these conditions.

4.2.5 Small peptide/RNA deletion library and ΦPpspA-LacZ activity in the SABRS

To complement the KEIO library screen, a short peptide and small regulatory RNA library was also screened. These deletions were generated in a similar fashion to the KEIO deletions and represent 80 small regulatory RNAs (sRNAs) and 60 short proteins (Hobbs, 2010). Most of these have unknown functions and a significant proportion were not part of the KEIO collection.

As seen in Fig 4.2.7, none of the selected mutants from the peptide/RNA library screen resulted in significantly increased SABT. Therefore none of these 130 single gene deletions from the small peptide/RNA library seem to alleviate the putative barriers inhibiting SABT *in vivo*.

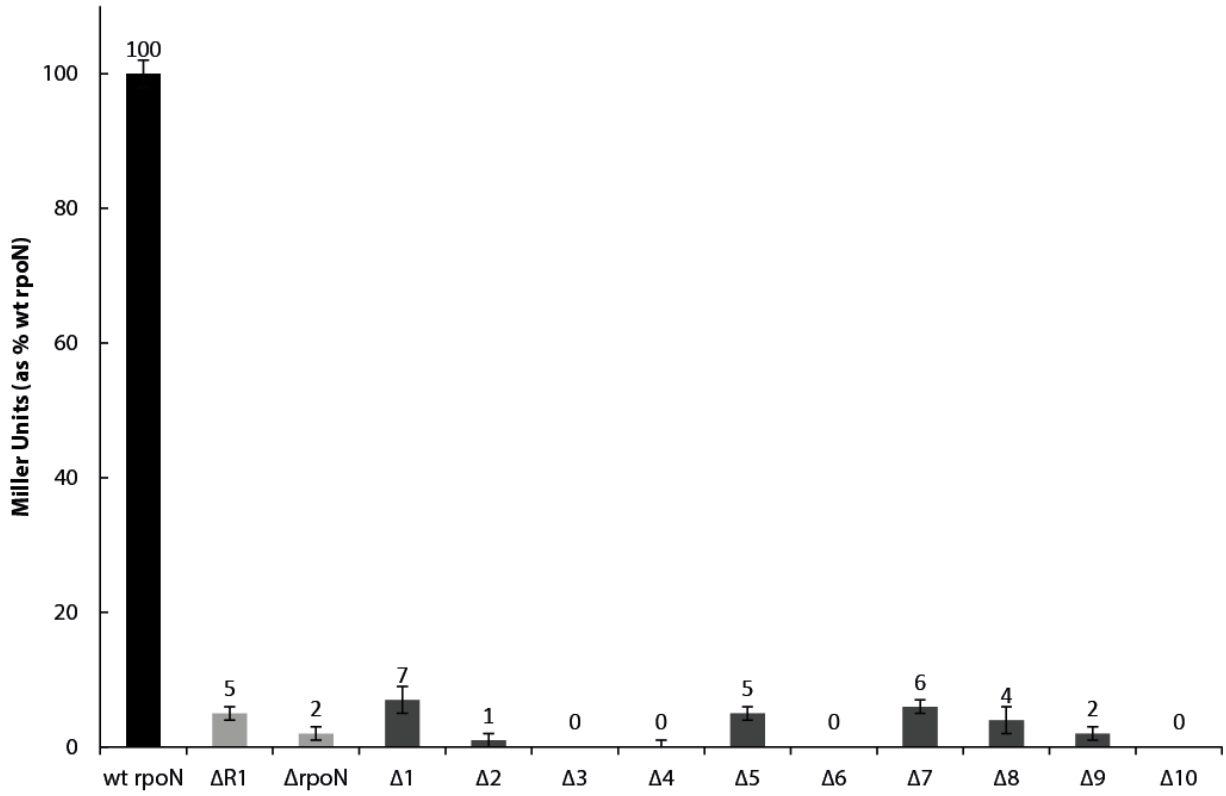


Fig 4.2.7: β-galactosidase activity of small peptide/RNA library mutants.

The SABRS was transduced with small peptide/RNA library lysate, and transductants screened for ΦP*pspA-lacZ* activity on plates with X-Gal or MacConkey. Mutant activity was quantified using β-galactosidase assays for 10 colonies with the highest apparent signal. However, none of the mutants show activity significantly higher than the *pspA*-SABRS negative control (labelled ΔR1).

4.2.6 *Salmonella Typhimurium LT2* overexpression clones and SABT

In 4.2.1-4.2.5, single gene deletion libraries were screened for mutants that could give rise to SABT *in vivo*. Here a complementary approach was used to find proteins that may influence SABT *in vivo*: an overexpression library was screened for SABT. While such overexpressed genes may not be present in the minimal *in vitro* conditions where bypass transcription is observed, overexpression clones may alter intracellular conditions or regulate transcription in a way that bypass transcription could be observed *in vivo*.

The ASKA library is a set of *E. coli* ORF clones but unfortunately was not viable to use. The pCA24N vector with the overexpression clones is Cam^R (Kitagawa et al, 2005), as is the pBAD plasmid with $\sigma^{54}_{\Delta R1}$ in the SABRS. Instead a similar library was used from *Salmonella Typhimurium LT2*, an intracellular pathogen very similar to *E. coli* that shares a large amount

of genetic material labelled the 'core genome' (Meysman, 2013). The *Salmonella Typhimurium LT2* DNA library was generated by partial digestion of chromosomal DNA with Sau3A, then cloned into the BamHI site of pBR328 (Bergman et al, 2014).

As seen in Fig 4.2.8, none of the overexpression clones examined result in activity considerably higher than the *pspA*-SABRS negative control. As with most single gene deletion screens, the overexpression library did not yield any promising candidate genes affecting bypass transcription *in vivo*.

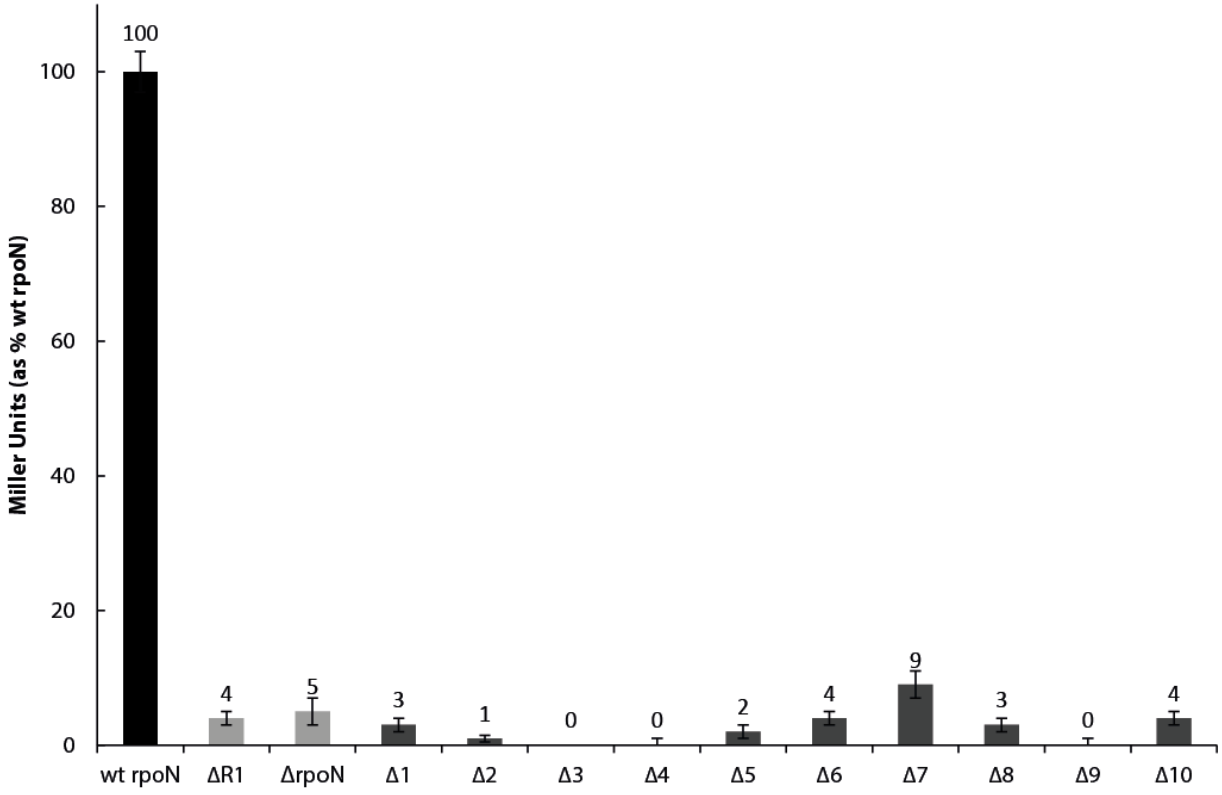


Fig 4.2.8: SABT of cells expressing Sau3A digested DNA from *Salmonella Typhimurium LT2*. SABRS was transformed with a pBR328 library expressing Sau3A fragments of *Salmonella Typhimurium* chromosome. Transformants were screened for ΦP*pspA-lacZ* activity on plates with X-Gal/MacConkey. Mutant activity was quantified using β-galactosidase assays for 10 colonies with the highest apparent signal. None of the mutants show activity significantly higher than the parental reporter strain (labelled ΔR1).

4.2.7 Selection of activator bypass mutants on minimal lactose media

Previous experiments involved screening for LacZ activity to identify mutants of interest (4.2.1-4.2.6). In this section *pspA*-SABRS mutants were selected if β-galactosidase levels (transcribed from the Φ*pspA-lacZ*) were sufficiently high to support growth on lactose, selecting for high levels of SABT rather than screening for low levels of SABT as done earlier.

Since the *pspA*-SABRS parental is *lacZ*⁻, background LacZ production is very low (Fig 3.2.4) and insufficient to support growth on minimal Gutnick media with lactose (Fig 4.2.9, 4). However the *pspA*-WTRS (wild type σ^{54}) does grow on minimal Gutnick media with lactose, indicating that σ^{54}-dependent transcription of ΦP_{pspA}-*lacZ* is both necessary and sufficient for a Lac^+ phenotype (Fig 4.2.9, 7).

This is a more strict approach than screening mutants and should only result in a Lac^+ phenotype if the ΦP_{pspA}-*lacZ* is transcribed sufficiently to support growth on lactose. This approach could also avoid potential shortcomings of the screening process such as bypass mutants possibly growing slower than other mutants on the screening plates. For instance, a Δ*ttdR* strain (identified as a putative bypass candidate in 4.2.1) is known to grow 50% slower than a wild type strain (Teramoto, 2010).

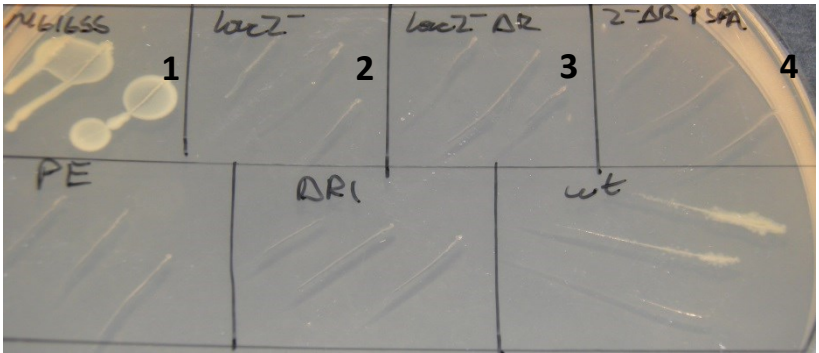


Fig 4.2.9: Reporter strain growth phenotypes on minimal Gutnick media with lactose.

Cells were grown overnight on minimal Gutnick media plates, supplemented with 5mM L-Glutamine and 1% (w/v) lactose. Strains include a MG1655 control (1), LacZ$^-$ (2), LacZ$^-$ *rpoN*208::Tn*10* (3) and LacZ$^-$ *rpoN*208::Tn*10* ΦP_{pspA}-*lacZ* (4). *pspA*-DRRS (5), *pspA*-SABRS (6) or *pspA*-WTRS (7).

Growth on minimal Gutnick media with lactose appears to require wild chromosomal *lacZ* (1), or functional Φ*pspA-lacZ* (7). Growth under these conditions is very poor or non-existent (2, 3, 4, 5 and 6) in the absence of chromosomal *lacZ* or transcribed Φ*pspA-lacZ*.

5 mutants were generated by selecting for *pspA*-SABRS KEIO mutants with a Lac$^+$ phenotype. The Lac$^+$ phenotype of these mutants was reproducible in 5ml overnight cultures with minimal media and 1% lactose (Fig 4.2.10, ΔA-ΔE). Moreover, overnight cultures confirmed that the *pspA*-WTRS was Lac$^+$, while the *pspA*-DRRS and *pspA*-SABRS were Lac$^-$ (Fig 4.2.10). This is in line with the hypothesis that σ^{54}-dependent transcription of ΦP_{pspA}-*lacZ* is necessary for a Lac$^+$ phenotype.

Colony PCR confirmed that the genetic markers of these 5 mutants remained intact (not shown).

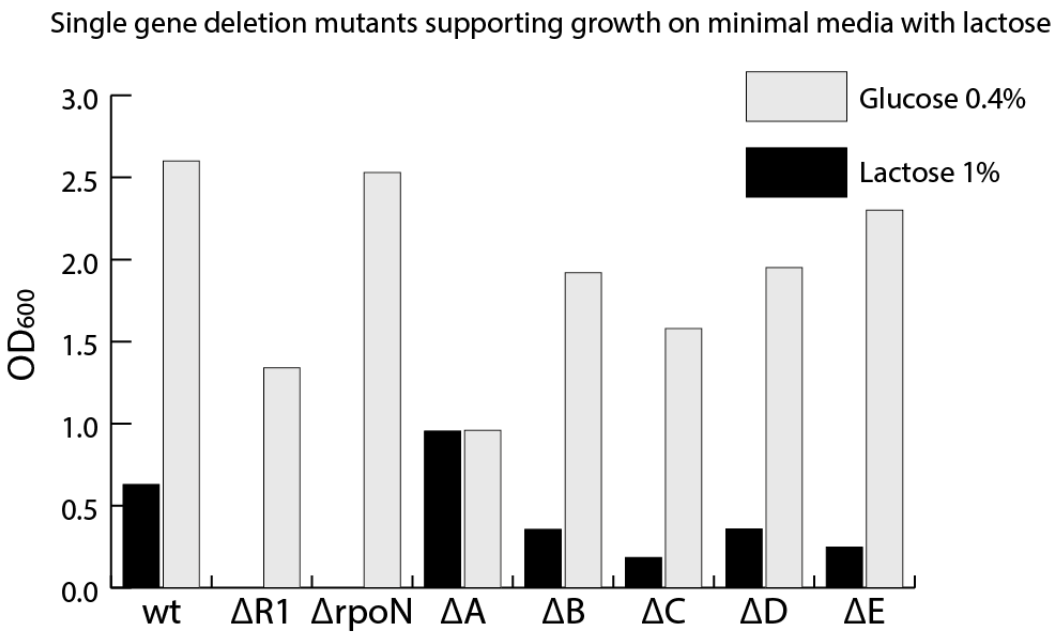


Fig 4.2.10: Reconstituted lactose growth phenotype in unidentified KEIO deletions of reporter strain.

Mutants (ΔA-ΔE) were generated by transducing the *pspA*-SABRS with KEIO library lysate. Transductants were grown overnight on minimal Gutnick media plates (with 1% lactose and 4mM L-glutamine) to select for a Lac^+ phenotype (ie. ΦPpspA-*lacZ* transcription). All resulting colonies were grown as 5ml overnight cultures in the same media with 1% lactose (black) or 0.4% glucose (grey), and OD_{600} measured.

To verify that the Lac^+ phenotype was mediated by $\sigma^{54}{}_{\Delta R1}$-dependent transcription ΦP_{pspA}-*lacZ*, β-galactosidase activity of these mutants was measured. However, little to no activity was detected in any of the mutants, regardless of whether they were grown on lactose or glucose (Fig 4.2.11). This is in contrast to the *pspA*-WTRS which was active, suggesting that none of the mutants developed a Lac^+ phenotype via transcriptional activation of ΦP_{pspA}-*lacZ*.

The mutants could be using alternative carbon sources such as alanine or proline, which has been described before in *E. coli* (Liu, 2005). Selective pressure has also been known to give rise to mutations allowing lactose metabolism in *Klebsiella* (Hall, 1979). Irrespective of the cause for this Lac^+ phenotype, the selected mutants did not give rise to SABT, and were not examined further.

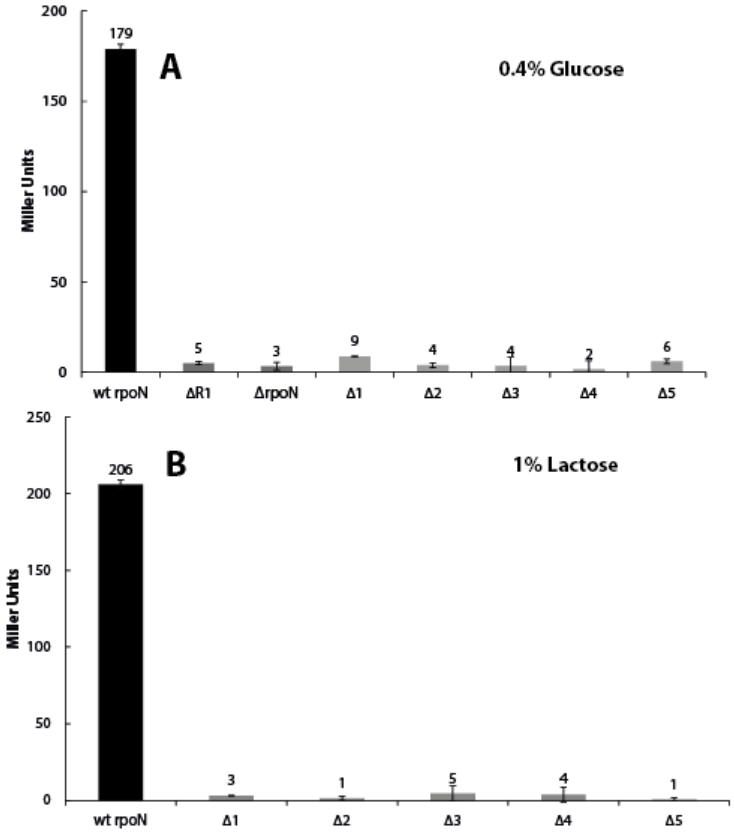


Fig 4.2.11: ΦPpspA-LacZ activity of Lac$^+$ mutants.

Lac$^+$ mutants were grown overnight on minimal media with 4mM L-glutamine and 0.4% glucose (A) or 1% lactose (B). β-galactosidase assays were done to measure ΦPpspA-*lacZ* activity. While earlier results suggesting that a Lac$^+$ phenotype was mediated through ΦPpspA- *lacZ* (Fig. 4.2.9), no significant activity was observed in the mutants regardless of whether they were grown on glucose (A) or lactose (B).

4.3 Discussion

4.3.1 Function of genes involved in putative repression of SABT *in vivo*

Genome wide screening and selection for bypass mutants (4.2.1-4.2.7) yielded 6 single gene deletions that may give rise to SABT *in vivo* (Δ*hldE*, Δ*yhbX*, Δ*nanT*, Δ*ttdR*, Δ*asnA* and Δ*greA*). All of these mutants were generated in the initial KEIO library screen (Fig 4.2.2).

In 3 of those deletion mutants (Δ*yhbX*, Δ*nanT* and Δ*greA*) the wild type *rpoN* locus appears to have been co-transduced back into these mutants instead of their retaining the starting *rpoN*208::Tn*10* (Fig 4.2.3), potentially accounting for the observed activity.

In the remaining 3 single gene deletion mutants (Δ*hldE*, Δ*ttdR*, and Δ*asnA*) the *rpoN*208::Tn*10* was retained (not shown). This suggests their ΦP_{pspA}-*lacZ* activity is $\sigma^{54}{}_{\Delta R1}$-dependent transcription, and therefore SABT. If SABT is observed in these deletion mutants, their function could indicate how SABT is inhibited when those genes are expressed.

The function of all 6 genes are summarised below in Table 4.2.1.

Gene	Function
hldE	Heptose 7-P kinase/heptose 1-P adenyltransferase; LPS core prescursor synthesis
yhbX*	Putative EptAB family phosphoethanolamine transferase, inner membrane
nanT*	Sialic acid transporter
ttdR	Transcriptional activator of ttdABT, tartrate-inducible, anaerobiosis nucleoid protein
asnA	Asparagine synthase A; aspartate-ammonia ligase
greA*	Transcript cleavage factor

Table 4.2.1. Function of putative repressive genes identified from screening KEIO library.

---

* These genes contain a wild type *rpoN* locus and are no longer considered bypass candidates (Fig 4.2.3).

4.3.2 Δ*asnA* as a candidate for putative bypass transcription

AsnA is an asparagine synthase and its transcription is σ^{70}-dependent. However, *asnA* is further regulated by the σ^{54}-dependent nitrogen assimilation factor (*nac*), via AsnC (Poggio, 2002). Nac acts as an adaptor between the σ^{54}-dependent and σ^{70}-dependent transcription and is activated by NtrC, the main regulatory protein under nitrogen-limiting conditions (Zimmer, 2000).

Nac is also thought to regulate the transcription of several proteins involved in membrane structure (*betA, pagP, mreB, yabI, ybjP, yohD*) and protein export (*ArgA, AroP, dppA, mdtC,*

oppA, secY, ydcS, ydgQ, *yhfM, yjcD*) (Frisch, 2010). Membrane stress or defects in protein exports are known to induce the psp operon (Adams, 2003).

It is possible that a deletion of a nac target gene (like *asnA*) could result in feedback on nac activity. This may have adverse effects on the transcription of other nac target genes, many of which are involved in membrane structure and protein export. This in turn may result in activation of the psp response and the ΦP_{pspA}-*lacZ* reporter (Adams, 2003), but not through PspF. The psp activator PspF needs (as a bEBP) $\sigma^{54}{}_{R1}$ in order to activate the *pspA* promoter, so any induction of the *pspA* promoter through the *asnA* deletion is PspF-independent and not SABT**.** MreB, one of the nac target genes, has already been linked directly to PspA function and may influence expression of *pspA* somehow (Jovanovic, 2014). While an *asnA* deletion might not result in SABT, it may increase σ^{54}-independent background transcription of ΦP_{pspA}-*lacZ,* and appear as SABT *in vivo*. σ^{54}-independent *pspA* transcription has not been observed in *E. coli,* but has been reported in organisms like *Yersinia enterocolitica* (Maxson and Darwin, 2006).

Based on *asnA*'s function as an asparagine synthase, the deletion was not assumed to cause SABT *in vivo*. The apparent SABT phenotype *in vivo* was assumed to be a false positive, likely due to an increase in some σ^{54}-independent background transcription of the Φ*pspA-lacZ* reporter. As such, this mutant was not examined further. Mapping the transcriptional start site could help in determining the likely activated promoter. In 4.3.3, a putative σ^{70}-dependent extended -10 promoter sequence was identified in the psp regulatory region (see 4.3.3 below).

4.3.3 Low levels of σ^{70}-dependent transcription of the *pspA* promoter in *E. coli*

Section 4.3.2 raised the possibility of low levels of σ^{54}-independent ΦP_{pspA}-*lacZ* transcription occurring in *E. coli*. While RNASeq data was not available for mutants displaying putative σ^{54}-independent *pspA* promoter transcription, data was available for the WTRS, SABRS and DRRS. Should σ^{70}-dependent *pspA* promoter transcription occur in some strains (eg. Δ*asnA* mutant), one might expect to find some evidence of this in the parental strains sent for RNA Sequencing.

Indeed, from RNAseq work there is some evidence of transcriptional activity upstream of the σ^{54}-dependent *pspA* transcription start site in the WTRS, SABRS and DRRS (Fig 4.3.3, A). Given that these transcripts are upstream of the σ^{54} promoter, it is very likely that they are driven by σ^{70}-like promoters. σ^{54}-independent *pspA* transcription has been reported previously in *Yersinia enterocolitica*, and suggested to be σ^{70}-dependent (Maxson and Darwin, 2006).

To investigate whether this would be a possibility in *E. coli* too, the ΦP_{pspA}-*lacZ* promoter region was examined for σ^{70} promoter elements. The ΦP_{pspA}-*lacZ* promoter region contains two sequences resembling the σ^{70} -10 (TATAAT) element (Fig. 4.3.3, B: green and blue), but no -35 (TTGACA) element was observed.

Interestingly, a -15 element (Fig 4.3.3, B: red) appears to be present upstream of one of these putative -10 promoter elements ($^{-15}$AG^{-14}). While such a -15 element typically has a TG sequence, the T has been reported to be degenerate and the G conserved (Djordjevic, 2011). -15 elements ($^{-15}$TG^{-14}) are found one base upstream of the ($^{-12}$TATAAT^{-7}) -10 promoter element. Promoters with -15 elements typically show fewer matches to the consensus –35 element, and contain short runs of T residues (Fig 4.3.3, underlined) in the spacer region (Mitchell, 2003). The −35 and −15 elements are directly related to σ^{70}-dsDNA interactions, and a -15 element is thought to be able to partially compensate for the lack of a -35 element (Djordjevic, 2011).

The transcription observed is about 8 bases downstream of the second σ^{70} -10 promoter sequence (blue), consistent with it driving its transcription. The σ^{70} -10 promoter sequence with an extended -10 promoter (red and green) could potentially also be partially active, as transcription appears to increase slightly 7 bases downstream of it (GCAA…). Combined, these indicate that σ^{70}-dependent transcription of the *pspA* promoter is plausible, although this data is by itself insufficient as evidence to prove that this is the case.

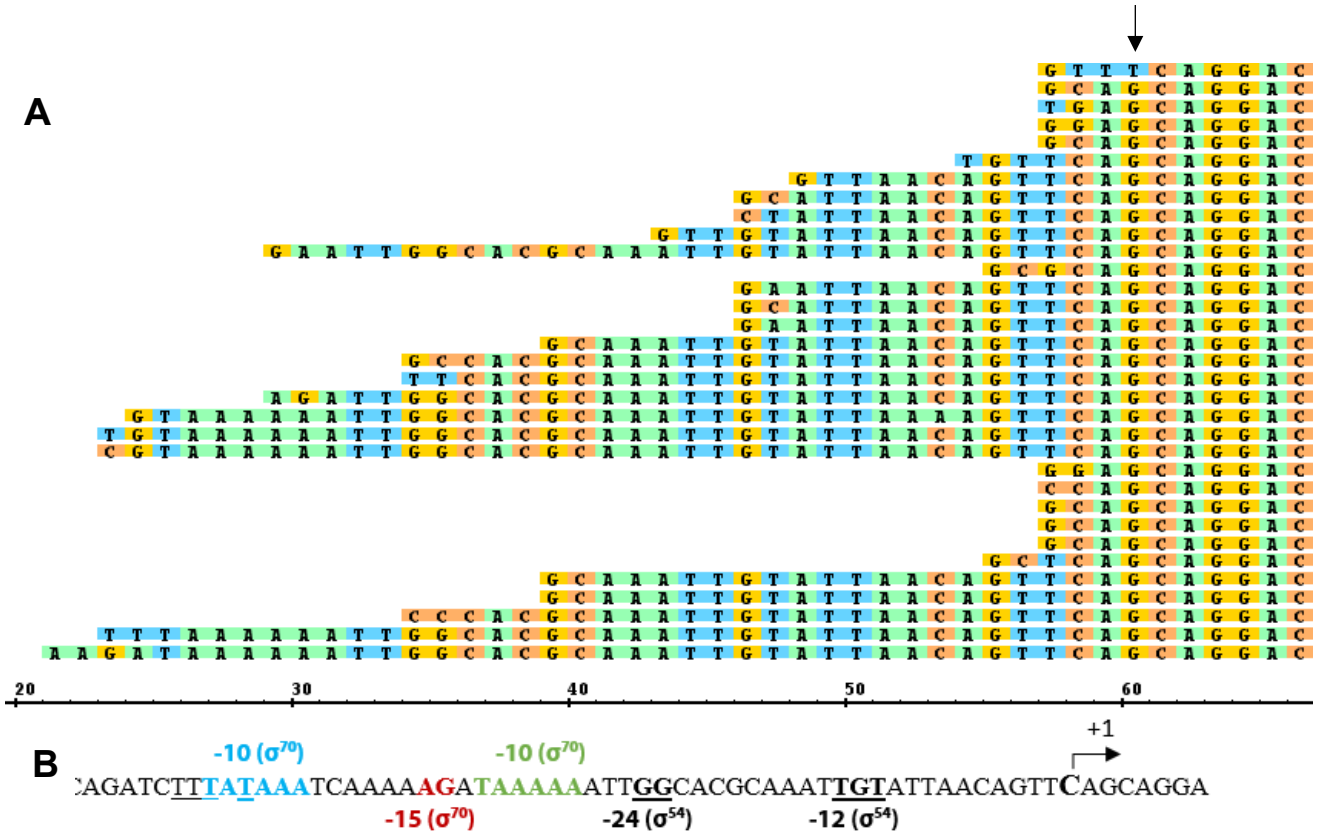


Fig. 4.3.3: Region upstream of the σ^{54}-dependent *pspA* transcription start site is transcribed, and contains putative σ^{70} promoter sequences.

(A) Detectable levels of transcriptional activity are observed upstream of the σ^{54}-dependent *pspA* transcription start site (arrowed), in the WTRS (wild type σ^{54}), SABRS ($\sigma^{54}{}_{\Delta R1}$), and DRRS (no σ^{54}). (B) The promoter region upstream of the *pspA* transcription start site (+1) is also annotated, including known σ^{54} -12 and -24 consensus sequences (bold underlined). Sequences resembling σ^{70}-dependent -10 promoter (TATAAT) are also annotated (green and blue), as well as a putative extended -10 promoter sequence (red). A short run of Ts is also observed in the spacer region, located near the 5' end of the sequence shown (underlined).

In *Yersinia*, no -35 or -15 element was detected in the *pspA* promoter region (Maxson and Darwin, 2006). However, the -10 element in *Yersinia* was shown to be sufficient for low levels of σ^{70}-dependent expression, which was abolished upon disruption of the -10 sequence (Maxson and Darwin, 2006). Therefore it is plausible that the putative σ^{70} -10 promoter elements in the *E. coli pspA* promoter (and a possible -15 element) could account for the apparent σ^{54}-independent transcription of ΦP_{pspA}-*lacZ* in the Δ*asnA* mutant (4.3.2). However this possibility was not directly tested since observing SABT was the main objective.

4.3.4 Δ*yhbX and* Δ*nanT* as candidates for putative bypass transcription

YhbX is a relatively unknown inner membrane protein, whereas NanT is a sialic acid transporter. It is plausible that either or both of these gene deletions result in membrane stress, inducing the psp operon either independently or dependently of PspF actions at the pspA promoter and so increase a σ^{54}-independent or dependent transcription of the ΦP_{pspA}-*lacZ* reporter.

However, the wild type *rpoN* loci had been shown to be co-transduced back into these two mutants (Fig 4.2.3). Therefore it is most likely that wild type *rpoN* transcription via PspF is the cause of the apparent SABT observed, as membrane stress conventionally signals transcriptional activation of the *pspA* promoter. Similarly to the Δ*asnA* mutant, the Δ*yhbX* and Δ*nanT* mutant were not explored further.

4.3.5 Δ*greA* as a candidate for putative bypass transcription

GreA is involved in the release of unstable RNAP complexes, such as when the polymerase complex has stalled or has incorporated incorrect nucleotides (Toulme, 2000; Erie, 1993). $\sigma^{54}{}_{\Delta R1}$RNAP intermediate open complexes are known to be unstable and are heparin-sensitive *in vitro* (Wang, 1997), prior to making a short transcript upon which they become stabilised. Combined these highlight GreA as a potential protein for inhibiting SABT *in vivo*.

However *greA* is located within 20kbp of *rpoN* on the *E. coli* chromosome, and colony PCR showed that the wild type *rpoN* locus has been reintroduced into the Δ*greA* mutant. Therefore, the *pspA*-SABRS is not a suitable negative control for Φ*PspA-lacZ* activity with GreA. As such, the increased activity in the Δ*greA* mutant is quite possibly due to the native chromosomal *rpoN* rather than the *greA* deletion.

In an attempt to examine GreA's role in SABT, the *ΔgreA788::kan* KEIO mutant was ordered and transduced into the *pspA*-SABRS. However, transduction of *ΔgreA788::kan* into the *pspA*-SABRS always coincided with the co-transduction of native chromosomal *rpoN*. Successful growth on Kanamycin (Δ*greA*) resulted in a loss of Tetracycline resistance (*rpoN*::Tn*10*) in all transductants. No colonies formed upon selecting for transductants on a double antibiotic (Kan/Tet) plate, even at slightly lower antibiotic concentrations.

While a *greA* deletion could explain why SABT is not readily observed *in vivo*, this could not be confirmed thus far due to the co-transduction of the Δ*greA* locus and wild type *rpoN*.

4.3.6 Δ*hldE* as a candidate for putative bypass transcription

hldE is a bifunctional protein involved in the biosynthesis of ADP-L-glycero-D-manno-heptose, an LPS precursor (Valvano, 2000). Moreover, *hldE* is pivotal for maintaining the integrity of the outer membrane.

An *hldE* mutant would therefore have a compromised outer membrane structure and might somehow trigger activation of the psp operon, by an indirect effect on the inner membrane like when OM-IM spanning complexes are compromsised This induction may increase σ^{54}-independent background transcription of *ΦPpspA-lacZ*, as is thought to be the case for the Δ*asnA* mutant (4.3.2).

Out of the 6 putative bypass mutants identified the Δ*hldE* mutant had the highest β-galactosidase activity, 13x higher activity than the *pspA*-SABRS negative control and 57% activity of the wild type *rpoN* positive control strain (Fig 4.2.2).

This activity observed is relatively high (57% of wild type) compared to the Δ*asnA* mutant (11% of wild type), which is also thought to be a result of psp induction (4.3.2). If psp induction does cause the apparent SABT activity in both mutants, perhaps other factors may increase the signal with the *hldE* mutant.

Perhaps the *ΦPpspA-lacZ* has mutated to become σ^{54}-independent, however sequencing of the *ΦPpspA-lacZ* confirmed that it was intact (not shown). Or there may be σ^{54}-independent transcription of the psp operon under certain conditions, triggered by an *hldE* deletion. This σ^{54}-independent *pspA* transcription could be σ^{70}-dependent, as it has been suggested to be in *Yersinia enterocolitica*. (Maxson and Darwin, 2006). Indeed, a putative extended -10 promoter sequence has been identified in the *E. coli pspA* promoter region (section 4.3.3)

Another alternative is that the Δ*hldE* strain resulted in artificially high readings of *ΦPpspA-lacZ* activity by influencing the β-galactosidase assay. Recently, HldE has been the target of HldE kinase inhibitors in an attempt to permeabilise the outer membranes in order to increase the intracellular drug concentrations (Atamanyuk, 2013). Similarly, an *hldE* deletion may result in permeabilisation of the membrane and increase the β-galactosidase assay signal. To test the latter hypothesis, a MG1655 control strain and a MG1655 Δ*hldE* strain were constructed, and β-galactosidase activity from the native chromosomal *lac* operon was determined (Fig 4.3.6).

HldE has not been reported to influence native chromosomal *lacZ*. However, in MG1655 an *hldE* deletion did result in a 264% increase in Miller Units compared to the wild type MG1655 control strain (Fig 4.3.6). Therefore an *hldE* deletion appears to cause an increase in LacZ activity at two independent promoters (*pspA* and *lac* promoters). Since an *hldE* deletion increased β-galactosidase activity at two unrelated promoters, it suggests the deletion itself causes artificially high levels during the assays. This could be achieved by increasing membrane permeability, which could increase extracellular LacZ concentration. Alternatively or additionally, membrane permeability could also increase the intracellular ONPG concentration (the LacZ substrate in the assay).

Combined, psp operon induction and increased membrane permeability could account for the apparent SABT observed in the Δ*hldE pspA*-SABRS. Since HldE is a protein not directly involved in transcription, the apparent SABT observed with this mutant was concluded to be a foible of the β-galactosidase assay. Therefore this mutant was not explored further.

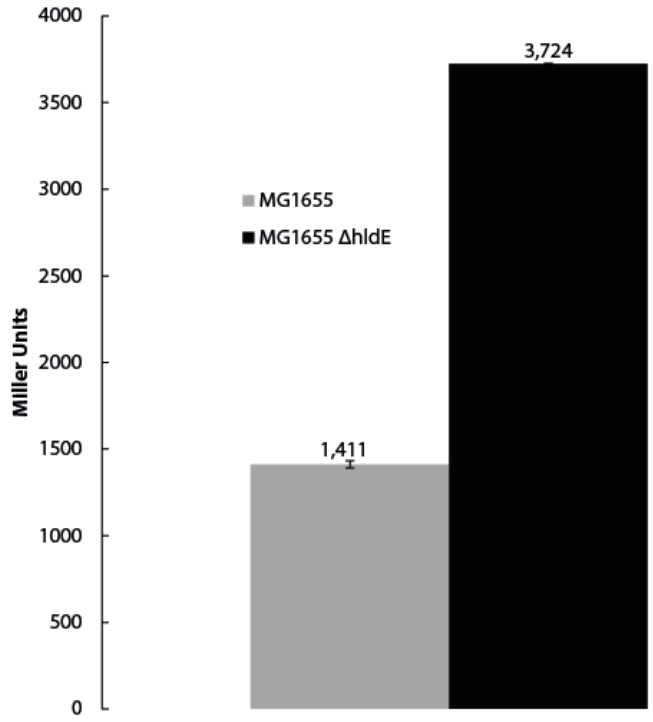


Fig 4.3.6: Effect of Δ*hldE* on native LacZ β-galactosidase activity.
MG1655 and MG1655 *ΔhldE745::kan* were grown to exponential phase and induced for 1 hour with 1mM IPTG. β-galactosidase activity arising from native *lac* operon was measured.

4.3.7 Δ*ttdR* as a candidate for inhibiting putative bypass transcription

TtdR is a transcriptional regulator involved in L-tartrate fermentation, and stimulates the expression of the *ttdABT* operon. It is expressed under anaerobic conditions and belongs to the same family of growth phase-specific nucleoid proteins as Fis, Dps, HU, IHF, and H-NS (Teramoto, 2010).

TtdR is repressed by the nitrate and nitrite two-component system, linking it to σ^{54} which is also involved in nitrogen regulation through NtrC (Kim et al, 2009). *tdtA* is the first gene in the operon activated by TtdR and is important in biofilm formation, a function which σ^{54} is also known to regulate (Herzberg, 2006). The second gene in the operon (*ttdB*) is involved in heat shock response, which σ^{54} is also involved in through IbpB (Kuczyńska-Wisńik, 2002). As such, there is some functional overlap between TtdR and σ^{54} target genes.

A *ttdR* deletion in the cell may result in the upregulation of alternative pathways involved in biofilm formation and heat shock response. Such alternative pathways may include the upregulation of σ^{54}–dependent genes, and indirectly the induction of σ^{54}-independent ΦP*pspA*-*lacZ* expression.

TtdR also has dozens of putative binding sites in or near genes involved in membrane structure and protein export (Teramoto, 2010). If TtdR does activate the expression of these genes, it is likely that a *ttdR* deletion may result in membrane stress and strong induction of the psp operon (and also ΦPpspA-*lacZ* expression).

Finally, TtdR is also thought to play a role in nucleoid structure and DNA folding (Teramoto, 2010). TtdR may influence SABT *in vivo* by changing chromosome structure or local promoter structure. Notably, DNA supercoiling *in vitro* affects bypass transcription and TtdR could have an impact on superhelicity. However, the role of DNA topology in SABT inhibition *in vivo* was inconclusive thus far (3.3.6).

4.4 Summary

The KEIO deletion library was screened for putative SABT mutants, and yielded 6 genes of interest (*hldE*, *asnA*, *ttdR, yhbX*, *nanT* and *greA*). However, co-transduction of wild type *rpoN* together with the deletions was observed in 3 of these mutants: Δ*yhbX*, Δ*nanT* and Δ*greA* (Fig 4.2.3). This suggests that the apparent SABT observed is likely a result of the reconstituted *rpoN* locus rather than SABT.

The remaining 3 gene deletions of interest (*hldE*, *asnA*, *ttdR*) would likely result in membrane stress and defects in protein export. These are inducing conditions of the psp operon and may result in elevated levels of σ^{54}-independent background transcription of ΦP*pspA*-*lacZ*.

GreA and TtdR are potentially relevant regarding putative inhibition of SABT *in vivo*, based on their roles in transcriptional regulation. However, this could not be confirmed due to β-galactosidase readings potentially being influenced by native chromosomal *rpoN*, or indirect induction of the psp operon respectively.

Overall, despite the screening and selection of over 30 000 mutants, no promising candidates were found for the potential inhibition of SABT *in vivo*. Not even under conditions where SABT had been observed before (42°C), or where wild type σ^{54} was more active (0.001% arabinose).

It is therefore likely none of the KEIO single gene deletions (studied as KOs) are responsible for inhibiting bypass transcription, at least not on their own. In the case that multiple genes are involved in the repression of SABT, it is possible that deleting one of them would have resulted in partial detectable levels of bypass transcription. This was not observed, indicating that the barriers to bypass transcription *in vivo* are very tightly regulated. While it is possible multiple deletions may give rise to clear SABT *in vivo*, this was not explored further in these experiments.

It also suggests that if any proteinaceous barriers are present to prevent spontaneous SABT from being readily observed *in vivo*, they are more likely to be essential genes not represented in the KEIO collection. These may include essential genes encoding proteins that reduce the levels of ΦP_{pspA}-*lacZ* transcript, such as RNase E (mRNA degradation) or Eno (degradosome subunit).

Chapter 5: Novel role for σ^{54} as a transcriptional repressor

5.1 Introduction and objectives

5.1.1 Recap and objectives

In Chapter 3 it was shown that increased temperature (42°C) can give rise to SABT *in vivo* at the σ^{54}-dependent gln_{Ap2} promoter (Fig 3.3.4). However, SABT was not observed *in vivo* under the same conditions at the *pspA* promoter (Fig 3.3.5). Is SABT unique to the gln_{Ap2} promoter or is it also detectable at other σ^{54}-dependent promoters? In this chapter I present a global analysis of transcription patterns in *E. coli* under the elevated temperature conditions which favour SABT at the gln_{Ap2} promoter.

In Chapter 4 single-gene deletion and overexpression libraries were screened to identify specific genes which may be involved in repressing or activating SABT *in vivo*. However, no single non-essential genes were shown to be responsible for repression or activation of SABT *in vivo*. Possibly because its regulation is strictly controlled, potentially by the actions of several gene products. Even if SABT is tightly regulated *in vivo*, why has σ^{54}-dependent transcription not evolved to be activator independent like other σ factors, by for example through the loss of the regulatory Region I of σ^{54} in combination with losses of chromosomal repressive genes?

To the above end, a global transcriptomic analysis of SABT *in vivo* may shed some light on both of these questions by addressing whether or not changes in transcription are seen which depend on σ^{54} and its region I deleted bypass variant.

Regarding the first question, we may see SABT at other σ^{54}-dependent promoters besides gln_{Ap2}. If so, perhaps these promoter sequences or their chromosomal environment may indicate what makes them unique, and why we struggle to detect SABT at other promoters such as the *pspA* promoter *in vivo* under standard conditions.

For the second question, we may also be able to classify σ^{54} promoters into two groups: where SABT is observed and where it is not. If so, a functional analysis of these two groups of promoters may allow us to address the question as to why σ^{54}-dependent transcription is closely regulated at some promoters, and has not evolved towards activator independence or allowed SABT to be readily observed *in vivo*. In order to address these possibilities, this chapter will focus on global transcriptomic analysis of SABT *in vivo*.

5.1.2 Introduction to RNA Sequencing

Various technologies are available to study and quantify the transcriptome, including hybridisation- and sequence-based strategies. Hybridisation-based strategies typically involve incubating fluorescently labelled cDNA with custom-designed microarrays or commercial high-density oligo microarrays (Wang, 2009). Hybridisation-based strategies are generally high throughput and relatively cheap, with the exception of high resolution arrays examining large genomes. Unfortunately these techniques have several limitations including: high background due to cross-hybridisation (Okoniewski, 2006); limited dynamic range due to high background and signal saturation; and requiring existing knowledge about the genome sequence. Hybridisation-based strategies also have the disadvantage that they cannot accurately determine the ends of transcripts, and therefore transcriptional start sites.

Unlike microarrays, the more recently adopted sequence-based approaches directly determine the cDNA sequence. Historically, Sanger sequencing of cDNA libraries was done but this method is relatively low throughput, costly and typically not quantitative (Boguski, 1994). Tag-based sequencing techniques were developed to circumvent these disadvantages, and are high throughput and more quantitative. However, they still rely on expensive sequencing technology, and many short reads cannot be mapped accurately on the reference genome. Isoforms are also generally indistinguishable (Wang, 2009).

Developments in deep sequencing technologies gave rise to a new approach for quantifying and mapping transcriptomes: RNA sequencing. This method typically involves sequencing a cDNA library generated from purified mRNA or total RNA (which can be size fractionated prior to analysis, see below 5.1.3), and has clear advantages over existing approaches. Unlike hybridization techniques, RNASeq is not restricted to detecting mRNA transcripts that correspond to known genomic sequences which are arrayed. Additionally, RNASeq offers benefits through high sensitivity, a wide dynamic range, and single nucleotide-resolution which allow sequence variations to be detected.

Compared to other sequence-based approaches RNASeq offers higher throughput, only requires low amounts of RNA, and allows for mapping of transcriptomes for large genomes at a low cost. It is also sufficiently accurate and sensitive to allow precise identification of the 5'end of the RNA, hence robust inferences about start sites of transcription and RNA processing can be made.

5.1.3 rRNA depletion and RNA Sequencing in *E. coli*

The mRNA for RNASeq is typically purified from total RNA using the oligo-dT method, in which the mRNA poly-A tails hybridise with oligo-dT beads. Unlike eukaryotes, in *E. coli* the mRNA is not polyadenylated and can therefore not be isolated using oligo-dT selection. mRNA only accounts for a small portion of total RNA and needs to be separated from the vast amount of ribosomal RNA (rRNA), to prevent amplification of rRNA when generating a cDNA library.

Numerous rRNA depletion methods have been developed that target specific regions of rRNA, but with limited success (Wurzel, 2010). Ribo-Zero, a new hybrid-subtraction kit from Epicentre, promises to remove all species of rRNAs. A comparison of various rRNA depletion methods concluded that Ribo-Zero performs the best, removing virtually all rRNA including the difficult to remove 5S rRNA (Giannoukos, 2012).

Ribo-Zero rRNA depletion of total RNA from 3 strains intended for use in RNA Sequencing in this study also confirmed almost complete removal of 23S and 16S rRNA (Fig 5.1.3, arrowed). This rRNA depletion method has an advantage over standard oligo-dT selection in that it also allows for the purification of many other small regulatory RNAs (sRNA) lacking a poly-A tail. These sRNA species are typically 50-250 bp in length and sufficiently abundant for detection (Mira, 2009). RNAs of specific sizes can also be enriched by recovering RNA from sizing gels.

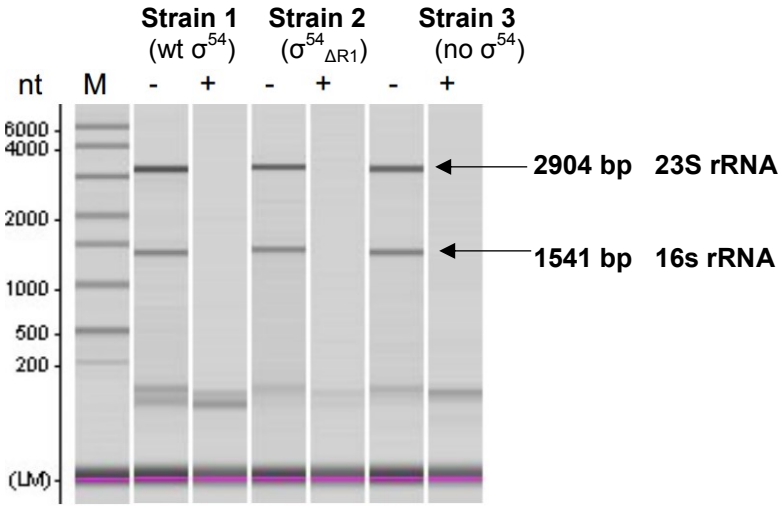


Figure 5.1.3: Analysis of the total RNA samples before and after Ribo-Zero rRNA depletion. RNA isolates from RNASeq samples were analysed on a Shimadzu MultiNA microchip electrophoresis system, before (-) and after (+) Ribo-Zero rRNA depletion. M = RNA marker.

5.1.4 First strand priming and synthesis

In order to create a cDNA library for sequencing, the RNA samples must be converted into DNA, and currently all protocols use a reverse transcriptase which uses RNA as a template. Like other polymerases, reverse transcriptase only initiates DNA synthesis from primers with 3' hydroxyl ends annealed to the template stand. This first strand priming is typically achieved in one of 3 ways (see Fig 5.1.4):

1) Oligo-dT primers annealed to the poly-A tail of mRNA
2) Random primers that anneal to random positions on the RNA molecule
3) Adapters attached to the 3' ends of the RNA.

Oligo-dT priming cannot be used in bacteria as they typically do not have poly-A tails. Further drawbacks of this method include that it only targets mRNA with a poly-A tail, and typically results in the enrichment of 3' ends as reverse transcriptase are prone to premature termination.

Random priming has the advantage of recovering non coding RNAs without poly-A tails, and also allowing the RNA to be subject to fragmentation. Fragmentation is often required as most current sequencing methods have short reads (~40-400bp), and fragmentation allows for improved coverage of the transcriptome. Therefore a further advantage of random priming is that the 3' end bias detected in oligo-dT priming is not observed. However, the disadvantage is that random priming does not appear to be completely random and bias towards amplifying certain sequences has been observed (Roberts, 2011; Hansen, 2010).

Finally, it is also possible to ligate a known primer to the 3' end of the RNA using T4 RNA ligase. This has the advantage over using random primers in that it does not result in priming bias. To avoid 3' enrichment as seen with oligo-dT priming, RNA can be fragmented prior to the addition of 3' reverse transcriptase primers. Combined, this method is the most powerful for RNASeq in *E. coli* as it avoids priming bias and 3' enrichment, while allowing for both RNA fragmentation and the isolation of non-coding RNAs.

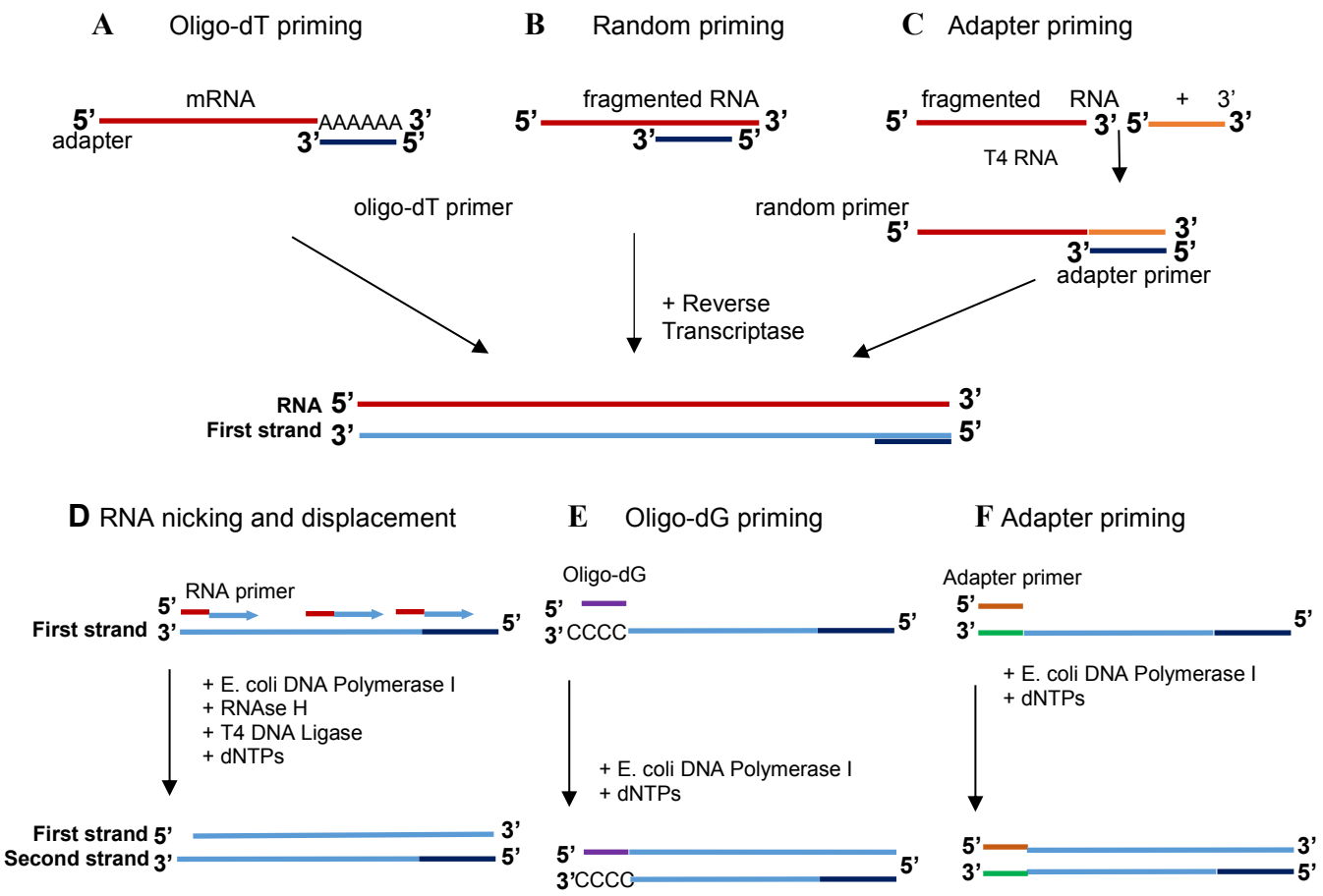


Fig 5.1.4: First and second strand synthesis for making a cDNA library from an RNA template.
First strand synthesis: Oligo-dT priming (A) involves using oligo-dT primers (dark blue) complementary to the 3' poly-A tails of mRNA. Random priming (B) includes the addition of random primers (dark blue) which anneal to previously fragmented RNA. For adapter priming (C) known RNA adapters are ligated to the 3' end of fragmented RNA, which can then be bound by complementary adapter primers (dark blue). For all 3 methods the RNA and primers are then incubated with Reverse Transcriptase and dNTPs to generate a DNA strand complementary to the RNA template, this is known as first strand synthesis.

Second strand synthesis: Several options exist to create a second strand using the first strand. One is RNA nicking and displacement (D), in which RNAse H is used to nick the original RNA template. These RNA fragments (red) are then used as primers for second strand synthesis, using the first strand as a template. T4 ligase repairs the nicks from the initial priming, but some 5' end information is lost. For oligo-dG priming (E), Moloney Murine Leukemia Virus Reverse Transcriptase is used to generate the first strand, adding several non-template dC nucleotides at the 3' end. Oligo-dGs primers (purple) can bind to the poly-C tail and act as a primer for second strand cDNA synthesis. For adapter priming (F), the first strand is synthesised after adding a known adapter to the 5' end of the RNA template. As

a result, the 3’ sequence of the first strand (green) is known and adapter primers (brown) are used to anneal to this sequence for second strand synthesis.

5.1.5 Second strand priming and synthesis

After generating the first strand, the second strand is generated using a DNA polymerase which requires a 3’ hydroxyl group and a primer for DNA synthesis. Multiple options exist regarding second strand synthesis, including:

1) RNA nicking and displacement
2) Oligo-dG primers (Zhu, 2001).
3) Primer complementary to 5’ adapter ligated to RNA template

RNA displacement synthesis uses a combination of *E. coli* DNA polymerase I, *E. coli* RNase H, and T4 DNA ligase. After first strand synthesis RNase H nicks the original RNA template, these RNA fragments are then used as primers for second strand synthesis (using first strand as a template). The 5’ to 3’ exonuclease activity of DNA Polymerase I degrades incoming RNA. Finally, T4 DNA ligase restores the nicks generated after priming. The main disadvantage of this method is that 5’ terminal information is lost, which can make it hard to characterise transcription start sites.

Moloney Murine Leukemia Virus Reverse Transcriptase *(*M-MLV RT) adds a number of non-template dC nucleotides at the end of the synthesized DNA (irrespective of template). This is the basis of oligo-dG priming for second strand synthesis (Zhu, 2001). Following first strand synthesis with M-MLV RT, oligo-dG primers are annealed to the 3’ oligo-dC tails generated during this process. These oligo-dGs subsequently act as a primer for second strand cDNA synthesis. The advantage of oligo-dG priming is that it accurately amplifies the 5’ terminal of the RNA, and has the potential to amplify entire transcripts when combined with oligo-dT first strand synthesis.

A third option is to ligate adapters to the 5’ end of the RNA template (before first strand synthesis). Oligos complementary to these 5’ adapters can later be used as primers for second strand synthesis. This technique also allows for the recovery of 5’ ends (like oligo-dG priming).

5.1.6 5’ end processing and sequencing adapters

In order to sequence cDNA, sequencing adapters must be present. These vary depending on the sequencing platform and may contain other functional elements. Such functional elements may include barcodes used for sequencing pooled samples, which is done when the number of reads obtained in a single run exceeds the number required for one sample.

Adapters can be ligated to the 5’ end of the RNA molecules prior to first strand synthesis. However, RNAs typically have a 5’ triphosphate that needs to be removed with a phosphatase, in order to allow T4 DNA ligase to ligate the adapters to a 5’ monophosphate.

In *E. coli*, primary transcripts that maintain a 5′ triphosphate are moderately resistant to degradation because the major endonuclease RNase E is sensitive to 5′ structure, strongly preferring degradation of processed RNAs with a 5′-monophosphate (Celesnik, 2007). 5’ end processing therefore plays an important role in the degradation of mRNA, and enzymes like RppH can cleave 5’ phosphates and reduce the stability of the mRNA (Deana, 2008).

This 5’ end processing can be detected in RNASeq by attaching 5’ adapters without prior phosphatase treatment of the 5’ RNA ends. Adapters for cDNA amplification will only ligate to processed RNAs with a 5’ monophosphate, thereby enriching the processed RNAs in the amplified cDNA.

5.1.7 Illumina Next Generation Sequencing

Once a cDNA library with sequencing adapters is formed, the library can be sequenced. Currently the sequencing platform of choice for RNASeq is Illumina sequencing, which uses solid-phase amplification and reversible terminator sequencing (see Fig 5.1.7). This platform allows for deep sequencing with long reads and high throughput of up to 600 Gb per run. Illumina sequencing also has sufficiently low error rates for transcriptome mapping and reference genome assembly (Glenn, 2011).

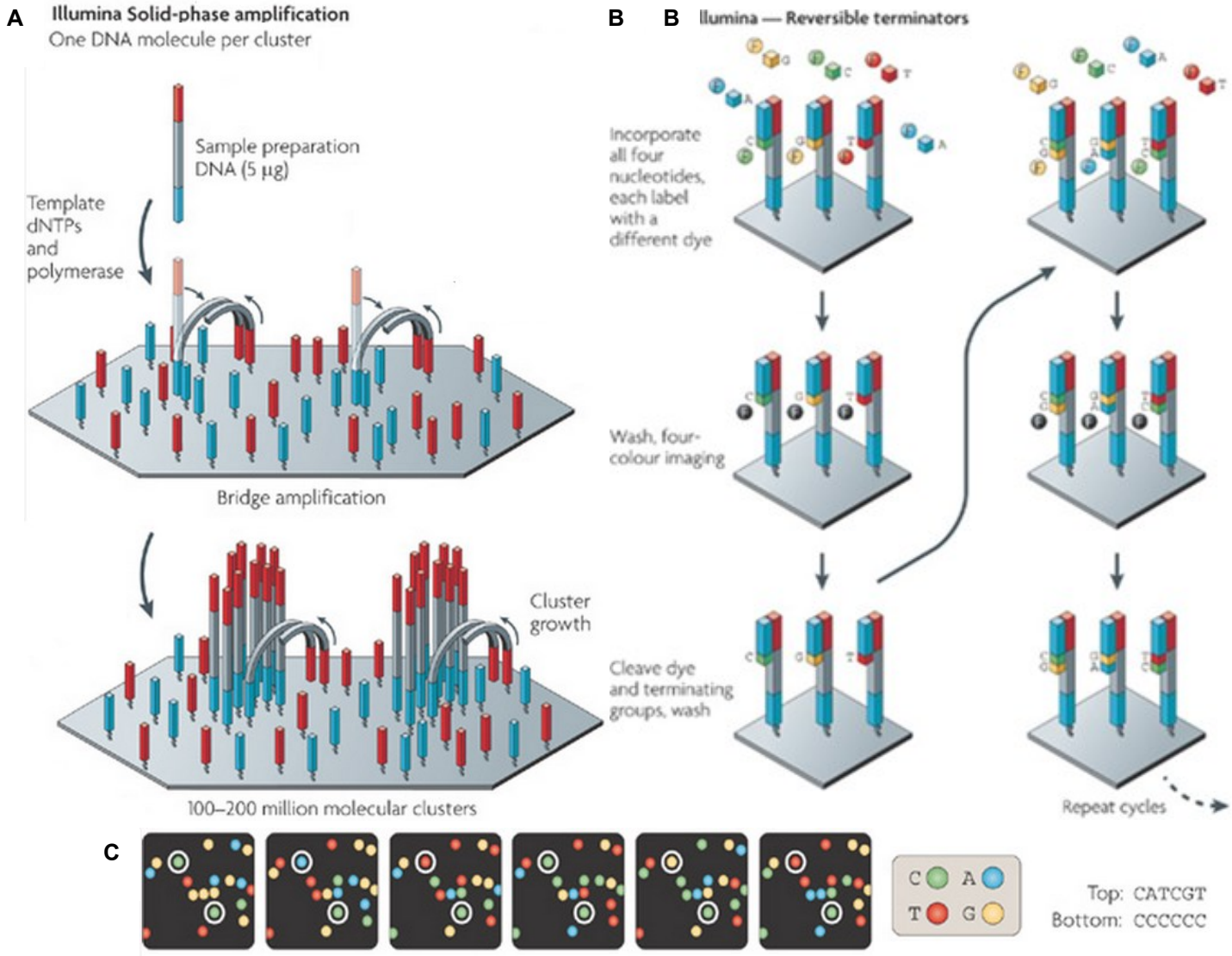


Fig 5.1.7: Illumina solid-phase amplification and reversible terminator sequencing.

Illumina sequencing templates are immobilised on a flow cell surface, with a dense lawn of amplification primers (A). Illumina solid-phase amplification (A) has two steps: 1) priming and extension of single-stranded template, and 2) bridge amplification of immobilised templates with adjacent primers to form clusters. Denaturation leaves single-stranded templates anchored to the substrate.

Solid-phase amplification creates millions of molecular clusters, with individual clusters having the same sequences. Next fluorescently labelled reversible terminators are added (each nucleotide a different dye), along with primers and DNA Polymerase (B). Once the first base is added, laser exciting allows for the first base in each cluster to be identified (C). The reversible terminators (3′-*O*-azidomethyl 2′-deoxynucleoside triphosphates) are incubated with tris(2-carboxyethyl)phosphine to remove the fluorescent dye and side arm from a linker, regenerating the 3′ hydroxyl group. The second base can now be determined, and the process repeated until the full sequence is known (adapted from Metzker, 2010).

5.1.8 A typical RNA Sequencing experiment for *E. coli*

Here a typical RNA Seq experiment with *E. coli* is described for the Illumina sequencing platform, using the novel Ribo-Zero rRNA depletion method described in 5.1.3.

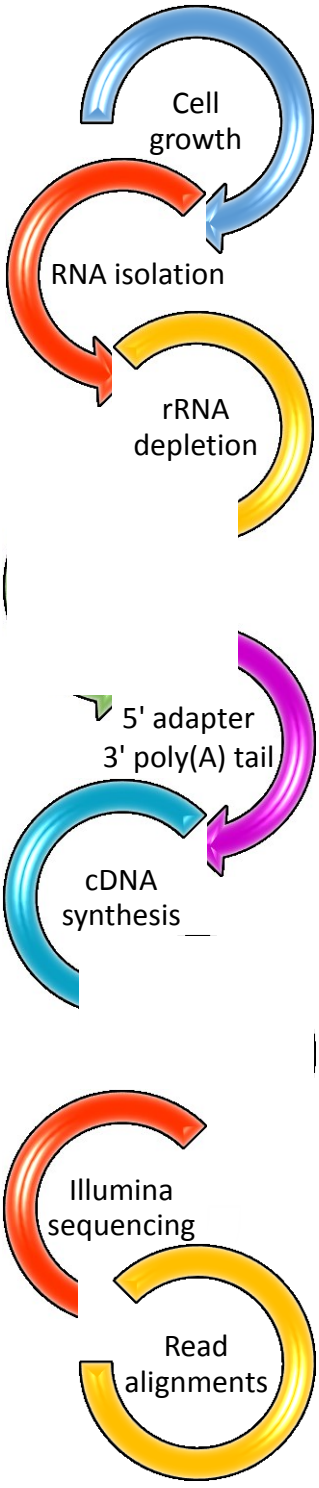


Fig 5.1.8: Flow-through of *E. coli* RNASeq experiment.

Cell cultures are grown to the desired conditions and centrifuged, leaving a cell pellet that includes total RNA of the cell.

Bacterial pellets are disrupted in a bead mill and total RNA is extracted with a mirVana RNA isolation kit (Ambion). Isolated RNAs range from long mRNAs to short 10-mers, and include miRNAs, siRNAs, and/or snRNAs. DNAse treatment removes any DNA contaminants.

5S, 16S and 23S rRNA are subsequently removed from the total RNA using Ribo-Zero rRNA depletion kits (Fig 5.1.4, arrowed).

rRNA depleted RNA is then fragmented by RNAse III to produce 60-100 nucleotide fragments (later used as cDNA templates). The 5' triphosphate ends are removed by 5' RNA polyphosphatase, and replaced with RNA adapters ligated to the 5' end. Poly(A) polymerase is subsequently used to added poly-A tails to the 3'- ends.

First-strand cDNA synthesis is performed using M-MLV reverse transcriptase and oligo(dT)-adapter primers. The resulting cDNA is PCR-amplified using a high fidelity DNA polymerase, and primers complementary to the 5' adapters.

The cDNA is then purified and size fractionated in the 200-450 bp size range (including 148 bp flanking sequences). The cDNA is then sequenced using 50 bp read length on an Illumina HiSeq 2000 system.

Then the adapters are trimmed from the sequenced cDNAs, aligned to the *E. coli* K-12 reference genome (NC_000913), and quantified. Transcription for each gene can then be

expressed in Reads Per Kb exon (gene) per Million mapped reads (RPKM).

5.2 Results

RNA next generation sequencing (by Vertis Biotechnologie AG) was used to study genes involved in σ^{54}-dependent transcription *in vivo*, and SABT in particular. Expression profiles were compared from 3 groups of *E. coli* strains: 1) WTRS (wild type σ^{54}); 2) SABRS ($\sigma^{54}{}_{\Delta R1}$ activator bypass mutant); 3) DRRS control (lacking any σ^{54}). In addition, cells were grown to exponential phase in minimal Gutnick media at 42°C, the same growth conditions under which SABT had been observed previously at the gln_{Ap2}-*gfp* (Fig 3.3.4).

WTRS:	Lac*Z*$^-$	*rpoN208::Tn10*	ΦP_{pspA}-*lacZ*	pBAD18-*rpoN*
SABRS:	Lac*Z*$^-$	*rpoN208::Tn10*	ΦP_{pspA}-*lacZ*	pBAD18-$rpoN_{\Delta R1}$
DRRS:	Lac*Z*$^-$	*rpoN208::Tn10*	ΦP_{pspA}-*lacZ*	pBAD18-empty

RNASeq reads were mapped to the *E. coli* K-12 reference genome (NC_000913), and the data was screened for upregulation and down regulation in the 4496 mapped genes. Transcription in the WTRS and SABRS was quantified relative to the DDRS (represented as fold change). Changes in transcription relative to the control strain (DDRS) were scored as significant whenever they crossed a certain threshold (3 standard deviations from the mean). Results are summarised in Fig 5.2.2, Fig 5.2.3, Fig 5,2.4 and Fig 5.2.5.

5.2.1 SABT is detectable at the *gln*$_{Ap2}$

Fig 5.2.1 shows that transcription is detectable at the *gln*$_{AP2}$ promoter in the SABRS ($\sigma^{54}{}_{\Delta R1}$ activator bypass mutant), and 4-fold higher than the DRRS (lacking σ^{54}). The transcription start site is the same as in the WTRS (arrowed), which is known to be active and σ^{54}-dependent (Ninfa, 1987; Retzer, 1989). This suggests that SABT at the *gln*$_{AP2}$ promoter does occur under these conditions *in vivo* at 42°C, albeit at low levels. This RNASeq result is in line with previous experiments, where SABT was reported *in vivo* at 42°C at the *gln*$_{AP2}$ promoter (Fig 3.3.4).

In the RNASeq data *gln*$_{Ap2}$ SABT is detectable at 3% of WTRS (Fig 5.2.1), whereas ΦP$_{glnAP2}$-*gfp* bypass signal *in vivo* was significant (40%) compared to the WTRS (Fig 3.3.4). While both experiments report activity at the *gln*$_{AP2}$ promoter, they are not directly comparable in terms of quantification. They reflect somewhat different aspects of gene expression, although each has an mRNA component.

The ΦP$_{glnAP2}$-*gfp* reporter measures a protein (GFP) whereas the RNASeq data quantifies mRNA, and in this case specifically mRNA at the *gln*$_{Ap2}$ promoter. mRNA is a lot less stable than GFP, which has a half-life of 28 hours (Corish, 1999). This means GFP signal could be able to accumulate throughout the course of the overnight experiment. If the GFP signal in the WTRS was rapidly saturated (ΦP$_{glnAP2}$-*gfp* on high copy plasmid), it would allow the SABT strain to catch up throughout the course of the overnight experiment.

However, this would not be the case with the RNASeq experiment where mRNA is rapidly degraded *in vivo*, and only a single chromosomal *gln*$_{Ap2}$ copy was present. Cells for RNA seq were also grown for a shorter time (into exponential phase).

Combined, these could explain why SABT was readily observed at the ΦP$_{glnAP2}$-*gfp* earlier (Fig 3.3.4), but only just detectable at the *gln*$_{Ap2}$ promoter in the RNASeq data (Fig 5.2.1). While SABT signal at the *gln*$_{Ap2}$ promoter is merely detectable in the RNASeq, it is still higher than the DRRS lacking any σ^{54}. This suggests that the signal is indeed $\sigma^{54}{}_{\Delta R1}$-dependent SABT.

As such, RNAseq appears to be a suitable method to quantify any putative SABT signal *in vivo*.

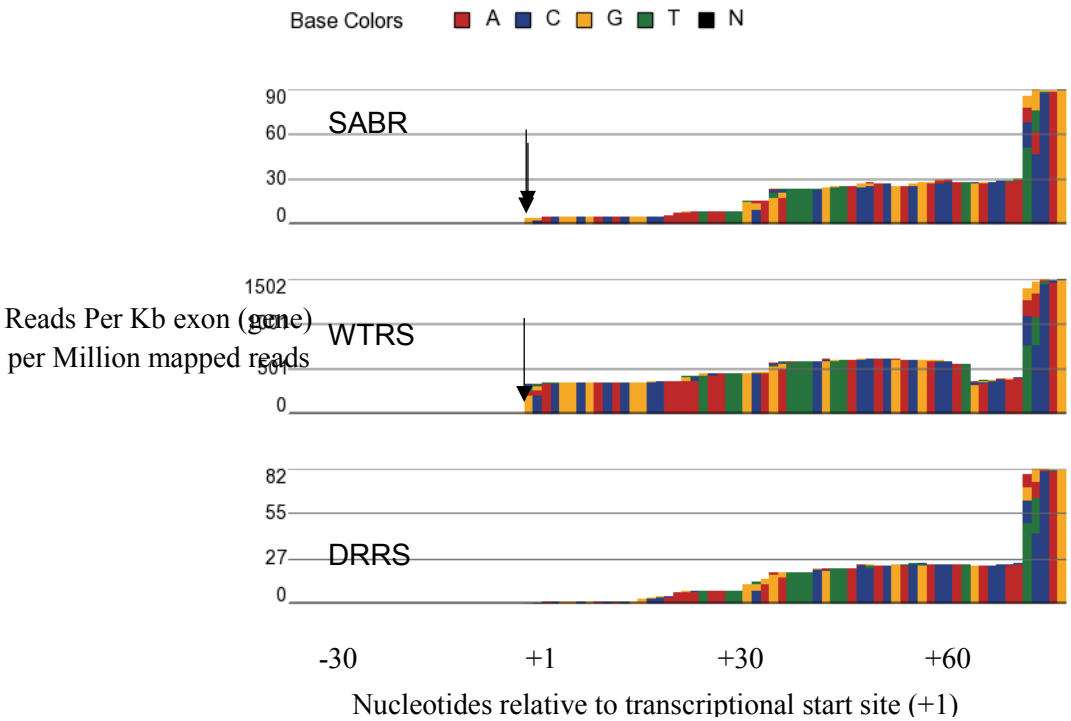


Fig 5.2.1: SABT at *gln*$_{AP2}$ promoter *in vivo* in minimal media at 42°C

Cells were grown to exponential phase (OD_{600} 0.4) in minimal media at 42°C before being harvested and submitted for RNASeq. Transcription at the *gln*$_{Ap2}$ promoter start site is observed in the WTRS and detectable in the SABRS (arrowed) but not in the DRRS (Δ*rpoN*). Transcription detected in the SABRS is 4x higher at the *gln*$_{AP2}$ promoter (arrowed) than the basal level of transcription in the DRRS.

5.2.2 Gene repression in the SABRS

RNASeq data for all genes in the WTRS (wt σ^{54}) and SABRS ($\sigma^{54}{}_{\Delta R1}$ bypass mutant) was quantified as a fold change relative to the DRRS (lacking σ^{54}). Transcription of 28 genes was more than 3 standard deviations below the mean in the SABRS, and these genes were scored as repressed in the SABRS (Fig 5.2.2).

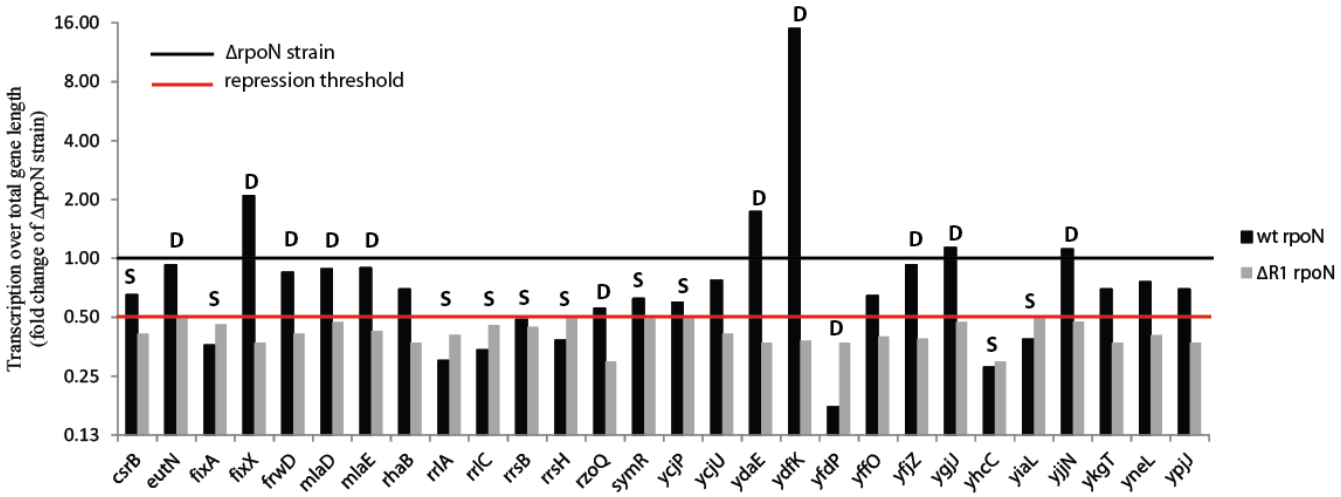


Fig 5.2.2: Genes repressed in the SABRS

In the *E. coli* genome 28 genes showed 2-fold or more decrease in transcription in the SABRS compared to the DRRS (lacking any σ^{54}). A 2-fold change between the SABRS and the DRRS is the equivalent of about 3 standard deviations from the mean.

Most of the downregulated genes in the SABRS (grey) showed around 40-50% activity of their DRRS counterparts, with no genes showing extreme inhibition. Many of these repressed genes also display similar levels of repression in the WTRS (black), indicating the phenotype may be related to both forms of σ^{54} –the full length protein and that lacking RegionI (labelled S). In some cases the genes repressed in the SABRS display dissimilar levels of expression in the WTRS (D). Transcription of these genes WTRS ranges from strong upregulation (*ydfK*), to very strong inhibition (*ydfP*).

28 genes were scored as repressed in the SABRS, 9 of which display similar levels of expression in the WTRS. Therefore their expression maybe be linked to σ^{54} activity and were labelled S (Fig 5.2.2, S). Their functions are summarised in Table 5.2.2a below:

Gene	**Gene function**
fixA	part of fix operon, anaerobic carnitine reduction (Eichler, 1995).
rrlA and *rrlC*	23S rRNA
rrsB and *rrsH*	16S rRNA
symR	small regulatory RNA repressing translation of *symE*, member of AbrB superfamily of transcriptional regulators (Kawano, 2007)
ycjP	putative component of ATP-dependent sugar transporter (Saurin, 1999)
yffO	regulated by the FlhDC flagellar transcriptional regulator (Stafford, 2005)
yhcC	Not well characterised, identified as iron-sulfur protein (Estellon, 2014)
yiaL	poorly studied, potential role in carbohydrate metabolism (Ibanez, 2000)

Table 5.2.2a: Function of genes with similar levels of repression in SABRS and WTRS.

Of the 28 genes scored as repressed in the SABRS, 12 show dissimilar levels of transcription in the WTRS (Fig 5.2.2, D). Transcription levels of these genes in the WTRS include: 3 upregulated genes, 1 strongly repressed gene (*ydfP*), and 7 genes with transcription levels similar to the DRRS. These genes and their functions are summarised in Table 5.2.2b:

Gene	**WTRS**	**Function**
eutN	= DRRS	possible transporter (Forouhar, 2007), *eut* operon involved in the ability to utilise ethanolamine as the sole source of carbon and nitrogen (Kofoid, 1999).
frwD	= DRRS	predicted PTS permease subunit (Tchieu, 2001)
mlaD	= DRRS	integral membrane component of phospholipid transporter (Malinverni, 2009)
mlaE	= DRRS	integral membrane component of phospholipid transporter (Malinverni, 2009)
rzoQ	= DRRS	putative Rz1-like protein, Qin prophage
yfjZ	= DRRS	affects mqsR toxicity and RNAse activity, which regulates cspD (Kim, 2010)
ygjJ	= DRRS	not characterised, hypothetical protein
yjjN	= DRRS	predicted L-galactonate oxidoreductase (Reed, 2006)
fixX	= DRRS	involved in anaerobic metabolism of carnitine (Eichler, 1995)
ydaE	> DRRS	not characterised
ydfK	>> DRRS	hypothetical membrane protein (Baranova, 1999)
ydfP	< SABRS	Qin prophage; conserved protein

Table 5.2.2b: Function of genes repressed in SABRS with dissimilar expression in the WTRS.

5.2.3 Gene upregulation in the SABRS

RNASeq data for all genes in the WTRS (wt σ^{54}) and SABRS ($\sigma^{54}_{\Delta R1}$ bypass mutant) was quantified as a fold change relative to the DRRS (lacking σ^{54}). Transcription of 49 genes was more than 3 standard deviations above the mean in the SABRS, and these genes were scored as upregulated in the SABRS (Fig 5.2.3).

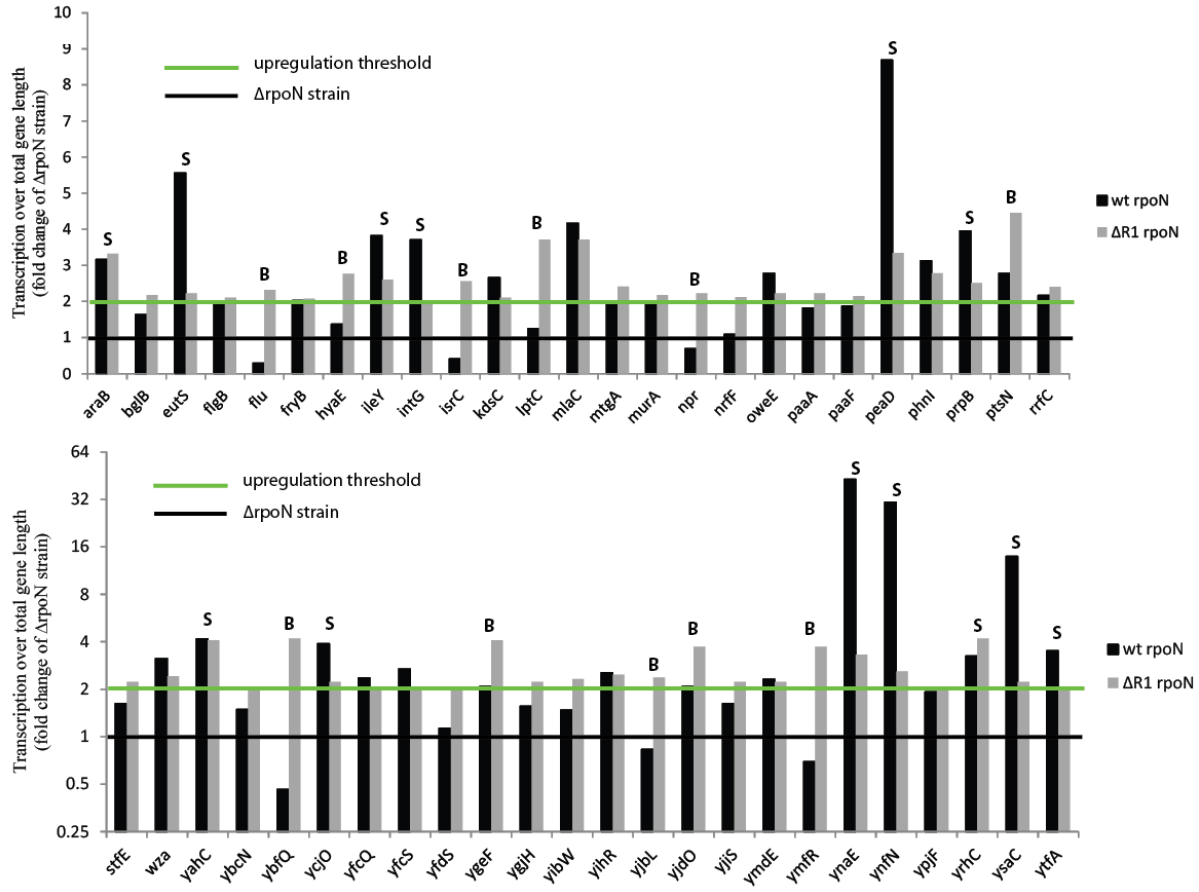


Fig 5.2.3: Genes upregulated in the SABRS.

Throughout the *E. coli* genome, 49 genes showed 2-fold or more increase in transcription for the SABRS (grey) compared to the DRRS (lacking any σ^{54}). A 2-fold change between the SABRS and DRRS strains is the equivalent of about 3 standard deviations from the mean.

Multiple genes were scored as upregulated in the SABRS (grey) compared to the DDRS control, but no genes show more than 4-fold increase in transcription. Several of these upregulated genes were also activated in the WTRS (black). Genes were classified into two groups: upregulation in the SABRS and WTRS (eg *peaD* and *yahC*), and upregulation in the SABRS only (eg. *flu* and *ybfQ*). Where genes were upregulated in both strains, their expression may be linked to σ^{54} and they were labelled S. Genes displaying activation only in the SABRS could be potential candidates for SABT, and labelled B.

Of the 49 genes upregulated in the SABRS, 14 are also upregulated in the WTRS. These genes are upregulated in the presence of σ^{54} or $\sigma^{54}_{\Delta R1}$, but not in the absence of σ^{54} (DDRS). Therefore their expression is potentially σ^{54}-dependent irrespective of the RegionI sequences and were labelled S (Fig 5.2.3, S). Their functions are summarised in Table 5.2.3a.

Gene	**Function**
araB	ribulokinase, catalyses the phosphorylation of L-ribulose to L-ribulose-5-phosphate
eutS	predicted structural protein, ethanolamine utilisation microcompartment
ileY	one of five isoleucine tRNAs
intG	*intG* clustered with chemotaxis, flagellum, and fimbria genes (Poleno, 2003)
peaD	pseudogene, predicted replication protein fragment
prpB	2-Methylisocitrate lyase, σ^{54} promoter
yahC	unknown inner membrane protein
ycjO	membrane component of a predicted ATP-dependent sugar transporter
ynaE	5' UTR of *ynaE* is similar to that of *ydfK* (Raghavan, 2011), expression of both genes is upregulated by cold shock (Polissi, 2003).
ynfN	predicted protein
yrhC	pseudogene
ysaC	uncharacterised gene
ytfA	Transcriptional regulator, overexpression causes a filamentous phenotype (Burke, 2013)

Table 5.2.3a: Function of genes upregulated in SABRS and WTRS.

Of the 49 genes upregulated in the SABRS, 13 are not upregulated in the WTRS (Fig 5.2.3). Expression of these genes upregulated in the SABRS but not WTRS or DRRS, indicating these genes are potential SABT candidates, labelled B (Fig 5.2.3, B). Their functions are summarised in Table 5.2.3b.

Gene	**Function**
flu	encodes 2 membrane proteins, important in biofilm formation (Danese, 2000).
hyaE	chaperone for HyaA, small subunit of hydrogenase 1 repressed by nitrate (Richard, 1999)
isrC	Small RNA transcribed from the region between the *yeeP* and *flu* genes (Chen, 2002)
lptC	component of the Lpt lipopolysaccharide transport system.
npr	*npr* forms part of the σ^{54} operon in *E. coli* K-12 (Powell, 1995) and is implicated in the regulation of nitrogen metabolism (Lee, 2013a)
nrfF	activator of formate-dependent nitrite reductase complex
ptsN	regulates the activity or expression of potassium transporters, part of σ^{54} operon in *E. coli* . May have a role in the regulation of nitrogen metabolism (Lee, 2013a)
ybfQ	conserved protein, rhs-like
ygeF	predicted protein, located in remnant of ETT2 pathogenicity island (Ren, 2004b)
yjbL	Conserved protein, associates with 50S ribosome subunit (Jiang, 2006)
yjdO	toxin of novel type V toxin-antitoxin system (Wang, 2012d). One of a small set of coding regions that do not contain cleavage sites for the MqsR toxin (Wang, 2013)
ymfR	predicted protein

Table 5.2.3b: Function of genes upregulated in SABRS but not the WTRS.

5.2.4 Gene repression in the WTRS

RNASeq data for all genes in the WTRS (wt σ^{54}) and SABRS ($\sigma^{54}{}_{\Delta R1}$ bypass mutant) was quantified as a fold change relative to the DRRS (lacking σ^{54}). Transcription of 21 genes was more than 3 standard deviations below the mean in the WTRS, and these genes were scored as repressed in the WTRS (Fig 5.2.3).

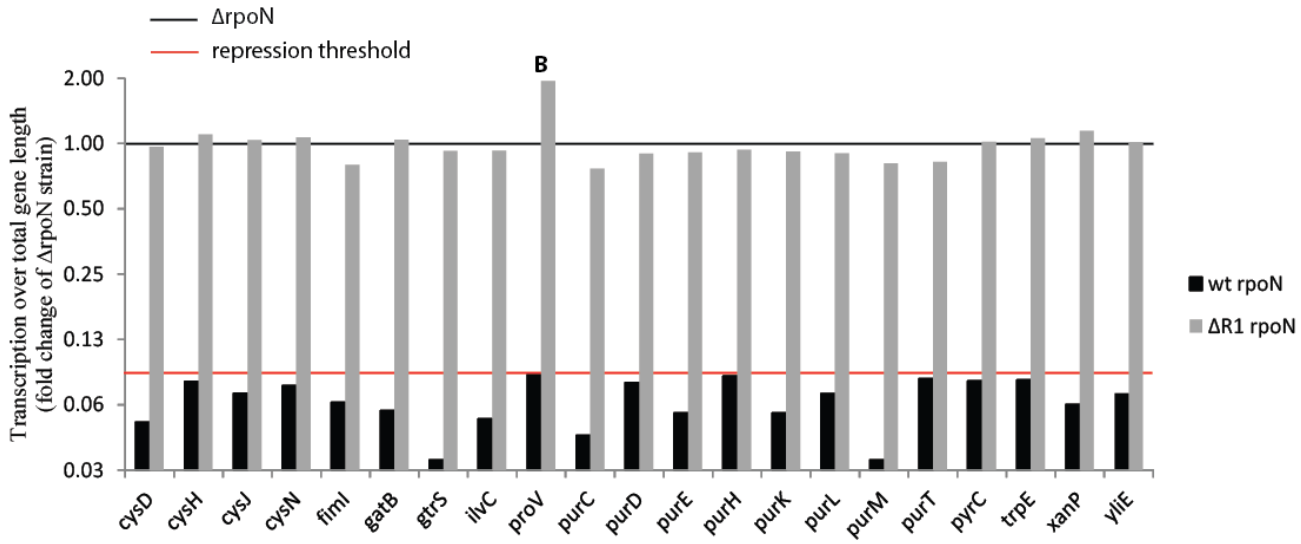


Fig 5.2.4: Repressed genes in the WTRS

Throughout the *E. coli* genome, RNASeq data showed repression for 21 genes in the WTRS (black) compared to the DRRS control. For scoring significant upregulation, a cutoff of 3 standard deviations was used (12-fold change). Repression in screened genes varies from the 12-fold repression (*cysH* and *purH*) to 30-fold inhibition (*gtrS* and *purM*).

For those same genes, there was little variation in transcription between the SABRS (grey) and the DRRS control. Except *ProV*, which showed ≈2-fold upregulation in the SABRS but repression in the WTRS, therefore a potential bypass promoter exists controlling *proV* expression (B).

21 genes were downregulated in the WTRS strain, but none showed downregulation in the SABRS and DRRS strains (Fig 5.2.4). Expression of those genes was almost identical in the SABRS and DRRS. With the exception of *proV* which was upregulated in the SABRS, highlighting it as a potential SABT candidate (labelled B). Functions of repressed genes in the WTRS are summarised in Table 5.2.4.

Gene	**Function**
cysD and cysN	sulfate adenylyltransferase subunits
cysH and cysJ	3'-phospho-adenylylsulfate reductase and sulfate reductase subunit
fimI	deletion in *fimI* produces a piliation-negative phenotype, (Valenski, 2003)
gatB	galacticol PTS permease subunit, part of sugar transporting PTS system
gtrS	serotype-specific glucosyl transferase, CPS-53 (KpLE1) prophage
ilvC	acetohydroxy acid isomeroreductase, biosynthesis isoleucine and valine
proV	ATP-binding subunit of glycine betaine/proline ABC transporter
purCDEHKLMT	*pur* operon is involved in *de novo* purine biosynthesis
pyrC	Dihydroorotase subunit, pathway for *de novo* pyrimidine nucleotide biosynthesis
trpE	Athranilate synthase subunit.
xanP	XanP is a member of the Nucleobase-Cation Symport-2 family of transporters
yliE	predicted c-di-GMP-specific phosphodiesterase. Overexpression of *yliE* reduces biofilm formation (Boehm, 2009)

Table 5.2.4: Function of genes repressed in the WTRS.

5.2.5 Gene upregulation in the WTRS

RNASeq data for all genes in the WTRS (wt σ^{54}) and SABRS ($\sigma^{54}{}_{\Delta R1}$ bypass mutant) was quantified as a fold change relative to the DRRS (lacking σ^{54}). Transcription of 26 genes was more than 3 standard deviations above the mean in the WTRS, and these genes were scored as upregulated in the WTRS (Fig 5.2.3).

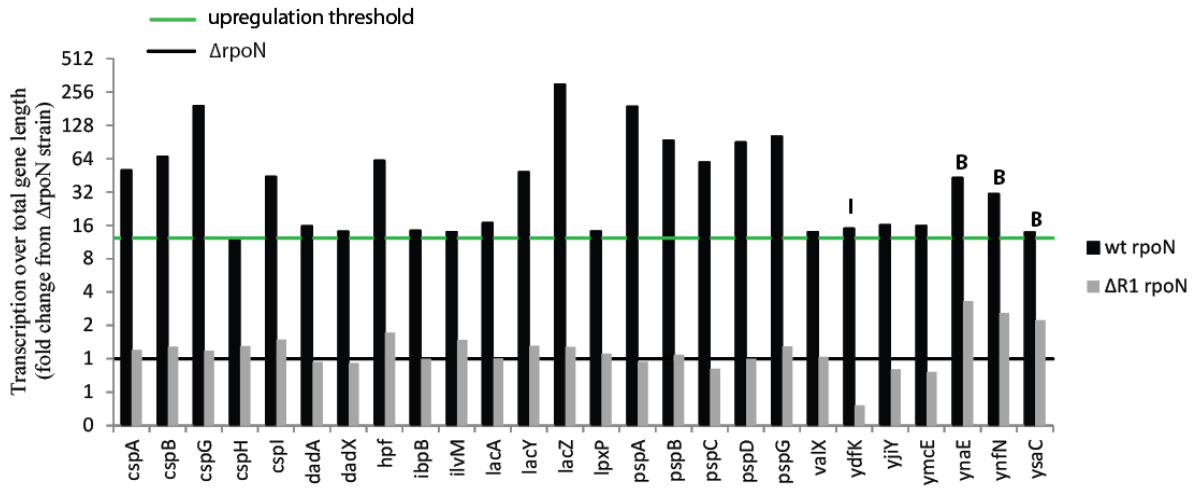


Fig 5.2.5: Upregulated genes in the WTRS

Over the *E. coli* genome, RNASeq data showed upregulation of 26 genes in the WTRS (**black**) compared to a DRRS control (lacking any σ^{54}). For scoring significant upregulation, a cutoff of 3 standard deviations was used (12-fold change for WTRS and DRRS). Upregulation in the WTRS varies from 12-fold upregulation (*cspH* and *dadX*) to over 100-fold upregulation (*cspG*, *LacZ* and *pspA*).

3 genes also showed upregulation in the WTRS and the SABRS (*ynaE*, *ynfN* and *ysaC*), making them candidates for SABT (**B**). One gene also showed upregulation for the WTRS (*ydfK*), but inhibition in the SABRS (**I**).

26 genes were upregulated in the WTRS strain, but most of them were not upregulated in the SABRS (Fig 5.2.5). 3 genes were also upregulated in the WTRS and the SABRS (*ynaE, ynfN*

and *ysaC*), highlighting them as potential genes for SABT (labelled B). *ydfK* was the only gene upregulated in the WTRS and inhibited in the SABRS (Fig 5.2.5). These gene functions are summarised in Table 3.6.

Gene	Function
cspA	major cold shock protein and detected during early-log-phase growth at 37°C (Jones, 1987). Positive regulator of H-NS (Brandi, 1994)
cspB, cspG, cspI	cold shock proteins
cspH	toxin inhibiting replication, activated by H-NS and mqsR but inhibited by mqsA (Kim, 2010c).
dadA, dadX	D-amino acid dehydrogenase and alanine racemase
hpf	hibernation promoting factor, part of σ^{54}operon in *E. coli*
ibpB	Small heat shock protein, σ^{54} promoter
ilvM	Acetohydroxyacid synthase II, biosynthesis of α-aceto-α-hydroxybutyrate and α-acetolactate
lacA, lacY	*lac* operon, lactose metabolism
lacZ	β-galactosidase, in this case also part of ΦP*pspA-lacZ*
lpxP	Palmitoleoyl acyltransferase, catalyses the incorporation of palmitoleate into lipid A
pspABCDG	part of psp operon, σ^{54} promoter
valX	Valine tRNA,
ydfK	predicted DNA-binding transcriptional regulator . Similar 5' UTR to that of *ynaE,* expression of both genes is upregulated by cold shock (Raghavan, 2011).
yjjY	Predicted protein
ymcE	Predicted protein
ynaE	Unknown function. Similar 5' UTR to that of *ydfK,* both genes are upregulated by cold shock (Raghavan, 2011).
ynfN	Predicted protein
ysaC	uncharacterised gene

Table 5.2.5: Function of genes upregulated in the WTRS

5.2.6 Comparison of RNASeq data and σ^{54} ChIP binding study

RNASeq was performed to report the effects of σ^{54} and $\sigma^{54}_{\Delta R1}$ on global gene expression. Genes of interest were obtained by screening genes for upregulation or downregulation (5.2.2-5.2.5). However, limiting target genes to strong upregulation or repression may overlook genes of interest with a less obvious phenotype. RNASeq produces a large amount of data for 4500 genes which cannot all be analysed individually, at least not without invoking additional informatics and computational help (not possible due to constraints on completing the thesis). Therefore additional candidate genes were selected to examine how their expression is affected in each of the strains.

These candidates were genes where σ^{54} is known to bind, selected from a recent study mapping the binding of σ^{54} across the *E. coli* genome (Bonocora, unpublished; Appendix C). This ChIP study identifies 145 σ^{54} binding sites in or near *E. coli* genes. This includes 18 of the 20 experimentally confirmed σ^{54} promoters, and 15 of 76 predicted σ^{54} promoters (Bonocora, unpublished). These predicted promoters were identified based on 3 criteria: i) mapped transcriptional start sites, ii) genetic evidence (mutation or heterologous gene expression), and iii) putative promoters based on similarity to -12/-24 consensus sequences (Barrios, 1999). In this study cells were grown in minimal M9 media at 30°C, to exponential phase. As such, different media and temperatures may account for some of the discrepancies between the ChIP study and RNASeq data from this chapter.

In the RNASeq data, 56 out of these 145 gene candidates were found to be upregulated or downregulated in minimal media at 42°C (in WTRS or SABRS, or both) (Fig 5.2.7, Fig 5.2.8, Fig 5.2.9). Regulation was scored as significant if it was more than 2 standard deviations from the mean. A less stringent threshold than 3 standard deviations was used because this gene set contained many genes already known to be σ^{54}-dependent, including genes with very strong upregulation (eg *LacZ* from ΦP_{pspA}-*lacZ*) or downregulation (*purK*) which skew the data set. Upregulation was scored as 200% or more in the WTRS and 125% or more in the SABRS. Downregulation was scored as 40% or more repression in the WTRS, hardly any repression was observed in the DRRS (see RNASeq supplementary data).

5.2.7 Putative SABT in SABRS in genes with σ^{54} binding sites.

19 out of 145 genes with σ^{54} binding sites were upregulated 125% or more in the SABRS (Fig 5.2.7). *GlnH, hydN, prpB, pspG, norV* and *rutA* are the genes where the σ^{54} binding sites are located in the conventional promoter region upstream of the genes' coding sequences. These are genes with known σ^{54} promoters and also upregulated in the SABRS. Therefore their transcription is likely to be mediated by σ^{54}-dependent gene expression via $\sigma^{54}{}_{\Delta R1}$, highlighting them as potential genes for SABT *in vivo*. *CspH*, *prpB* and *pspG* were studied later for putative SABT *in vitro* (Fig 6.2.2, 6.2.5, 6.2.7.)

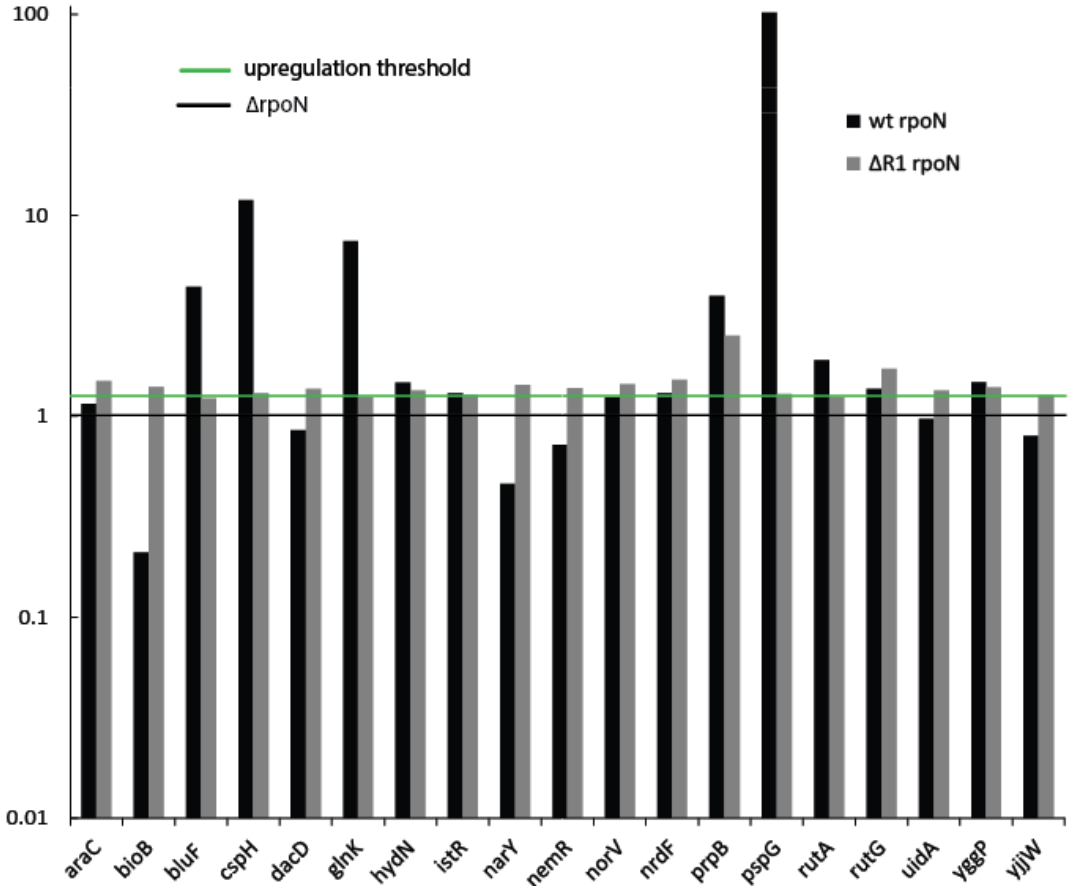


Fig. 5.2.7: Upregulation of genes with σ^{54} binding sites in the SABRS.

19 genes upregulated 25% or more in the SABRS have ChIP σ^{54} binding motifs. 6 of these binding sites are in the promoter region of the genes and could therefore help in promoter recognition and transcriptional activation (eg. SABT) of these genes (*glnH, hydN, prpB, pspG, norV, rutA*). *cspH* had a σ^{54} binding site but no consensus sequence was identified. The remaining ChIP binding sites are intragenic (eg. *bioB*) or downstream of the gene (eg *rutG*), and thus less likely to contribute to SABT.

5.2.8 Upregulation of genes with σ^{54} binding sites in WTRS

14 out of 145 genes with σ^{54} ChIP binding sites were found to be upregulated, 9 being genes with known σ^{54} promoters (Fig 5.2.8). This indicates σ^{54}-dependent expression of these genes occurs in the WTRS. Of particular interest are the cold shock genes (*cspA* and *cspG*) which have σ^{54} binding sites. They are strongly upregulated in the WTRS but have no consensus σ^{54} binding sequences in their promoter region (Bonocora, unpublished).

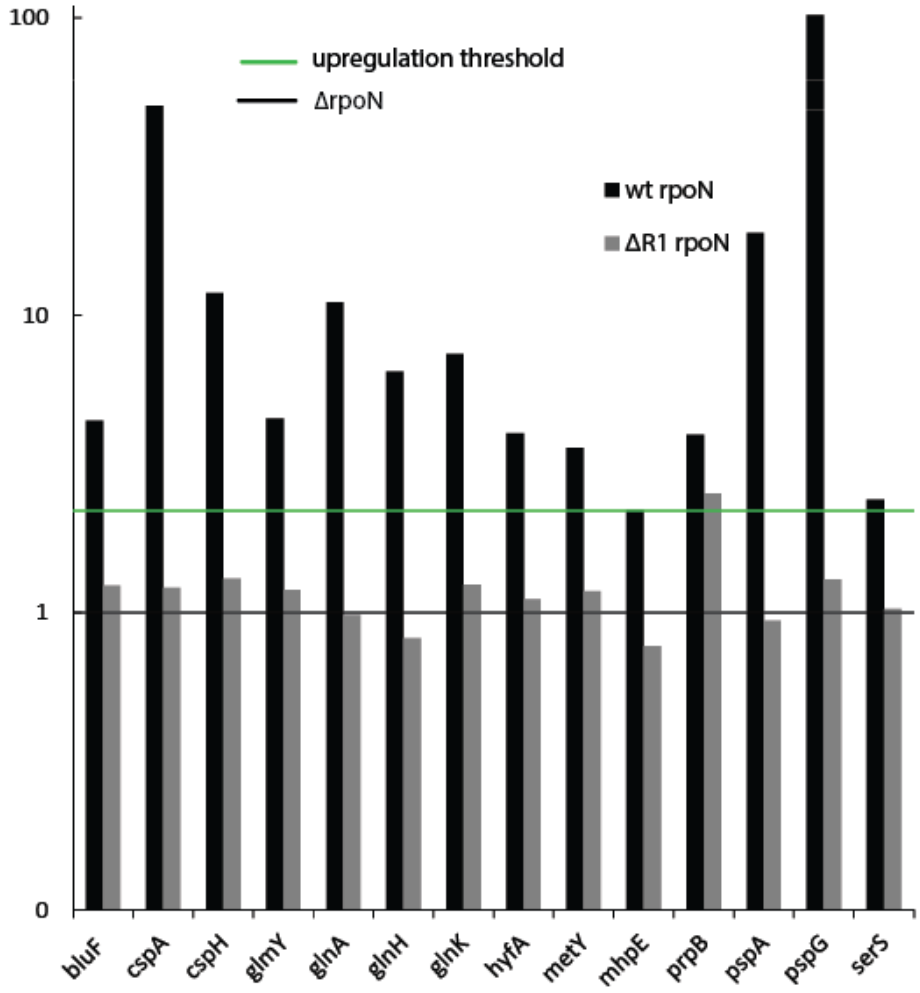


Fig 5.2.8: Upregulation of genes with σ^{54} binding motifs in WTRS.

14 genes with σ^{54} ChIP binding motifs are upregulated 2-fold or more in the WTRS. 9 of these have known σ^{54} promoters (*glmY, glnA, glnH, glnK, hyfA, metY, prpB, pspA, pspG*). *CspA*, *cspH* and *metY* have σ^{54} binding sites but no consensus binding sequence. The binding motif for *bluF* is intragenic and for *mphE* is intergenic antisense.

5.2.9 Repression of genes with σ^{54} binding sites in WTRS

While σ^{54} is typically involved in the transcriptional activation of genes, 28 genes with σ^{54} binding sites were downregulated 40% or more in the WTRS but not in the SABRS (Fig 5.2.9). Since the expression is higher in the SABRS and DRRS, their transcription is presumably σ^{54}-independent (eg. σ^{70}-like transcription). This may suggest a potential role of σ^{54} in transcriptional repression.

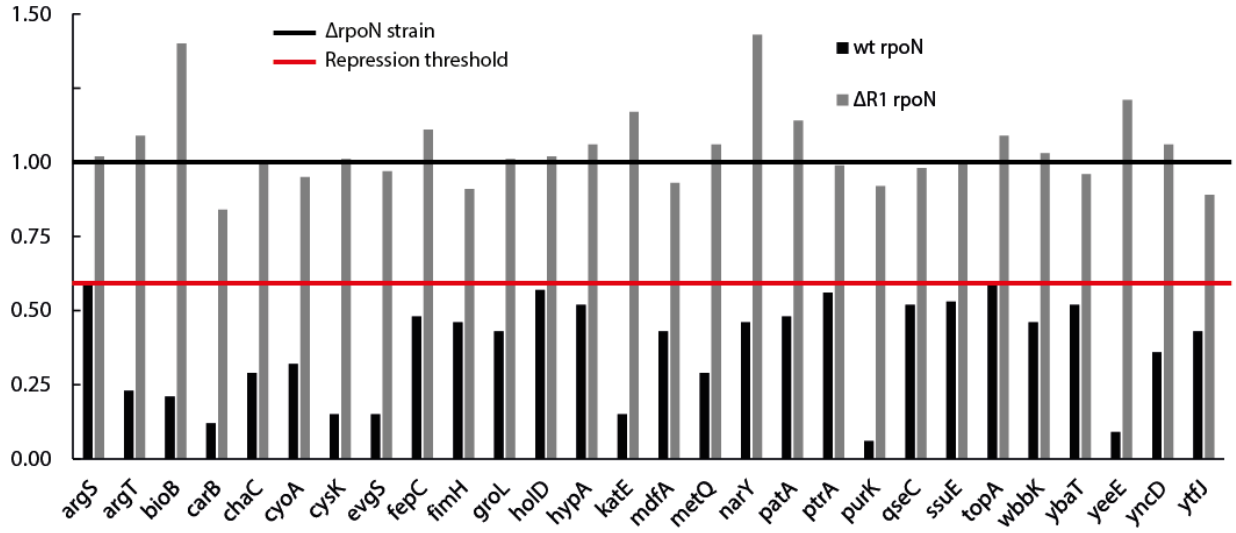


Fig 5.2.9: Repression of genes with σ^{54} binding sites in WTRS.

28 genes with σ^{54} binding sites are repressed 40% or more in the WTRS, but the same genes are not repressed in the SABRS. 7 of these genes have binding motifs in the upstream region of the gene (*argT, chaC, hypA, mdfA, patA, topA, ytfJ*). 3 genes have binding sites but no consensus sequences (*cyoA, ssuE* and *yeeE*). The remaining genes have intragenic motifs, except *katE* where the binding site is downstream of the gene.

5.2.10 Repression in WTRS of genes with σ^{54} binding sites upstream of gene

The 28 genes with σ^{54} ChIP binding sites were found to be downregulated in the presence of wild type σ^{54} (Fig 5.2.9). For 7 of those downregulated genes, their respective σ^{54} ChIP binding sites mapped very close to the transcriptional start site (Fig 5.3.6). In most of these genes, transcriptional repression was seen at both the transcription start site and throughout the entire gene (Fig 5.2.10). This suggests σ^{54} binding to the promoter region of some genes may result in transcriptional inhibition.

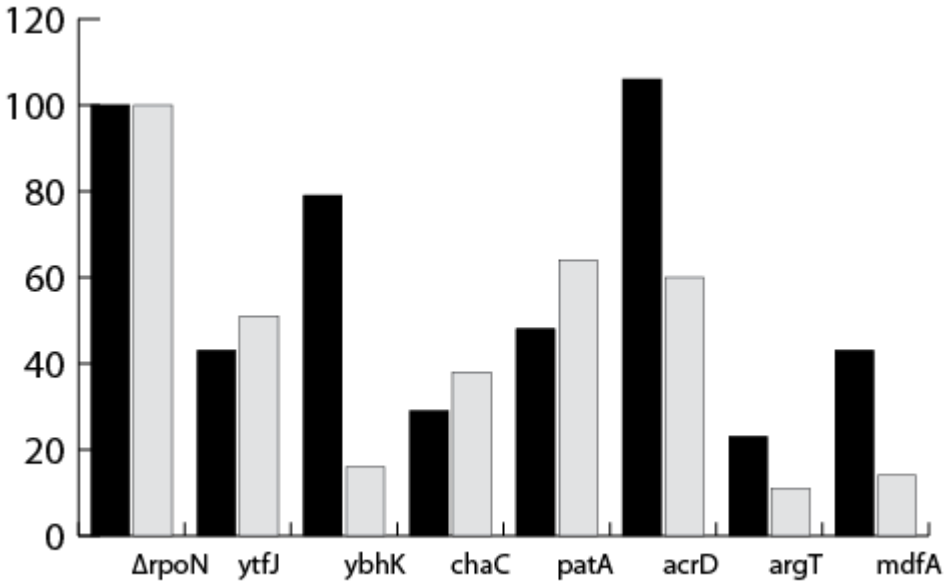


Fig 5.2.10: Putative transcriptional inhibition by σ^{54} binding near transcription start site.
Several genes with σ^{54} ChIP binding motifs upstream of the translation start site display transcriptional repression in the WTRS, but not SABRS or DRRS (see RNASeq data, Appendix B). Reduced transcriptional activity was observed over the entire length of the gene (black) and at the experimental transcription start site (grey). While most of these genes display inhibition both locally and over the entire gene length (*ytfJ, chaC, patA, argT and mdfA*), in some cases this repression appears to be mostly localised to the transcriptional start site (*ybhK* and *acrD*).

5.2.11 Putative repressive σ^{54} binding sites and local repression of transcription

The putative inhibitory σ^{54} binding sites upstream of genes (identified in 5.2.10) map to within 30 bases upstream (Fig 5.2.11a, Fig 5.2.11b) or downstream (Fig 5.2.11c) of the experimental transcription start sites. Moreover, the RNASeq data also displays a correlation between the presence of these putative inhibitory σ^{54} binding sites and repression of proximal experimental transcription start sites (Fig 5.2.11a, Fig 5.2.11b, Fig 5.2.11c, RNASeq supplementary data).

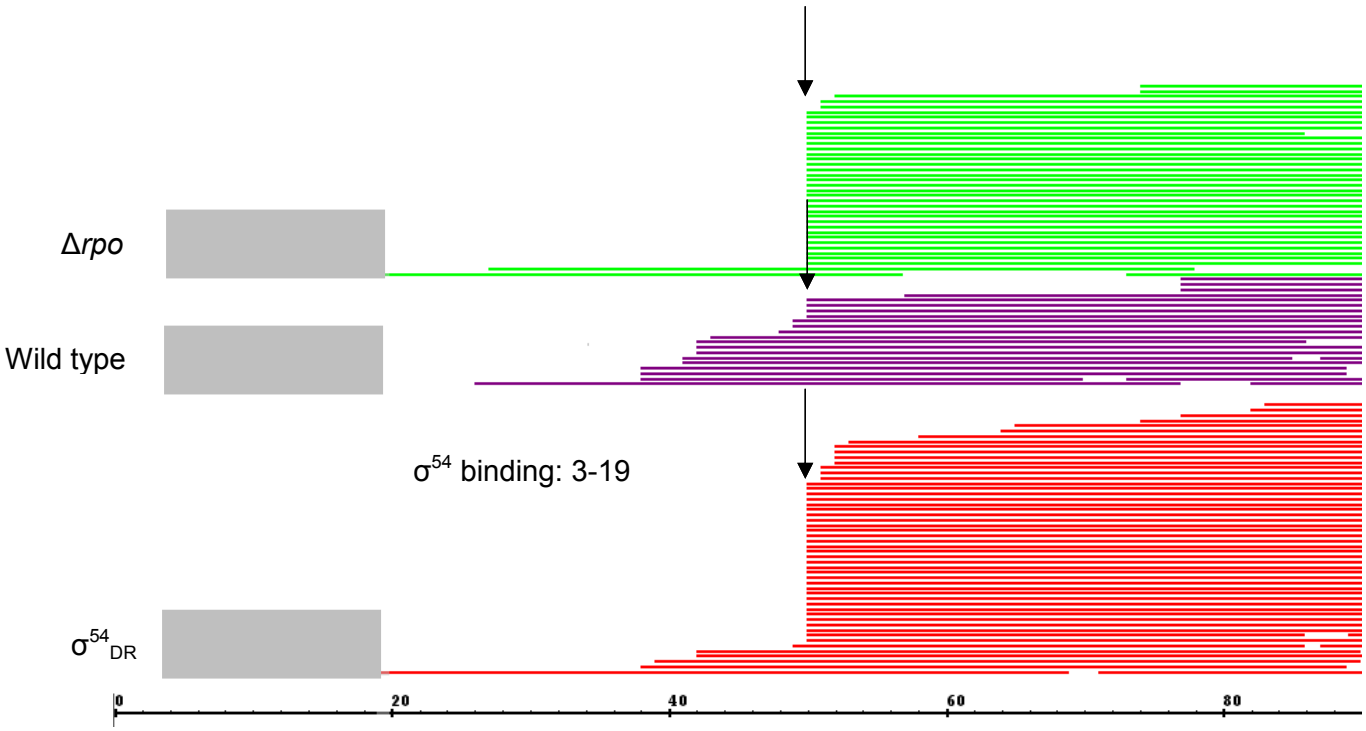


Fig 5.2.11a: Local repression of transcription at σ^{54} binding site in *ytfJ* promoter region.

Transcription of the promoter region of *ytfJ* (100 bases upstream of translational start site) in WTRS (purple), SABRS (red) and DRRS (green). There is a decrease in the transcription directly downstream of the σ^{54} motif (grey) in the WTRS, but not in the DRRS and SABRS (arrowed). The experimental transcription start site maps to ±coordinate 50 (arrowed), therefore the σ^{54} binding would cover the -35 promoter region of σ^{70}-like transcription factors (coordinate 15).

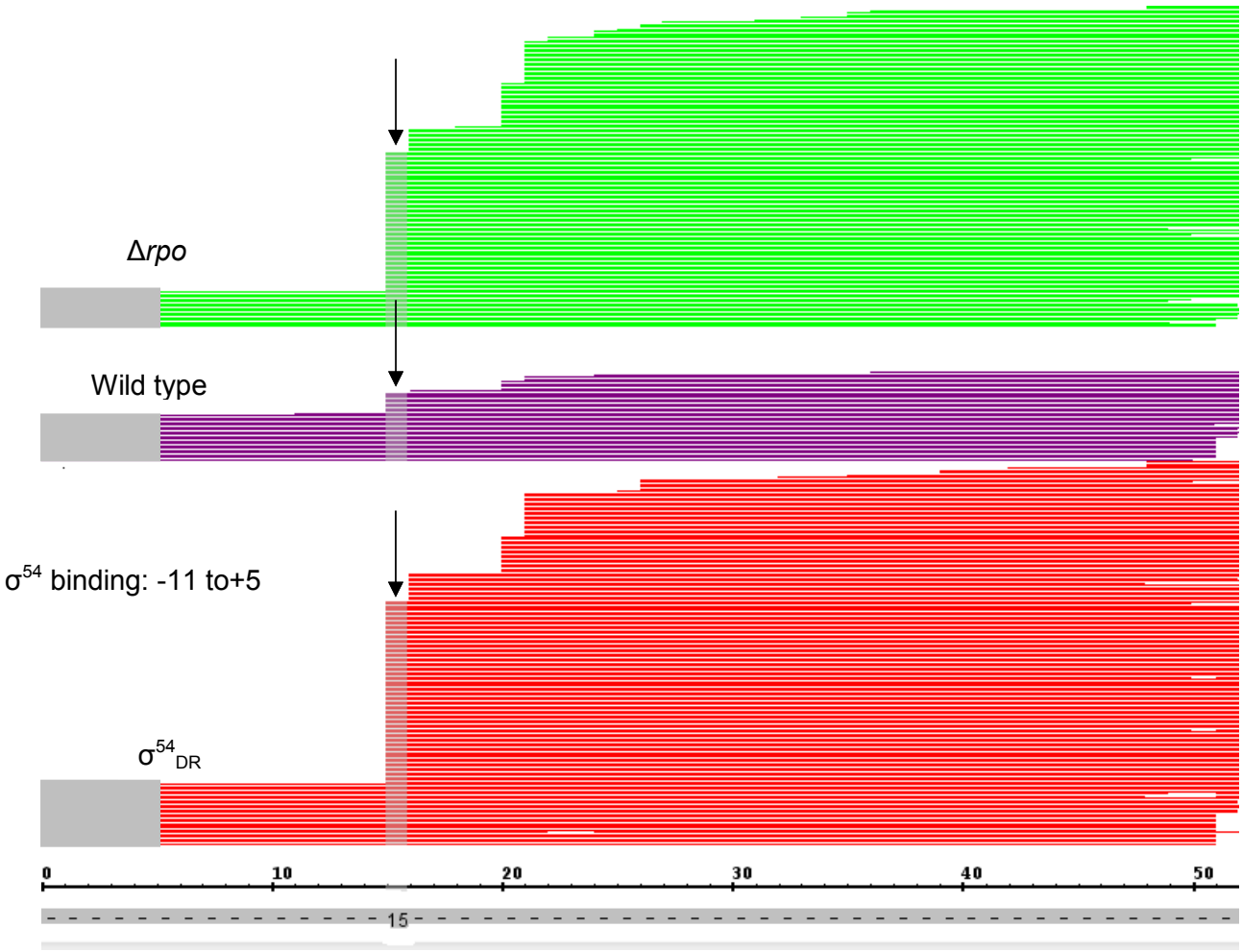


Fig 5.2.11b: Local repression of transcription at σ^{54} binding site in *chaC* promoter region.

Transcription of the promoter region of *chaC* (100 bases upstream of translational start site) in WTRS (purple), SABRS (red) and DRRS (green). There is a decrease in the transcription directly downstream of the σ^{54} motif (grey) in the WTRS, but not in the DRRS and SABRS (arrowed). The experimental transcription start site maps to ±coordinate 15 (arrowed), therefore the σ^{54} binding would cover the -10 promoter region of σ^{70}-like transcription factors (coordinate 5). It may also cover the -35 promoter sequence (coordinate -20) since σ^{54} is usually bound to RNAP as a larger Eσ^{54} complex.

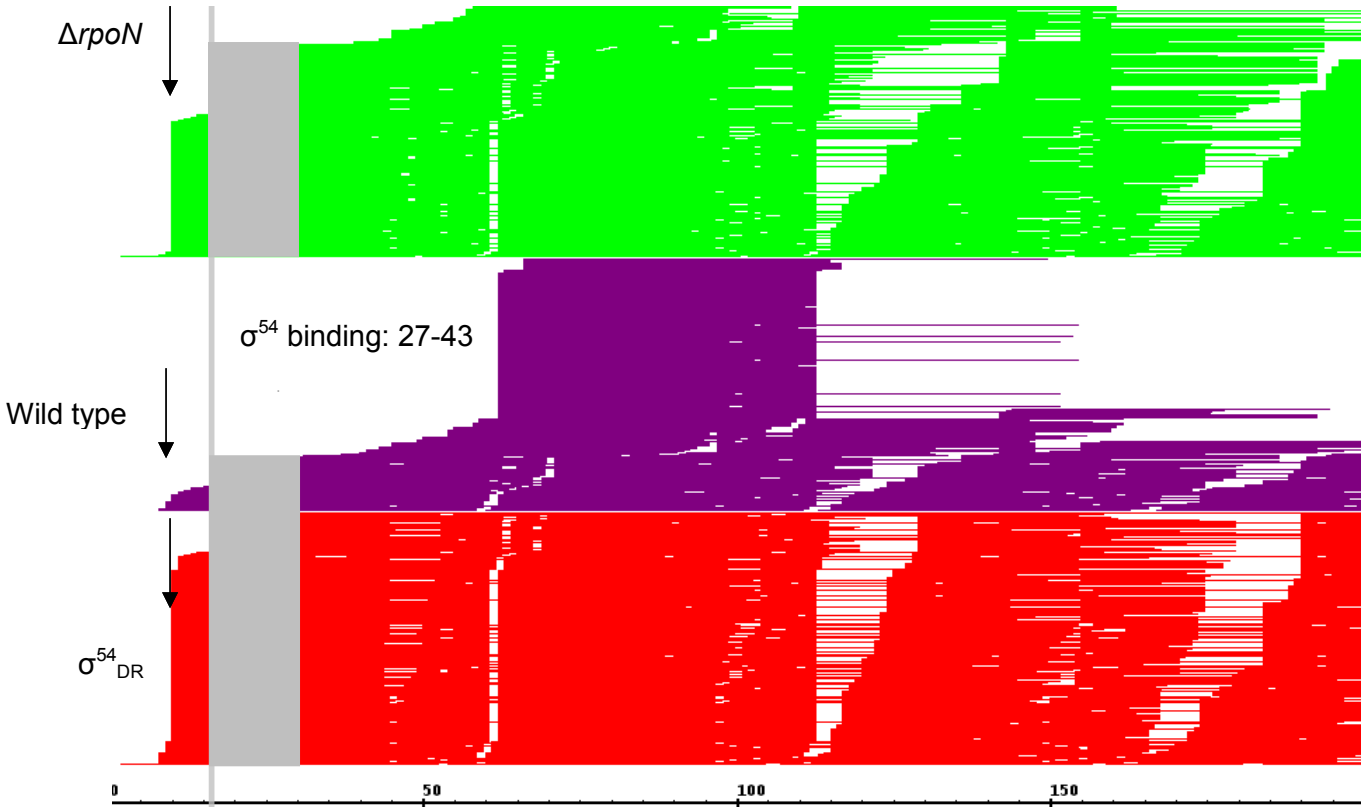


Fig 5.2.11c: Local repression of transcription at σ^{54} binding site in *argT* promoter region.

Transcription of the promoter region of *argT* (100 bases upstream of translational start site) in WTRS (purple), SABRS (red) and DRRS (green). There is a slight decrease in the transcription directly upstream of the σ^{54} motif (grey) in the WTRS, but not in the DRRS and SABRS (arrowed). The experimental transcription start site maps to ±coordinate 10 (arrowed), therefore the σ^{54} binding would cover the region of DNA directly upstream of the transcription start site.

This repression of proximal transcription start sites is observed for all 7 genes with putative inhibitory sites upstream of genes (*acrD*, *mdfA*, *patA*, and *yhbK* not shown; see RNASeq supplementary data). Since this repression can only be scored when the genes are activated in the SABRS and DRRS (and repressed in WTRS), these 7 genes represent a lower limit of the number of genes that could be a target of σ^{54}-dependent repression of σ^{70}-like gene expression.

5.3 Discussion

5.3.1 Gene upregulation by wild type σ^{54} and activator bypass $\sigma^{54}_{\Delta R1}$

σ^{54} is known to regulate the expression of many genes, including those involved in: nitrogen regulation, carbon source utilisation, fermentation pathways, flagellar synthesis, biofilm formation, and bacterial virulence (Reitzer, 2001). Here, several genes from these pathways were found to be upregulated (*intG, yfcQ, yfcS, ytfA*) or downregulated (*csrB, yffO, yiaL*) in the presence of both wild type σ^{54} and $\sigma^{54}_{\Delta R1}$.

11 of the 18 known σ^{54} promoters (Zhao, 2010) were upregulated in the presence of wild type σ^{54}, under these RNASeq conditions. These include: *astC*, *gln* operon, *hydN*, *ibpB*, nac, *prpB*, *psp* operon, *rtcB*, *zraP*, and the ΦP_{pspA}-*lacZ* fusion. *prpB* was the gene with a known σ^{54} promoter that was most upregulated in the presence of σ^{54} and activator bypass $\sigma^{54}_{\Delta R1}$, highlighting *prpB* as a candidate for SABT. To confirm *prpB*'s potential role in SABT, the *prpB* promoter was later assayed for SABT *in vitro* using purified components (Fig 6.2.2).

RNASeq also confirmed the σ^{54}-dependent upregulation of 7 from 21 genes with putative σ^{54} promoters, based on microarray data and primer extension assays (Zhao, 2010). These include *abgB, aslB, emrD, rhaD, sfmF, yaiS* and *ycdM*. RNASeq data for one of these predicted promoters (*yahE*) showed almost 2-fold upregulation in the presence of activator bypass $\sigma^{54}_{\Delta R1}$. To verify putative SABT of this gene *in vivo*, the *yahE* promoter was later assayed for SABT *in vitro* (Fig 6.2.4).

5.3.2 Putative repressive σ^{54} binding motifs around transcription start sites

Section 5.3.1 covered some of the known gene pathways upregulated by σ^{54} *in vivo*. However, earlier RNASeq data (Fig 5.2.10) also pointed towards a potentially novel role of σ^{54} in gene repression. Namely, it appears that 7 genes (*ytfJ, chaC, patA, argT, mdfA ybhK,* and *acrD*) can be identified where σ^{54} binding motifs at transcriptional start sites coincided with gene repression in the WTRS only (Fig. 5.2.10). There appear to be two classes: 1) binding upstream of the transcription start site (Fig 5.3.2, A) and 2) binding downstream of the start site (Fig. 5.3.2, B).

$E\sigma^{54}$ is known to be able to bind very tightly to its consensus sequence to form a closed transcriptionally silent complex which persists in the absence of its cognate activators (Buck, 2000; Guo, 2000). As such, binding of wild type $E\sigma^{54}$ to its consensus sequences could physically block σ^{54}-independent expression of genes. Additionally, σ^{54} binding motifs for some of these repressed genes overlap with the -35/-10 recognition sequences of σ^{70}-like sigma factors (Fig 5.3.2, A). Therefore $E\sigma^{54}$ binding could potentially mask this region from $E\sigma^{70}$. In other cases the sites are directly downstream of the transcription start site, where binding of $E\sigma^{54}$ may obstruct unstable $E\sigma^{70}$ promoter complexes as they extend their interactions with the promoter DNA prior to full initiation and beginning transcription elongation (Fig 5.3.2, B).

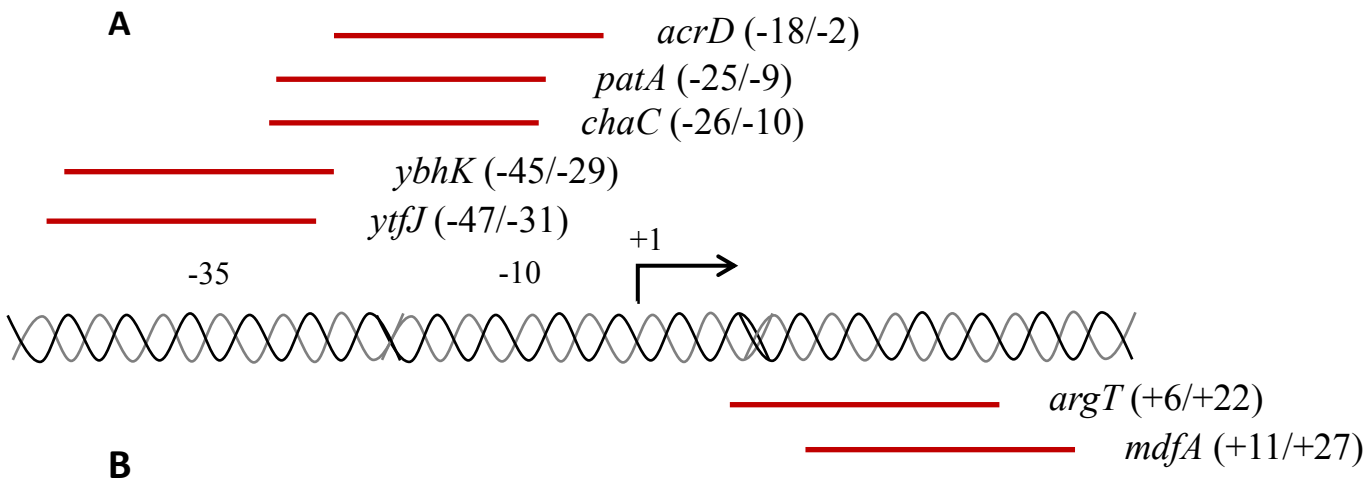


Fig 5.3.2: Relative position of repressive σ^{54} binding sites.

Location of putative inhibitory σ^{54} ChIP binding motifs relative to the experimental transcriptional start site of their respective genes, as determined from RNASeq. Motifs are classified into two groups: downstream of the transcriptional start site (A), and upstream of the transcriptional start site (B).

To verify whether this novel putative inhibitory function of σ^{54} is plausible, an analysis of the transcription of these genes in the literature was carried out. Ideally, σ^{54}-dependent transcriptional activation from these putative inhibitory sites would not be observed. However, σ^{54} could conceivably have a dual function as an activator and a repressor at the same locus, depending on the conditions (potentially mediated by the cognate bEBPs). Secondly, σ^{54}-dependent repression of any of these genes in the literature would strongly support the idea that these binding sites are inhibitory in nature.

ytfJ is not a well characterized gene, and its σ^{54}-dependent transcription has not been reported. However, it has been shown to be σ^{24}-dependent (Dartigalongue, 2001). This is line with a hypothesis that *ytfJ* transcription is σ^{54}-independent and that σ^{54} binding could repress σ^{24}-dependent transcription.

A *chaC* promoter *gfp* fusion construct was shown to display activity *in vivo* in rich media, but was inhibited under nitrogen limiting conditions (Zaslaver, 2006). RNASeq data also displayed *chaC* repression in the minimal Gutnick media (nitrogen limiting). If *chaC* inhibition is restricted to nitrogen conditions (ie. an NtrC-dependent inducing condition for σ^{54}), it would strengthen the hypothesis that σ^{54} could result in repression of some genes (eg. *chaC*).

PatA has been reported to be transcribed by σ^{38} and was shown to compete with σ^{54} for binding to promoter DNA (Schneider, 2013). Additionally, a loss of σ^{54} increased expression four-fold in nitrogen-excess medium (Schneider, 2013). Both a loss of σ^{54} or nitrogen rich conditions would reduce σ^{54} activity, and thereby limit any putative repression. Given that a four-fold increase in *patA* transcription is observed under these conditions, it supports the hypothesis that σ^{54} binding may inhibit *patA* expression.

Despite *argT* having a predicted σ^{54} promoter, thus far no experimental evidence has been reported to confirm σ^{54}-dependent transcription of *argT* (Barrios, 1999). This is in agreement with the idea that this σ^{54} binding site functions as a transcriptional repressor rather than an activator. No σ^{54}-independent transcriptional activation of *argT* has been reported thus far,

however the RNSeq data suggests this does occur since *argT* is transcribed in the absence of σ^{54} (see Appendix B).

While the *mdfA* transcription start site has been confirmed experimentally, no σ factor promoters have been identified as of yet (Eguchi, 2003). The promoter region of *mdfA* is not well studied, as such the literature does not confirm nor disprove the potential role of σ^{54} binding in *mdfA* inhibition. However, the σ^{54} binding site is downstream of the transcription start site suggesting that its expression is not σ^{54}-dependent.

acrD has a known σ^{70} promoter (Huerta, 2003) and also a predicted σ^{54} promoter (Zhao, 2010). The expression of *acrD* was shown to double upon induction of plasmid-borne σ^{54} (Zhao, 2010). However, it should be noted that σ^{54} overexpression in this study was induced with anhydrotetracycline. AcrD is part of a multidrug exporter and as such it is quite possible that upregulation of *acrD* was due to addition of anhydrotetracycline rather than σ^{54} expression (Aires, 2005). Therefore, while σ^{54}-dependent expression of *acrD* has been proposed in the literature once, additional factors would suggest that this may not be σ^{54}-dependent after all. As such, the putative role of σ^{54} binding in local *acrD* repression remains plausible.

ybhK has been identified as an FtsK-binding protein, yet little is known about its transcription (Butland, 2005; Arifuzzaman, 2006; Hu, 2009). Like *acrD*, *ybhK* transcription has been suggested to be upregulated upon σ^{54} induction, and in the same study (Zhao, 2010). It is now known that *ybhK* activated is induced by AcrA, which is part of the acrAD multidrug efflux system (Federowicz, 2014). As such, the induction of σ^{54} with anhydrotetracycline in the Zhao 2010 study has likely resulted in incorrectly scoring the transcription of both of these genes as σ^{54}-dependent. No other studies have been able to confirm the expression of either of these genes as being σ^{54}-dependent. Therefore it remains reasonable that that these σ^{54} binding sites could be repressive, rather than increasing transcription.

Overall, σ^{54}-dependent transcription has not been shown conclusively at any of the genes with putative inhibitory σ^{54} binding sites near their transcription start sites. In some cases, the

literature has reported a positive correlation between conditions inducing σ^{54} activity and repression of genes with putative inhibitory σ^{54} binding sites (*chaC* and *patA*). Combined, these findings support the initial hypothesis that the putative σ^{54} binding motif for certain genes is repressive.

5.3.3 Alignment of σ^{54} consensus sequences involved in activation, SABT and repression

RNASeq data was examined at 145 σ^{54} ChIP binding sites (5.2.6-5.2.9). 56 out of 145 genes displayed upregulation or repression of transcription in the WTRS, SABRS, or both. Several genes with known σ^{54} binding sites were upregulated in the WTRS (Fig 5.2.8) and SABRS (Fig 5.2.7 and Fig 5.2.8).

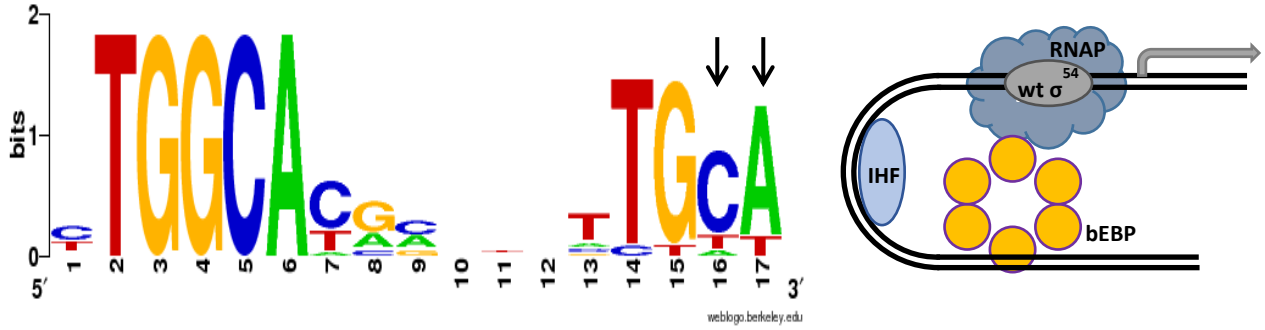
However, no such correlation was observed for repression in the WTRS. The genes inhibited in the WTRS were not also downregulated in the SABRS (Fig 5.2.9). If any repression is caused by tight binding of $E\sigma^{54}$ (Fig 5.2.11a; 5.2.11b; 5.2.11c), this is not observed with $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant. σ^{54} activator bypass mutants often disrupt the interaction between σ^{54} and -12 promoter element (Wang, 1997). This may weaken the tight binding of $E\sigma^{54}$ to DNA and thereby impact its putative role in repression σ^{70}-like transcription.

If σ^{54} binding does inhibit transcription it would suggest a novel role for σ^{54} in transcription, as a transcriptional repressor of σ^{70}-like transcription. While σ^{54} has been reported to result in repression of σ^{70}-like transcription (Zafar, 2014), this would be the first report of it doing so by physically blocking σ^{70}.

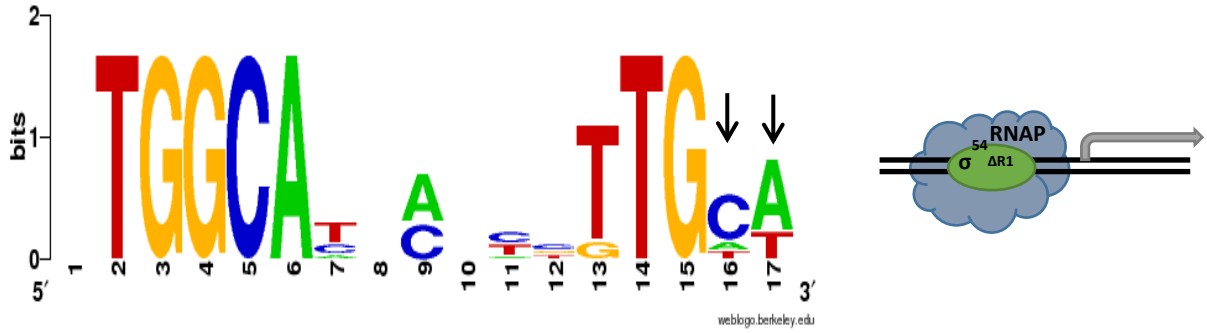
It has long been a subject of debate as to why σ^{54} binds tightly as a transcriptionally silent complex, requiring activator and ATP to initiate transcription. This is in stark contrast to other sigma factors which spontaneously result in open transcriptionally active complex. This dual role as a potential transcriptional repressor may shed some light as to why σ^{54}-dependent transcription has not evolved to be spontaneous for making open promoter complexes like other σ factors. Indeed it is possible that σ^{54} was originally a repressor of transcription and only later became a subject of activation.

Alignment of σ^{54} consensus sequences was analysed to determine whether the transcriptional activation and repression are targets of the same σ^{54} consensus sequences (Fig 5.3.3).

A **Consensus sequence at sites with wild type $E\sigma^{54}$ activity**



B **Consensus sequence at sites with $E\sigma^{54}{}_{\Delta R1}$ bypass activity**



C **Consensus sequence at sites showing inhibition by $E\sigma^{54}$**

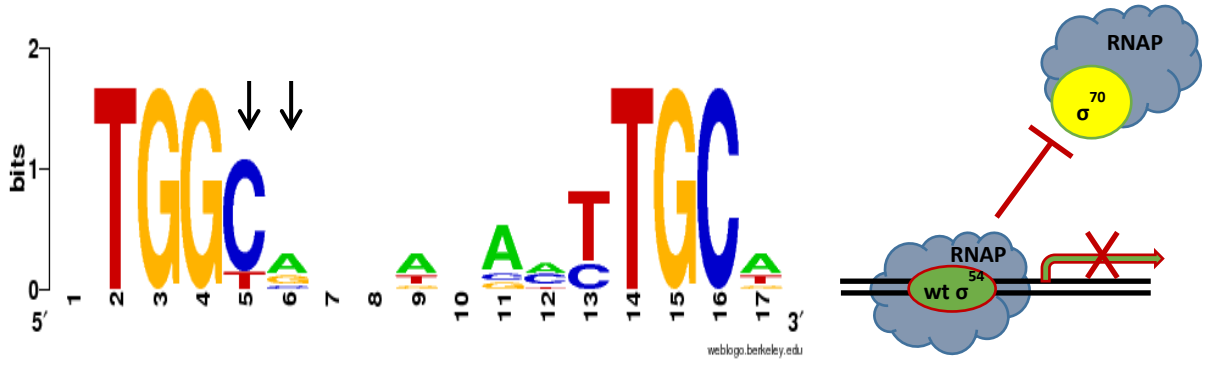


Fig 5.3.3: Alignment of σ^{54} binding sites.

Alignment of consensus sequences for σ^{54} binding at: σ^{54}-dependent activated promoters (**A**), activator bypass promoters (**B**) and promoters repressed by σ^{54} (**C**). For (**A**) and (**B**) vs. (**C**), the consensus sequences are distinct at the -12 and -24 elements, arrowed. Number of sequences aligned were 14 (**A**), 7 (**B),** and 7 (**C**) respectively.

The consensus sequences for activation of WTRS (Fig 5.3.3, A) and SABRS (Fig 5.3.3, B) are virtually identical, with very strongly conserved -24 regions but less conserved -12 sequences (arrowed). Therefore, the -24 region may play a more important role in transcriptional activation via σ^{54}, since it is more conserved where activation is detected. Indeed, activator dependent σ^{54}-dependent transcription can still be seen where the -12 region has been disrupted (Wang, 1997).

The consensus sequences preserved in repression are slightly different (Fig 5.3.3, C), with more conserved -12 sequences but less conserved -24 sequences (arrowed). This may suggest the -12 promoter sequence is more important for repression as it is more conserved at sites displaying inhibition. Indeed this sequence is known to be important for tight binding of $E\sigma^{54}$ to form a transcriptionally silent complex (Buck, 200). Disruption of this motif results in σ^{54} activator bypass transcription (Wang, 1997). Additionally, substitutions in the -12 TTGC tetranucleotide indicated the consensus C is important for keeping basal transcription in check (Wang, 1999).

5.3.4 σ^{54} and its potential role in the transcription of cold shock genes

The RNASeq experiment linked σ^{54} to the upregulation of multiple genes with known σ^{54} promoters, as well as confirming some predicted σ^{54} promoters (5.3.1). Some σ^{54} binding sites also coincided with σ^{54}-dependent repression of genes, suggesting a possible inhibitory role for σ^{54} (5.3.2). Furthermore, σ^{54} expression also resulted in the upregulation of several genes involved in the cold shock response (Fig 5.2.5 and 5.2.8). Here, σ^{54} and its potential role in the transcription of cold shock genes is analysed further.

Bacteria express a well-defined set of proteins after a rapid decrease in temperature, which differ from those expressed under heat shock conditions. The effects of the cold shock response is seen at several levels: decrease in membrane fluidity, stabilising secondary DNA and RNA structure, slow and inefficient protein folding and adaptation of ribosome to function at low temperatures (Phadtare, 2004).

Unlike the heat shock response which is regulated by σ^{32}, as of yet no cold-shock specific sigma factor has been identified (Weber, 2003). However, σ^{54} has been shown to be required for cold shock response in *Bacillus subtilis* (Wiegeshoff, 2006).

The major cold shock response protein in *E. coli* (CspA) and its homologous proteins (cspB, cspG, cspI) appear to be significantly upregulated in the WTRS, compared to the SABRS and a DRRS (Table 5.3.4). These 4 upregulated cold shock genes are cold shock inducible (Ivancic, 2013), unlike cspC, cspD and cspE which are not cold shock-inducible and not upregulated, Table 5.3.4 (Lee, 1994; Etchegaray, 1996). *ydfK* and *ynaE* are two further genes typically upregulated by cold shock (Polissi, 2003) and also upregulated in the WTRS (Fig

5.2.5). Combined, this strongly suggests that the upregulation of these cold shock genes observed in the RNASeq is indeed cold shock-dependent.

Gene	wt σ^{54} strain	$\sigma^{54}_{\Delta R1}$ strain
cspA	↑↑↑	=
cspB	↑↑↑	=
cspC	↓	=
cspD	=	=
cspE	=	=
cspF	↑	=
cspG	↑↑↑↑	=
cspH	↑↑	=
cspI	↑↑↑	=

Table 5.3.4: Transcription of cold shock response genes in presence of σ^{54} and $\sigma^{54}_{\Delta R1}$.

While a functional link between σ^{54} and the cold shock response could explain why the operon is upregulated in the WTRS, strains for the RNASeq experiment were grown in minimal media at 42°C. However, the RNASeq sample preparation protocol may provide an explanation for the apparent cold shock induction. Samples were grown at 42°C for several hours, then placed on ice and immediately lysed with a cold solution (95% methanol, 5% phenol) for phenol extraction prior to centrifugation. While this step took less than 5 minutes and all samples were treated equally, it is possible that this step in the RNASeq protocol may have transiently and inadvertently induced the cold shock response. This could result in the brief and strong upregulation of cold shock genes in the RNASeq strains able to mount a cold shock response.

Not only is the cold shock response upregulated in the presence of wild type σ^{54} (shown to be required for its expression in *Bacillus subtilis*), but global ChIP analysis of σ^{54} binding

showed σ^{54} binding at sites for *cspA* and *cspH* (Bonocora, unpublished). However consensus sequences are not identified, and σ^{54} binding could potentially be somewhat indirect and involve other factors.

Combined, these suggest that σ^{54} could be involved in the transcriptional activation of the cold shock response. To verify to putative σ^{54}-dependent expression *cspA* and *cspH*, the *cspA* and *cspH* promoters were assayed for σ^{54}-dependent transcription *in vitro* in Chapter 6.

5.4 Summary

Global transcriptomic analysis of σ^{54}-dependent gene expression identified 6 genes (Fig 5.2.7) with σ^{54}-dependent promoters displaying putative SABT *in vivo* (*GlnH, hydN, prpB, pspG, norV* and *rutA*). One further SABT candidate, *cspH*, was also identified. While *cspH* has a σ^{54} ChIP binding motif, no σ^{54} consensus sequence has been characterised.

RNASeq data also pointed towards a potential novel role for σ^{54} as a transcriptional repressor. Here 7 genes (*ytfJ, chaC, patA, argT, mdfA ybhK,* and *acrD*) were identified with σ^{54} binding motifs in their promoter region and displayed repression in the presence of σ^{54}, but not with $\sigma^{54}{}_{\Delta R1}$ or without σ^{54} (Fig 5.2.9). Moreover, these putative repressive σ^{54} binding sites resulted in local transcriptional repression for these genes (5.2.10). Furthermore, these findings are compatible with the current literature regarding the transcriptional regulation of these genes (5.3.4).

Alignment of σ^{54} motifs (Fig 5.3.3) revealed a more conserved -12 binding sequence for σ^{54} promoters that corresponded to proposed sites for transcriptional inhibition, suggesting this motif may be important for the putative repressive function of σ^{54}. The -24 motif was more conserved for σ^{54} promoters that corresponded to σ^{54}-dependent upregulation of genes, suggesting the -24 motif may play a more important role in transcriptional activation. Little difference was found between promoters known to be involved in σ^{54}-dependent upregulation of genes and those involved in putative SABT *in vivo*.

Finally, upregulation of several genes involved in the cold shock response was observed in the presence of wild type σ^{54}, but not $\sigma^{54}{}_{\Delta R1}$ or without σ^{54} (Fig 5.2.5). Two of these genes (*cspA* and *cspH*) by ChIP seq have σ^{54} binding activity but lack consensus sequences,

suggesting a possibly novel role for σ^{54} in the σ^{54}-dependent transcription of cold shock response genes.

Chapter 6: Transcription of cold shock genes and SABT *in vitro*

6.1 Introduction and objectives

6.1.1 Recap and objectives

In 5.2.6, increased transcription was detected in the presence of $\sigma^{54}{}_{\Delta R1}$ at several genes with known σ^{54} promoters (*glnH, hydN, prpB, pspG, norV, rutA*). This elevated expression was seen in the presence of the region I (but not always with the full length σ^{54}). The loss of region I is known to increase SABT *in vitro* (Chapter 1), suggesting that these genes may be displaying SABT *in vivo*. These genes have not yet been examined for SABT *in vitro*, raising the question as to whether SABT can be observed at these promoters *in vitro*.

In 5.3.4, several cold shock related genes were upregulated in the strain with wild type σ^{54}. Moreover, two of these cold shock genes (*cspA* and *cspH*) appear to have σ^{54} ChIP binding sites, which could potentially be σ^{54}-dependent promoters driving the transcription of these genes (Bonocora, unpublished). However, as of yet no σ^{54}-dependent transcription of *cspA* or *cspH* has been reported in *E. coli*.

Chapter 6 will focus on examining σ^{54}-dependent transcription (and SABT) *in vitro* at some the aforementioned gene promoters. For the genes displaying putative SABT *in vivo*, these assays could potentially confirm those findings by reproducing them *in vitro*.

For the cold shock genes these assays may show that wild type σ^{54} activates their expression *in vitro*, through a putative σ^{54}-dependent promoter (supporting *in vivo* RNASeq and ChIP data). σ^{70}-dependent transcription of *cspA* has already been shown in *E. coli* (Goldstein,

1990; Tanabe, 1992), yet overlapping and active σ^{54} and σ^{70} promoters have been reported (Reichenbach, 2009).

Alternatively, these assays could show that σ^{54}-dependent transcription of *cspA* and *cspH* does not occur *in vitro*. This would suggest σ^{54} upregulates σ^{70}-dependent transcription of cold shock genes in some other way. Positive regulation of σ^{70} from a divergent σ^{54} promoter has been described in *Pseudomonus putida* (Johansson, 2008). On the other hand σ^{54}'s upregulation of σ^{70}-like transcription could be indirect, by influencing expression proteins like primary heat shock sigma factor (σ^{32}), or a transcriptional regulator like NtrC which is thought to regulate as much as 2% of the genome (Zimmer, 2000). Thus it is also quite possible some key components are missing *in vitro*.

6.1.2 Abortive initiation assays

The synthesis of a short RNA transcript (a type of abortive initiation assay) is a technique commonly used to study σ^{54}-dependent transcription *in vitro*, including SABT (Tintut, 1995; Gralla, 1996; Burrows, 2010; Wiesler, 2012; Lawton, 2014). It is slightly different from the typical abortive assay in that not all 4 rNTPs are not provided, forcing a short transcript to be produced.

In this assay σ^{54} and RNAP are incubated with promoter DNA to form a transcriptionally inactive closed complex. The tight interaction between σ^{54} and the DNA just downstream of the GC promoter element at -12 typically prevents access of single-stranded DNA to the active site of core RNAP (Chapter 1).

Next ATP, $PspF_{1\text{-}275}$, RNA dinucleotide (-1, +1) primers and radioactive nucleotides (+2) are added to the reaction. ATP and bEBPs (eg PspF) catalyse the steps required for stable association of the holoenzyme with melted DNA to produce an open, stable and transcriptionally active complex (Wedel, 1995). Should stable open and transcriptionally active promoter complexes form under these conditions, short radioactive transcripts are produced which can subsequently be visualised and quantified on small denaturing gels, resolved from the radioactive nucleotide (Fig 6.1.1). Consequently, abortive initiation assays can be a good indicator of σ^{54}-dependent transcription initiation.

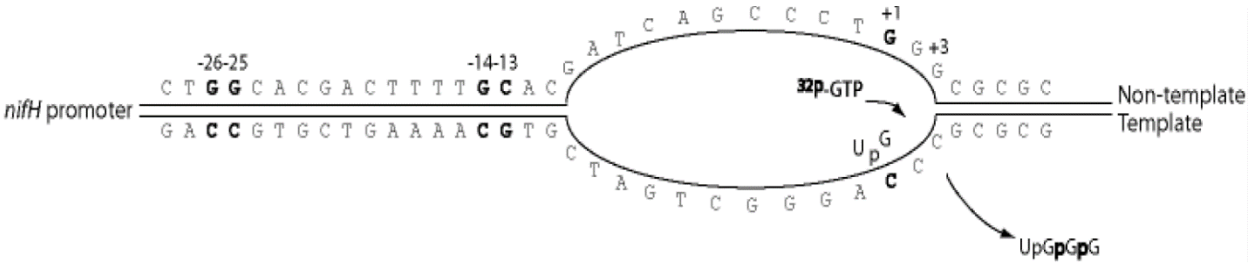


Fig 6.1.1: Abortive initiation assay with *nifH* promoter DNA.

The σ^{54} consensus GG and GC elements for the σ^{54}-dependent *nifH* promoter are highlighted, at -26/-25 and -14/-13 respectively. The transcription start site (G), is in bold. The initiation UpG dinucleotide primer and [α-^{32}P] GTP are added, both complementary to the template strand (-1/+1 and +2 nucleotides respectively). σ^{54}RNAP oscillates between the -1 and +3 sites, generating UpG**p**G**p**G (^{32}P highlighted in bold). The transcript products are then run on a 20% urea gel at 300V. The gels are then exposed to an imaging plate for 30 minutes, visualised on a phosphoimager, and quantified on AIDA Image Analyzer.

6.2 Results

6.2.1 Abortive initiation assay with *nifH* promoter DNA

A σ^{54}-dependent *nifH* promoter *(Sinorhizobium meliloti*) is known to form closed complexes with σ^{54}RNAP, which are unable to isomerise into open complexes without the assistance of activators (Buck, 1992). Moreover, the *nifH* promoter is known to result in high levels of SABT *in vitro* with some σ^{54} activator bypass mutants (Buck, unpublished; Chaney, 1999). Therefore abortive initiation assays using the *nifH* promoter should show that this assay can readily detect: 1) SABT with $\sigma^{54}{}_{\Delta R1}$ and 2) $PspF_{1\text{-}275}$-dependent activation of a σ^{54} promoter with wild type σ^{54}.

Abortive assays for the *nifH* promoter were run with heparin to reduce background activity. Heparin is a highly negatively charged molecule that acts as a competitor to DNA and prevents transcription reinitiation by disrupting unstable open promoter complexes (Sümegi, 1977). The *nifH* promoter DNA test duplex also contains a mismatched -10/-1 non-template strand (see Chapter 2). This mismatch creates a -10/-1 bubble when forming a heteroduplex with the wild type template strand, mimicking the late-melted promoter during open complex conformation (Sharma, 2014). Typically this late-melted form of promoter DNA facilitates open complex formation (Heilemann, 2007). DNA opening on supercoiled plasmid DNA (and bubble templates) is greater than on linear DNA, and SABT has been found to be higher

with supercoiled DNA templates compared to linear templates (Wang, 1997), and so the -10 to -1 bubble templates typically yield higher levels of SABT.

In line with previous studies, wild type σ^{54}-dependent *nifH* transcription initiation *in vitro* was observed in the presence of $PspF_{1\text{-}275}$ activator but not without it (Fig 6.2.1). Partial activity is observed in the $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant with and without $PspF_{1\text{-}275}$ (Fig 6.2.1), in agreement with the idea that this mutant form of sigma bypasses the activator requirement. Combined, these results show that σ^{54} and $\sigma^{54}{}_{\Delta R1}$ function as expected *in vitro*, and that this assay is a suitable indicator for both SABT and wild type σ^{54}–dependent transcription.

Strikingly, not adding $PspF_{1\text{-}275}$ resulted in a 3x fold increase in σ^{54}–dependent transcription with the $\sigma^{54}{}_{\Delta R1}$ bypass mutant. Thus PspF may interact to drive closed complex formation when an open complex is not going to be made due to defects in $\sigma^{54}{}_{R1}$. The basis of this effect is unknown, but could reside in interactions PspF makes with promoter DNA stabilising RPo (Zhang, 2012; Bose, 2008).

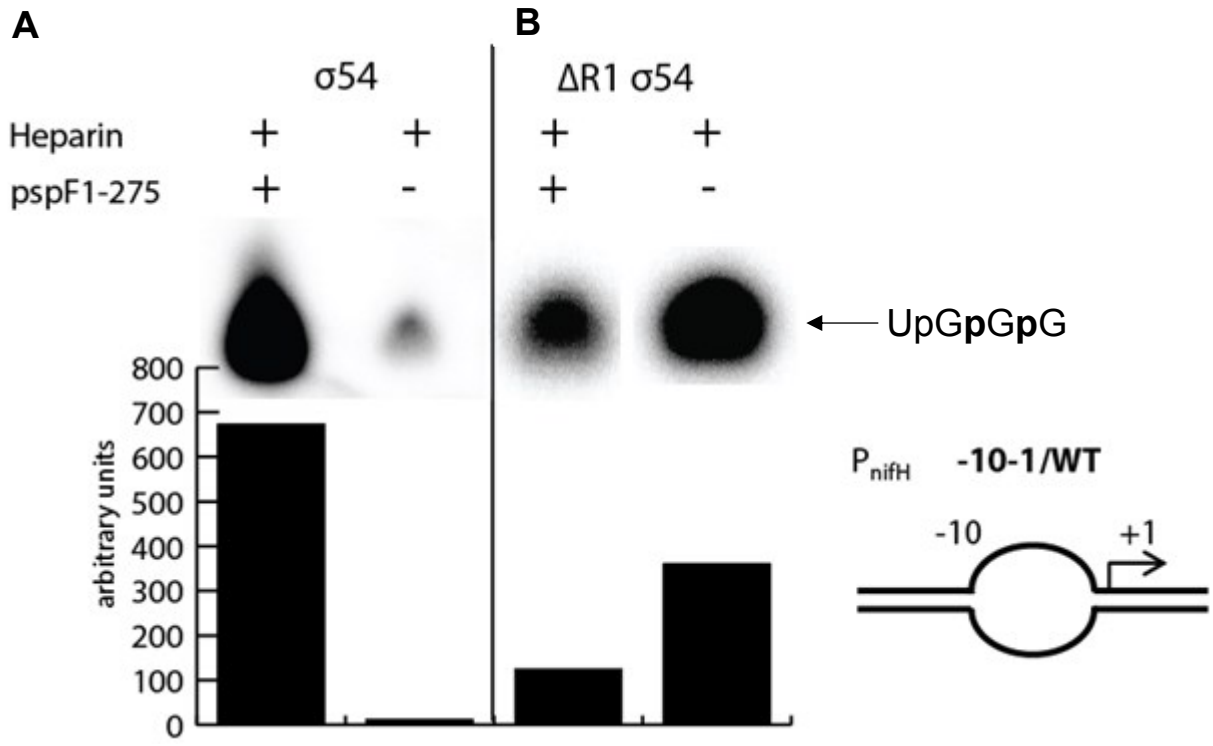


Fig 6.2.1: Abortive initiation assay *in vitro* at the *nifH* promoter with wild type σ^{54} and $\sigma^{54}{}_{\Delta R1}$.

Linear probes covering the -60 to +28 region around the *nifH* transcription start site were used (A), and a variant with a -10/-1 bubble. (B). Experiments were run with wild type σ^{54} and $\sigma^{54}_{\Delta R1}$, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 90 mins with heparin. The UG**GG** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.2 Abortive initiation assay at *prpB* promoter

In 5.2.6, several genes with σ^{54} promoters were highlighted as candidate genes for SABT *in vivo*. *PrpB* was the gene with the highest % increase in transcription with $\sigma^{54}_{\Delta R1}$ (251%), relative to the control strain lacking σ^{54} (Fig 5.2.7).

To confirm whether SABT was indeed being observed at the *prpB* promoter *in vivo* (as seen in the RNASeq data), abortive initiation was studied at the *prpB* promoter *in vitro*. Abortive transcription was studied using linear wild type homoduplex DNA (Fig 6.2.2, A), and heteroduplex DNA containing a -10/-1 mismatch mimicking the late-melted promoter (Fig 6.2.2, B). Abortive initiation assays were performed in the absence of heparin, as activator bypass mutant open complexes have typically been found to be sensitive to heparin (Gralla and Wang, 1996).

With homoduplex DNA low levels of transcription were observed for wild type σ^{54} with activator, but no significant activity was seen without activator or with the $\sigma^{54}_{\Delta R1}$ activator bypass mutant (Fig 6.2.2, A). This suggests detectable levels of σ^{54}-dependent transcription

do occur in the presence of $PspF_{1-275}$ activator, but no SABT is observed with homoduplex *prpB* promoter DNA. However, since activity of wild type σ^{54} and $PspF_{1-275}$ is already low under these conditions, significant levels of SABT would not be expected, and so may occur in living cells but may not be readily detected with purified components and a linear fully base paired DNA template.

With the heteroduplex DNA a 2-fold increase in abortive transcript levels was observed with wild type σ^{54} and $PspF_{1-275}$ compared to no $PspF_{1-275}$ (Fig 6.2.2, B). Significant SABT was also observed with the $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant without $PspF_{1-275}$, exceeding activity of the wild type σ^{54} with activator (Fig 6.2.2, B). These findings support the earlier RNASeq findings suggesting SABT could be occurring at the *prpB* promoter *in vivo*. Interestingly, SABT activity at the *prpB* promoter also appears to be downregulated upon addition of $PspF_{1-275}$ (as observed with the *nifH* promoter DNA).

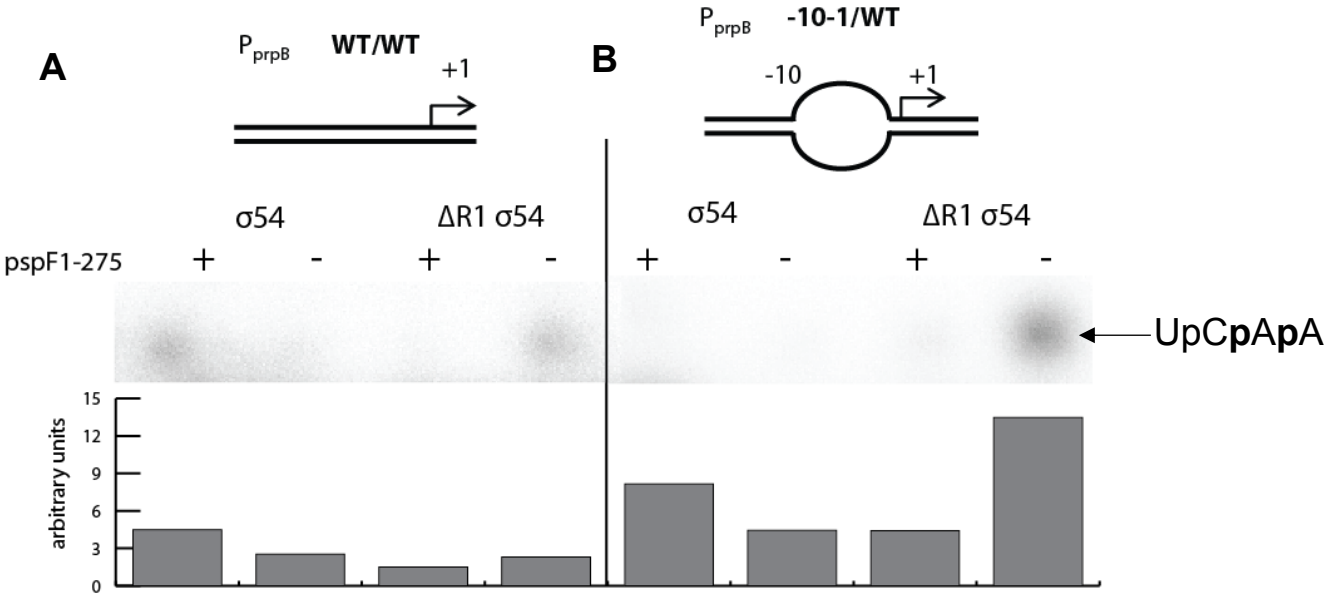


Fig 6.2.2 Abortive initiation assay *in vitro* at the *prpB* promoter with wild type σ^{54} and $\sigma^{54}{}_{\Delta R1}$.
Linear probes covering the -60 to +28 region around the *prpB* transcription start site were used (A), and a variant with a -10/-1 bubble (B). Experiments were run with wild type σ^{54} and $\sigma^{54}{}_{\Delta R1}$, with and without $PspF_{1-275}$. Reactions were incubated for 90 mins without heparin. The UC**AA** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.3 Abortive initiation assay at *hypA* promoter

In 6.2.2 we observed SABT at the *prpB* promoter *in vitro* (but with heteroduplex promoter DNA only). *hypA* is a gene with the same -24 and -12 σ^{54} binding motifs as *prpB* in its promoter region (TGGCACNNNNNNTGCTT), and there is experimental evidence to suggest that *hypA* has a σ^{54}-dependent promoter (Hopper, 1994; Zhao, 2010). Therefore if σ^{54}-dependent SABT was observed the *prpB* promoter, we might also expect to see SABT at the *hypA* promoter *in vitro*.

This hypothesis was tested, and indeed we observe very similar results for the abortive assays for the *prpB* and *hypA* promoters (Fig 6.2.2 and 6.2.3). SABT is observed with the *hypA* promoter heteroduplex DNA (Fig 6.2.3 B). However, simply because the *hypA* and *prpB* promoters behave similarly *in vitro* it does not follow that their transcription will also be similar *in vivo*, as they may be subject to further negative or positive regulation *in vivo*.

In fact, the RNASeq data shows their transcription is not so similar *in vivo* (RNASeq supplementary data). *hypA* is downregulated in the presence of wild type σ^{54} (50%), compared to $\sigma^{54}{}_{\Delta R1}$ (106%) or no σ^{54} (100%). This is in contrast to *prpB*, which is upregulated in both the wild type σ^{54} (396%) and $\sigma^{54}{}_{\Delta R1}$ (251%) strains, relative to the control lacking σ^{54} (100%).

Combined the results in 6.2.2 and 6.2.3 describe two σ^{54}-dependent promoters (*prpB* and *hypA*) with the same -24 and -12 motifs and with similar SABT activity *in vitro*, yet behave very differently *in vivo*. This implies that while SABT activity observed *in vivo* can thus far be replicated *in vitro* (*prpB*), this comparison does not necessarily hold true in the opposite direction (*hypA*). This observation supports the underlying hypothesis of the PhD project that SABT is subject to further regulation *in vivo*. More specifically, these results suggest that this repression can be independent of the -24 and -12 promoter sequences, perhaps involving other distant or proximal regulatory elements.

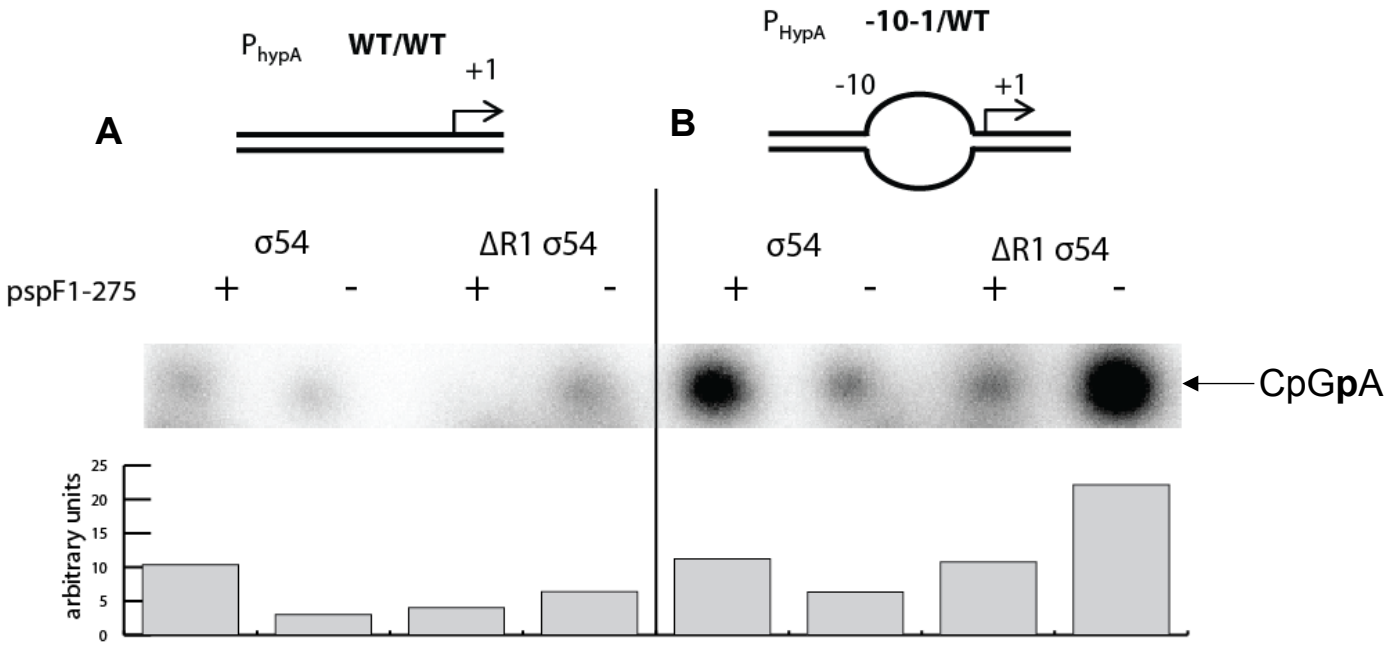


Fig 6.2.3: Abortive initiation assay *in vitro* at the *hypA* promoter with wild type σ^{54} and $\sigma^{54}{}_{\Delta R1}$. Linear probes covering the -60 to +28 region around the *hypA* transcription start site were used (A), and a variant with a -10/-1 bubble. (B). Experiments were run with wild type σ^{54} and $\sigma^{54}{}_{\Delta R1}$, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 90 mins without heparin. The CG**A** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.4 Abortive initiation assay at *yahE* promoter

yahE is a relatively unknown gene thought to have a σ^{54}-dependent promoter based on ChIP and microarray data (Barrios, 1999; Zhao, 2010). RNASeq data indicated a 170% increase in transcription of this gene with $\sigma^{54}{}_{\Delta R1}$ compared to no σ^{54} (see RNASeq supplementary data). Therefore an abortive transcription assay with the *yahE* promoter *in vitro* may be able to confirm $\sigma^{54}{}_{\Delta R1}$-dependent SABT at this promoter *in vivo*.

Moreover, *yahE* has the same σ^{54} -12 element as *prpB* and *hypA* (**TGCTT**), but a distinct -24 sequence (**TGGAAG** instead of **TGGCAC**). Earlier findings suggested that the σ^{54} -24 element may play a role in the transcriptional activation of σ^{54}-dependent promoters, and the -12 element in transcriptional repression through repressive interactions with the fork junction structure downstream of the GC (5.3.3). Should this hypothesis be true one might expect to see slightly different SABT *in vitro* with *yahE* compared to *hypA* and *prpB*, since they contain the same -12 elements but distinct -24 elements. On the other hand if the -12

sequence is key for SABT *in vitro* (and not the -24 element), one would expect to see similar activity across the 3 gene promoters as they all share the TGCTT -12 promoter element.

To examine study these possibilities, abortive initiation was examined at the *yahE* promoter *in vitro*. With homoduplex promoter DNA only basal levels of activity were observed wild type σ^{54} and the $\sigma^{54}_{\Delta R1}$ activator bypass mutant, regardless of $PspF_{1-275}$ addition (Fig 6.2.4, A). This is not unexpected as abortive initiation with homoduplex promoter DNA is typically quite low (Figs. 6.2.2, 6.2.3. 6.2.5).

With the heteroduplex *yahE* promoter DNA, high levels of transcript are observed throughout the experiment (Fig 6.2.4, B). Since high activity is also observed regardless of $PspF_{1-275}$ addition, significant background activity appears to be present with heteroduplex *yahE* promoter DNA (Fig 6.2.4, B). Despite high background activity, wild type σ^{54} with $PspF_{1-275}$ displays higher activity than without $PspF_{1-275}$, suggesting some σ^{54}-dependent transcription occurs. Additionally, activity is higher with the $\sigma^{54}_{\Delta R1}$ activator bypass mutant in the absence of $PspF_{1-275}$, a condition thought to favour SABT in earlier experiments (Fig 6.2.1, Fig 6.2.2, Fig 6.2.3, and Fig 6.2.5). This suggests that, despite high background activity, some $\sigma^{54}_{\Delta R1}$–dependent SABT can be detected at the *yahE* promoter. As such these findings support the RNASeq data suggesting SABT could be occurring at the *yahE* promoter *in vivo.*

Furthermore, these results confirm that SABT *in vitro* at the *yahE* promoter is distinct to what was observed for *hypA* and *prpB* (Fig 6.2.2 and Fig 6.2.3). This is in line with the idea that the σ^{54} -24 promoter element is key for SABT *in vivo*, as variations in this sequence affect SABT *in vitro* (with an invariable -12 element). The GG -24 motif is conserved in 99.5% of σ^{54} promoters, highlighting its importance in σ^{54}-dependent transcription (Barrios, 1999). However, these results do not exclude the possibility that both motifs are involved in the transcriptional activation of σ^{54}-dependent SABT *in vitro*, as is currently the widely accepted view in the literature.

The -12 GC element is preserved in 96% of σ^{54} promoters, exceptions include the *pspA, pspG* and *glnH* promoters. These -12 motif variants result in very high levels of σ^{54}-dependent transcription *in vivo* (Fig 5.2.8), this is in line with the idea that the GC -12 motif is involved in transcriptional repression. It has been reported that disruption of the C base in -12 TTGC tetranucleotide is responsible for keeping basal transcription in check (Wang, 1999), providing further support to this hypothesis.

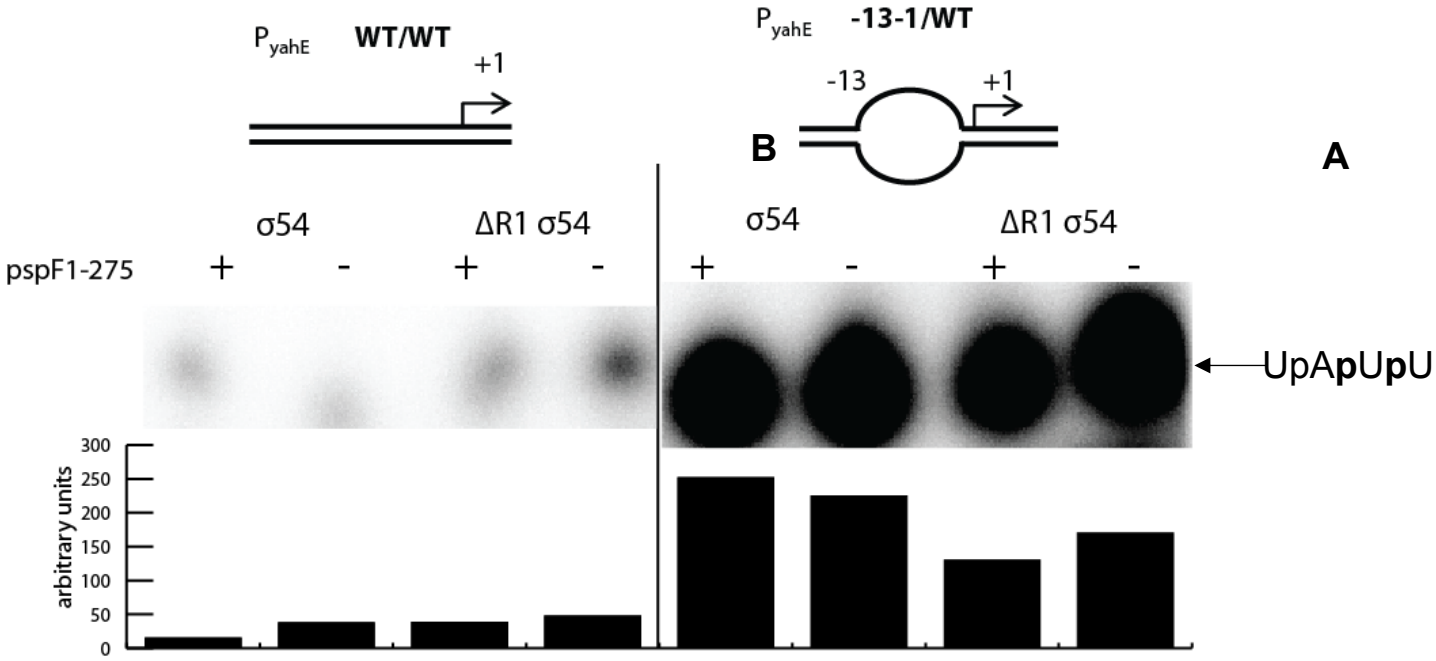


Fig 6.2.4: Abortive initiation assay *in vitro* at the *yahE* promoter with wild type σ^{54} and $\sigma^{54}_{\Delta R1}$.

Linear proves covering the -60 to +28 region around the *yahE* transcription start site were used (A), and a variant with a -10/-1 bubble. (B). Experiments were run with wild type σ^{54} and $\sigma^{54}_{\Delta R1}$, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 90 mins without heparin. The UA**U**U transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.5 Abortive initiation assay at *pspG* promoter

The final promoter examined for SABT *in vitro* was the *pspG* promoter. In the RNASeq data *pspG* was the gene with the highest upregulation in the wild type σ^{54} strain (>100 fold). As such, wild type σ^{54}–dependent *pspG* transcription appears to be highly upregulated *in vivo* despite a having a variant of the canonical -12 GC motif in the σ^{54} promoter consensus sequence (TT**GT**A). Moreover, the *pspG* promoter also showed potential SABT in the $\sigma^{54}_{\Delta R1}$ strain (Fig 5.2.6). However, the *pspA* promoter did not display potential SABT *in vivo* (Fig 3.3.5, 5.2.6) despite having the same -12/-24 promoter sequences as *pspG*

(T**TGGC**ACGNNNN**TTGT**A). The *pspA* promoter was not examined for SABT *in vitro* as this has already been shown to occur (Zhang, unpublished). Therefore, abortive initiation assays *in vitro* showing that the *pspG* promoter results in SABT could confirm the RNASeq observation that SABT can occur at this promoter *in vivo*. It would also support earlier findings (6.2.3) that SABT *in vivo* could be subject to further negative regulation independent of the promoter sequence, since *pspA* and *pspG* have the same -12 and -24 elements but only *pspG* displays SABT *in vivo*.

With the homoduplex *pspG* promoter DNA, barely detectable levels of activity were observed with wild type σ^{54}, even with $PspF_{1\text{-}275}$ (Fig 6.2.5, A). Interestingly, SABT with $\sigma^{54}{}_{\Delta R1}$ was 2.5x higher than wild type σ^{54}-dependent transcription (Fig 6.2.5, A). The latter finding supports the RNASeq data suggesting SABT could be occurring *in vivo* at the *pspG* promoter (Fig 6.2.5, A). It is possible that the non-canonical -12 promoter sequence of *pspG* (TT**GT**A) results in a weaker closed complex interaction between σ^{54}, RNAP and DNA; thereby facilitating open complex formation and SABT even with homoduplex promoter DNA.

With heteroduplex *pspG* promoter DNA, σ^{54}-dependent transcription is observed with wild type σ^{54} and $PspF_{1\text{-}275}$. 50% less activity is observed without $PspF_{1\text{-}275}$ activator, supporting the notion that this transcription is σ^{54}-dependent (Fig 6.2.5, B). This supports the notion that an intact -24 σ^{54} promoter sequence is perhaps sufficient for the transcription of σ^{54}-dependent promoters. In fact, later σ^{54}-dependent transcription of the *cspH* promoter (lacking a -12 motif) was observed, indicating a -24 motif is indeed sufficient for detectable levels of σ^{54}-dependent transcription (Fig 6.2.8)

The transcription levels of heteroduplex *pspG* promoter DNA with wild type σ^{54} and $PspF_{1\text{-}275}$ are very similar to SABT with $\sigma^{54}{}_{\Delta R1}$ without $PspF_{1\text{-}275}$ (Fig 6.2.5, B). This is in line with RNASeq findings suggesting SABT occurs at the *pspG* promoter in *vivo*. SABT is higher in the absence of $PspF_{1\text{-}275}$, which has been observed in other abortive assays too (Figs: 6.2.1, 6.2.2, 6.2.3, and 6.2.4).

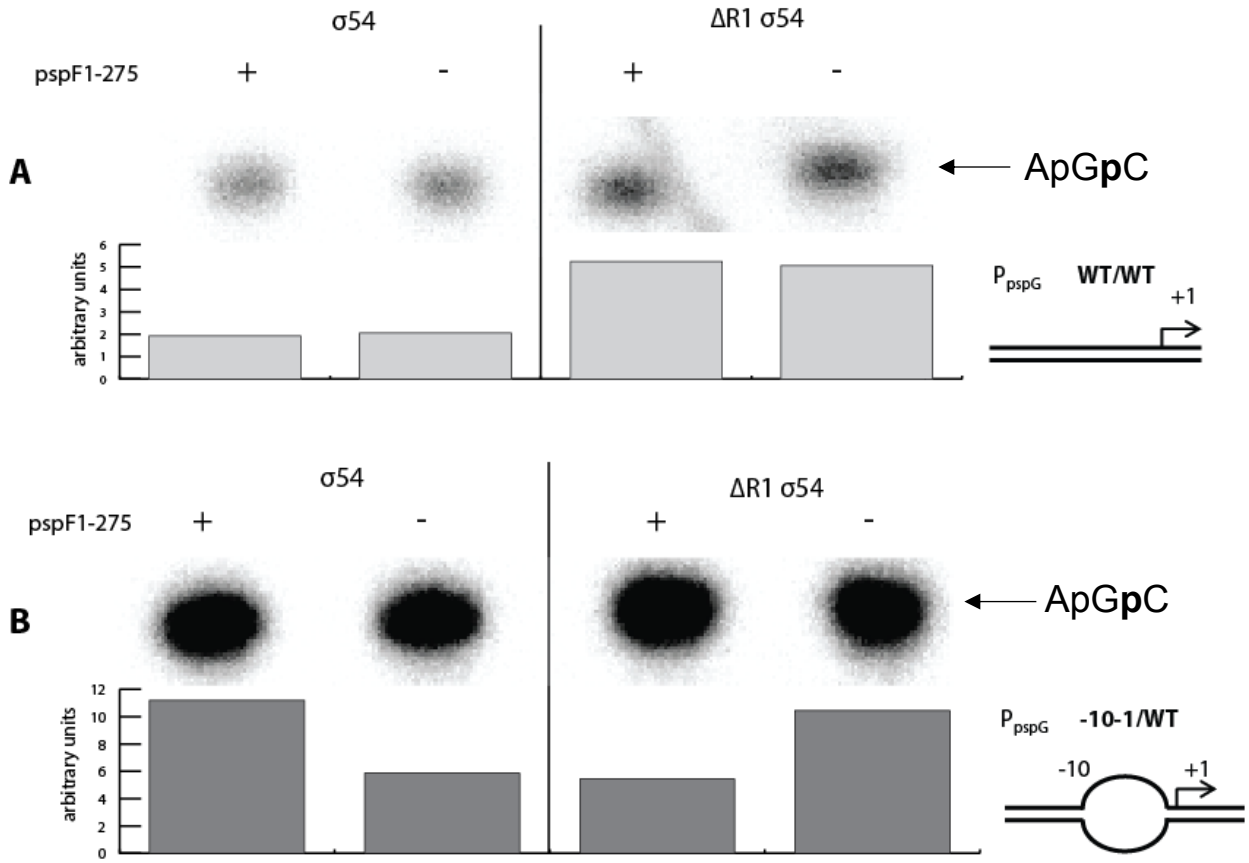


Fig 6.2.5: Abortive initiation assay at the *pspG* promoter with σ^{54} and $\sigma^{54}{}_{\Delta R1}$

Linear probes covering the -60 to +28 region around the *pspG* transcription start site were used (**A**), in addition to a mismatched -10 to -1 sequence (**B**). Experiments were done with σ^{70} and σ^{54}, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 30 mins, with and without heparin. The AG**C** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.6 Abortive initiation assay at *cspA* promoter

Earlier RNASeq data indicated that the major cold shock gene (*cspA*) promoter may be transcribed via σ^{54} (5.3.4). *CspA* has a σ^{54} ChIP binding site which could potentially be a σ^{54}-dependent promoter driving the transcription of *cspA* (Bonocora, unpublished). However, as of yet no σ^{54}-dependent transcription of *cspA* has been reported in *E. coli*. Here *in vitro* transcription of the *cspA* promoter is examined by analysing σ^{54}-dependent transcription arising from the putative σ^{54}-dependent promoter (start site known from RNASeq data). This activity is then compared to the known σ^{70}-dependent promoter (Goldstein, 1990; Tanabe, 1992).

Significant σ^{54}-dependent transcription of the *cspA* promoter is not detected under any of conditions assayed (Fig 6.2.6). While RNASeq data (5.2.7) suggested a link between σ^{54} activity and *cspA* transcription, no evidence for a direct link was seen *in vitro*. This is in line with the absence of a clear -24 and -12 binding motif for σ^{54} in this region of the *cspA* promoter DNA. However, σ^{54}-dependent expression of *cspA* may require additional factors not present in the minimal *in vitro* setting. Alternatively, it is possible that σ^{54} affects σ^{70}-dependent *cspA* transcription, potentially by regulation σ^{70}-like transcription through a thus far uncharacterised mechanism. This hypothesis was not tested.

σ^{70}-dependent transcription at the *cspA* promoter was observed with the homoduplex DNA but appears to be heparin-sensitive (Fig 6.2.6, A). With the heteroduplex DNA, σ^{70}-dependent transcription was readily detected, and was heparin-resistant (Fig 6.2.6, B). This is in line with previous studies suggesting *cspA* transcription is σ^{70}-dependent (Goldstein, 1990; Tanabe, 1992).

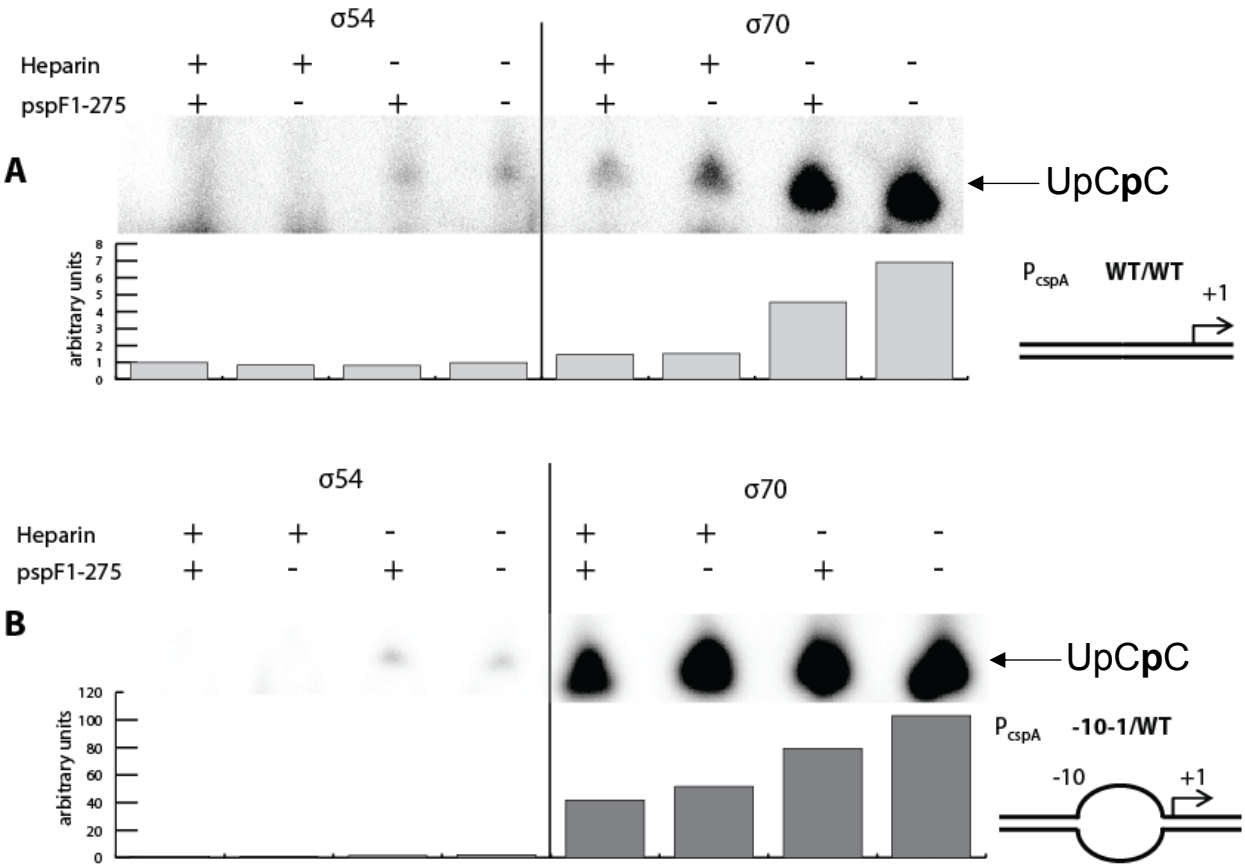


Fig 6.2.6: Abortive initiation assay at the *cspA* promoter with σ^{54} and σ^{70}.

Linear probes covering the -60 to +28 region around the *cspA* transcription start site were used (**A**), in addition to a mismatched -10 to -1 sequence (**B**). Experiments were done with σ^{70} and σ^{54}, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 30 mins, with and without heparin. The UC**C** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.2.7 Abortive initiation assay at *cspH* promoter

cspH is another cold shock gene upregulated in the presence of wild type σ^{54}, but not $\sigma^{54}{}_{\Delta R1}$ or without σ^{54} (5.2.7). Like *cspA*, *cspH* has a σ^{54} ChIP binding site without a clear-24/-12 σ^{54} consensus motifs (Bonocora, unpublished). While a -24 σ^{54} motif (CT**GG**CT) is present 28 bases upstream of the transcription start site, a -12 GC element is lacking in this promoter (Fig 6.2.7).

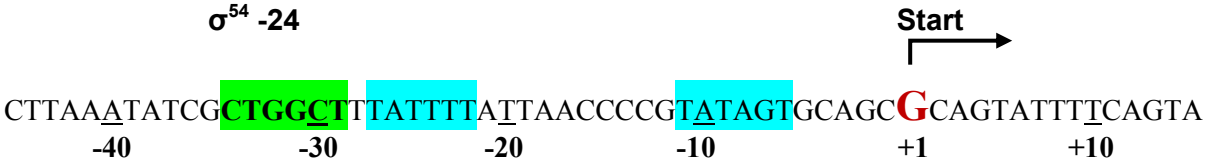


Fig 6.2.7: σ^{54} binding motif upstream of *cspH* transcription start site.
The promoter region of *cspH* contains a putative -24 binding sequence for σ^{54} (green), -28 bases upstream of the transcription start site (red). Putative σ^{70} -10 promoters (TATAAT) are also annotated (blue).

While the ChIP, RNASeq data and a -24 motif suggest a σ^{54}-dependent transcription at the *cspH* promoter could occur, such a promoter remains to be characterised. One way to examine the existence of such a σ^{54} promoter it so examine σ^{54}-dependent transcription of the *cspH in vitro*.

σ^{54}-dependent abortive transcription of the *cspH* promoter was studied *in vitro*, and compared to σ^{70}-dependent transcription (Fig 6.2.8). Virtually no heparin-resistant transcription was detected with σ^{54} or σ^{70} (Fig 6.2.8). σ^{70}-dependent transcription was detected in the absence of heparin with homoduplex and heteroduplex DNA. Compared to σ^{70}-dependent transcription without heparin, only ≈25% activity was observed with σ^{54} (Fig 6.2.8). The σ^{54}-dependent transcription appears to be partially independent of $PspF_{1\text{-}275}$, suggesting that a proportion of this activity is likely to be background activity mediated by the lack of heparin (Fig 6.2.8). Without heparin transcription reinitiation is not inhibited, and background transcription is higher (Mennes, 1992).

Basal levels of σ^{54}-dependent transcription were detected for *cspH* (unlike for *cspA*). It is possible that the low levels of σ^{54}-dependent transcription observed with *cspH* are mediated by the -24 motif, a region previously thought conserved in transcriptional activation of σ^{54} promoters (5.3.8). However, this basal level of transcription is unlikely to account for the 12x increase in transcription (in RNA Seq) observed for *cspH* with σ^{54} compared to no σ^{54} (Fig 5.2.5). As with *cspA*, it is likely that σ^{54} somehow positively regulates a σ^{70}-dependent *cspH* promoter, either directly or indirectly.

The abortive initiation assays suggest that *cspH* transcription is σ^{70}-dependent (like *cspA*), however no σ^{70}-dependent promoter for *cspH* has been characterised as of yet. Two AT-rich sequences are observed in the promoter region and could be putative σ^{70} -10 promoter elements (Fig. 6.2.7, blue).

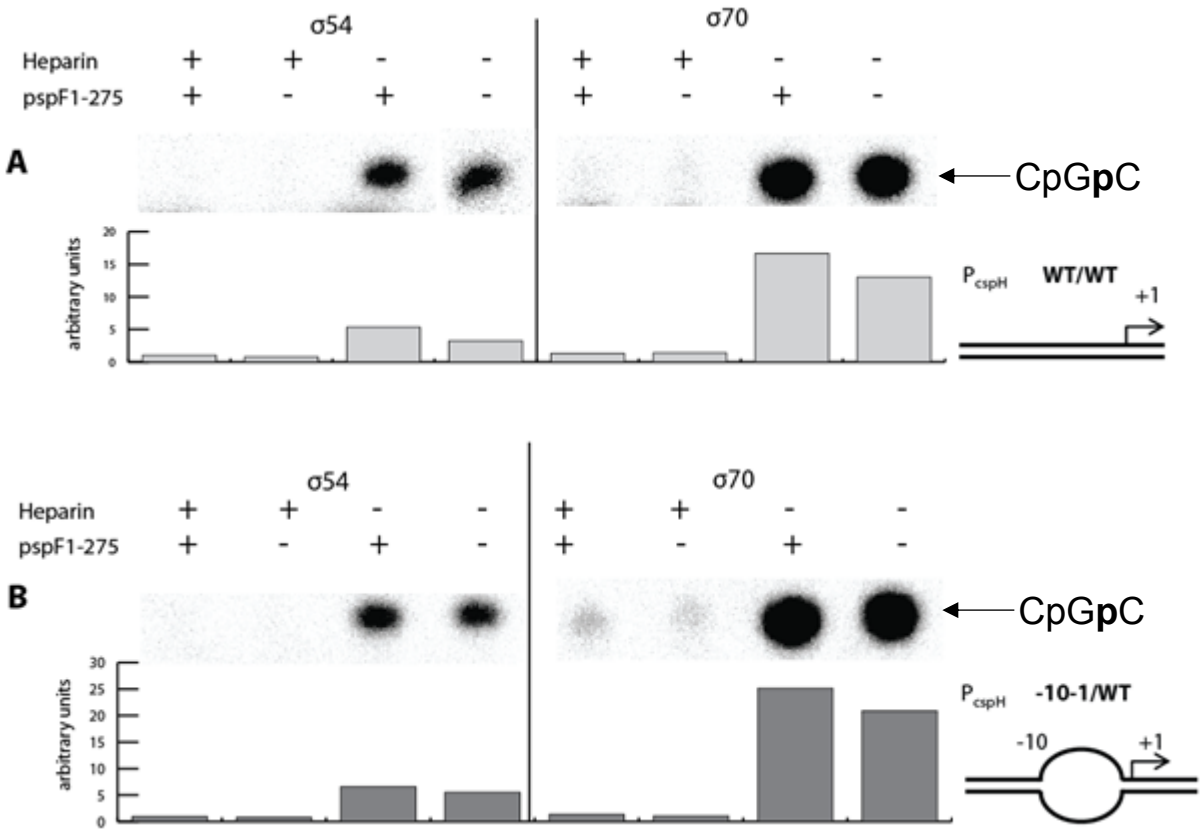


Fig 6.2.8: Abortive initiation assay at the *cspH* promoter with σ^{54} and σ^{70}.

Linear probes covering the -60 to +28 region around the *cspH* transcription start site were used (**A**), in addition to a mismatched -10 to -1 sequence (**B**). Experiments were done with σ^{70} and σ^{54}, with and without $pspF_{1\text{-}275}$. Reactions were incubated for 30 mins, with and without heparin. The CG**C** transcript product (arrowed) was quantified via incorporated radioactive nucleotides (^{32}P highlighted in bold). Reactions were performed in triplicates, and results obtained followed the same global pattern and were within a ±10% error range.

6.3 Summary

Abortive assays were able to detect SABT *in vitro* for *prpB*, *yahE* and *pspG* promoters, in line with RNASeq data *in vivo*. However, these promoters appear to have little in common that distinguishes them from other known σ^{54}-dependent promoters that do not display SABT *in vivo*.

HypA also results in SABT *in vitro,* and has the same -24/-12 σ^{54} consensus motif in its promoter region as *prpB*. Unlike *prpB*, this gene only results in very low levels of SABT *in vivo* (106%) compared to the strain lacking any σ^{54} (100%). This suggests that expression of σ^{54}-dependent genes *in vivo* is subject to further negative regulation independent of the σ^{54} promoter sequence. The fact that SABT is observed at the *pspG* promoter, but not the *pspA* promoter (same -12/-24 elements) further support this hypothesis. The degree of SABT, at least *in vitro,* seems to depend on how easily the DNA might melt out to form an open complex. In the cell factors which decrease the spontaneous opening of DNA might reduce or prevent SABT and vice versa. Such factors are absent *in vitro*, and so a strict correlation between *in vivo* and *in vitro* observations might be hard to achieve. Local dynamic changes in DNA superhelicity may well be reflected in the *in vivo* data sets where SABT was sought.

Moreover, $PspF_{1\text{-}275}$ addition results in reduced $\sigma^{54}{}_{\Delta R1}$-dependent SABT *in vitro* at the *nifH*, *prpB*, *yahE*, *pspG* and *hypA* promoters (Fig 6.2.1, 6.2.2, 6.2.3, 6.2.4, 6.2.5). The highly conserved GAFTGA motif in L1 loop of $PspF_{1\text{-}275}$ is known to interact with residues 18-25 and 33-39 of σ^{54}, located in region I of σ^{54} (Zhang, 2012). This interaction is thought to communicate changes associated with ATP hydrolysis, leading to conformational rearrangements in the RNA polymerase closed complex, thereby promoting open complex formation (Zhang, 2009). However the $\sigma^{54}{}_{\Delta R1}$ activator bypass mutant lacks region I of σ^{54}, therefore the $PspF_{1\text{-}275}$'s role in transcriptional activation could be disrupted. As a result $PspF_{1\text{-}275}$ addition might interfere with existing SABT *in vitro* without enhancing open complex formation, thereby reducing overall SABT activity via some unexpected contacts with the closed complex or intermediate complexes.

Abortive assays also did not detect significant σ^{54}-dependent transcription at the *cspA* and *cspH* promoter regions *in vitro*. Since RNASeq data suggests that expression of these genes is strongly upregulated in the presence of wild type σ^{54}, it is likely that this effect is largely indirect through other proteins as opposed to σ^{54}-dependent transcription of *cspA* or *cspH* promoters.

However, basal levels of σ^{54}-dependent transcription were observed (Fig 6.2.8) at the *cspH* promoter (lacking a -12 motif), suggesting a -24 motif is sufficient for basal levels of σ^{54}-dependent transcription. This supports earlier findings (section 5.3.3) indicating that the -24 motif in the σ^{54} promoter is important for transcriptional activation, and the -12 element for transcriptional repression.

Overview

The main objective of this project was to investigate the activator bypass problem *in vivo* and *in vitro*. The activator bypass problem is the unpublished observation that SABT is detected at several σ^{54} promoters *in vitro*, but not readily observed *in vivo* at the same promoters. In order to address this problem, the underlying hypothesis was tested (ie. that SABT is not readily observed *in vivo* under standard growth conditions).

Previous studies have demonstrated SABT occurs *in vitro* at the σ^{54}-dependent *pspA* and gln_{Ap2} promoters (Wang, 1997). However, in sections 3.2.5 and 3.3.1 I showed that SABT is not readily observed *in vivo* under standard conditions at the σ^{54}-dependent *pspA* and gln_{Ap2} promoters respectively. Since SABT is observed under minimal conditions *in vitro*, these findings suggest that additional inhibitory factors may be present *in vivo* to prevent σ^{54}-dependent transcription from occurring without activators.

One key difference between the *in vivo* and *in vitro* work is that the *in vivo* assays quantify a full length reporter gene product (LacZ or GFP) whereas the *in vitro* assays measure short abortive transcripts. Therefore, one could argue these results indicate that SABT allows for abortive products to be formed from an unstable promoter complex, but not full length mRNA which is thought to require a stable promoter complex (see Fig 1.4.4).

However, SABT was also shown to be observed *in vivo* (ie. full length proteins formed) under certain conditions like high temperature (Fig 3.3.4), suggesting SABT is not limited to only forming short abortive initiation products. Moreover, full length transcription assays *in vitro* showed that $\sigma^{54}{}_{\Delta R1}$–dependent transcription can also give rise to larger products of several hundred bps (Fig 3.3.6). Combined, these results indicate that SABT is not observed *in vivo* due to repressive barriers present *in vivo*, and which are thus far uncharacterised (Fig 7.1.1).

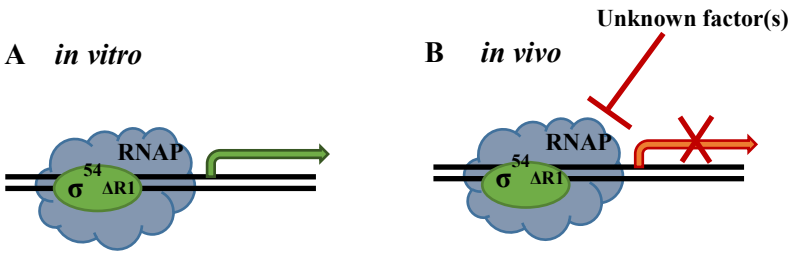


Fig 7.1.1: Inhibition of SABT in vivo by unknown factor(s).

SABT is detected *in vitro* (A) in a minimal setting but is not readily observed *in vivo* (B), suggesting additional unknown factor(s) inhibit SABT *in vivo*.

Elevated temperature (42°C) increased SABT *in vivo* at the gln_{Ap2} promoter (3.3.4), possibly due to a loss of function from a protein negatively regulating SABT *in vivo*. Chapter 4 focused on trying to identify such putative inhibitory proteins by screening several tens of thousands of non-essential single-gene deletions for SABT *in vivo*. However, no promising candidates were found. This suggests that if such repressive proteins are present, they are likely to be essential genes (not represented in single-gene deletions) or involve more than one protein.

To obtain a more thorough understanding of the effects of $\sigma^{54}{}_{\Delta R1}$ on transcription *in vivo*, global transcriptomic analysis was performed in Chapter 5 by doing RNA Sequencing. Several genes with σ^{54} promoters were found to display slightly elevated levels of $\sigma^{54}{}_{\Delta R1}$-dependent transcription (*GlnH, hydN, prpB, pspG, norV* and *rutA*). *In vitro* assays confirmed SABT at the two selected promoters which were tested *in vitro* (*prpB* and *pspG*). However, these promoters displaying SABT appear to have little in common that distinguishes them from other known σ^{54}-dependent promoters that do not display SABT *in vivo*. As such, the factors thought to negatively regulate SABT *in vivo* still remain unknown. Since these presumed repressive factors do not seem to be influenced by *rpoN*-inducing or inhibiting conditions, they are thought to be general repressors of transcription (section 3.3.1)

The RNASeq data yield another interesting observation: several genes with σ^{54} promoter sequences appeared to be inhibited in the presence of σ^{54} but expressed in the absence of σ^{54} (Fig 5.2.10). Moreover, these σ^{54} binding sites correlated with local transcriptional inhibition in the presence of σ^{54}, indicating this repression may be directly related to σ^{54} binding (5.2.10). As such, it was hypothesised σ^{54}RNAP binding may function as a repressor of σ^{70}-like transcription by binding to the promoter region of these genes (Fig 7.1.2).

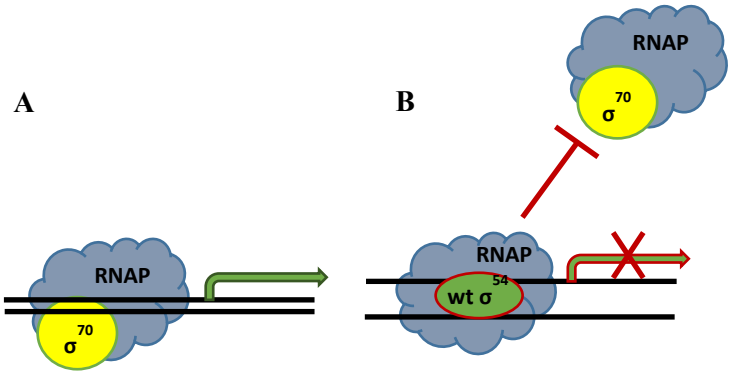


Fig 7.1.2: Repression of σ^{70}-like transcription by σ^{54}RNAP binding.

Genes transcribed by σ^{70}-dependent transcription (A) can do so freely in the absence of σ^{54}. In the presence of σ^{54}RNAP and where such genes have σ^{54} binding sites in their promoter regions (B), the binding of σ^{54}RNAP is thought to obstruct σ^{70}-dependent transcription.

Future Work

The findings of this project thus far indicate that σ^{54} binding sites in the promoters of certain genes recruit σ^{54}RNAP to obstruct σ^{70}-like transcription of the gene. This is a completely novel role for σ^{54}, which thus far has only ever been described as a transcriptional activator in the literature. Further experiments will be required to confirm whether σ^{54} does indeed have a dual role in transcription, as an activator of σ^{54}-dependent gene expression and also a repressor of σ^{70}-like transcription.

To examine this hypothesis, abortive assays will be done *in vitro* with the promoters containing putative repressive σ^{54} binding sites. First, the assays will be performed in the presence of σ^{70} to show their expression is σ^{70}-like dependent. Subsequently these experiments will also be run in the presence of various concentrations of σ^{54}, to verify whether σ^{54} binding does indeed interfere with the σ^{70}-like transcription. Should this prove to be the case, these findings will be further tested by scrambling the σ^{54} promoter sequences, as this should at least partially alleviate the inhibitory effects of σ^{54} binding. Additionally, conservation studies of these σ^{54} binding sites across different bacterial species could indicate whether they are evolutionarily conserved. Should this be the case, it may provide further support to the notion that these sites are functional.

Combined, the data from this project may end up providing sufficient evidence for a novel role for what is arguably the most mysterious sigma factor in bacteria. Doing so could also provide a valid explanation as to why σ^{54}RNAP forms inactive closed promoter complexes upon promoter DNA binding (unlike other sigma factors), and also why σ^{54}-dependent transcription has not yet evolved towards activator independence over time.

Bibliography

Aires JR, Nikaido H. (2005) Aminoglycosides are captured from both periplasm and cytoplasm by the AcrD multidrug efflux transporter of Escherichia coli. J Bacteriol. 187(6):1923-9.

Akira Ishihama (2000). Functional modulation of E. coli RNA polymerase. Annual Review of Microbiology 54, 499-518.

Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A. (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol. 181(20):6361-70.

Andrews SC, Berks BC, McClay J, Ambler A, Quail MA, Golby P, Guest JR. (1997) A 12-cistron Escherichia coli operon (hyf) encoding a putative proton-translocating formate hydrogenlyase system. Microbiology. 143 (Pt 11):3633-47.

Arthur, T. M., L. C. Anthony & R. R. Burgess, (2000) Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase. J Biol Chem 275: 23113-23119.

Atamanyuk D, Faivre F, Oxoby M, Ledoussal B, Drocourt E, Moreau F, Gerusz V. (2013) Vectorization Efforts To Increase Gram-Negative Intracellular Drug Concentration: A Case Study on HldE-K Inhibitors. J Med Chem. [Epub ahead of print]

Austin, S., and R. Dixon. (1992). The prokaryotic enhancer binding protein NTRC has an ATPase activity which is phosphorylation and DNA dependent. EMBO J 11,2219-2228.

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

Baranova NN, Danchin A, Neyfakh AA. (1999) Mta, a global MerR-type regulator of the Bacillus subtilis multidrug-efflux transporters. Mol Microbiol. 31(5):1549-59.

Bar-Nahum, G. & E. Nudler, (2001) Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. Cell 106: 443-451.

Barrios H, Valderrama B, Morett E. (1999) Compilation and analysis of sigma(54)-dependent promoter sequences. Nucleic Acids Res. 27(22):4305-13.

Belik AS, Tarasova NN, Khmel' IA. (2008) [Regulation of biofilm formation in Escherichia coli K12: effect of mutations in HNS, StpA, lon, and rpoN genes]. Mol Gen Mikrobiol Virusol. 4, 3-5.

Bertin P, Terao E, Lee EH, Lejeune P, Colson C, Danchin A, Collatz E. (1994) The H-NS protein is involved in the biogenesis of flagella in Escherichia coli. J Bacteriol. 176(17):5537-40.

Bissette D L, Anderson R L. (1974) Lactose and D-galactose metabolism in group N streptococci: presence of enzymes for both the D-galactose 1-phosphate and D-tagatose 6-phosphate pathways. J Bacteriol. 117:318–320

Boguski MS, Tolstoshev CM, Bassett DE Jr. (1994) Gene discovery in dbEST. Science. 265(5181):1993-4.

Bose D, Pape T, Burrows PC, Rappas M, Wigneshweraraj SR, Buck M, Zhang X. (2008) Organization of an activator-bound RNA polymerase holoenzyme. Mol Cell. 32(3), 337-46.

Bradley MD, Beach MB, de Koning AP, Pratt TS, Osuna R. (2007) Effects of Fis on Escherichia coli gene expression during different growth stages. Microbiology. 153(Pt 9):2922-40.

Brahmachary, P., M. G. Dashti, J. W. Olson, and T. R. Hoover (2004) Helicobacter pylori FlgR is an enhancer-independent activator of sigma54-RNA polymerase holoenzyme. J Bacteriol 186:4535-4542.

Brandi A, Pon CL, Gualerzi CO. (1994) Interaction of the main cold shock protein CS7.4 (CspA) of Escherichia coli with the promoter region of hns. Biochimie. 76(10-11):1090-8.

Brissette JL, Russel M, Weiner L, Model P. (1990) Phage shock protein, a stress protein of Escherichia coli. Proc Natl Acad Sci U S A. 87(3):862-6.

Brissette, J.L., Weiner, L., Ripmaster, T.L., and Model, P. (1991). Characterization and sequence of the Escherichia coli stress-induced psp operon. Journal of molecular biology 220, 35-48.

Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol. 2(1):57-65.

Buck M, Cannon W. (1992) Activator-independent formation of a closed complex between sigma 54-holoenzyme and nifH and nifU promoters of Klebsiella pneumoniae. Mol Microbiol. 6(12):1625-30.

Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD. (2000) The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J Bacteriol. 182(15), 4129-36.

Buck, M (1986) Deletion analysis of the Klebsiella pneumoniae nitrogenase promoter: importance of spacing between conserved sequences around positions -12 and -24 for activation by the nifA and ntrC (glnG) products. J Bacteriol 166: 545-551.

Burrows PC, Joly N, Buck M. (2010) A prehydrolysis state of an AAA+ ATPase supports transcription activation of an enhancer-dependent RNA polymerase. Proc Natl Acad Sci U S A. 107(20):9376-81.

Burrows, P. C., S. R. Wigneshweraraj & M. Buck, (2008) Protein-DNA interactions that govern AAA+ activator-dependent bacterial transcription initiation. J Mol Biol 375: 43-58.

Bush, M., and Dixon, R. (2012). The role of bacterial enhancer binding proteins as specialized activators of sigma54-dependent transcription. Microbiology and molecular biology reviews : MMBR 76, 497-529.

C.E. Vrentas (2005) Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. Genes Dev. 19: 2378–2387

Cannon W, Gallegos MT, Buck M. (2001) DNA melting within a binary sigma(54)-promoter DNA complex. J Biol Chem. 276(1):386-94.

Cannon, W., M. Chaney, and M. Buck. (1999). Characterisation of holoenzyme lacking sigmaN regions I and II. Nucleic Acids Res 27, 2478-2486.

Carbone A, Fioretti FM, Fucci L, Ausió J, Piscopo M. (2012) High efficiency method to obtain supercoiled DNA with a commercial plasmid purification kit. Acta Biochim Pol. 59(2):275-8.

Carmona M, Rodríguez MJ, Martínez-Costa O, De Lorenzo V. (2000) In vivo and in vitro effects of (p)ppGpp on the sigma(54) promoter Pu of the TOL plasmid of Pseudomonas putida. J Bacteriol. 182(17):4711-8.

Carpousis, A. J. & J. D. Gralla, (1980) Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. Biochemistry 19: 3245-3253.

Celesnik H, Deana A, Belasco JG (2007) Initiation of RNA Decay in Escherichia coli by 5′ Pyrophosphate Removal. 27(1): 78-90.

Chaney M, Buck M. (1999). The sigma 54 DNA-binding domain includes a determinant of enhancer responsiveness. Mol Microbiol. 33(6): 1200-9.

Chaney, M., R. Grande, S. R. Wigneshweraraj, W. Cannon, P. Casaz, M. T. Gallegos, J. Schumacher, S. Jones, S. Elderkin, A. E. Dago, E. Morett, and M. Buck (2001) Binding of transcriptional activators to sigma 54 in the presence of the transition state analog ADP-aluminum fluoride: insights into activator mechanochemical action. Genes Dev 15:2282-2294.

Cheetham, G. M. & Steitz, T. A. (2000) Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. Curr. Opin. Struct. Biol. 10, 117–123.

Chen, B., M. Doucleff, D. E. Wemmer, S. De Carlo, H. H. Huang, E. Nogales, T. R. Hoover, E. Kondrashkina, L. Guo, and B. T. Nixon (2007) ATP ground- and transition states of bacterial enhancer binding AAA+ ATPases support complex formation with their target protein, sigma54. Structure 15:429-440.

Chen, B., T. A. Sysoeva, S. Chowdhury, L. Guo, S. De Carlo, J. A. Hanson, H. Yang, and B. T. Nixon (2010) Engagement of arginine finger to ATP triggers large conformational changes in NtrC1 AAA+ ATPase for remodeling bacterial RNA polymerase. Structure 18:1420-1430.

Cherepanov, PP, W Wackernagel (1995). Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158(1):9-14.

Cho BK, Knight EM, Barrett CL, Palsson BØ. (2008) Genome-wide analysis of Fis binding in Escherichia coli indicates a causative role for A-/AT-tracts. Genome Res. 18(6):900-10.

Choudhary, M., C. Mackenzie, N. J. Mouncey & S. Kaplan, (1999) RsGDB, the Rhodobacter sphaeroides Genome Database. Nucleic Acids Res 27: 61-62.

Corish P, Tyler-Smith C (1999) Attenuation of green fluorescent protein half-life in mammalian cells. Protein Eng. 12 (12): 1035-1040.

Dame RT, Wyman C, Goosen N. (2000) H-NS mediated compaction of DNA visualised by atomic force microscopy. Nucleic Acids Res.28(18):3504-10.

Darst, S. A., N. Opalka, P. Chacon, A. Polyakov, C. Richter, G. Zhang & W. Wriggers, (2002) Conformational flexibility of bacterial RNA polymerase. Proc Natl Acad Sci U S A 99: 4296-4301.

Dartigalongue C, Missiakas D, Raina S. (2001) Characterization of the Escherichia coli sigma E regulon. J Biol Chem. 276(24):20866-75.

Daughdrill GW, Chadsey MS, Karlinsey JE, Hughes KT, Dahlquist FW. (1997) The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28. Nat Struct Biol. 4(4):285-91.

Deana A, Celesnik H, Belasco JG. (2008) The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. Nature. 451(7176):355-8.

Delgado, A., R. Salto, S. Marques, and J. L. Ramos (1995) Single amino acids changes in the signal receptor domain of XylR resulted in mutants that stimulate transcription in the absence of effectors. J Biol Chem 270:5144-5150.

Deredge DJ, Baker JT, Datta K, Licata VJ. (2010) The glutamate effect on DNA binding by pol I DNA polymerases: osmotic stress and the effective reversal of salt linkage. J Mol Biol. 401(2):223-38

Deutscher J, Francke C, Postma PW (2006) How phosphotransferase systemrelated protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol Mol Biol Rev. 70(4):939-1031.

Djordjevic, M. (2011). Redefining Escherichia coli σ70 Promoter Elements: −15 Motif as a Complement of the −10 Motif . Journal of Bacteriology, 193(22), 6305–6314.

Doherty, A. J., B. A. Connolly, and A. F. Worrall. (1993) Overproduction of the toxic protein, bovine pancreatic DNaseI, in Escherichia coli using a tightly controlled T7-promoter-based vector. Gene 136:337–340.

Donato GM, Kawula TH. (1999) Phenotypic analysis of random hns mutations differentiate DNA-binding activity from properties of fimA promoter inversion modulation and bacterial motility. J Bacteriol. 181(3):941-8.

Doucleff M, Malak LT, Pelton JG, Wemmer DE (2005) The C-terminal RpoN domain of sigma54 forms an unpredicted helix-turn-helix motif similar to domains of sigma70. J Biol Chem. 280(50):41530-6.

Doucleff, M., Pelton, J.G., Lee, P.S., Nixon, B.T., and Wemmer, D.E. (2007). Structural basis of DNA recognition by the alternative sigma-factor, sigma54. Journal of molecular biology 369, 1070-1078.

Drolet M. (2006) Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. Mol Microbiol. 59(3):723-30.

Duprey A, Reverchon S, Nasser W. (2014) Bacterial virulence and Fis: adapting regulatory networks to the host environment. Trends Microbiol. 2014 Feb;22(2):92-9.

Ebright RH. (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. J Mol Biol. 304(5):687-98.

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

Eguchi Y, Oshima T, Mori H, Aono R, Yamamoto K, Ishihama A, Utsumi R. (2003) Transcriptional regulation of drug efflux genes by EvgAS, a two-component system in Escherichia coli. Microbiology. 149(Pt 10):2819-28.

Eichler K, Buchet A, Bourgis F, Kleber HP, Mandrand-Berthelot MA. (1995) The fix Escherichia coli region contains four genes related to carnitine metabolism. J Basic Microbiol. 35(4):217-27.

Elderkin, S., Jones, S., Schumacher, J., Studholme, D., and Buck, M. (2002). Mechanism of action of the Escherichia coli phage shock protein PspA in repression of the AAA family transcription factor PspF. Journal of molecular biology 320, 23-37.

ENGLESBERG E, ANDERSON RL, WEINBERG R, LEE N, HOFFEE P, HUTTENHAUER G, BOYER H. (1962) L-Arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of Escherichia coli. J Bacteriol. 84:137-46.

Erie DA, Hajiseyedjavadi O, Young MC, von Hippel PH. (1993) Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. Science. 262(5135):867-73.

Etchegaray JP, Jones PG, Inouye M. (1996) Differential thermoregulation of two highly homologous cold-shock genes, cspA and cspB, of Escherichia coli. Genes Cells. 1(2):171-8.

Federowicz, S., Kim, D., Ebrahim, A., Lerman, J., Nagarajan, H., Cho, B., Palsson, B. (2014). Determining the Control Circuitry of Redox Metabolism at the Genome-Scale. PLoS Genetics, 10(4), e1004264.

Feklistov A, Darst SA. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. Cell. 147(6), 1257-69.

Feklistov A, Sharon B, Darst S, Gross C (2012). Bacterial Sigma Factors: A Historical, Structural, and Genomic Perspective. Annual Review of Microbiology. 68: 357-376

Finkel SE, Johnson RC. (1992) The Fis protein: it's not just for DNA inversion anymore. Mol Microbiol. 6(22):3257-65.

Fisher MA, Grimm D, Henion AK, Elias AF, Stewart PE, Rosa PA, Gherardini FC: Borrelia (2005) burgdorferi sigma54 is required for mammalian infection and vector transmission but not for tick colonization. Proc Natl Acad Sci USA. 102(14):5162-5167.

Fouqueau T, Zeller ME, Cheung AC, Cramer P, Thomm M. (2013) The RNA polymerase trigger loop functions in all three phases of the transcription cycle. Nucleic Acids Res. 41(14):7048-59.

Francke C, Groot Kormelink T, Hagemeijer Y, Overmars L, Sluijter V, Moezelaar R, Siezen RJ. (2011) Comparative analyses imply that the enigmatic Sigma factor 54 is a central controller of the bacterial exterior. BMC Genomics. 1;12:385.

Friedman LJ, Gelles J. (2012a) Mechanism of transcription initiation at an activator-dependent promoter defined by single-molecule observation. Cell. 148(4):679-89.

Friedman LJ, Mumm JP, Gelles J (2012b) RNA polymerase approaches its promoter without long-range sliding along DNA. Proc Natl Acad Sci U S A. 110(24):9740-5.

Frisch RL, Bender RA. (2010) Expanded role for the nitrogen assimilation control protein in the response of Klebsiella pneumoniae to nitrogen stress. J Bacteriol. 192(19):4812-20.

Gallegos, M.T., Cannon, W.V., and Buck, M. (1999). Functions of the sigma(54) region I in trans and implications for transcription activation. The Journal of biological chemistry 274, 25285-25290.

Garcia, E., S. Bancroft, S. G. Rhee, and S. Kustu. (1977). The product of a newly identified gene glnF, is required for synthesis of glutamine synthetase in Salmonella. Proc. NatI. Acad. Sci. USA 74:1662-1666.

Genschik P, Drabikowski K, Filipowicz W (1998) Characterization of the Escherichia coli RNA 3'-terminal phosphate cyclase and its sigma54-regulated operon. J Biol Chem. 1998 Sep 25;273(39):25516-26.

Geszvain, K., T. M. Gruber, R. A. Mooney, C. A. Gross & R. Landick, (2004) A hydrophobic patch onthe flap-tip helix of E.coli RNA polymerase mediates sigma(70) region 4 function. J Mol Biol 343: 569-587.

Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A. (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol. 2012;13(3):R23.

Glenn T.C. (2011) Field guide to next-generation DNA sequencers, Molecular ecology resources. 11, 759-769.

Goldman, S. R., R. H. Ebright & B. E. Nickels, (2009) Direct detection of abortive RNA transcripts in vivo. Science 324: 927-928.

Goldstein J, Pollitt NS, Inouye M. (1990) Major cold shock protein of Escherichia coli. Proc Natl Acad Sci U S A. 87(1):283-7.

Graumann P, Schröder K, Schmid R, Marahiel MA. (1996) Cold shock stress-induced proteins in Bacillus subtilis. J Bacteriol. 178(15):4611-9.

Grigoroudis AI, Panagiotidis CA, Lioliou EE, Vlassi M, Kyriakidis DA (2007) Molecular modeling and functional analysis of the AtoS-AtoC two-component signal transduction system of Escherichia coli. Biochim Biophys Acta. 2007 Aug;1770(8):1248-58.

Gross, C. A., C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy & B. Young, (1998) The functional and regulatory roles of sigma factors in transcription. Cold Spring Harb Symp Quant Biol 63: 141-155.

Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol. 57:441-66.

Gunnelius L, Hakkila K, Kurkela J, Wada H, Tyystjärvi E, Tyystjärvi T. (2014) The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. Nucleic Acids Res. 2014 Apr;42(7):4606-14.

Guo Y, Lew CM, Gralla JD. (2000) Promoter opening by sigma(54) and sigma(70) RNA polymerases: sigma factor-directed alterations in the mechanism and tightness of control. Genes Dev. 14(17), 2242-55.

Gutnick D., Calvo J. M., Klopotowski T., Ames B. N. (1969). Compounds which serve as the sole source of carbon or nitrogen for Salmonella typhimurium LT-2. J. Bacteriol. 100:215–219

Guzman LM, Belin D, Carson MJ, Beckwith J. (1995) Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol. 177(14):4121-30.

Hall, BG (1979) Lactose metabolism involving phospho-beta-galactosidase in Klebsiella. J Bacteriol.138(3): 691–698.

Hansen et al. (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic acids research. 38, e131.

Hashimoto-Gotoh T, Franklin FC, Nordheim A, Timmis KN. (1981) Specific-purpose plasmid cloning vectors. I. Low copy number, temperature-sensitive, mobilization-defective pSC101-derived containment vectors. Gene. 16(1-3):227-35.

Haugen, S. P., W. Ross & R. L. Gourse, (2008) Advances in bacterial promoter recognition and its control by factors that do not bind DNA. Nat Rev Microbiol 6: 507-519.

He S, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG, Wang Z, Chen F, Lindquist EA, Sorek R, Hugenholtz P. (2010) Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods. 7(10):807-12.

Heilemann M, Lymperopoulos K, Wigneshweraraj SR, Buck B, Kapanidis AN (2007) Studying σ54-dependent transcription at the single-molecule level using alternating-laser excitation (ALEX) spectroscopy. Optics InfoBase > Conference Papers > ECBO > 2007 > PS > Page 6633_92.

Helmann JD. (2002) The extracytoplasmic function (ECF) sigma factors. Adv Microb Physiol. 2002;46:47-110.

Hengge-Aronis R (2002) Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. Microbiol Mol Biol Rev. 2002 Sep;66(3):373-95.

Ho, Y. S., L. M. Burden, and J. H. Hurley (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. EMBO J. 19:5288-5299.

Hobbs EC, Storz G. (2012) Competition assays using barcoded deletion strains to gain insight into small RNA function. Methods Mol Biol. 905:63-72.

Hong, E., Doucleff, M., and Wemmer, D.E. (2009). Structure of the RNA polymerase core-binding domain of sigma(54) reveals a likely conformational fracture point. Journal of molecular biology 390, 70-82.

Hoover, T. R., E. Santero, S. Porter, and S. Kustu. (1990). The integration host factor stimulates interaction of RNA polymerase with NIFA, the transcriptional activator for nitrogen fixation operons. Cell 63:11-22.

Hopper S, Babst M, Schlensog V, Fischer HM, Hennecke H, Böck A. (1994) Regulated expression in vitro of genes coding for formate hydrogenlyase components of Escherichia coli. J Biol Chem. 269(30):19597-604.

Hopper S, Böck A (1995) Effector-mediated stimulation of ATPase activity by the sigma 54-dependent transcriptional activator FHLA from Escherichia coli. J Bacteriol. 177(10):2798-803.

Hsieh, M. & J. D. Gralla, (1994) Analysis of the N-terminal leucine heptad and hexad repeats of sigma 54. J Mol Biol 239: 15-24.

Huerta AM, Collado-Vides J. (2003) Sigma70 promoters in Escherichia coli: specific transcription in dense regions of overlapping promoter-like signals. J Mol Biol. 333(2):261-78.

Hughes KT, Mathee K. (1998) The anti-sigma factors. Annu Rev Microbiol.52:231-86.

Ibañez E1, Campos E, Baldoma L, Aguilar J, Badia J. (2000) Regulation of expression of the yiaKLMNOPQRS operon for carbohydrate utilization in Escherichia coli: involvement of the main transcriptional factors. J Bacteriol. 182(16):4617-24.

Ikeda H, Tomizawa JI. (1965) Transducing fragments in generalized transduction by phage P1. 3. Studies with small phage particles. J Mol Biol. 14(1): 120-9.

Jensen KF. (1993) The Escherichia coli K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. J Bacteriol. 175(11):3401-7.

Jiang M, Datta K, Walker A, Strahler J, Bagamasbad P, Andrews PC, Maddock JR. (2006) The Escherichia coli GTPase CgtAE is involved in late steps of large ribosome assembly. J Bacteriol. 188(19):6757-70.

Johansson LUM, Lisandro DS, Bernardo MD, Moscoso JA and Shingler V (2008) s54-RNA polymerase controls s70-dependent transcription from a non-overlapping divergent promoter. Molecular Microbiology. 70(3), 709–723

Joly, N., and M. Buck (2010) Engineered interfaces of an AAA+ ATPase reveal a new nucleotide-dependent coordination mechanism. J Biol Chem 285:15178-15186.

Joly N, Zhang N, Buck M, Zhang X (2012) Coupling AAA protein function to regulated gene expression. Biochim Biophys Acta. 1823(1):108-16.

Jones PG, VanBogelen RA, Neidhardt FC. (1987) Induction of proteins in response to low temperature in Escherichia coli. J Bacteriol. 169(5):2092-5.

Jones S (2009) Bacterial transcription: Sigma 54 minds the gap. Nature Reviews Microbiology 7, 3.

Jovanovic G, Mehta P, Ying L, Buck M. (2014) Anionic lipids and the cytoskeletal proteins MreB and RodZ define the spatio-temporal distribution and function of membrane stress controller PspA in Escherichia coli. Microbiology. 160(Pt 11):2374-86.

Jovanovic M, Burrows PC, Bose D, Cámara B, Wiesler S, Zhang X, Wigneshweraraj S, Weinzierl RO, Buck M. (2011) Activity map of the Escherichia coli RNA polymerase bridge helix. J Biol Chem. 286(16):14469-79.

Jovanovic M, James EH, Burrows PC, Rego FG, Buck M, Schumacher J (2011) Regulation of the co-evolved HrpR and HrpS AAA+ proteins required for Pseudomonas syringae pathogenicity. Nat Commun. 2:177.

Jovanovic G, Weiner L, Model P (1996) Identification, nucleotide sequence, and characterization of PspF, the transcriptional activator of the Escherichia coli stress-induced psp operon. J Bacteriol. 178(7):1936-45.

Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science. 314(5802):1144-7.

Kapanidis, A.N., Margeat, E., Laurence, T.A., Doose, S., Ho, S.O., Mukhopadhyay, J., Kortkhonjia, E., Mekler, V., Ebright, R.H., and Weiss, S. (2005). Retention of transcription initiation factor sigma70 in transcription elongation: single-molecule analysis. Molecular cell 20, 347-356.

Kawano M, Aravind L, Storz G. (2007) An antisense RNA controls synthesis of an SOS-induced toxin evolved from an antitoxin. Mol Microbiol. 64(3):738-54.

Keseler, I.M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I.T., Peralta-Gil, M. and Karp, P.D. (2005) EcoCyc: a comprehensive database resource for Escherichia coli. Nucleic Acids Res., 33, D334-337.

Khlebnikov A, Skaug T, Keasling JD (2002). Modulation of gene expression from the arabinose-inducible araBAD promoter. Journal of Industrial Microbiology and Biotechnology. 29(1): 34-37.

Kill K, Binnewies TT, Sicheritz-Ponten T, Willenbrock H, Hallin PF, Wassenaar TM, Ussery DW (2005) Genome update: sigma factors in 240 bacterial genomes. Microbiology. 151(Pt 10):3147-3150.

Kim Y, Wang X, Zhang XS, Grigoriu S, Page R, Peti W, Wood TK. (2010) Escherichia coli toxin/antitoxin pair MqsR/MqsA regulate toxin CspD. Environ Microbiol. 12(5):1105-21.

Kim, E.S., Song, J.Y., Kim, D.W., Chater, K.F., and Lee, K.J. (2008). A possible extended family of regulators of sigma factor activity in Streptomyces coelicolor. Journal of bacteriology 190, 7559-7566.

Kontur WS, Saecker RM, Davis CA, Capp MW, Record MT Jr. (2006) Solute probes of conformational changes in open complex (RPo) formation by Escherichia coli RNA polymerase at the lambdaPR promoter: evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RPo. Biochemistry 45(7):2161-77.

Korea CG, Badouraly R, Prevost MC, Ghigo JM, Beloin C. (2010) Escherichia coli K-12 possesses multiple cryptic but functional chaperone-usher fimbriae with distinct surface specificities. Environ Microbiol. 2010 Jul;12(7):1957-77.

Korzheva N, Mustaev A, Kozlov M, Malhotra A, Nikiforov V, Goldfarb A, Darst SA. (2000) A structural model of transcription elongation. Science. 289(5479):619-25.

Kullik, I., S. Fritsche, H. Knobel, J. Sanjuan, H. Hennecke & H. M. Fischer, (1991) Bradyrhizobium japonicum has two differentially regulated, functional homologs of the sigma 54 gene (rpoN). J Bacteriol 173: 1125-1138.

Kumar R, Shimizu K (2010) Metabolic regulation of Escherichia coli and its gdhA, glnL, gltB, D mutants under different carbon and nitrogen limitations in the continuous culture. Microbial Cell Factories. 9:8

Kustu S, Santero E, Keener J, Popham D, Weiss D (1989) Expression of sigma 54 (ntrA)-dependent genes is probably united by a common mechanism. Microbiol Rev. 53(3):367-76.

Kuznedelov, K., Korzheva, N., Mustaev, A., & Severinov, K. (2002). Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA–DNA hybrid length. The EMBO Journal, 21(6), 1369–1378.

La Teana A, Brandi A, Falconi M, Spurio R, Pon CL, Gualerzi CO. (1991) Identification of a cold shock transcriptional enhancer of the Escherichia coli gene encoding nucleoid protein H-NS. Proc Natl Acad Sci U S A. 1;88(23):10907-11.

Lang, B., Blot, N., Bouffartigues, E., Buckle, M., Geertz, M., Gualerzi, C. O., Travers, A. (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.

Laskos L., Dillard J. P., Seifert H. S., Fyfe J. A. M., Davies J. K. (1998) The pathogenic neisseriae contain an inactive rpoN gene and do not utilize the pilE ς^{54} promoter. Gene 208, 95–102.

Lawton E, Jovanovic M, Joly N, Waite C, Zhang N, Wang B, Burrows P, Buck M. (2014) Determination of the self-association residues within a homomeric and a heteromeric AAA+ enhancer binding protein. J Mol Biol. 426(8):1692-710.

Leang C, Krushkal J, Ueki T, Puljic M, Sun J, Juárez K, Núñez C, Reguera G, DiDonato R, Postier B (2009) Genome-wide analysis of the RpoN regulon in Geobacter sulfurreducens. BMC Genomics. 10:331.

Lee CR, Park YH, Kim M, Kim YR, Park S, Peterkofsky A, Seok YJ. (2013) Reciprocal regulation of the autophosphorylation of enzyme INtr by glutamine and α-ketoglutarate in Escherichia coli. Mol Microbiol. 88(3):473-85.

Lee SJ, Xie A, Jiang W, Etchegaray JP, Jones PG, Inouye M. (1994) Family of the major cold-shock protein, CspA (CS7.4), of Escherichia coli, whose members show a high sequence similarity with the eukaryotic Y-box binding proteins. Mol Microbiol. 11(5):833-9.

Liu L., Li Y., Li S., et al. (2012) Comparison of next-generation sequencing systems, Journal of biomedicine & biotechnology. 2012, 251364.

Lonetto M1, Gribskov M, Gross CA. (1992) The sigma 70 family: sequence conservation and evolutionary relationships. J Bacteriol. 1992 Jun;174(12):3843-9.

Metzker M.L. (2009) Sequencing technologies - the next generation, Nature reviews Genetics. 11, 31-46.

Malinverni JC1, Silhavy TJ. (2009) An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. Proc Natl Acad Sci U S A. 106(19):8009-14.

Mariani, R. & S. I. Maffioli, (2009) Bacterial RNA polymerase inhibitors: an organized overview of their structure, derivatives, biological activity and current clinical development status. Curr Med Chem 16: 430-454.

Markov D, Naryshkina T, Mustaev A, Severinov K. (1999) A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. Genes Dev. 13(18):2439-48.

Mathew R, Chatterji D. (2006) The evolving story of the omega subunit of bacterial RNA polymerase. Trends Microbiol. 2006 Oct;14(10):450-5.

Maxson ME, Darwin AJ. (2006) Multiple promoters control expression of the Yersinia enterocolitica phage-shock-protein A (pspA) operon. Microbiology. 152(Pt 4):1001-10.

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

Mecsas J, Cowing DW, Gross CA (1991) Development of RNA polymerase-promoter contacts during open complex formation. J Mol Biol. 220(3):585-97.

Mennes AM, Quint A, Gribnau JH, Boot CJ, van der Zaal EJ, Maan AC, Libbenga KR. (1992) Specific transcription and reinitiation of 2,4-D-induced genes in tobacco nuclei. Plant Mol Biol. 18(1):109-17.

Merrick MJ (1993) In a class of its own--the RNA polymerase sigma factor sigma 54 (sigma N). Mol Microbiol. 10(5):903-9.

Merrick, M., and Chambers, S. (1992). The helix-turn-helix motif of sigma 54 is involved in recognition of the -13 promoter region. Journal of bacteriology 174, 7221-7226.

Metzker M (2010). Sequencing technologies — the next generation. *Nature Reviews Genetics* **11**, 31-46

Michiels, J., M. Moris, B. Dombrecht, C. Verreth & J. Vanderleyden, (1998) Differential regulation of Rhizobium etli rpoN2 gene expression during symbiosis and free-living growth. J Bacteriol 180: 3620-3628.

Mitchell, J. E., Zheng, D., Busby, S. J. W., & Minchin, S. D. (2003). Identification and analysis of "extended –10" promoters in Escherichia coli. Nucleic Acids Research, 31(16), 4689–4695.

Mooney, R. A., I. Artsimovitch & R. Landick, (1998) Information processing by RNA polymerase: recognition of regulatory signals during RNA chain elongation. J Bacteriol 180: 3265-3275.

Morett E, Segovia (1993) The sigma 54 bacterial enhancer-binding protein family: mechanism of action and phylogenetic relationship of their functional domains. J Bacteriol. 175(19):6067-74.

Mukhopadhyay, J., A. N. Kapanidis, V. Mekler, E. Kortkhonjia, Y. W. Ebright & R. H. Ebright, (2001) Translocation of sigma(70) with RNA polymerase during transcription: fluorescence resonance energy transfer assay for movement relative to DNA. Cell 106: 453-463.

Murakami KS, Darst SA (2003) Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol. 13(1):31-9.

Murakami, K. S., S. Masuda & S. A. Darst, (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. Science 296: 1280-1284.

Naryshkina, T., Kuznedelov, K., and Severinov, K. (2006). The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. Journal of molecular biology 361, 634-643.

Nechaev S, Chlenov M, Severinov K. (2000) Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. J Biol Chem. 275(33):25516-22.

Nedialkov YA, Opron K, Assaf F, Artsimovitch I, Kireeva ML, Kashlev M, Cukier RI, Nudler E, Burton ZF. (2013) The RNA polymerase bridge helix YFI motif in catalysis, fidelity and translocation. Biochim Biophys Acta. 1829(2):187-98.

Ninfa AJ, Reitzer LJ, Magasanik B (1987) Initiation of transcription at the bacterial glnAp2 promoter by purified E. coli components is facilitated by enhancers. Cell. 50(7):1039-46.

Nudler, E., (2009) RNA polymerase active center: the molecular engine of transcription. Annu Rev Biochem 78: 335-361.

Oguiza JA, Gallegos MT, Chaney MK, Cannon WV, Buck M. (1999) Involvement of the sigmaN DNA-binding domain in open complex formation. Mol Microbiol. 33(4):873-85.

Ogura, T. & A. J. Wilkinson, (2001) AAA+ superfamily ATPases: common structure--diverse function. Genes Cells 6: 575-597.

Okoniewski MJ, Miller CJ. (2006) Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. BMC Bioinformatics. 7:276.

Opalka N, Brown J, Lane WJ, Twist KA, Landick R, Asturias FJ, Darst SA. (2010) Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. PLoS Biol. 8(9). pii: e1000483.

Palacios S, Escalante-Semerena JC (2004) 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of Salmonella enterica by the PrpR protein. Microbiology. 150(Pt 11):3877-87.

Periago PM (2002) Identification of proteins involved in the heat stress response of Bacillus cereus ATCC 14579. Appl. Environ. Microbiol. 68: 3486–3495

Paget, M. S., & Helmann, J. D. (2003). The σ70 family of sigma factors. Genome Biology, 4(1), 203.

Perederina A, Svetlov V, Vassylyeva MN, Tahirov TH, Yokoyama S, Artsimovitch I, Vassylyev DG. (2004) Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. Cell. 118(3):297-309.

Poggio S, Domeinzain C, Osorio A, Camarena L. (2002) The nitrogen assimilation control (Nac) protein represses asnC and asnA transcription in Escherichia coli. FEMS Microbiol Lett. 206(2):151-6.

Polen T, Rittmann D, Wendisch VF, Sahm H. (2003) DNA microarray analyses of the long-term adaptive response of Escherichia coli to acetate and propionate. Appl Environ Microbiol. 69(3):1759-74.

Polissi A, De Laurentis W, Zangrossi S, Briani F, Longhi V, Pesole G, Dehò G. (2003) Changes in Escherichia coli transcriptome during acclimatization at low temperature. Res Microbiol. 154(8):573-80.

Popham DL, Szeto D, Keener J, Kustu S. (1989) Function of a bacterial activator protein that binds to transcriptional enhancers. Science. 243(4891), 629-35.

Poulsen P, Jensen KF. (1991) Three genes preceding pyrE on the Escherichia coli chromosome are essential for survival and normal cell morphology in stationary culture and at high temperature. Res Microbiol. 142(2-3):283-8.

Powell BS, Court DL, Inada T, Nakamura Y, Michotey V, Cui X, Reizer A, Saier MH Jr, Reizer J. (1995) Novel proteins of the phosphotransferase system encoded within the rpoN operon of Escherichia coli. Enzyme IIANtr affects growth on organic nitrogen and the conditional lethality of an erats mutant. J Biol Chem. 270(9):4822-39.

Puhler, G., H. Leffers, F. Gropp, P. Palm, H. P. Klenk, F. Lottspeich, R. A. Garrett & W. Zillig (1989) Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci U S A 86: 4569-4573.

Raghavan R, Sage A, Ochman H. (2011) Genome-wide identification of transcription start sites yields a novel thermosensing RNA and new cyclic AMP receptor protein-regulated genes in Escherichia coli. J Bacteriol. 193(11):2871-4.

Rao L, Ross W, Appleman JA, Gaal T, Leirmo S, Schlax PJ, Record MT Jr, Gourse RL. (1994) Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. J Mol Biol. 235(5):1421-35.

Rappas M, Schumacher J, Beuron F, Niwa H, Bordes P, Wigneshweraraj S, Keetch CA, Robinson CV, Buck M, Zhang X (2005) Structural insights into the activity of enhancer-binding proteins. Science. 307(5717):1972-5.

Rappas, M., J. Schumacher, H. Niwa, M. Buck, and X. Zhang (2006) Structural basis of the nucleotide driven conformational changes in the AAA+ domain of transcription activator PspF. J Mol Biol 357:481-492.

Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, Bui OT, Knight EM, Fong SS, Palsson BO. (2006) Systems approach to refining genome annotation. Proc Natl Acad Sci U S A. 103(46):17480-4.

Reichenbach B, Göpel Y and Görke B (2009) Dual control by perfectly overlapping s54- and s70-promoters adjusts small RNA GlmY expression to different environmental signals. Molecular Microbiology 74Linda U. M. Johansson, Dafne Solera,†

Reitzer LJ, Movsas B, Magasanik B. (1989) Activation of glnA transcription by nitrogen regulator I (NRI)-phosphate in Escherichia coli: evidence for a long-range physical interaction between NRI-phosphate and RNA polymerase. J Bacteriol. 171(10):5512-22.

Reitzer, L. & B. L. Schneider, (2001) Metabolic context and possible physiological themes of sigma(54)-dependent genes in Escherichia coli. Microbiol Mol Biol Rev 65: 422-444.

Ren CP, Chaudhuri RR, Fivian A, Bailey CM, Antonio M, Barnes WM, Pallen MJ. (2004) The ETT2 gene cluster, encoding a second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition. J Bacteriol. 186(11):3547-60.

Revyakin A, Liu C, Ebright RH, Strick TR (2006) Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science. 314(5802):1139-43.

Rezuchova B, Miticka H, Homerova D, Roberts M, Kormanec J. (2003). New members of the Escherichia coli sigmaE regulon identified by a two-plasmid system. FEMS Microbiol Lett. 225(1):1-7.

Richard DJ, Sawers G, Sargent F, McWalter L, Boxer DH. (1999) Transcriptional regulation in response to oxygen and nitrate of the operons encoding the [NiFe] hydrogenases 1 and 2 of Escherichia coli. Microbiology. 145 (Pt 10):2903-12.

Rick Astley (1987) Never gonna give you up. We're no strangers to love, you know the rules and so do I. A full commitment's what I'm thinking of, you wouldn't get this from any other guy. Anal. Biochem. 1(3):3-7.

Roberts (2011) Improving RNA-Seq expression estimates by correcting for fragment bias. Genome biology. 12,R22.

Royce TE, Rozowsky JS, Gerstein MB. (2007) Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. Nucleic Acids Res. 2007;35(15):e99.

Saecker, R.M., Record, M.T., Jr., and Dehaseth, P.L. (2011). Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. Journal of molecular biology 412, 754-771.

Sallai, L., and P. A. Tucker (2005) Crystal structure of the central and C-terminal domain of the sigma(54)-activator ZraR. J Struct Biol 151:160-170.

Sangita Phadtare (2004) Recent Developments in Bacterial Cold-Shock Response Curr. Issues Mol. Biol. 6: 125-136.

Sasse-Dwight S, Gralla JD. (1990) Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor sigma 54. Cell. 62(5):945-54.

Saurin W, Hofnung M, Dassa E. (1999) Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters. J Mol Evol. 48(1):22-41.

Schneider BL, Hernandez VJ, Reitzer L. (2013) Putrescine catabolism is a metabolic response to several stresses in Escherichia coli. Mol Microbiol. 88(3):537-50.

Schneider DA, Ross W, Gourse RL. (2003) Control of rRNA expression in Escherichia coli. Curr Opin Microbiol. 6(2):151-6.

Schneider R, Lurz R, Lüder G, Tolksdorf C, Travers A, Muskhelishvili G. (2001) An architectural role of the Escherichia coli chromatin protein FIS in organising DNA. Nucleic Acids Res. 15;29(24):5107-14.

Schumacher J, Joly N, Rappas M, Bradley D, Wigneshweraraj SR, Zhang X and Buck M (2007) Sensor I Threonine of the AAA+ ATPase Transcriptional Activator PspF Is Involved in Coupling Nucleotide Triphosphate Hydrolysis to the Restructuring of σ54-RNA Polymerase. The Journal of Biological Chemistry, 282, 9825-9833.

Schumacher, J., N. Joly, M. Rappas, X. Zhang, and M. Buck. (2006). Structures and organisation of AAA+ enhancer binding proteins in transcriptional activation. J Struct Biol 156, 190-199.

Severinov, K., (2000) RNA polymerase structure-function: insights into points of transcriptional regulation. Curr Opin Microbiol 3: 118-125.

Sharma CM, Vogel J (2009) Experimental approaches for the discovery and characterization of regulatory small RNA. Current Opinion in Microbiology. 12(5): 536–546.

Sharma, A., Leach, R. N., Gell, C., Zhang, N., Burrows, P. C., Shepherd, D. A., Tuma, R. (2014). Domain movements of the enhancer-dependent sigma factor drive DNA delivery into the RNA polymerase active site: insights from single molecule studies. Nucleic Acids Research, 42(8), 5177–5190.

Sheikh J, Hicks S, Dall'Agnol M, Phillips AD, Nataro JP. (2001) Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli. Mol Microbiol. 41(5):983-97.

Shingler, V., and H. Pavel (1995) Direct regulation of the ATPase activity of the transcriptional activator DmpR by aromatic compounds. Mol Microbiol 17:505-513.

Sokalingam S, Raghunathan G, Soundrarajan N, Lee S-G (2012) A Study on the Effect of Surface Lysine to Arginine Mutagenesis on Protein Stability and Structure Using Green Fluorescent Protein. PLoS ONE 7(7): e40410.

Spassky A, Rimsky S, Buc H, Busby S (1988) Correlation between the conformation of Escherichia coli –10 hexamer sequences and promoter strength: use of orthophenanthroline cuprous complex as a structural index. EMBO J. 7:1871-1879.

Stafford GP, Ogi T, Hughes C. (2005) Binding and transcriptional activation of non-flagellar genes by the Escherichia coli flagellar master regulator FlhD2C2. Microbiology. 151(Pt 6):1779-88.

Steitz, T. A. (2009) The structural changes of T7 RNA polymerase from transcription initiation to elongation. Curr. Opin. Struct. Biol. 19, 683–690.

Studholme, D.J., and Dixon, R. (2003). Domain architectures of sigma54-dependent transcriptional activators. Journal of bacteriology 185, 1757-1767.

Sümegi J, Udvardy A, Venetianer P (1977) In vitro transcription of the ribosomal RNA genes of E. coli DNA. Molecular and General Genetics MGG 16.III. 151(3): 305-312.

Sweetser D, Nonet M, Young RA (1987) Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. Proc Natl Acad Sci U S A. 84(5): 1192–1196.

Syed A, Gralla JD (1997) Isolation and properties of enhancer-bypass mutants of sigma 54. Mol Microbiol. 23(5):987-95.

Syed A, Gralla JD. (1998) Identification of an N-terminal region of sigma 54 required for enhancer responsiveness. J Bacteriol. 180(21):5619-25.

Tanabe H, Goldstein J, Yang M, Inouye M. (1992) Identification of the promoter region of the Escherichia coli major cold shock gene, cspA. J Bacteriol. 174(12):3867-73.

Tchieu JH, Norris V, Edwards JS, Saier MH Jr. (2001) The complete phosphotransferase system in Escherichia coli. J Mol Microbiol Biotechnol. 3(3):329-46.

Tian ZX, Li QS, Buck M, Kolb A, Wang YP. (2001) The CRP-cAMP complex and downregulation of the glnAp2 promoter provides a novel regulatory linkage between carbon metabolism and nitrogen assimilation in Escherichia coli. Mol Microbiol. 41(4):911-24.

Tina Ivancic, Polona Jamnik and David Stopar (2013) Cold shock CspA and CspB protein production during periodic temperature cycling in Escherichia coli. BMC Research Notes, 6:248.

Tintut Y, Wang JT and Gralla JD (1995) Abortive Cycling and the Release of Polymerase for Elongation at the σ54-dependent glnAp2 Promoter. The Journal of Biological Chemistry, 270, 24392-24398.

Toulmé F, Mosrin-Huaman C, Sparkowski J, Das A, Leng M, Rahmouni AR. (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. EMBO J. 19(24):6853-9.

Toulokhonov I, Artsimovitch I, Landick R. (2001) Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science. 292(5517):730-3.

Valvano MA, Marolda CL, Bittner M, Glaskin-Clay M, Simon TL, Klena JD. (2000) The rfaE gene from Escherichia coli encodes a bifunctional protein involved in biosynthesis of the lipopolysaccharide core precursor ADP-L-glycero-D-manno-heptose. J Bacteriol. 182(2):488-97.

Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. Nature 417, 712-719.

Vecerek B, Rajkowitsch L, Sonnleitner E, Schroeder R, Bläsi U. (2008) The C-terminal domain of Escherichia coli Hfq is required for regulation. Nucleic Acids Res. 36(1):133-43.

Verger D, Carr PD, Kwok T, Ollis DL (2007) Crystal structure of the N-terminal domain of the TyrR transcription factor responsible for gene regulation of aromatic amino acid biosynthesis and transport in Escherichia coli K12. J Mol Biol. 367(1):102-12.

Vrentas C, Gaal T, Berkmen MB, Rutherford ST, Haugen SP, Vassylyev DG, Ross W, Gourse RL (2008). Still looking for the magic spot: the crystallographically defined binding site for ppGpp on RNA polymerase is unlikely to be responsible for rRNA transcription regulation. J Mol Biol. 377(2):551-64.

Vrentas CE, Gaal T, Ross W, Ebright RH, Gourse RL (2005). Response of RNA polymerase to ppGpp: requirement for the omega subunit and relief of this requirement by DksA. Genes Dev. 19(19):2378-87.

Wang JT, Gralla JD (1996) The transcription initiation pathway of sigma 54 mutants that bypass the enhancer protein requirement. Implications for the mechanism of activation. J Biol Chem. 271(51):32707-13.

Wang JT, Syed A, Hsieh M, Gralla JD (1995) Converting Escherichia coli RNA polymerase into an enhancer-responsive enzyme: role of an NH2-terminal leucine patch in sigma 54. Science. 270(5238):992-4.

Wang L, Gralla JD (2001) Roles for the C-terminal region of sigma 54 in transcriptional silencing and DNA binding.

Wang L, Guo Y, Gralla JD (1999) Regulation of Sigma 54-Dependent Transcription by Core Promoter Sequences: Role of −12 Region Nucleotides. J. Bacteriol. 181(24): 7558-7565.

Wang X, Lord DM, Cheng HY, Osbourne DO, Hong SH, Sanchez-Torres V, Quiroga C, Zheng K, Herrmann T, Peti W, Benedik MJ, Page R, Wood TK. (2012) A new type V toxin-antitoxin system where mRNA for toxin GhoT is cleaved by antitoxin GhoS. Nat Chem Biol. 8(10):855-61.

Wang X, Lord DM, Hong SH, Peti W, Benedik MJ, Page R, Wood TK. (2013) Type II toxin/antitoxin MqsR/MqsA controls type V toxin/antitoxin GhoT/GhoS. Environ Microbiol. 15(6):1734-44.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10, 57-63.

Wang JT, Syed A, Gralla JD. (1997) Multiple pathways to bypass the enhancer requirement of sigma54 RNA polymerase: roles for DNA and protein determinants. Proc Natl Acad Sci U S A. 94(18), 9538-43.

Wang, Y. K., J. H. Lee, J. M. Brewer, and T. R. Hoover. (1997). A conserved region in the sigma54-dependent activator DctD is involved in both binding to RNA polymerase and coupling ATP hydrolysis to activation. Mol Microbiol 26:373-386.

Weber MH, Marahiel MA. (2003) Bacterial cold shock responses. Sci Prog. 86(Pt 1-2):9-75.

Wedel A, Kustu S. (1995). The bacterial enhancer-binding protein NTRC is a molecular machine: ATP hydrolysis is coupled to transcriptional activation. Genes Dev. 9(16):2042-52.

Wiegeshoff F, Beckering CL, Debarbouille M, Marahiel MA (2006) Sigma L is important for cold shock adaptation of Bacillus subtilis. J Bacteriol. 188(8):3130-3133.

Wiesler SC, Burrows PC, Buck M. (2012) A dual switch controls bacterial enhancer-dependent transcription. Nucleic Acids Res. 40(21):10878-92.

Wigneshweraraj SR, Burrows PC, Severinov K, Buck M. (2005) Stable DNA opening within open promoter complexes is mediated by the RNA polymerase beta'-jaw domain. J Biol Chem. 280(43):36176-84.

Wigneshweraraj SR, Kuznedelov K, Severinov K, Buck M. (2003) Multiple roles of the RNA polymerase beta subunit flap domain in sigma 54-dependent transcription. J Biol Chem. 278(5):3455-65..

Wigneshweraraj, S. R., P. C. Burrows, et al. (2005). The second paradigm for activation of transcription. Progress in Nucleic Acid Research and Molecular Biology, Academic Press. 79: 339-369.

Wigneshweraraj, S., D. Bose, P. C. Burrows, N. Joly, J. Schumacher, M. Rappas, T. Pape, X. Zhang, P. Stockley, K. Severinov, and M. Buck. (2008). Modus operandi of the bacterial RNA polymerase containing the sigma54 promoter-specificity factor. Mol Microbiol 68, 538-546.

Wolfe AJ, Millikan DS, Campbell JM, Visick KL: Vibrio fischeri sigma54 controls motility, biofilm formation, luminescence, and colonization. Appl Environ Microbiol 2004, 70(4):2520-2524.

Wong, C., and J. D. Gralla. (1992). A role for the acidic trimer repeat region of transcription factor sigma 54 in setting the rate and temperature dependence of promoter melting in vivo. J Biol Chem 267, 24762-24768.

Wong, C., Y. Tintut & J. D. Gralla, (1994) The domain structure of sigma 54 as determined by analysis of a set of deletion mutants. J Mol Biol 236: 81-90.

Yamamoto K, Hirao K, Oshima T, Aiba H, Utsumi R, Ishihama A (2005) Functional characterization in vitro of all two-component signal transduction systems from Escherichia coli. J Biol Chem. 2005 Jan 14;280(2):1448-56.

Yang, C. C., and H. A. Nash. (1989). The interaction of E. coli IHF protein with its specific binding sites. Cell 57, 869-880.

Young BA, Gruber TM, Gross CA. (2004) Minimal machinery of RNA polymerase holoenzyme sufficient for promoter melting. Science. 303(5662):1382-4.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U. (2006) A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. Nat Methods. 3(8):623-8.

Zaychikov E, Martin E, Denissova L, Kozlov M, Markovtsov V, Kashlev M, Heumann H, Nikiforov V, Goldfarb A, Mustaev A. (1996) Mapping of catalytic residues in the RNA polymerase active center. Science. 273(5271):107-9.

Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA. (1999) Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. Cell. 98(6):811-24.

Zhang N, Joly N, Buck M. (2012) A common feature from different subunits of a homomeric AAA+ protein contacts three spatially distinct transcription elements. Nucleic Acids Res. 40(18):9139-52.

Zhang N, Joly N, Burrows PC, Jovanovic M, Wigneshweraraj SR, Buck M. (2009) The role of the conserved phenylalanine in the sigma54-interacting GAFTGA motif of bacterial enhancer binding proteins. Nucleic Acids Res. 37(18):5981-92.

Zhang, X., M. Chaney, S. R. Wigneshweraraj, J. Schumacher, P. Bordes, W. Cannon, and M. Buck. (2002). Mechanochemical ATPases and transcriptional activation. Mol Microbiol 45:895-903.

Zhao B, Yeo CC, Poh CL. (2005) Proteome investigation of the global regulatory role of sigma 54 in response to gentisate induction in Pseudomonas alcaligenes NCIMB 9867. Proteomics. 5(7):1868-76.

Zhao K, Liu M, Burgess RR. (2010) Promoter and regulon analysis of nitrogen assimilation factor, sigma54, reveal alternative strategy for E. coli MG1655 flagellar biosynthesis. Nucleic Acids Res. 38(4):1273-83.

Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques 30: 892-897.

Zimmerman SB. (2006) Cooperative transitions of isolated Escherichia coli nucleoids: implications for the nucleoid as a cellular phase. J Struct Biol. 153(2):160-75.

Zou C, Fujita N, Igarashi K, Ishihama A. (1992) Mapping the cAMP receptor protein contact site on the alpha subunit of Escherichia coli RNA polymerase. Mol Microbiol. 6(18):2599-605.

Appendix A (primers)

Name	Sequence (5' to 3')	Description
GreA F	GTGAATTGTAGCTGACCTGG	Amplify *greA* gene
GreA R	ACAATACATCAACATCTTGAG	Amplify *greA* gene
AsnA F	GAATGATTATTGCATGTGT	Amplify *asnA* gene
AsnA R	CGTCACGTCGTCAGGCAAC	Amplify *asnA* gene
TtdR F	GTGAATATGTAATTAACAG	Amplify *ttdR* gene
TtdR R	GACAACATCCAACGTCTGTG	Amplify *ttdR* gene
HldE F	GAACTGGTGCGGGCAAGCTG	Amplify *hldE* gene
HldE R	ATGTACGAAGCGAGATCTGTG	Amplify *hldE* gene
NanT F	GGCGCTGGCCCAGCAGTTG	Amplify *nanT* gene
NanT R	CCGGAACCGGCTGGCAGGAG	Amplify *nanT* gene
Tn10 F	GAATTGGTAAAGAGAGTCGTG	Amplify *tn10* transposon

Tn10 R	GGGAAGACTCACTCGTGGCTG	Amplify *tn10* transposon
PspA UAS	CATGATGAAATTCGCCACTTG	Amplify ΦP_{pspA}-*lacZ*
LacZ F	ATGGGTATTTTTTCTCGC	Amplify ΦP_{pspA}-*lacZ*
LacZ Int F	CTGACCGATCTGATTAAGTCC	Amplify ΦP_{pspA}-*lacZ*
LacZ R	CCGCCACATATCCTGATCTTC	Amplify ΦP_{pspA}-*lacZ*
pBAD F	ATG CCA TAG CAT TTT TAT CC	Sequencing pBAD18 inserts
pBAD R	GAT TTA ATC TGT ATC AGG	Sequencing pBAD18 inserts
Down A	TACCCGTGATATTGCTGAAGAGC	Inverse PCR identification
Down B	AGCCGGCCACAGTCGATGAATCC	Inverse PCR identification
Up A	TTGTCAAGACCGACCTGTCC	Inverse PCR identification
Up B	CAGTCATAGCCGAATAGCCTCTCC	Inverse PCR identification
Kan F	ATCTGATCCTTCAACTCAGC	Amplify Kanamycin cassette
Kan R	TTAGAAAAACTCATCGAGCATC	Amplify Kanamycin cassette

Appendix B (RNASeq)

RNASeq data for selected genes (greater than 4-fold change in expression between any two of the 3 strains), expressed in Reads Per Kb exon (gene) per Million mapped reads (RPKM).

				Wild type *rpoN*		*rpoN*$_{\Delta R1}$		No *rpoN*	
Gene	Gene length	Chromosome start region	Chromosome end region	Unique reads	RKPM	Unique reads	RKPM	Unique reads	RKPM
adhE	2776	1294669	1297444	1986	49	14761	390	10058	358
alaA	1318	2405483	2406800	1104	58	6499	362	4367	327
alaE	550	2797086	2797635	503	63	82	11	45	8
alaT	176	4035183	4035358	24489	9557	3219	1342	0	1429
alaU	176	3424980	3425155	24529	9573	3233	1348	0	1453
alaV	176	225400	225575	24345	9501	3186	1328	0	1477
argA	1432	2947164	2948595	1659	80	7542	386	4758	328
argD	1321	3486982	3488302	486	25	3051	169	2285	171
argT	883	2425031	2425913	1636	127	7196	598	4919	550
argU	177	563846	564022	5613	2178	1533	636	597	400
aroA	1384	957935	959318	1823	90	7851	416	5705	407
artJ	832	899067	899898	677	56	6209	548	3984	473
artP	829	902229	903057	377	31	1960	174	1395	171
asnW	176	2056051	2056226	4908	1915	1328	554	173	364
bioA	1390	807191	808580	501	25	2005	106	1129	85

bioB	1141	808467	809607	399	24	2480	160	1189	114
bioC	856	810645	811500	357	29	2114	181	975	120
bioD	778	811393	812170	282	25	1391	131	851	116
bioF	1255	809504	810758	522	29	3021	177	1634	132
carA	1249	29551	30799	375	21	2618	154	2244	179
carB	3322	30717	34038	1146	24	7310	161	6464	193
cirA	2092	2242800	2244891	839	28	3428	120	2127	100
codB	1360	354046	355405	237	12	1623	88	1155	88
csgF	517	1100934	1101450	33	4	127	18	84	16
cspA	313	3717972	3718284	106796	23436	2383	559	1466	463
cspB	316	1639363	1639678	61464	13360	1099	255	635	199
cspG	313	1050584	1050896	27042	5934	156	37	97	31
cspH	313	1050186	1050498	615	135	63	15	36	11
cspI	313	1636479	1636791	3049	669	96	23	48	15
cvpA	589	2428297	2428885	674	79	6229	776	4553	769
cysA	1198	2537739	2538936	1552	89	15306	938	9100	853
cysC	706	2871409	2872114	399	39	3674	382	2462	375
cysD	1009	2873443	2874451	529	36	9224	671	6923	695
cysH	835	2885600	2886434	322	26	4153	365	2802	332
cysI	1813	2886409	2888221	1472	56	14636	592	10440	595
cysJ	1900	2888121	2890020	1384	50	19091	737	13096	710
cysK	1072	2530331	2531402	13001	833	82068	5618	60141	5543
cysM	1012	2536694	2537705	778	53	2311	168	1844	180
cysN	1528	2872014	2873541	990	45	12943	622	8542	582
cysP	1117	2540534	2541650	571	35	5584	367	3934	365
cysU	934	2539701	2540634	528	39	3705	291	2321	267
cysW	976	2538826	2539801	633	45	5667	426	3275	398
dadA	1399	1236694	1238092	6019	296	333	17	263	19
dadX	1171	1238002	1239172	6561	385	397	25	321	27
dtpD	1582	740298	741879	299	13	1745	81	1037	65
elaB	406	2378744	2379149	500	85	2730	493	2115	515
entB	958	626817	627774	82	6	623	48	361	39
entC	1276	624008	625283	274	15	1950	112	1138	90
entE	1711	625193	626903	221	9	1692	73	1000	59
entF	3982	613280	617261	509	9	2168	40	1250	31
envR	763	3410825	3411587	12	1	46	4	39	5
essD	316	576521	576836	106	23	30	7	13	5
eutS	436	2573492	2573927	16	3	6	1	2	0
evgS	3694	2482296	2485989	629	12	3871	77	2953	79
fbaB	1153	2175534	2176686	866	52	3751	239	2599	223
fes	1225	611938	613162	139	8	593	36	405	33
fimA	649	4541038	4541686	1454	154	6834	773	7640	1163
fimC	826	4542227	4543052	61	5	332	29	303	36
fimI	640	4541651	4542290	5574	598	64906	7442	60356	9317
fixX	388	45363	45750	6	1	1	0	2	1
flu	3220	2069463	2072682	539	11	4024	92	1284	39
gadE	628	3656289	3656916	22	2	195	23	150	24
gapA	1096	1860695	1861790	11608	727	64826	4340	47766	4306
gatA	553	2172619	2173171	373	46	4000	531	2637	524
gatB	385	2172304	2172688	213	38	3536	674	2125	648
gatC	1456	2170945	2172400	1018	48	6948	350	5213	366
gatD	1141	2169857	2170997	338	20	2553	164	1997	174
gatZ	1363	2173081	2174443	3041	153	13809	743	10201	779
gcvB	306	2940618	2940923	7893	1772	853	205	519	168

gcvH	490	3047182	3047671	164	23	776	116	687	142
gcvP	2974	3044190	3047163	351	8	1691	42	1322	44
glgS	301	3189761	3190061	170	39	516	126	522	171
glmY	284	2689179	2689462	7244	1752	1800	465	1124	391
glnA	1510	4054648	4056157	8149	371	676	33	513	34
glnH	847	846481	847327	7818	634	925	80	843	98
glnK	439	471722	472160	32	5	5	1	3	1
glnP	760	845683	846442	2590	234	407	39	314	43
glnQ	823	844964	845786	2957	247	487	43	409	51
glnU	175	696088	696262	29662	11642	5462	2290	2673	3050
glyA	1354	2682276	2683629	7010	356	28484	1544	21927	1600
gsiD	1012	871013	872024	281	19	1088	79	775	80
gtrS	1432	2467053	2468484	131	6	3250	167	2519	180
guaB	1567	2630626	2632192	1017	45	5039	236	4102	259
hdeA	433	3654431	3654863	163	26	877	149	742	169
hdeB	427	3653989	3654415	30	5	183	31	115	27
hisC	1171	2090322	2091492	528	31	2250	141	1546	146
hisJ	883	2424028	2424910	1404	109	7836	651	6285	703
hisP	874	2421758	2422631	203	16	710	60	615	75
hokA	253	3718471	3718723	51	14	10	3	0	3
hpf	388	3344095	3344482	1681	298	44	8	11	5
hsdS	1495	4578091	4579585	162	7	802	39	599	40
hyfA	718	2599123	2599840	35	3	9	1	6	1
iaaA	1066	865691	866756	908	59	3251	224	2655	253
ibpA	514	3865032	3865545	3485	466	627	90	409	79
ibpB	529	3864492	3865020	10642	1382	687	95	515	96
ilvA	1645	3953254	3954898	12092	505	1720	77	1316	80
ilvB	1789	3849119	3850907	810	31	6226	255	3207	180
ilvC	1576	3955893	3957468	1398	61	22522	1049	18080	1133
ilvD	1951	3951401	3953351	12365	435	1662	63	1167	59
ilvE	1030	3950407	3951436	13268	885	1805	129	1259	135
ilvG	1745	3948483	3950227	46572	1833	7194	303	4718	277
ilvM	364	3950124	3950487	1105	209	109	22	41	15
ilvN	391	3848825	3849215	99	17	1099	206	563	152
insA	376	20233	20608	128	23	16	3	4	3
insA	376	290295	290670	7	1	1	0	0	1
insC	466	1466808	1467273	51	8	177	28	3	60
insE	400	314415	314814	45	8	14	3	0	1
insH	1117	687220	688336	869	53	110	7	6	8
iraM	424	1210903	1211326	20	3	123	21	80	19
isrC	304	2069239	2069542	12	3	69	17	19	6
katE	2362	1811791	1814152	227	7	1654	51	1048	44
kilR	322	1416032	1416353	12	3	5	1	1	1
lacA	712	360473	361184	4537	438	252	26	179	26
lacY	1354	361150	362503	26662	1353	670	36	381	28
lacZ	3175	362455	365629	246873	5341	974	23	568	18
leuA	1672	81958	83629	19509	801	3585	157	2115	129
lit	994	1197818	1198811	83	6	463	34	330	33
livJ	1204	3596578	3597781	1158	66	4167	254	4385	360
lpxP	1021	2493567	2494587	3693	248	271	19	181	18
lysP	1570	2245085	2246654	2061	90	5732	268	5740	361
mcrB	1480	4575981	4577460	343	16	1361	67	1018	70
mcrC	1147	4574935	4576081	44	3	309	20	223	19
mdtK	1474	1741381	1742854	3607	168	876	44	611	41

metE	2362	4010976	4013337	4206	122	34051	1058	29800	1247
micF	193	2311006	2311198	993	353	6012	2286	3929	2011
mlaC	736	3335278	3336013	6	1	5	0	1	0
mokA	285	3718471	3718755	44	11	8	2	0	2
mprA	631	2808692	2809322	794	86	3514	409	2776	435
nrdD	2239	4458545	4460783	2004	61	494	16	345	15
ompC	1204	2309668	2310871	37728	2152	160440	9779	111684	9167
peaD	412	565499	565910	25	4	9	2	1	0
pheA	1261	2735667	2736927	3891	212	19091	1111	9004	706
poxB	1819	908554	910372	229	9	1109	45	610	33
prfH	601	253602	254202	78	9	27	3	12	2
proV	1303	2802737	2804039	668	35	14039	791	5288	407
proW	1165	2803932	2805096	322	19	6139	387	2349	202
proX	1093	2805054	2806146	288	18	4771	320	1935	175
pspA	769	1366003	1366771	47047	4202	219	21	172	22
pspB	325	1366725	1367049	6985	1476	76	17	42	16
pspC	460	1366949	1367408	4469	667	57	9	45	11
pspD	322	1367317	1367638	1553	331	16	4	10	4
pspE	415	1367613	1368027	7751	1283	566	100	524	125
pspG	343	4260763	4261105	4992	1000	59	13	34	10
purB	1471	1189839	1191309	1978	92	11098	554	7884	545
purC	814	2594927	2595740	429	36	6811	614	6523	803
purD	1390	4202665	4204054	422	21	4490	237	3668	264
purE	610	551814	552423	192	22	2859	344	2259	378
purF	1618	2426743	2428360	1030	44	9549	433	7354	450
purH	1690	4203966	4205655	439	18	4518	196	3557	210
purK	1168	550750	551917	208	12	3121	196	2481	213
purL	3988	2689678	2693665	1230	21	14756	272	12148	301
purM	1138	2619119	2620256	183	11	3992	257	3552	318
purN	739	2620156	2620894	505	47	2476	246	1760	245
purR	1126	1735768	1736893	1785	109	6975	455	4864	427
purT	1279	1928805	1930083	364	20	3398	195	3073	237
pyrB	1036	4469483	4470518	109	7	627	44	649	67
pyrC	1147	1120784	1121930	528	32	6258	400	4571	394
pyrD	1111	1003891	1005001	365	23	1997	132	1459	130
pyrE	742	3813150	3813891	147	14	865	86	822	109
pyrI	562	4469009	4469570	136	17	606	79	508	97
pyrL	235	4470422	4470656	86	25	528	165	264	126
rbsC	1066	3933211	3934276	257	17	1052	72	777	77
rfaL	1360	3794871	3796230	341	17	1707	92	1224	90
rpoN	1534	3342639	3344172	761	34	617	30	84	7
rrfF	220	3421445	3421664	1000	312	162	54	70	66
serC	1189	956776	957964	8485	490	34391	2123	23777	1976
slp	667	3651884	3652550	71	7	302	33	201	30
sodA	721	4098733	4099453	2166	206	11258	1146	7695	1054
rrfF	220	3421445	3421664	1000	312	162	54	70	66
serC	1189	956776	957964	8485	490	34391	2123	23777	1976
slp	667	3651884	3652550	71	7	302	33	201	30
sodA	721	4098733	4099453	2166	206	11258	1146	7695	1054
sufA	469	1762042	1762510	122	18	555	87	360	77
talA	1051	2576588	2577638	304	20	1287	90	801	76
thrT	176	4173677	4173852	12516	4885	1814	756	926	1480
tktB	2104	2577558	2579661	449	15	2159	75	1316	62
tnaC	175	3886358	3886532	3	1	14	6	8	5

trpB	1294	1315246	1316539	1193	63	5021	285	3686	309
trpD	1696	1317813	1319508	628	25	5677	246	3791	230
trpE	1663	1319408	1321070	398	16	4857	214	3355	203
trpL	145	1321062	1321206	418	198	1727	874	1515	1036
tyrP	1312	1987605	1988916	170	9	744	42	504	38
valX	176	2518973	2519148	2048	799	143	60	75	57
xanP	1492	3826868	3828359	262	12	4478	220	2911	193
yagP	439	296605	297043	275	43	81	14	34	9
yahC	598	333749	334346	12	1	11	1	2	0
ybdZ	319	613065	613383	40	9	265	61	143	48
ybfQ	340	735568	735907	2	0	17	4	2	1
ycaC	727	944154	944880	201	19	1131	114	677	92
yceI	676	1117124	1117799	248	25	1360	148	690	123
ycgF	1312	1213487	1214798	1775	93	465	26	280	21
yciE	607	1312742	1313348	109	12	445	54	267	43
yciF	601	1313294	1313894	18	2	205	25	99	16
yciG	280	1313880	1314159	7	2	79	21	37	13
yciW	1228	1347004	1348231	297	17	3421	204	2202	177
yciX	268	1333215	1333482	67	17	176	48	211	80
ycjM	1780	1368140	1369919	251	10	22	1	15	1
yceI	676	1117124	1117799	248	25	1360	148	690	123
ycgF	1312	1213487	1214798	1775	93	465	26	280	21
yciE	607	1312742	1313348	109	12	445	54	267	43
yciF	601	1313294	1313894	18	2	205	25	99	16
yciG	280	1313880	1314159	7	2	79	21	37	13
yciW	1228	1347004	1348231	297	17	3421	204	2202	177
yciX	268	1333215	1333482	67	17	176	48	211	80
ycjM	1780	1368140	1369919	251	10	22	1	15	1
ydaE	271	1415862	1416132	5	1	1	0	0	1
ydcO	1276	1502929	1504204	128	7	541	31	302	24
ydfK	334	1630996	1631329	1013	208	24	5	7	14
ydiI	511	1763246	1763756	264	35	1322	190	960	198
ydiJ	3157	1763653	1766809	1238	27	7340	171	5671	179
ydiK	1213	1766998	1768210	3873	219	1616	98	1174	96
ydjN	1492	1808858	1810349	4390	202	42211	2076	26968	1786
ydjO	904	1810353	1811256	76	6	548	44	388	42
yeaD	985	1861774	1862758	301	21	1245	93	1106	111
yebV	337	1919704	1920040	237	48	1002	218	589	173
yeeD	328	2082250	2082577	527	110	5279	1181	2212	1129
yeeE	1159	2082491	2083649	663	39	8338	528	4300	435
yeeN	817	2054782	2055598	334	28	1888	170	1342	162
yfbN	817	2385732	2386548	10	1	49	4	39	5
yfdP	463	2471526	2471988	3	0	6	1	8	3
yfjW	1804	2771240	2773043	110	4	874	36	704	39
ygeF	590	2988476	2989065	6	1	11	1	2	0
yhdU	280	3410543	3410822	37	9	182	48	131	46
yiaG	391	3717401	3717791	275	48	1088	204	646	163
yjcB	382	4272783	4273164	1683	303	200	38	155	40
yjcD	1450	4276402	4277851	348	16	1587	80	1146	78
yjgL	1915	4473360	4475274	771	28	3948	151	2795	144
yjgZ	430	4499183	4499612	31	5	6	1	3	1
yjhB	1318	4501981	4503298	43	2	199	11	116	9
yjiY	2251	4587152	4589402	2763	84	129	4	119	5
yliE	2449	872102	874550	198	6	2665	80	1916	79

yliL	370	851794	852163	460	85	92	18	54	15
ymcE	331	1050970	1051300	3255	675	145	32	56	42
ymdF	274	1067204	1067477	36	9	178	48	119	43
ymfD	766	1196090	1196855	71	6	342	33	227	30
ymfE	805	1196756	1197560	27	2	167	15	136	17
ymfL	658	1202391	1203048	1578	165	353	39	20	24
ymfR	283	1204672	1204954	1	0	5	1	1	0
ymjD	166	1388892	1389057	6	2	23	10	8	15
ynaE	334	1432015	1432348	1172	241	85	19	2	6
ynfM	1354	1667623	1668976	1027	52	4808	261	3233	238
ynfN	256	1635978	1636233	177	47	14	4	4	2
yqcC	430	2921806	2922235	49	8	198	34	105	30
yqeJ	583	2987226	2987808	9	1	63	8	28	5
yrhC	451	3621810	3622260	14	2	17	3	3	1
ysaC	202	3748836	3749037	20	7	3	1	0	0
ytcA	376	4302151	4302526	3	1	14	3	7	2

Appendix C (σ^{54} ChIP study)

Unpublished confidential data obtained from Dr. Joseph Wade (University of Albany, NY).

IA. Intergenic Sense					Downstream		
ID[a]	Peak Center	FAT[b]	Motif[c]	Motif Strand[d]	b#[e]	Gene[e, f]	Strand[g]
OS01	347850	17	G**TGGC**ACACCCC**TTGC**T	+	b0331	*prpB*	+
OS02	455741	5	T**TG**T**C**ATGAATT**TTGC**A	+	b0437	*clpP*	+
OS03	471768	18	C**TGGC**ACACCGC**TTGC**A	+	b0450	*glnK*	+
OS04	688441	21	T**TGGC**ACATCTA**TTGC**T	-	b0656	*insH3*	-
OS05	815921	4	T**TGGC**AGGTTAA**TTGC**T	-	b0780	*ybhK*	-
OS06	847290	6	C**TGGC**ACGATTT**TT**T**C**A	-	b0811	*glnH*	-
OS07	882891	10	T**TGGC**GAAGAAA**TTGC**A	+	b0842	*mdfA*	+
OS08	892944	2	TAT**GC**ACGTTTA**TTGC**A	+	b0854	*potF*	+
OS09	1014824	4	G**TGGC**GTGAATT**TTGC**G	+	b0953	*rmf*	+
OS10	1073263	6	C**TGGC**ATCCGCT**TTGC**A	-	b1012	*rutA*	-
OS11	1165190	4	T**TGG**TATGACCAA**TGC**A	+	b1109	*ndh*	+
OS12	1271626	2	G**TGG**TCCGTGGA**TTGC**A	+	b1218	*chaC*	+
OS13	1308544	22	AG**GGC**ACGGTTT**TTGC**A	-	b1250	*kch*	-
OS14	1328870	3	C**TGGC**AATAGAT**TTGC**T	+	b1274	*topA*	+
OS15	1366050	21	T**TGGC**ACGCAAA**TTG**TA	+	b1304	*pspA*	+
OS16	1561149	10	A**TGGC**ATGAGATC**TGC**A	-	b1488	*ddpX*	-
OS17	1830088	7	C**TGGC**ACGAACCC**TGC**A	-	b1748	*astC*	-
OS18	1864820	6	A**TGGC**ATGAGAG**TTGC**T	+	b1783	*yeaG*	+
OS19	2060018	24	C**TGGC**AAGCATC**TTGC**A	-	b1988	*nac*	-
OS20	2176637	1	C**TGGC**CCGCCTT**TTGC**G	+	b2098	*yegT*	+
OS21	2321425	5	C**TGGC**ACTCCCC**TTGC**T	+	b2221	*atoD*	+
OS22	2425886	12	A**TGGC**ATAAGACC**TGC**A	-	b2310	*argT*	-
OS23	2599175	16	A**TGGC**ATCCTTTA**TGC**A	+	b2481	*hyfA*	+
OS24	2689367	200	T**TGGC**ACAGTTAC**TGC**A	-	b4441	*glmY*	-
OS25	2830446	11	C**TGGC**ACGCAATC**TGC**A	+	b2710	*norV*	+
OS26	2836164	11	C**TGGC**ATGATTTG**TG**AA	-	b2713	*hydN*	-
OS27	2848494	91	T**TGGC**ACAAAAAA**TGC**T	-	b2725	*hycA*	-
OS28	2848633	201	C**TGGC**ACAATTA**TTGC**T	+	b2726	*hypA*	+
OS29	2998270	11	C**TGGC**GTAAATC**TTGC**C	+	b2866	*xdhA*	+
OS30	3004190	3	C**TGGC**ACACTTA**TTG**TT	+	b2870	*ygeW*	+
OS31	3014019	26	G**TGG**TGCGATTG**TTGC**T	+	b2878	*ygfK*	+
OS32	3029005	3	TAT**GC**CCGTTTA**TTGC**A	-	b2887	*ygfT*	-
OS33	3217457	10	G**TGGC**GCAATCCC**TGC**A	+	b3073	*patA*	+
OS34	3416975	29	C**TGGC**ACTACTT**TTGC**T	+	b3268	*yhdW*	+
OS35	3556144	17	C**TGGC**ACGACGG**TTGC**A	-	b3421	*rtcB*	-
OS36	3598850	47	C**TGGC**ACAGTTG**TTGC**T	-	b3461	*rpoH*	-
OS37	4056144	21	T**TGGC**ACAGATT**T**C**GC**T	-	b3870	*glnA*	-
OS38	4199750	39	T**TGGC**ACGGAAGA**TGC**A	-	b4002	*zraP*	-
OS39	4199910	9	A**TGGC**ATGATTTC**TGC**T	+	b4003	*zraS*	+
OS40	4260832	15	T**TGGC**ATGATTC**TTG**TA	+	b4050	*pspG*	+
OS41	4297443	41	G**TGGC**ATAAAAGA**TGC**A	-	b4079	*fdhF*	-
OS42	4437375	34	C**TGGC**ATCACAC**TTGC**G	-	b4216	*ytfJ*	-

IB. Intergenic Antisense					Downstream		
ID[a]	Peak.Ctr	FAT[b]	Motif[c]	Motif Strand[d]	b#[e]	Gene[e]	Strand[f]
OA01	374140	12	TG**GGC**ATACAAAA**TGC**A	-	b0352	*mhpE*	+
OA02	1067669	2	CC**GGC**ATGAACAA**TGC**G	+	b1006	*rutG*	-
OA03	1814362	2	GC**GGC**GTGAACC**TTGC**A	-	b1732	*katE*	+
OA04	2802707	12	G**TGGC**ATGAATA**TTG**AT	-	b2676	*nrdF*	+
OA05	3144330	61	C**TGGC**ATATATT**TTGC**C	+	b2998	*yghW*	-
OA06	3370638	3	T**TGG**TATGAAAA**TTG**TA	+	b3225	*nanA*	-
OA07	3809831	1	G**TGGC**GTAGTATAC**GC**T	+	b3638	*yicR*	-
OA08	3889556	3	A**TGGC**TGGCTTC**TTG**AA	-	b3709	*tnaB*	+

IIA. Intragenic Sense					Intragenic		
ID[a]	Peak	FAT[b]	Motif[c]	Motif Strand[d]	b#[e]	Gene[e, f]	Strand[g]
IS01	33273	2	C**TGGC**CTTCGAA**TTGC**A	+	b0033	*carB*	+
IS02	512538	7	C**TGGC**ACTGGTT**TTGC**T	+	b0486	*ybaT*	+
IS03	534346	2	GC**GGC**ACAAATGC**TGC**A	+	b0507	*gcl*	+
IS04	551087	1	C**TGGC**ACCGCGTG**TGC**A	-	b0522	*purK*	-
IS05	567582	2	G**TGG**TGCAATAC**TTGC**A	+	b0543	*emrE*	+
IS06	655861	1	T**TGG**TAAAGTTT**TTGC**T	+	b0622	*pagP*	+
IS07	769197	2	CC**GG**TATGGAATA**TGC**T	+	b0732	*mngB*	+
IS08	773450	1	TG**GG**AACGCTTC**TTGC**C	+	b4515	*cydX*	+
IS09	808913	1	C**TGG**AACAAATGG**TGC**A	+	b0775	*bioB*	+
IS10	939112	1	C**TGGC**CTCGACT**TTGC**A	+	b0893	*serS*	+
IS11	1037161	2	TC**GG**TATCAATT**TTGC**T	+	b0978	*appC*	+
IS12	1177855	1	G**TGG**AACAAAAA**TTGC**G	+	b1119	*nagK*	+
IS13	1213765	6	T**TGGC**GCAGGTT**TTGC**T	-	b1163	*bluF*	-
IS14	1247330	3	TCA**GC**ATGAACA**TTGC**A	-	b1198	*dhaM*	-
IS15	1252922	1	T**TGGC**TCAACACA**TGC**A	-	b1202	*ycgV*	-
IS16	1462952	5	T**TGGC**ATGGAAAAA**GC**A	+	b1400	*paaY*	+
IS17	1464737	2	CC**GG**TACGGAAA**TTGC**T	+	b1401	*ydbA_1*	+
IS18	1519002	23	TC**GGC**ATGAATA**TTGC**G	-	b1451	*yncD*	-
IS19	1535850	2	C**TGGC**ACTACCG**TTGC**A	-	b1467	*narY*	-
IS20	1615388	6	T**TGG**TGTGGCTT**TTGC**A	+	b1528	*ydeA*	+
IS21	1662480	2	TA**GG**AATGGCTA**TTGC**A	+	b1590	*ynfH*	+
IS22	1679126	1	A**TGG**ACTGATTAA**TGC**A	+	b1606	*folM*	+
IS23	1838205	3	G**TGGC**GCAGATTA**TGC**T	+	b1757	*ynjE*	+
IS24	1958599	3	TC**GG**TATGCTGA**TTGC**A	+	b1876	*argS*	+
IS25	2079713	7	AC**GG**TGCAAATT**TTGC**A	-	b2010	*dacD*	-
IS26	2101814	2	G**TGG**TACAGAAAA**TGC**G	-	b2032	*wbbK*	-
IS27	2360527	2	A**TGGC**ACTGAATA**TGC**T	-	b2249	*yfaY*	-
IS28	2370326	2	C**TGGC**ATGGAGCC**TGC**A	+	b2257	*arnT*	+
IS29	2484404	11	C**TGGC**ATACATTA**TGC**A	+	b2370	*evgS*	+

IS30	2526524	6	T**TGGC**ATTGTCG**TTGC**A	-	b2411	*ligA*	-
IS31	2846034	1	G**TGGC**GCGTTTG**TTGC**C	-	b2723	*hycC*	-
IS32	2912348	2	C**TGG**AACGCTTT**T**C**GC**A	-	b2785	*rlmD*	-
IS33	2954738	2	C**TGGC**ACGCGATG**TGC**A	-	b2821	*ptrA*	-
IS34	3012660	7	TGA**GC**ACGAACC**TTGC**A	-	b2876	*yqeC*	-
IS35	3074948	13	C**TGGC**GGCAATA**TTGC**A	-	b4465	*yggP*	-
IS36	3169588	4	T**TGG**TGCGAAAT**TTGC**T	+	b3026	*qseC*	+
IS37	3178185	1	GC**GGC**GCGGGAT**TTGC**A	+	b3037	*ygiB*	+
IS38	3206700	2	A**TGGC**ACCAAAC**TTGC**T	+	b3063	*ttdT*	+
IS39	3241894	1	T**TGG**TGCCGAATA**TGC**A	-	b3092	*uxaC*	-
IS40	3330429	6	T**TGGC**ATGATGG**TTGC**C	-	b3184	*yhbE*	-
IS41	3449718	1	C**TGGC**ATGATTCG**TG**AA	-	b3319	*rplD*	-
IS42	3538782	1	G**TGGC**ACTGAACA**TGC**T	+	b3409	*feoB*	+
IS43	3565544	49	AA**GGC**ATGTTTTA**TGC**A	-	b3429	*glgA*	-
IS44	3683845	1	T**TGGC**ACGGCAA**TTG**AT	-	b3530	*bcsC*	-
IS45	3803524	7	A**TGG**TGCGTAAAA**TGC**A	-	b3630	*waaP*	-
IS46	3966932	10	C**TGG**TGCTCTTT**TTGC**T	+	b3785	*wecA*	+
IS47	4079719	3	C**TGGC**GCGAATTC**TGC**A	-	b3892	*fdoI*	-
IS48	4131399	3	T**TGGC**GCGAATA**TTGC**C	+	b3941	*metF*	+
IS49	4342113	8	TG**GG**TATGGCTC**TTGC**T	+	b4120	*melB*	+
IS50	4370368	2	TG**GG**TATCAAAG**TTGC**A	+	b4143	*groL*	+
IS51	4445887	1	CA**GGC**ACTGGAT**TTGC**T	+	b4221	*tamB*	+
IS52	4457537	2	AA**GG**AACTATTC**TTGC**A	+	b4236	*cybC*	+
IS53	4547262	1	C**TGGC**TCATTAA**TTGC**C	+	b4320	*fimH*	+
IS54	4561465	3	AC**GGC**AAAGAAA**TTGC**A	-	b4333	*yjiK*	-
IS55	4606188	5	A**TGGC**AACAAAT**TTGC**A	+	b4373	*holD*	+
IS56	4612117	3	C**TGGC**ATCGTTA**TTGC**T	+	b4378	*yjjV*	+
IS57	4613489	17	C**TGGC**TCTGTTT**TTGC**A	-	b4379	*yjjW*	-
IS58	4627943	2	C**TGG**AACGCTTCC**TGC**A	-	b4391	*ettA*	-

IIB. Intragenic Anti Sense					**Intragenic**		
ID[a]	**Peak**	**FAT**[b]	**Motif**[c]	**Motif Strand**[d]	**b#**[e]	**Gene**[e]	**Strand**[f]
IA01	71210	1	CC**GGC**ACGAAAC**T**C**GC**T	-	b0064	*araC*	+
IA02	220685	1	T**TGGC**GTCGATA**T**C**GC**C	+	b0197	*metQ*	-
IA03	261344	12	AC**GGC**ACAGTTTA**TGC**A	-	b0243	*proA*	+
IA04	453263	1	TC**GGC**ACCATTAA**TGC**T	+	b0434	*yajG*	-
IA05	485009	2	T**TGGC**GCGTTTC**TTGC**G	-	b0464	*acrR*	+
IA06	619311	1	T**TGGC**CCGATAA**TTGC**C	+	b0588	*fepC*	-
IA07	674001	1	GC**GG**TATTGCTC**TTGC**A	+	b0642	*leuS*	-
IA08	702665	8	C**TGGC**CTGCTTTA**TGC**A	+	b0678	*nagB*	-
IA09	797417	4	CCA**GC**ACGGTTT**TTGC**A	+	b0766	*ybhA*	-
IA10	1098490	6	A**TGGC**TTATTATA**TGC**A	-	b1034	*ycdX*	+
IA11	1272586	2	TC**GG**TACAGGTT**TTGC**A	+	b1219	*ychN*	-
IA12	1628490	2	C**TGG**TGGGGATT**TTGC**A	+	b1542	*ydfI*	-
IA13	1693966	2	C**TGGC**ACAGCAA**TTGC**C	+	b1617	*uidA*	-
IA14	1724329	2	C**TGG**TTCAGTGT**TTGC**T	-	b1649	*nemR*	+
IA15	1781364	6	GC**GGC**ACGGAAAC**TGC**A	-	b1701	*fadK*	+

IA16	2060666	1	C**TGG**TCGATAAT**TTGC**A	+	b1990	*erfK*	-
IA17	2210856	1	CG**GGC**GCAGTTTA**TGC**A	+	b2125	*yehT*	-
IA18	2531372	8	C**TGGC**ATTACTG**TTGC**A	-	b2414	*cysK*	+
IA19	2584411	3	AC**GG**TACAATTTA**TGC**A	-	b2469	*narQ*	+
IA20	2730510	1	A**TGG**TGCAGTTC**TTGC**T	+	b2592	*clpB*	-
IA21	2795717	2	G**TGG**AATATAAT**TTGC**T	-	b2668	*ygaP*	+
IA22	2960070	1	C**TGG**AACAGTTT**T**C**GC**T	+	b2822	*recC*	-
IA23	3089038	1	C**TGGC**AAGCGCG**TTGC**A	-	b2945	*endA*	+
IA24	3110785	1	C**TGGC**TGATTAA**TTGC**A	+	b2970	*yghF*	-
IA25	3152925	1	G**TGGC**ATAGGTT**T**C**GC**A	+	b3010	*yqhC*	-
IA26	3440661	1	T**TGGC**GCTGTTTA**TGC**T	+	b3299	*rpmJ*	-
IA27	3851258	2	TC**GGC**ACGAATT**TTG**AC	+	b4616	*istR*	-

IIIA. No Motif			**Closest**		
ID[a]	**Peak**	**FAT**[b]	**b#**[e]	**Gene**[e]	**Strand**[f]
O1	451032	2	b0432	*cyoA*	-
O2	996793	1	b0937	*ssuE*	-
O3	1049913	1	b0989	*cspH*	-
O4	1488778	1	b4493	*gapC*	-
O5	2083593	1	b2013	*yeeE*	-
O6	3316371	1	b3171	*metY*	-
O7	3717904	1	b3556	*cspA*	+
O8	4177649	1	b3985	*rplJ*	+
O9	4423061	1	b4200	*rpsF*	+
O10	4617597	1	b4383	*deoB*	+

a: OA = Outside of a gene in the Antisense orientation, OS = Outside of a gene in the Sense orientation

b: FAT = Fold Above Threshold score (indication of ChIP-seq occupancy)

c: Associated gene from Table 1

d: Genomic orientation of associated gene from Table 1

e: Closest downstream gene

f: Underlined genes indicate conservation with identified s^{54} ChIP-seq peaks from Salmonella enterica LT2

g: Genomic orientation of closest downstream gene