2012

rRNA Promoters as Targets for Transcription Factors: Structural and Functional Studies of PhERI and CarD

Joseph S. Osmundson

rRNA PROMOTERS AS TARGETS FOR TRANSCRIPTION FACTORS: STRUCTURAL AND FUNCTIONAL STUDIES OF PhERI AND CarD

A Thesis Presented to the Faculty of
The Rockefeller University
In Partial Fulfillment of the Requirement for
the degree of Doctor of Philosophy

by
Joseph S. Osmundson
June 2012

rRNA PROMOTERS AS TARGETS FOR TRANSCRIPTION FACTORS: STRUCTURAL AND FUNCTIONAL STUDIES OF PhERI AND CarD

Joseph S. Osmundson, Ph.D.

The Rockefeller University 2012

Transcription, the process of copying information encoded in DNA into RNA, to facilitate the expression of encoded proteins, is a central process in all living organisms. The expression and repression of subsets of genes allows different cell types in an organism to maintain diverse physiological roles and permits individual cells to respond to various environmental stimuli. Transcription in prokaryotic cells is performed by a single macromolecular complex, RNA polymerase. In rapidly growing cells with abundant resources, prokaryotic RNA polymerase is mostly located at ribosomal RNA (rRNA) promoters, actively transcribing the large, structured RNAs required for protein translation. As resources become more scarce, RNA polymerase directly responds to cellular signals that lead to the repression of rRNA transcription. This regulation has long been thought to be driven primarily by small molecule effectors that signal scarcity.

In this thesis, I will report work done on two transcription factors in prokaryotes that regulate RNA polymerase activity at rRNA promoters. The *Staphylococcus aureus* (Sau) Phage G1 protein PhERI (previously ORF67), was previously described as a general RNA polymerase inhibitor. PhERI expression in Sau cells inhibits cell growth, which could have therapeutic potential against this deadly pathogen.

I describe the structure of PhERI bound to Sau σ^{A}_{4}, the region of RNA polymerase to which it binds. While PhERI interacts with RNAP through σ, I show that RNA polymerase activity at most -10/-35 promoters is not affected. Structural, biochemical and genomic approaches demonstrate that PhERI interacts with σ^{A}_{4} near the -35 element of all promoters, but blocks the binding of an additional RNA polymerase subunit, the α-CTD, to UP-element DNA sequences. PhERI therefore only inhibits RNA polymerase activity at promoters requiring UP-element activation, most notably the rRNA promoters. This work not only delineates the mechanism of PhERI but also describes novel -10/-35 promoters in *Staphlococcus aureus*, defines rRNA promoters in this organism for the first time, and shows UP-element activation is required for rRNA transcription.

The mycobacterial protein CarD is known to interact with RNA polymerase, but its impact on transcription directly at promoters has not been described. The structure of the *Thermus Thermophilus* CarD was solved in the lab, allowing a model for the interaction between CarD and RNA polymerase to be built. I show that CarD stimulates RNA polymerase activity at rRNA promoters, but not all promoters, by directly stabilizing the RNA polymerase open complex on promoter DNA. These two proteins both exploit unique parameters of RNA polymerase at rRNA promoters to specifically regulate RNA polymerase activity at these functionally important promoters.

ACKNOWLEDGEMENTS

I would like to start by thanking my PhD mentor Dr Seth Darst, without whom this thesis would not have been possible. The work I present on the phage protein PhERI required the use of novel techniques that are well beyond my expertise, and I am grateful for Seth's support in developing the methods ultimately required to address the fundamental mechanistic questions we were after. I appreciate Seth's patience in teaching me X-ray crystallography, in particular the independence he left for me to make my own mistakes, and learn from them, while providing the support necessary to never fall off the deep end of the pool. Thanks, Seth.

I would like to thank all the members of the Darst lab whose input was critical to the ultimate success of these projects. In particular, Dr Lars Westblade provided the initial intellectual spark for the PhERI project and was helpful in discussing data and hypotheses throughout, even after he left the lab. Dr Elizabeth Campbell provided a sounding board for ideas, from the mundane cloning strategies to the most ambitious crystallization trials. The work on CarD was initiated by Dr Campbell and Dr Devendra Srivastava, who solved the CarD crystal structure. All my colleagues, including Matt Bick, Andrey Feklistov, Albert Weixlbaumer, Natasha Opalka, Brian Bae, and Valerie Lamour, made the Darst lab both an intellectually engaging and a fun work environment.

I would also like to thank all of the summer students with whom I have worked on these projects. Thanks to Leigh Harris, Fatmata Bah, Lizzy Hubin, Mya Thandar, and Ronnie Almonte for your dedication to my crazy projects and ideas.

This thesis benefited from many critical collaborations. Ramesh Wigneshwearaj at Imperial College London was helpful in our initial efforts to purify *Staphylococcus aureus* RNA polymerase and provided the expression vectors and strains required for *in vivo* expression of PhERI.

Ann Hochschild's lab at Harvard Medical School, and in particular a graduate student Cristina Montero-Diaz, performed genetic experiments to identify σ mutants that proved useful for our biochemical studies. Ann and Cristina also were intellectual contributors throughout the project and Ann suggested UP-element implication in the PhERI story years before we had the wisdom to test it. This work is the culmination of intellectual input from both Ann's lab and the Darst lab.

The work on CarD was done in collaboration with Christina Stallings' lab at Washington University and Michael Glickman at Sloan Kettering. The ChIP-Seq data we present in Chapter 7 were produced in her lab. Again, Christina was a crucial collaborator and provided thoughtful insight throughout the project.

The RNA-seq experiments were a collaboration with the genomics core facility here at the Rockefeller University. I appreciate their flexibility and in particular their willingness to try something unconventional. Connie Zhao was critical to our sample preparation efforts and Scott Dewell aided with all the data analysis. This work would never have reached its full potential without this analysis.

I owe gratitude to my committee: Brian Chait, Kirk Deitsch, and Charlie Rice. Your insight was invaluable and your support reassuring. I genuinely looked forward to our meetings every year and will miss your input and encouragement. In particular, I

need to thank Kirk Deitsch for allowing me to spend time in his lab in a futile effort to product transgenic malaria parasites.

I absolutely must thank everyone in the Rockefeller University Dean's Office who helped make my experience here successful. Thanks to Dean Sid Strickland and Assistant Dean Emily Harms, who supported my science and even my extracurricular activities and were always willing to give a kind word when needed. Thanks are due as well to the entire Dean's Office staff, Cris Rosario, Marta Delgado, and Kristen Cullen, who tackle students' problems as though they were their own.

While the work presented here represents my graduate research, my teaching as a grad student was a significant portion of my education. Thanks to Emily Harms who is the driving force behind The Rockefeller University's summer undergraduate program for giving me the opportunity to mentor students in the lab and in the classroom. Special thanks to Katayoun Chamany at the Eugene Lang Liberal College at the New School for Liberal Arts for trusting me to develop and teach a course and supporting me through the process. Katayoun, and all my students at the New School, remain an absolute inspiration.

Finally, thanks to my family and friends who supported me through this process and managed to keep me (mostly) sane. Thanks are due to my parents for their unyielding support, kindness and love, even when I did not deserve it. To my fellow Rockefeller Grad students, particularly my roommates Matthew Meredith and Laura Macro, who put up with my mess. To Hala Iqbal for the qunches and rants that kept me sane on 3RRB. To Claire Atkinson for critical reading of this manuscript and my life, and for endless conversations about science, books, and the nature of happiness (if it exists). To my

extended New York City family, past and present, for helping me make my home here. In particular, I need to thank Whitney Richards-Calathes who lent me her family, which I have refused to give up. Anytime I wanted to quit, she was there with support and an encouraging word. This work is not mine, but belongs to everyone who has lent me a kind word in a frustrating moment.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADP	Adenosine di-phosphate
Bsub	Bacillus subtilis
cDNA	complimentary DNA
ChIP-seq	Chomatin immunoprecipitation coupled with DNA sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CTD	C-terminal domain
DNA	Deoxyribonucleic acid
Eco	Escherichia coli
EM	Electron Microscopy
FRET	Fluorescence resonance energy transfer
gDNA	genomic DNA
HTH	Helix-turn-helix
mRNA	messenger RNA
MRSA	Methicillin resistant Staphlyococcus aureus
Msm	Mycobacterium smegmatis
Mtb	Mycobacterium tuberculosis
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
NTP	Nucleoside triphosphate
OPC	Open promoter complex
ORF	Open reading frame

PCR	Polymerase chain reaction
Pf	Plasmodium falciparum
PEG	Polyethylene gylcol
PhERI	Phage Encoded rRNA Inhibitor
RID	RNA polymerase interacting domain
RNA	Ribonucleic acid
RNA-seq	RNA-sequencing
RNAP	RNA polymerase
Rpo	RNA polymerase open complex
rrn	rRNA promoter
rRNA	Ribosomal RNA
Sau	Staphylococcus aureus
SNP	Single nucleotide polymorphism
Taq	Thermus aquaticus
TRCF	Transcription replair coupling factor
tRNA	Transfer RNA
Tth	Thermus thermophilus
UTR	Untranslated region

Chapter 1:

Introduction

Staphylococcus aureus

Staphylococcus aureus (Sau) is a gram-positive bacterium that causes a wide variety of infections of the skin as well as pneumonia, meningitis, endocarditis, and sepsis (Lowy, 1998). Sau infections can range from mild to extremely serious and if left untreated can be lethal (Klein et al., 2007a). Interestingly, Sau persistently colonizes the nasal cavity of more than 20% of the human population and intermittent colonization occurs in more than half of healthy individuals (Burian et al., 2010). The switch from non-pathogenic to pathogenic growth requires multiple integrated signals and large-scale changes in gene expression (George and Muir, 2007). The molecular determinants of these life-style choices are largely undetermined.

While Sau infections are generally well controlled by antibiotics, resistance to penicillin was reported by the late 1940s (Benner and Morthlan.V, 1967; Tong and Rossmann, 1997). As additional classes of antibiotics were discovered and introduced into patients, Sau continued to gain resistance. Methicillin-resistant Sau (MRSA) arose by the late 1950s, leaving few treatment options for serious Sau infections (Tong et al., 2012). The glycopeptide drug vancomycin, which inhibits cell wall synthesis, has become the antibiotic of last resort for MRSA infections, although recently strains have arisen that are resistant to this drug as well (Howden et al., 2010; Nordmann et al., 2007). Because of the widespread antibiotic resistance, Sau has become an increasingly costly

infection to treat and mortality rates have risen drastically, particularly from hospital-acquired infections (Klein et al., 2007a).

New antibiotics, including linezolid and trigecycline, which target the bacterial ribosome, have been developed to address antibiotic resistant strains of Sau and other bacteria (Tong et al., 2012). Resistance to linezolid has already been described in patients, and can stem from point mutations at the drug-binding site or from horizontal gene transfer of the chloramphenicol resistance plasmid *cfr* (Rossolini et al., 2010). Given the rapid rise of Sau resistance to past therapies, novel antibiotics are likely to be necessary to treat resistant Sau infections in the near future. Additionally, the recently developed antibiotics against Sau inhibit cell growth by previously exploited mechanisms, allowing resistance to rapidly spread by horizontal gene transfer of previously existing gene clusters (Rossolini et al., 2010; Ruiz de Gopegui et al., 2012; Witte and Cuny, 2011). Unexploited targets for antimicrobial therapy could potentially lead to antibiotics that are less susceptible to resistance.

Bacteriophages and drug discovery

Bacteriophages (phages) are viruses that infect bacterial cells (1925; Bronfenbrenner and Korb, 1925). Phage infections can be lysogenic, wherein the phage genome is integrated into the host cell genome and the phage is transmitted vertically through cell division, or lytic, with the phage rapidly hijacking the host cell machinery to favor replication of the phage and production of phage proteins, ultimately leading to lysis of the host bacterium (Echols, 1972; Oppenheim et al., 2005). Upon the initial injection of the phage genome of a double stranded DNA phage into the host cell, the

host transcriptional machinery will transcribe early phage genes (Kadesch et al., 1982; Rosenberg et al., 1982; Siebenlist, 1979). While some phage encode a dedicated RNA polymerase for subsequent transcription of phage genes (Chamberlin et al., 1970), others will use the host machinery throughout the phage life-cycle (Oppenheim et al., 2005).

As early as the 1920s, scientists recognized the therapeutic potential of phage to treat bacterial infections to their ability to specifically and rapidly lyse bacterial cells (Mills, 1956; Smith, 1924). With the discovery of potent antibiotics, using phages as direct therapeutic agents fell out of favor in the West but continued to be used therapeutically in the former Soviet Union (Bacteriophages and their Implications on Future Biotechnology: A Review IN PRESS).

Phages are advantageous therapeutically because they can infect and lyse a specific bacterial species, whereas antibiotics tend to affect bacterial species indiscriminately. However, the potential for phages to provoke an immune response in the patient and the inherent issues with using active biological entities in patients has prevented their widespread use (Kelly et al., 2011; Paul et al., 2011). Phage lytic enzymes are the individual proteins that facilitate the lysis of the host cell and subsequent release of phage particles (Fischetti, 2010; Paul et al., 2011). Lytic enzymes alone have the ability to lyse bacteria in a species-specific fashion and several phage lytic enzymes are currently being developed for use in topical creams or for direct use in patients, particularly to clear colonizing Sau bacteria prior to surgery (Fischetti, 2008, 2010; Grandgirard et al., 2008; Koller et al., 2008; Pastagia et al., 2011). Because lytic enzymes are full-length proteins, which are inherently difficult to deliver to infected cells and tissues, their therapeutic use may be somewhat limited. Despite the difficulties, the

increasingly desperate need for novel anti-microbial agents has driven the recent reemergence of direct phage therapies (Kelly et al., 2011).

While phages have been studied for their therapeutic potential, they have also been used since the early days of modern molecular biology as simple models to study systems such as transcriptional regulation (Gribskov and Burgess, 1986; Kassavetis and Geiduschek, 1984; Minakhin and Severinov, 2005; Oppenheim et al., 2005; Orsini et al., 1993; Ptashne, 1992) and DNA replication (Allison et al., 1977; Black and Peng, 2006; Doublie et al., 1998; Ray et al., 1975; Yano and Rothman-Denes, 2011). Because of its small genome and binary life-style choices between lytic and lysogenic, the phage lambda (λ) provides a model for how gene regulation, mediated by direct interactions between RNA polymerase and protein factors, is related to a developmental switch (Oppenheim et al., 2005). Studies on the phage M13 have revealed the molecular mechanisms of DNA replication in *Escherichia coli* (Eco) cells (Allison et al., 1977; Ray et al., 1975). The M13 phage genome is still used in the majority of biochemical assays studying the *in vitro* activity of prokaryotic DNA polymerase (Yeeles and Marians, 2011). Work using phages as model systems provided the basis for our understanding of some of the most central tenants of modern biology, such as the central dogma, and provided hypotheses that were subsequently tested and validated in more complex model systems.

Phage are incredibly abundant and diverse, and the mechanisms through which they inhibit host cell growth may reveal currently unexploited drug targets. Phage are among the most abundant life forms on planet earth, estimated to number at 10^{31} total particles (Bergh et al., 1989; Whitman et al., 1998). Phage play critical roles in the

ecology, evolution (Bohannan and Lenski, 2000), and pathogenicity (Wagner and Waldor, 2002) of prokaryotic cells. The large number of phage particles also represents a immense and largely untapped source of biological diversity: phages that have been sequenced contain large numbers of putative open reading frames (ORFs) with no known homolog (Liu et al., 2004). Understanding how phage, in their vast diversity, arrest host cell growth is likely to reveal novel mechanisms of inhibition that could provide targets for the design of new small molecule inhibitors.

Recent work has mined phage genomes for proteins or peptides that inhibit cell growth in Sau. Kwan *et al.* (2005) undertook a massive sequencing project of Sau phages and created a library of more than 900 putative phage ORFs in an inducible vector that was transformed into Sau (Kwan et al., 2005; Liu et al., 2004). ORFs that inhibited cell growth were considered inhibitors of essential cell functions (Fig 1.1a). The work highlighted one phage protein, phage 77 ORF104, which inhibits DNA replication through an interaction with dnaI (Fig1.1b). Small molecule inhibitors of dnaI were shown to have the same effect on cell growth as ORF104, illustrating the potential to use phage genomics to identify essential cellular functions and aid in the design of small molecules to target these functions. The G1 phage protein PhERI (Phage Encoded rRNA Inhibitor; previously ORF67) was also identified in this screen. While ORF104 targeted DNA replication (Fig 1.1b), PhERI was described as a putative RNA polymerase (RNAP) inhibitor because its expression decreased the levels of RNA in a pulse-chase experiment (Fig 1.1c) (Liu et al., 2004).

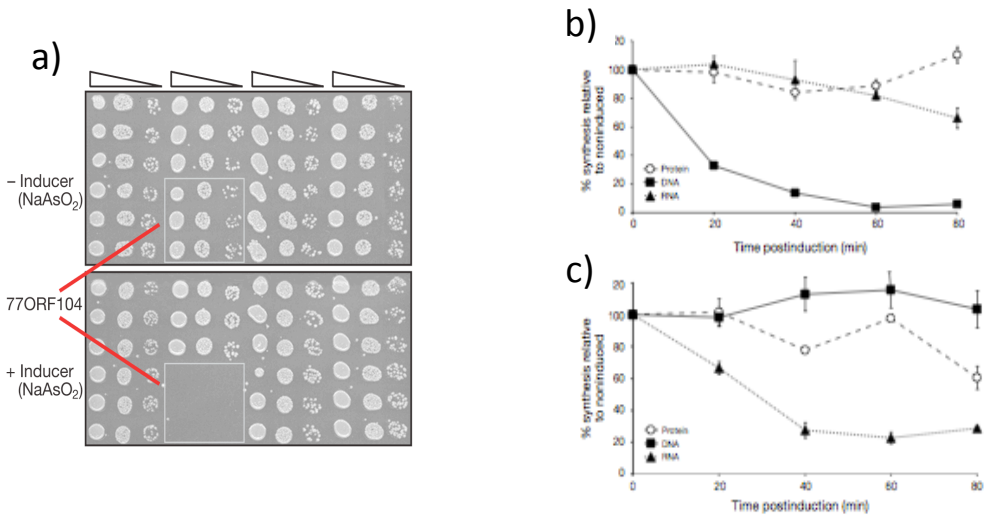


Figure 1.1: PhERI is a putative RNA polymerase inhibitor. a) Screen for ORFs that inhibit cell growth in Sau. ORFs were cloned into an inducible vector and cells were grown in the absence (upper panel) or the presence (lower panel) of inducer in Sau RN4220. ORF104, indicated with red arrows, shows strong inhibition of growth in Sau. b) ORF104 inhibits DNA synthesis. Cells were pulse labeled with radiolabeled DNA, RNA, and amino acids at different time points after the induction of ORF104. ORF104 inhibits DNA synthesis. c) PhERI inhibits RNA synthesis. Cells were pulse labeled with radiolabeled DNA, RNA, and amino acids at different time points after the induction of PhERI. PhERI inhibits synthesis of RNA. Adapted from (Liu et al., 2004).

Structure and function of prokaryotic RNAP

Transcription in prokaryotes is performed by the ~400 kDa core RNAP enzyme (subunit composition $\alpha_2\beta\beta'\omega$; (Darst, 2001)). Promoter recognition and initiation require an additional promoter-specificity subunit, σ, which binds to the core RNAP to form the holoenzyme (Murakami and Darst, 2003). Holoenzyme can interact with promoter DNA in a sequence specific manner through the σ factor. The initial interactions are with double stranded DNA to form a closed complex. The σ factor then facilitates melting of the transcription bubble around the -10 element (to form an open complex) and initiation of transcription.

Structural studies of RNAPs have been critical in our understanding of these complex molecular machines. Crystal structures of thermophillic RNAP enzymes (Murakami and Darst, 2003; Vassylyev et al., 2002; Zhang et al., 1999) and electron microscopy (EM) studies of the Eco RNAP (Darst et al., 1989; Finn et al., 2000; Polyakov et al., 1995) revealed the overall architecture of the enzyme (Fig 1.2). The active site, and its catalytic magnesium ion, is buried in a cleft at the interface between the β and β' subunits deep within the RNAP enzyme. The β and β' subunits form a clamp around the active site (Fig 1.2a). Opposite the cleft is the RNA-exit channel, through which the nascently transcribed RNA will be expelled (Fig 1.2b), and the secondary channel, through which substrate NTPs can enter the active site (Fig 1.2a). Structures of RNAP crystallized with elongation-complex substrates (Gnatt et al., 2001; Kettenberger et al., 2004; Mustaev and Korzheva, 2001) illuminated the molecular architecture of RNAP enzymes as it is actively transcribing genes (Fig 1.2c). DNA binds to RNAP through the open space provided by the clamp; the template strand then enters

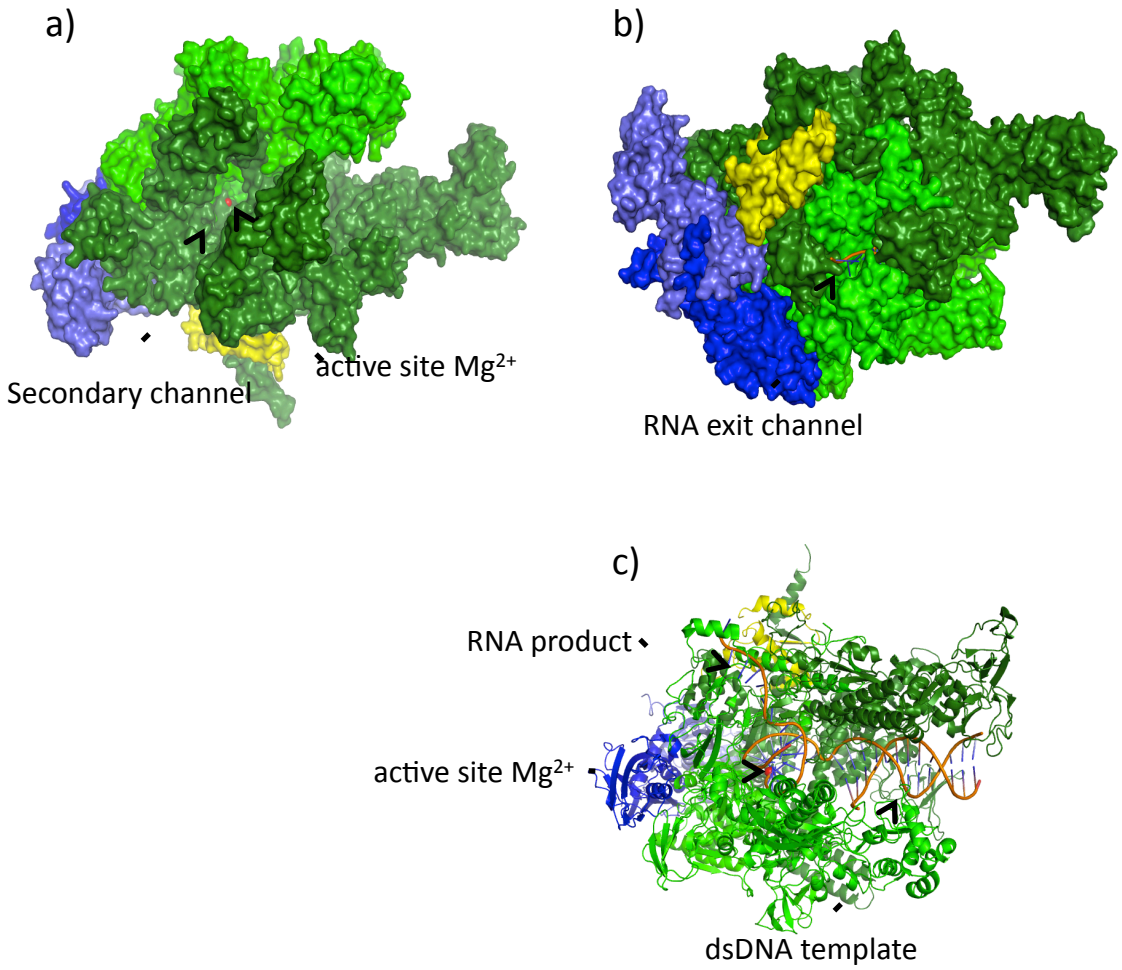


Figure 1.2: Xray crystal structures of prokaryotic core RNAP. a) Surface map of core RNAP. α is shown in blue, β is light green and β' dark green, and ω in yellow. This view is looking through the RNAP secondary channel at the active site magnesium (red sphere). b) Surface map of core RNAP, colored as above, showing the RNA-exit channel. c) Structure of core RNAP, colored as above, bound to transcription elongation complex (TEC) substrates. DNA (orange) enters the active site through the cleft in the RNAP clamp. RNA exits RNAP through the RNA-exit channel.

the active site through a series of interactions with positively charged RNAP residues (Fig 1.2c and d). Structural studies of RNAP at different states of the transcription cycle highlight the conformational changes required as RNAP binds to and melts promoter DNA, begins to initiate RNA synthesis, and clears the promoter to enter the stable elongation stage of transcription.

Structural studies of non-thermophilic RNAP

To date, all high-resolution structural data on prokaryotic RNAPs come from structural studies of thermophilic bacteria (Campbell et al., 2001; Darst, 2001; Murakami et al., 2002b; Vassylyev et al., 2002; Vassylyev et al., 2007). Eco RNAP has been extensively studied using biochemical and genetic techniques, but has thus far resisted high-resolution structural studies despite extensive effort (Twist et al., 2011a). Functional differences between Eco RNAP and its thermophilic homologs have led to the misinterpretation of high-resolution structures (Vrentas et al., 2008). Additionally, many protein factors that regulate RNAP function in Eco have no homologs in thermophilic bacteria, and therefore cannot be studied by crystallography without issues arising from cross-species artifacts. Eco RNAP is also modulated by many phage proteins, which have been extensively characterized by biochemical techniques (Geiduschek, 1991; Snyder et al., 1976; Stevens, 1977; Twist et al., 2011b), whereas RNAP from the thermophilic bacteria have fewer known phage-encoded regulators.

To study the structure of RNAP from non-thermophilic organisms such as Sau and Eco, hybrid methods have been necessary. Cryo-EM studies have led to density maps of RNAP at sub-atomic resolution (Darst et al., 1989; Finn et al., 2000; Polyakov et

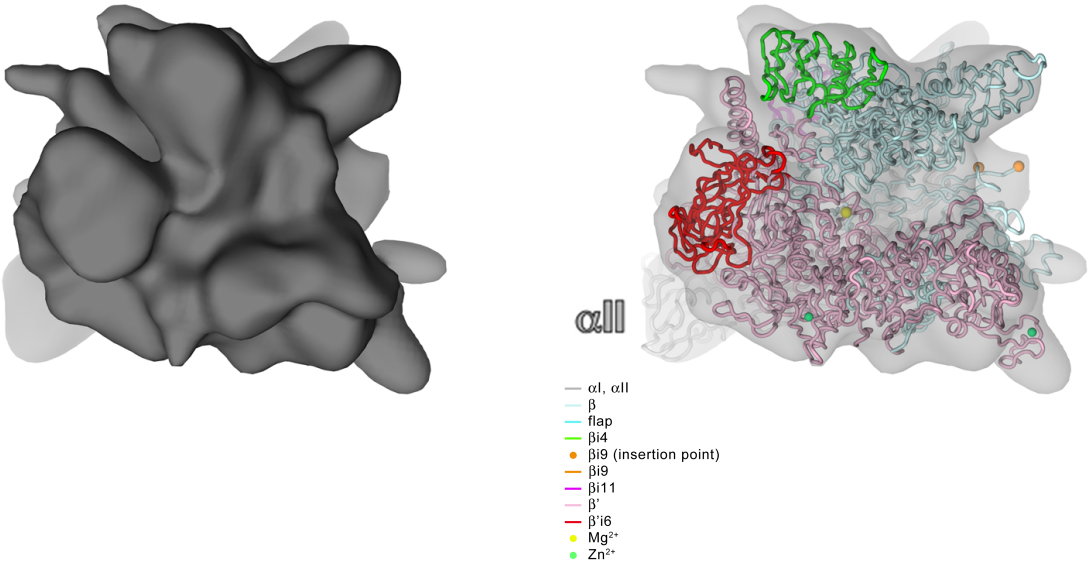


Figure 1.3: Single particle cryo-EM density map (at 14Å resolution, left panel,) and the fit of the Eco RNAP homology model into the EM density (EM density map is shown as the gray envelope, and the Eco atomic model is shown as a ribbon and colored according to the key shown in the inset, right panel). Adapted from (Opalka et al., 2010).

al., 1995). High-resolution structures and homology models from X-ray crystallography can be fit into the lower resolution EM data to provide molecular models. Recent work illustrates how this method can be used to build a model of Eco RNAP to molecular resolution (Opalka et al., 2010) (Fig 1.3). However, these hybrid methods do not allow for the study of small-molecule binding and do not easily elucidate small-scale conformational changes between different states in the transcription cycle. Structural studies on RNAP from other organisms may provide well-diffracting crystals that explain, in molecular detail, the functional differences and the discrepancies in small-molecule and protein factor binding to RNAP from thermophilic and non-thermophilic bacteria.

Promoter recognition by RNAP

The group 1, or primary, σ factors (σ^{70} in Eco, σ^{A} in Sau) are responsible for the bulk of transcription during log-phase growth and are essential for viability (Gruber and Gross, 2003). Prokaryotic housekeeping promoters generally consist of two DNA elements that can be recognized by σ only in the context of the RNAP holoenzyme. Sequences at position -10 (consensus: TATAAT) and -35 (consensus: TTGACA) relative to the transcription start site (+1) are directly recognized by σ domain 2 (σ_2) and domain 4 (σ_4) respectively (Fig 1.4). The spacer sequence between these two elements does not have a conserved sequence; however the length of the spacer (optimal spacing: 17bp) is critical to promoter binding and subsequent transcription (Fig 1.4) (Harley and Reynolds, 1987).

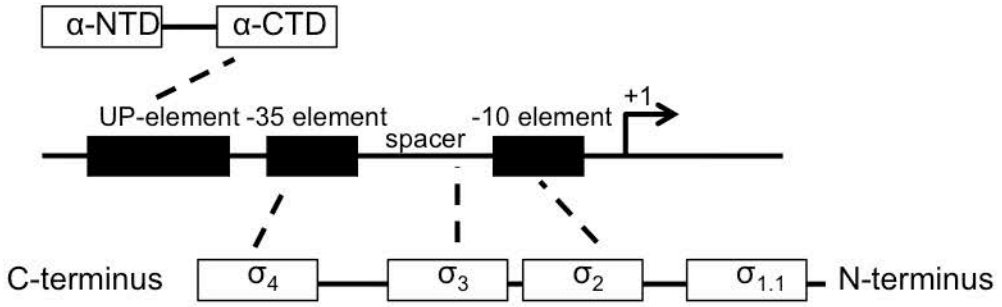
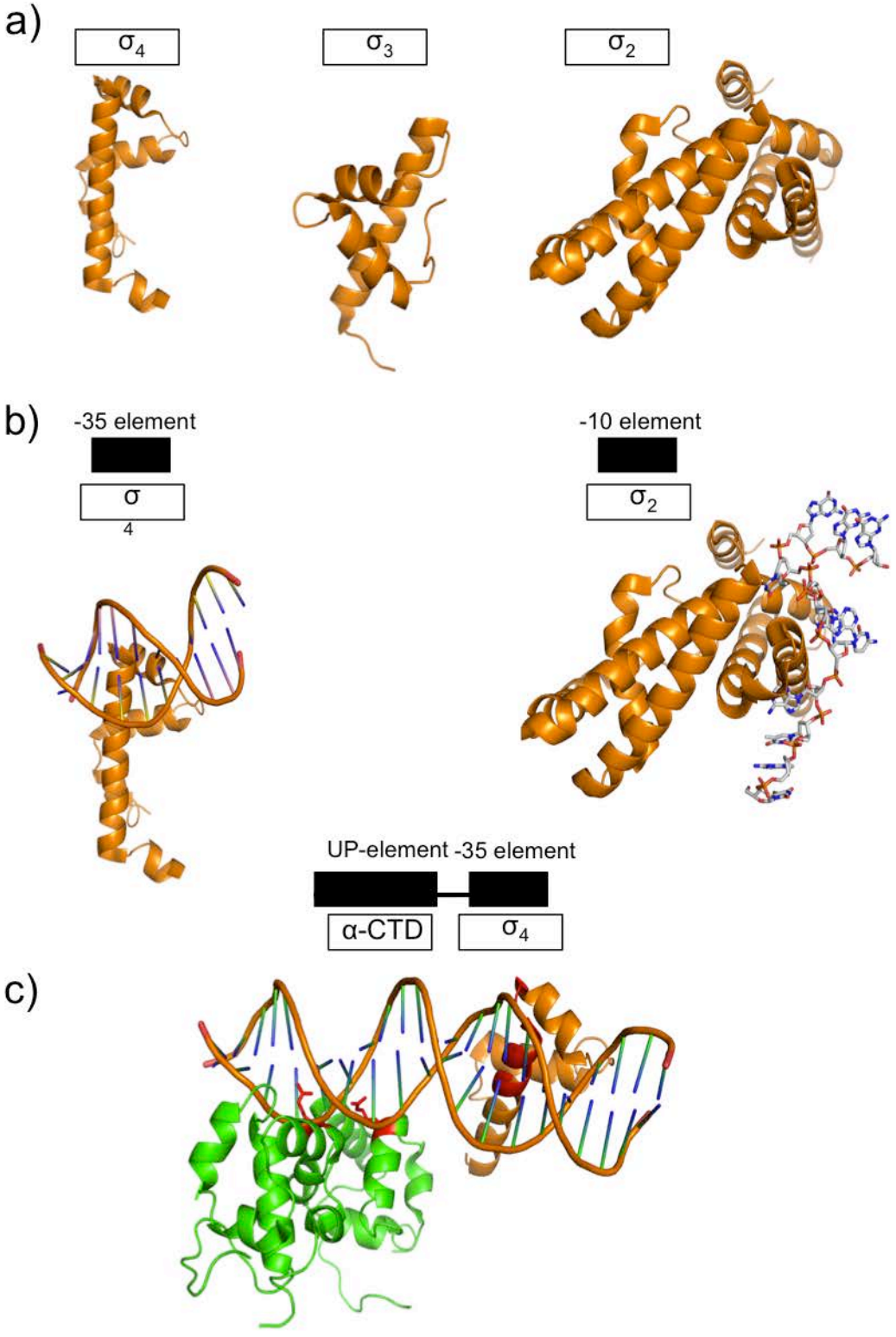


Figure 1.4: Schematic of group 1 σ factors and RNAP binding to promoters. Group 1 σ factors are multi-domain proteins connected by flexible linkers (bottom panel). The -10 (with respect to the start site, +1) and -35 DNA elements are recognized by σ region 2 (σ_2) and 4 (σ_4) respectively. An A/T rich UP-element upstream of the -35 element is recognized through an interaction with the RNAP core α-CTD domain. Extended -10 (TGn) elements and DNA in the spacer region are recognized by σ_2 and σ region 3 (σ_3).

σ^A is comprised of multiple domains connected by flexible linkers (Fig 1.4) (Paget and Helmann, 2003). The structures of independent domains have been solved (Fig 1.5a) (Campbell et al., 2002), and σ bound to RNAP has been crystallized (Fig 1.6c) (Murakami et al., 2002b; Vassylyev et al., 2002). Structures of σ_2 (Feklistov and Darst, 2011) and σ_4 (Campbell et al., 2002) have been solved in complex with DNA corresponding to the -10 and -35 promoter elements respectively (Fig 1.5b). σ_4 interacts with the double stranded -35 element through a conserved helix-turn-helix (HTH) DNA-binding motif that sits in the major groove of promoter DNA (Fig 1.5b) Full-length σ in solution is not able to bind to promoter DNA as the DNA binding determinants are not accessible (Callaci et al., 1999; Camarero et al., 2001). σ makes obligatory contacts with RNAP, including an interaction between σ_4 and a region of the β-subunit, the β-flap (Fig 1.6a and b) (Murakami et al., 2002b; Vassylyev et al., 2002). Only when σ is organized in the context of the RNAP holoenzyme (Fig 1.6c and d) are its DNA-binding domains are exposed and appropriately positioned to interact simultaneously with the -10 and -35 elements to facilitate DNA binding and subsequent melting and initiation of transcription (Fig 1.6) (Campbell et al., 2002; Murakami et al., 2002a).

Housekeeping transcription is performed by the σ^{70} class of σ factors, including the Sau σ^A, which recognize -10/-35 promoters as described above (Gruber and Gross, 2003). Alternative σ factors can bind to RNAP and direct transcription to promoters containing different sequence elements as a response to environmental stimuli (Ades, 2007; Feklistov and Darst, 2009; Testerman et al., 2002). Furthermore, subsets of -10/-35 promoters have additional sequence elements that facilitate for RNAP binding and transcription initiation. Extended -10 promoters, with the sequence TGn immediately

Figure 1.5 (adjacent page): Structural analysis of σ domains and σ / DNA interactions. a) Crystal structure of σ_4 (left panel), σ_3 (central panel) and σ_2 (right panel). b) Structures of σ_2 and σ_4 interacting with the -10 and -35 DNA elements. σ_4 (left panel) interacts with double stranded DNA (consensus: TTGACA) though its major groove. σ_2 (right panel) interacts with bases of the -10 element (consensus: TATAAT) as they unwind. c) Crystal structure of σ_4 interacting with the -35 element and the α-CTD bound to an upstream UP-element.


a)
σ4
σ3
σ2
b)
-35 element
σ4
-10 element
σ2
UP-element -35 element
α-CTD
σ4
c)

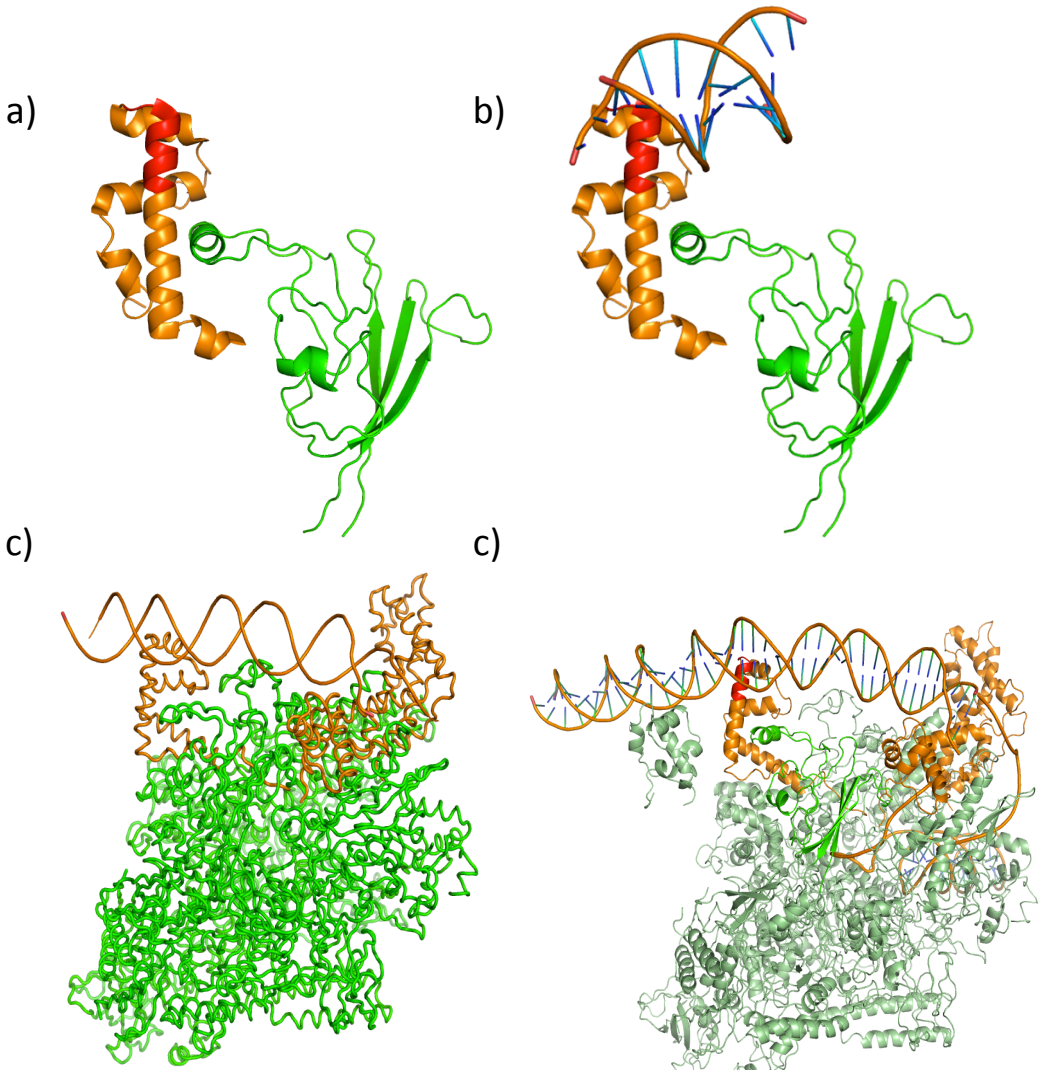


Figure 1.6: Structure of RNAP holoenzyme and Rpo. a) Structure of σ_4 (orange, the Helix-turn-helix responsible for DNA recognition is shown in red) interacting with the RNAP β-flap tip helix (green). b) Structure of σ_4 and the β-flap, as above, showing the interaction between σ_4 and the -35 element of promoter DNA. c) Structure of the RNAP holoenzyme bound to upstream fork DNA, mimicking the interactions of the closed promoter complex. RNAP is colored green and σ is colored orange. d) Model of the RNAP open promoter complex (Rpo). RNAP is shown in green and the β-flap is highlighted in bright green. σ is shown in orange and the promoter DNA is orange.

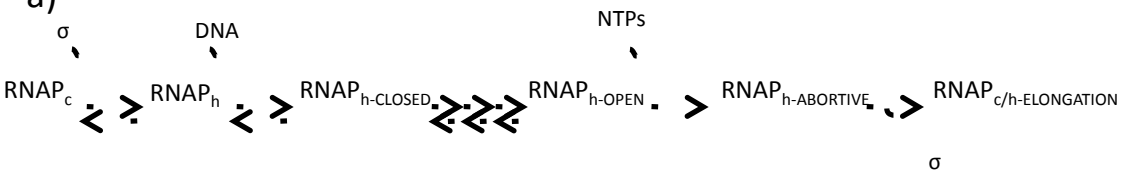
upstream of the -10 element, do not require the interaction between the -35 element and σ_4 (Bown et al., 1997). These promoters are recognized through an interaction between the extended -10 and σ regions 2 and 3 (Barne et al., 1997; Sanderson et al., 2003). The region between nucleotides -6 and the -3, known as the discriminator, is also an important determinant of the stability of promoter DNA bound to RNAP, and therefore of the strength of promoter binding, promoter stability and subsequent transcription at certain promoters (Feklistov et al., 2006; Haugen et al., 2008; Travers, 1984).

Kinetics of RNAP initiation at prokaryotic promoters

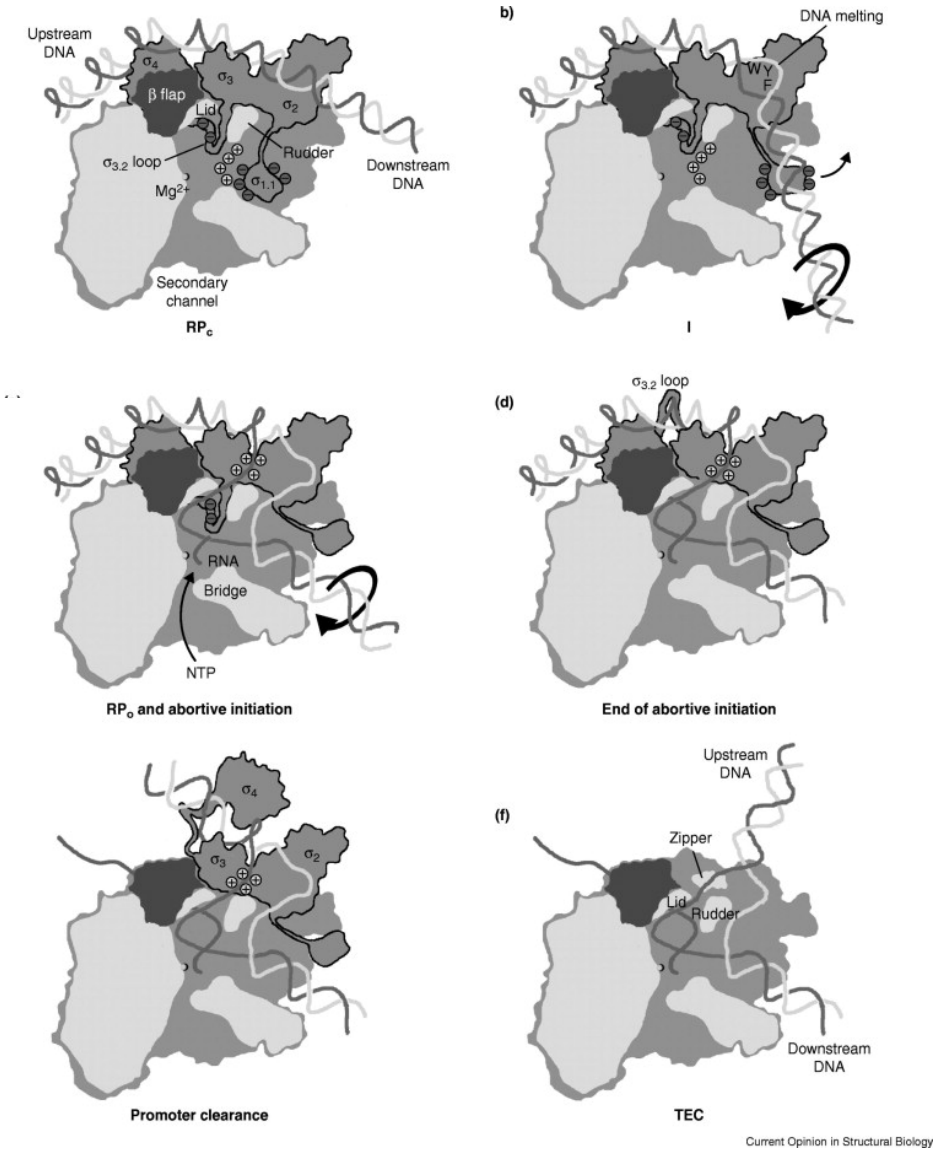
While structural biology has been critical to our detailed understanding of the interactions between RNAP and promoter DNA, the process of transcription initiation proceeds through many unstable, transient intermediates that are difficult to capture by structural studies. RNAP holoenzyme must first interact with DNA in its double stranded state though interactions well upstream of the promoter start site, including the UP-element (see below) and the -35 element (Saecker et al., 2011). In order to initiate transcription, RNAP must melt the promoter DNA so that the single-stranded template strand can enter the enzyme's active site. Kinetic and footprinting studies demonstrate that the transition between the closed RNAP promoter complex and the stable open promoter complex (Rpo) in which promoter DNA is melted and in the enzyme active site, proceeds via at least two short lived intermediates, and involves conformational changes in both RNAP and promoter DNA (Chen et al., 2010; Saecker et al., 2011). The kinetic steps involved in transcription initiation are outlined in Fig 1.7a.

Figure 1.7 (adjacent page): Kinetics and structural changes required for promoter recognition, DNA melting, transcription initiation and promoter clearance. a) Kinetic scheme of transcription initiation. Core RNAP (RNAPc) binds to a σ factor to form holoenzyme (RNAPh) which can then interact specifically with double stranded DNA elements such as the -35 element and UP-element to form the unstable closed promoter complex (RNAPh-closed). Several isomerization steps are required to form the open promoter complex (RNAPh-open) in which promoter DNA is melted and the template strand is in the RNAP active site. NTPs can then enter the active site through the secondary channel and produce abortive RNA products several nucleotides in length. Once an RNA product is made of sufficient length to push σ region 3.2 out of the RNA-exit channel and break the interactions between σ_2 and σ_4 with promoter DNA, RNAP will clear the promoter and enter the stable, processive, elongation phase. b) Overview of structural changes required for transcription initiation, adapted from (Murakami and Darst, 2003).

a)


σ
DNA
NTPs
RNAP$_c$
RNAP$_h$
RNAP$_{h\text{-CLOSED}}$
RNAP$_{h\text{-OPEN}}$
RNAP$_{h\text{-ABORTIVE}}$
RNAP$_{c/h\text{-ELONGATION}}$
σ

b)


Upstream DNA
σ4
σ3
β flap
σ2
Lid
σ3.2 loop
Mg2+
Rudder
σ1.1
Downstream DNA
Secondary channel
RPc
b)
DNA melting
WY F
I
RNA
Bridge
NTP
RPo and abortive initiation
σ3.2 loop
(d)
End of abortive initiation
σ4
σ3
σ2
Promoter clearance
Upstream DNA
(f)
Zipper
Lid
Rudder
Downstream DNA
TEC
Current Opinion in Structural Biology

Once Rpo is formed, RNA production can begin at the active center by the diffusion of nucleotide tri-phosphates (NTPs) through the secondary channel. However, the stable interactions between σ and DNA, and the fact that σ region 3.2 ($\sigma_{3.2}$) is blocking the RNA path to the RNA-exit channel, lead to the production of short transcripts only a few bases long (Munson and Reznikoff, 1981; Sen et al., 1998). RNAP remains bound to promoter DNA and produces abortive transcripts until an RNA product is made that is long enough to push $\sigma_{3.2}$ out of the exit channel and break RNAP contacts with σ and promoter DNA (Fig 1.7b) (Murakami and Darst, 2003). The force required to escape from abortive initiation is provided by "scrunching," as RNAP pulls downstream DNA into the active site, but remains bound to promoter DNA, producing a stressed intermediate (Revyakin et al., 2006). Once RNAP has escaped the promoter, it forms a transcription elongation complex (TEC), which is capable of transcribing to the end of the gene in a highly processive manner (Mustaev and Korzheva, 2001; Nudler, 1999; Nudler et al., 1996). The RNAP structural modulations required for DNA binding, melting, initiation and escape are outlined in Fig 1.7b.

Structural biology of promoter melting and Rpo formation

Structures of the RNAP holoenzyme (Fig 1.6c) and σ bound to promoter DNA fragments (Fig 1.6b) provide a basis for molecular models of RNAP promoter recognition, DNA melting and transcription initiation (Darst, 2001; Murakami et al., 2002a). Initial interactions between promoter DNA and RNAP holoenzyme form the RNAP closed complex, in which σ_4 is bound to the -35 promoter element but the DNA is double stranded (Saecker et al., 2011). A recently solved structure of σ_2 bound to single

stranded -10 element promoter DNA fragments (Fig 1.8) informs the mechanism of DNA melting by the RNAP holoenzyme. σ_2 contains pockets that specifically recognize the highly conserved A at position -11 (Fig 1.8a) and the T at -7 (Fig 1.8b) as they flip out of the DNA double helix. This structure not only demonstrates the molecular mechanism for sequence specific recognition at the -10 element but shows unambiguously that RNAP holoezyme can only recognize single-stranded DNA bases as they flip out of the double stranded helix, effectively coupling DNA melting and sequence-specific DNA recognition (Feklistov and Darst, 2011). DNA melting ultimately forms a roughly 11-base single stranded transcription bubble (Siebenlist, 1979) that places promoter DNA inside a protein enclosed, positively-charged channel leading to the RNAP active center.

Ribosomal RNA promoters exhibit unique kinetics that are exploited for tight regulation

During exponential growth, prokaryotic cells expend significant energetic resources producing the enzymatic machinery required for robust translation. The majority of transcription in growing cells is at ribosomal RNA (rRNA) promoters (rrn promoters) (Gourse et al., 1996). However, as cells approach saturation and resources become scarce, rRNA production is dramatically reduced and the transcriptional profile is rapidly altered. A series of direct and indirect signals are responsible for the regulation of transcription at rrn promoters. These combined signals affect the switch between rapidly dividing cells with ample resources and cells at stationary phase with limited resources, termed the stringent response (Chatterji and Ojha, 2001; Traxler et al., 2008).

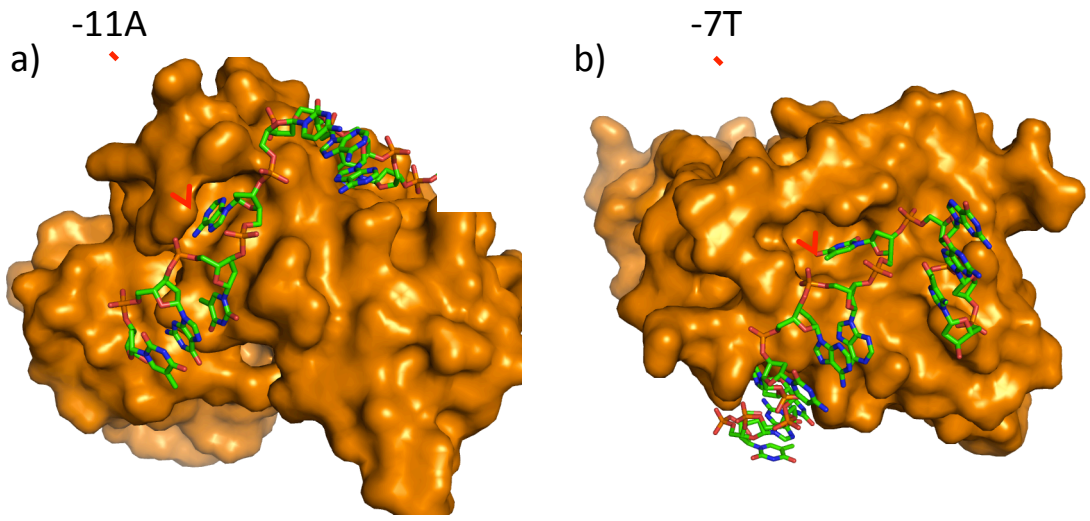


Figure 1.8: Structure of σ_2 bound to single stranded oligos with the -10 element sequence. a) The conserved -11 A base (tAtaat) is tightly recognized by a protein pocket in σ. b) The conserved -7T base (tataaT) is tightly recognized by a protein pocket in σ. The less conserved bases (TaTAAt) remain stacked in a helical-like conformation and are recognized through the phosphate backbone.

The stringent response is well studied in Eco. As cells reach saturating levels, the intracellular concentration of NTPs and amino acids, the substrates for transcription and translation respectively, decrease. Amino acid starvation activates a ribosomal associated protein, relA (Friesen et al., 1976), which converts GTP into the stress-signal ppGpp, also known as magic spot (Paul et al., 2004; Traxler et al., 2008). ppGpp has direct effects on RNAP activity, although the nature of its interaction with RNAP is still debated (Jishage et al., 2002; Laurie et al., 2003; Perederina et al., 2004; Vrentas et al., 2008).

rrn promoters are targeted by ppGpp activity through a decrease in the stability of open promoter complexes (Barker et al., 2001a; Barker et al., 2001b; Gourse et al., 1998). Most -10/-35 promoters tested are not rate-limited by the stability of the OPC; once OPCs form, they are generally a stable intermediate (Saecker et al., 2011). rrn promoters, however, have characteristically unstable OPCs due to G/C-rich sequences in the discriminator elements (Pemberton et al., 2000), and can therefore be effectively regulated by stabilization or destabilization of this intermediate in the transcription cycle (Barker et al., 2001b).

In Eco, the protein DksA potentiates the effect of ppGpp on rrn promoters (Paul et al., 2004; Perederina et al., 2004). DksA binds directly to RNAP and decreases activity at rrn promoters *in vitro* and DksA mutants lose the ability to regulate rRNA expression as the cells approach stationary phase (Paul et al., 2004). The structure of DksA reveals that the protein is an extended coiled-coil with a small globular domain (Fig 1.9a). Structural homology to the cleavage factor GreA (Fig 1.9b) suggests that DksA binds directly to the RNAP secondary channel and the coiled-coil may place DksA residues near the RNAP

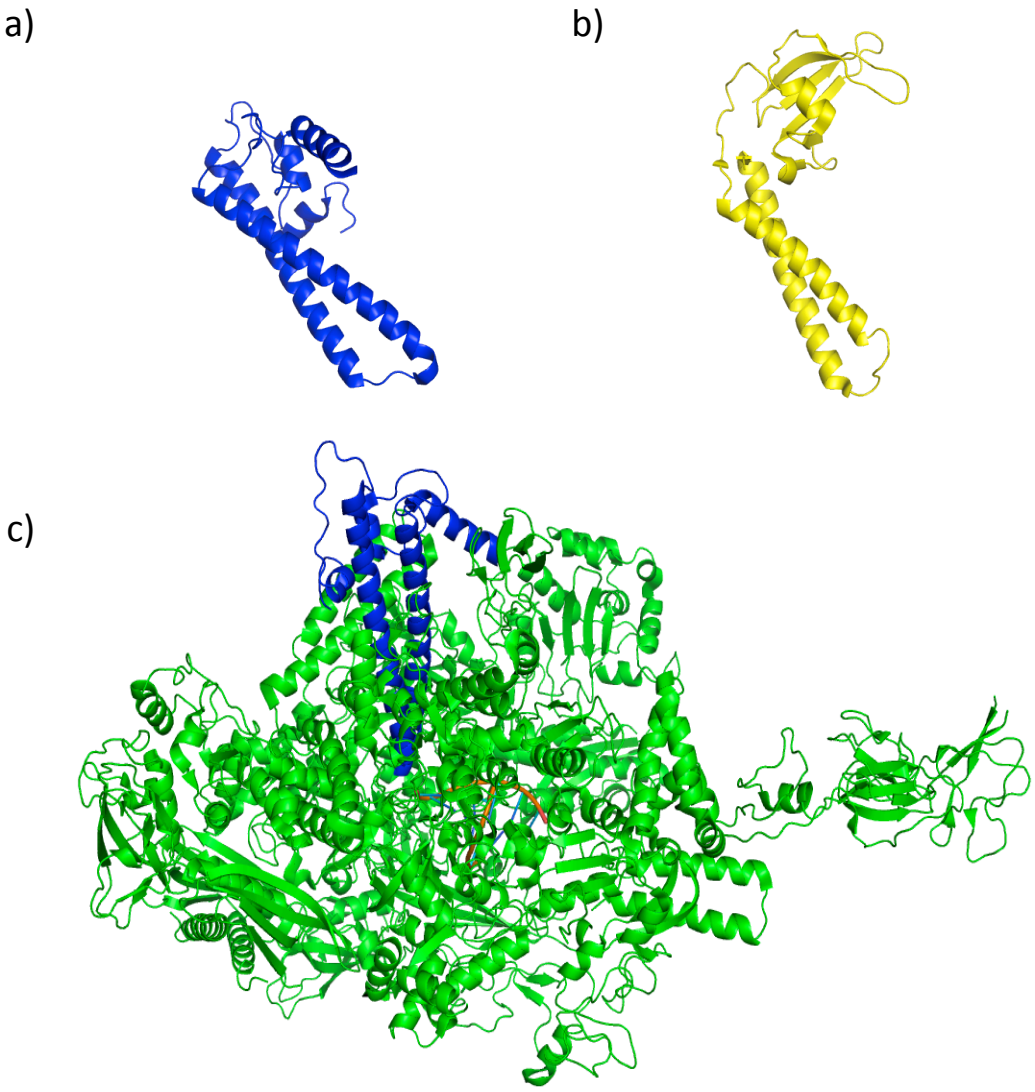


Figure 1.9: Structure of DksA a) Crystal structure of Dksa. DksA consists of a small globular domain and a long, helical coiled-coil. b) Crystal structure of Eco GreA, highlighting the structural homology with DksA. c) Model of DksA bound to the RNAP secondary channel. The co-crystal structure of *Thermus thermophilus* RNAP and the GreA Tth homolog Gfh1 was used to align the helical coil of DksA (blue) with the co-crystalized Gfh1 to produce a model of DksA bound in the secondary channel. Pdb codes: Dksa: 1TJL; GreA: 1GRJ; Tth RNAP/Gfh1: 3AOH.

active site (Fig 1.9c) (Perederina et al., 2004). *In vitro*, DksA and ppGpp act synergistically to inhibit RNAP activity at rrn promoters (Paul et al., 2004).

Recent work highlighted the differences between rrn promoters and rRNA regulation in gram-negative bacteria, such as Eco, and the gram-positive model organism Bacillus subtilus (Bsub) (Krasny and Gourse, 2004). Kransy and Gourse found no DksA homolog in Bsub, although RelA is conserved. In Bsub, RNAP is not regulated directly through an interaction with ppGpp. Rather, Bsub rrn promoters are sensitive to the concentration of initiating nucleotide, GTP, and therefore relA may function to reduce the intracellular GTP concentration and thereby inhibit synthesis of rRNAs (Krasny and Gourse, 2004), whereas in Eco ppGpp interacts directly with RNAP to modulate its activity.

These organisms clearly evolved different mechanisms to control the stringent response (Fig 1.10). Because rrn promoters have been studied in relatively few organisms, including Eco, Bsub and *Thermus thermophilus* (Tth) (Vrentas et al., 2008), it is difficult to evaluate whether these regulatory mechanisms are conserved or whether different bacterial families have evolved as yet unexplored mechanisms of controlling the stringent response. Sau, similar to Bsub, has no DksA homolog but has a clear RelA homolog. However, rRNA promoters and their regulation in this organism are entirely uncharacterized.

UP-Element function

In addition to the regulation described above, it was initially noted that mutations upstream of the -35 element in an Eco rrn promoter had a remarkable effect on RNAP

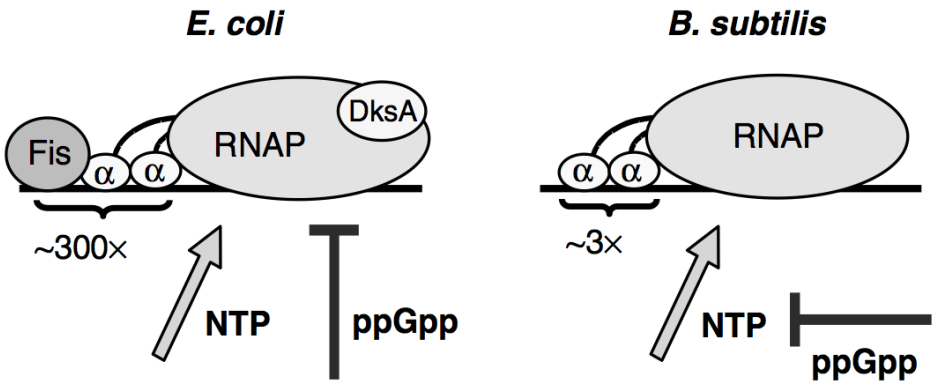


Figure 1.10: Model of rrn promoter regulation in E. coli and B. subtilis. In E. coli, the RNAP α-CTD binds the UP-element and interacts with the co-activator Fis. The small molecule ppGpp binds directly to RNAP and modulates the stability of the open promoter complex, which affects RNAP activity at these promoters. In B. subtilis, the α-CTD also interacts with promoter UP-elements, but not with a co-activator. ppGpp does not affect RNAP activity directly, but likely decreases the concentration of the initiating nucleotide at rrn promoters, GTP, and thereby decreases RNAP activity. Adapted from: (Krasny and Gourse, 2004).

activity (Gaal et al., 1989). Subsequent work showed that this effect was mediated by an A/T-rich DNA sequence upstream of the -35 element, termed the UP-element (Fig 1.5) (Estrem et al., 1998). The RNAP core α-subunit consists of two domains. While a dimer of the N-terminal domains is assembled in the core RNAP enzyme, the C-terminal domains (CTD) are linked to RNAP only by flexible linkers. The α-CTD can interact directly with DNA and is responsible for recognizing the upstream A/T rich DNA sequence (Fig 1.5) (Blatter et al., 1994; Estrem et al., 1999; Gaal et al., 1996; Ross et al., 1993). The α-CTD/UP-element interaction is required for robust transcription from rRNA promoters in both Eco and Bsub (Gaal et al., 1989; Krasny and Gourse, 2004; Newlands et al., 1991) although the requirement in Eco is more pronounced (Fig 1.10). A recent study solved the X-ray crystal structure of σ_4 and the α-CTD, which also interact with one another, bound to a -35 and UP-element containing DNA fragment (Fig 1.5c). UP-element activation provides an explanation for the extraordinary strength of the rrn promoters *in vivo* and is also a potential site of regulation for rRNA transcription.

***In vitro* studies of Sau RNAP**

Because of its medical relevance, Sau has been extensively studied. Many strains, including MRSA strains, have been fully sequenced (Baba et al., 2008; Holden et al., 2004; Iandolo et al., 2002; Nair et al., 2011; Ohta et al., 2004) and genetic approaches have been used in an attempt to understand the regulatory steps in the switch between non-pathogenic and pathogenic growth (Felden et al., 2011; Tuchscherr et al., 2011). Relatively fewer studies have attempted to use a purified, *in vitro* transcription system to study gene expression in Sau (Deora and Misra, 1996; Rao et al., 1995; Reynolds and

Wigneshweraraj, 2011). Most of the *in vitro* biochemistry has also focused on RNAP activity at pathogenicity promoters characterized by weak -10/-35 elements, suboptimal spacer length and binding sites for additional regulatory proteins (Rao et al., 1995; Reynolds and Wigneshweraraj, 2011). Mechanistic studies on RNAP activity at typical -10/-35 promoters have yet to be published. Because the genome of Sau is A/T-rich (Quail et al., 2012), promoter DNA may have different topology than in other organisms. The Sau RNAP is not active at all Eco promoters tested, but no sequence specificity was identified to explain the discrepancy in RNAP activity between these organisms (Rao et al., 1995). Furthermore, it is unclear how small molecules, such as ppGpp affect RNAP activity in Sau, and whether Sau RNAP is susceptible to inhibition by small molecules that bind to RNAP from other organisms.

PhERI binds directly to Sau RNAP

ORF67 (PhERI) was initially identified as a growth inhibitor in Sau cells. Further research attempted to identify the Sau binding partner of PhERI (Dehbi et al., 2008). PhERI was immobilized on a column and Sau cell lysates were allowed to bind. Only Sau RNAP was found to bind specifically to immobilized PhERI. To identify the RNAP subunit and domains to which PhERI binds, a yeast 2-hybrid approach was used. PhERI binds to Sau's group 1 σ factor, σ^A, and the interaction was localized to $\sigma^A{}_4$ (Fig 1.11a). By FRET, PhERI was found to disrupt the interaction between RNAP and promoter DNA. PhERI, and not a control ORF, was shown to inhibit purified Sau RNAP and Eco RNAP complexed with Sau σ^A *in vitro* on a -10/-35 promoter, λP_L (Fig 1.11b) (Dehbi et al., 2008). These results suggested that PhERI may be acting as a general anti-σ factor.

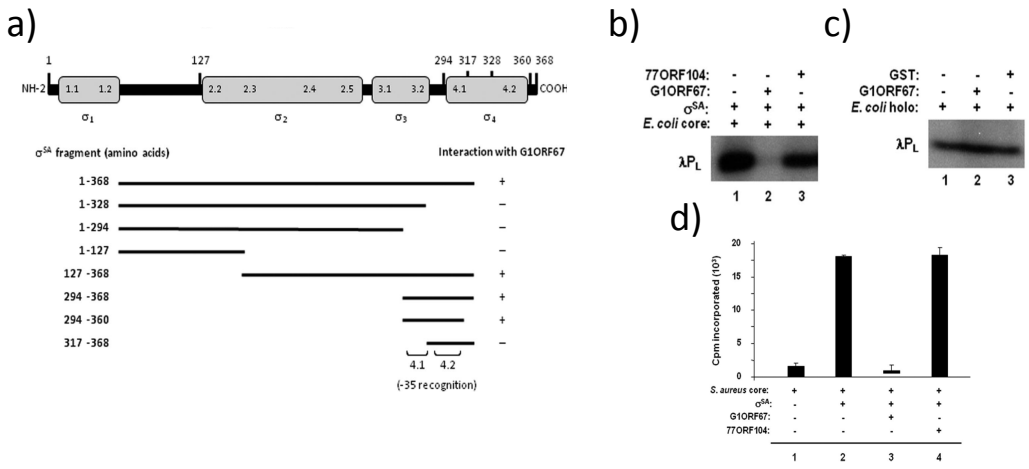


Figure 1.11: PhERI binds to Sau RNAP though an interaction with sigma region 4 and inhibits RNAP function. a) Yeast-two-hybrid showing the interaction between PhERI and Sau sigma region 4. b) PhERI inhibits the hybrid holoenzyme formed by Eco core and Sau σ^A at the -10/-35 λpL promoter. c) ORF77, a control protein does not inhibit the Eco RNAP/Sau σ^A hybrid RNAP enzyme. d) PhERI, but not ORF77, inhibits Sau RNAP at the λpL promoter. Adapted from (Dehbi et al., 2008).

Anti-σ factors bind to σ and modulate its activity

Because gene expression in prokaryotes is determined primarily at the step of transcription initiation (Young et al., 2002), the activity of σ factors is tightly regulated and can rapidly switch in response to diverse cellular stimuli (Kang et al., 1999; Raivio and Silhavy, 2001). Anti-σ factors are proteins of bacterial or phage origin that bind to σ and modulate its activity (Brown and Hughes, 1995; Helmann, 1999; Pene and Uzan, 2000). Anti-σ factors can be further regulated by anti-anti-σ factors and proteolysis, and these various signals play an important role in diverse cellular processes including sporulation, the biosynthesis of flagella, and switching to alternative σ factors (Campbell et al., 2008).

The T4 phage protein AsiA is the most well-understood phage encoded anti-σ factor (Orsini et al., 1993). AsiA is expressed early in T4 infection and is responsible for inhibiting Eco RNAP activity at Eco (Severinova et al., 1998), but not early phage, promoters (Orsini et al., 2004). AsiA binds to σ^{70}_4 and blocks the recognition of the -35 element (Lambert et al., 2004). The structure of AsiA bound to σ^{70}_4 was solved by NMR and revealed that AsiA reorganizes the structure of σ^{70}_4 such that the highly conserved HTH required for -35 recognition is converted into an extended helix, incapable of DNA binding (Fig 1.12a) (Lambert et al., 2004). Early phage promoters are able to compensate for the inability of σ^{70}_4 to recognize the -35 element through the presence of a strong UP-element or an extended -10 element (Orsini et al., 2004). AsiA forms a protein-protein interaction with a second phage protein, MotA, which can then recruit the RNAP/AsiA/MotA complex to phage promoters (Ouhammouch et al., 1995). AsiA is

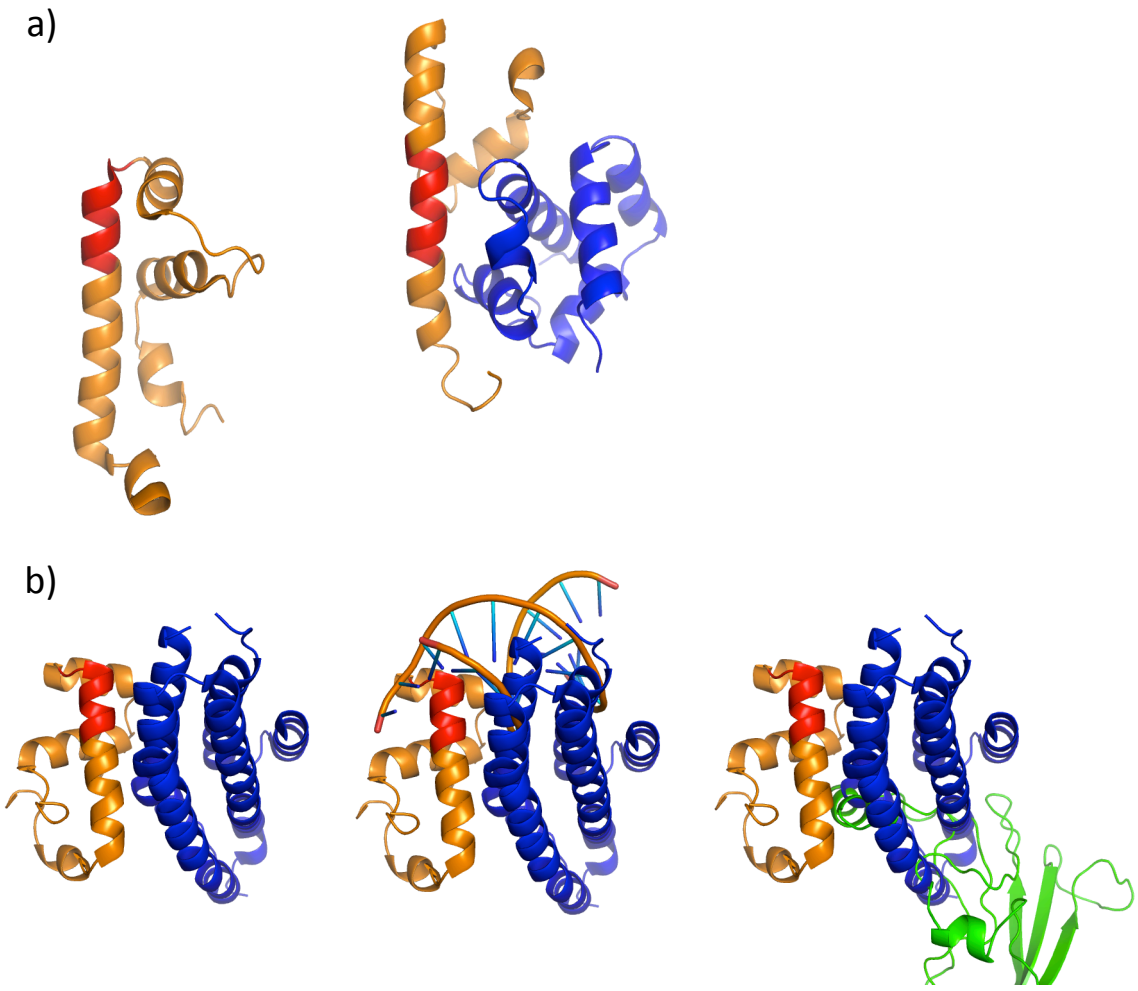


Figure 1.12: Structure of anti-σ factors reveals their mechanism. a) The conserved structure of σ_4 with helix-turn-helix (HTH) responsible for DNA recognition colored in red (left panel). Interaction with AsiA (blue) reorganizes the structure of σ_4 (orange) so that the HTH (red) becomes an extended helix incapable of interacting with promoter DNA. b) RSD blocks the DNA and RNAP binding surfaces of σ_4. Co-crystal structure of RSD (blue) and σ_4 (orange) with the HTH responsible for DNA-recognition colored red (left panel) The interaction with Rsd sterically blocks the DNA binding site of σ_4 (central panel) and the β-flap tip helix (green, left panel) required for proper interaction between σ and core RNAP.

therefore both an inhibitor of RNAP activity and, through its interaction with MotA, a co-activator of phage gene expression.

σ switching in Eco and sporuation in Bacillus subtillis are controlled by the activity of bacterially encoded anti-σ factors. The Eco protein RSD binds to σ^{70} through an interaction with region 4 and is responsible for sequestering σ^{70}, which has the highest affinity of cellular σ factors for RNAP (Maeda et al., 2000), to induce alternative transcriptional profiles (Jishage and Ishihama, 1999). RSD was crystallized in complex with $\sigma^{70}{}_4$. While the conformation of σ is maintained in the presence of RSD, RSD blocks both the DNA binding and the core binding surfaces of $\sigma^{70}{}_4$ (Fig. 1.12b) (Patikoglou et al., 2007). Bacterial and phage anti-σ factors have evolved various mechanisms to disrupt σ function and thereby alter RNAP activity. Structural studies have been central to our understanding of the detailed molecular mechanisms through which these anti-σ factors bind to and modulate RNAP activity.

The first section of this work will detail my attempts to understand the molecular mechanism of PhERI. Using X-ray crystallography, I solved the structure of PhERI bound to Sau $\sigma^{A}{}_4$. The structure described the interaction between the two proteins in atomic detail, yet cannot alone explain the inhibitory function of this protein on RNAP activity. Subsequent biochemical, genetic and bioinformatic work shows that PhERI specifically targets transcription from rRNA promoters by blocking UP-element activation by the RNAP α-CTD (PhERI: Phage Encoded rRNA Inhibitor). PhERI is the first example of a transcription factor that binds to σ but that modulates RNAP activity through an interaction with another region or RNAP.

Chapter 2:

Structural Studies of the

PhERI / σ^{A}_{4} Complex

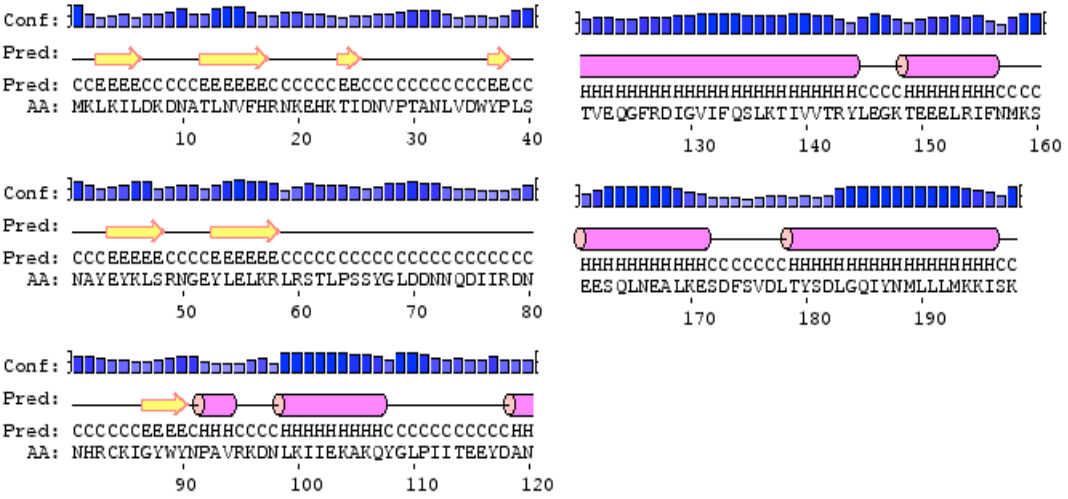
PhERI was identified as an inhibitor of cell growth in Sau (Liu et al., 2004). It was subsequently shown to bind directly to Sau RNAP through an interaction with σ^{A}_{4} and inhibit transcription *in vitro* (Dehbi et al., 2008). These results suggest that PhERI may function as a general anti-σ factor. Structural studies on this class of proteins have been particularly useful in determining the details of their interactions with σ and their mechanisms of action (Fig. 1.12) (Campbell et al., 2008). This chapter will describe my structural studies of PhERI and its interaction with σ^{A}_{4}. I solved X-ray crystal structures of PhERI bound to Sau σ^{A}_{4}. Unlike other anti-σ factors, PhERI does not appear to interact with the DNA or core binding determinants of σ^{A}_{4}. Additional biochemical and genetic data that validate the crystal structures will be presented. The bacterial 2-hybrid work identifying a σ^{A}_{4} mutant that is deficient in PhERI binding was performed by Cristina Montero-Diaz, a graduate student in the lab of Ann Hocschild's lab at Harvard Medical School.

3.0 Å Structure of the PhERI / σ^{A}_{4} complex

PhERI is a 25kDa protein with no sequence homology to any protein of known function (Fig 2.1a). Previous work had determined that PhERI does not interact with Eco σ^{70} (Dehbi et al., 2008); therefore PhERI is not toxic when expressed in Eco cells, as is

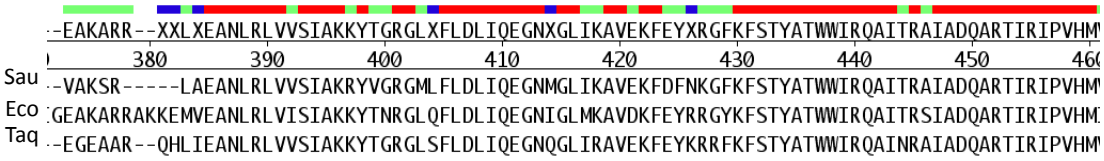
Figure 2.1 (adjacent page): Primary sequence of PhERI and Sau σ^A. a) Primary sequence and predicted secondary structure of PhERI. Secondary structure prediction was performed with PSIPRED. Predicted β-sheets are indicated with a yellow arrow and α-helices with a pink cylinder. b) Primary structure and alignment of the group 1 sigma factor from Sau, Eco and Taq. Regions 2, 3 and 4, responsible for DNA recognition, are shown. Secondary structure in region 4 is indicated as white boxes. Red indicates highly conserved residues.

a)

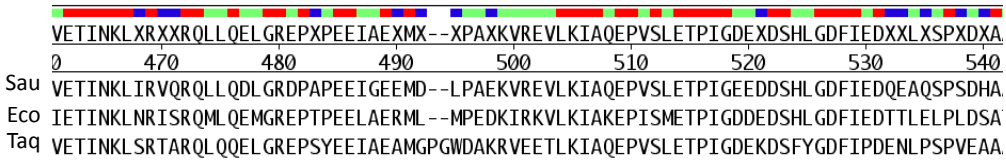


b)

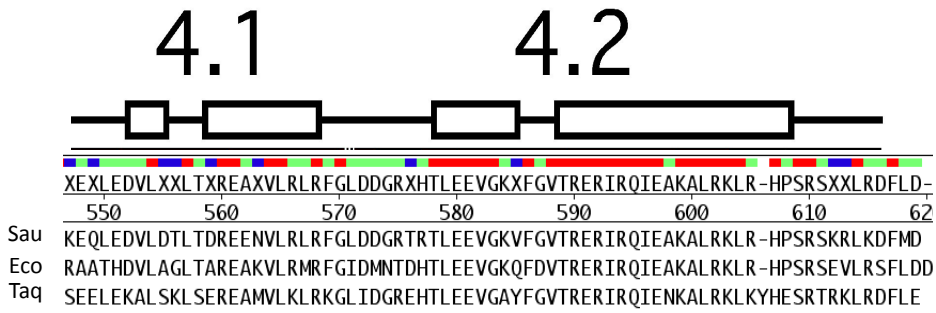
σ region 2



σ region 3



σ region 4



often the case with anti-σ factors. σ^A has been studied structurally in many contexts (Campbell et al., 2002; Murakami et al., 2002a; Murakami et al., 2002b; Patikoglou et al., 2007), but Sau domain 4 has not been crystallized. I used alignments with the Eco and *Thermus aquaticus* (Taq) group 1 σ factors (Fig 2.1a) to identify the Sau σ^A_4. I cloned PhERI into a co-expression cassette, as previously described (Campbell and Darst, 2000), with Sau σ^A_4 (Fig 2.2a). For protein purification, σ^A_4 was cloned with a precision protease cleavable N-terminal 6(his) tag. Initially, I tested the solubility of PhERI alone, σ^A_4 alone, and the two proteins together. While PhERI alone was largely insoluble (Fig 2.2b, lane 8), it was solublized when expressed with σ^A_4 (Fig 2.2b, lane 11). This is indicative of a stable biochemical interaction between the two proteins.

The PhERI / σ^A_4 complex was purified (Fig 2.2c) and screened for crystallization conditions. Large crystals formed under several conditions, and one condition (Fig 2.3a) (JCSG+ condition 59: 0.16 M Ca acetate, 0.08 M Na cacodylate pH 6.5, 14.4% PEG 8000, 20% Glycerol) produced crystals that diffracted to near 3.0 Å (Fig 2.3b). Although the conserved structure of σ^A_4 has been solved and could have been used as a molecular replacement search model, anti-σ factors have been shown to alter the conformation of σ. Furthermore, σ^A_4 is small, at ~10kD, and therefore may not have been sufficient to phase the structure of the complex. We purified seleno-methionine substituted protein complex to determine initial phase information experimentally.

Native PhERI /σ^A_4 crystals diffracted to 3.0Å and selenomethionine substituted protein crystals diffracted to 3.4Å with sufficient anomalous signal to locate the selenium sites and determine initial electron density maps after density modification (Fig 2.3c). Selenium sites were found with Shake-and-bake and the initial phases were determined

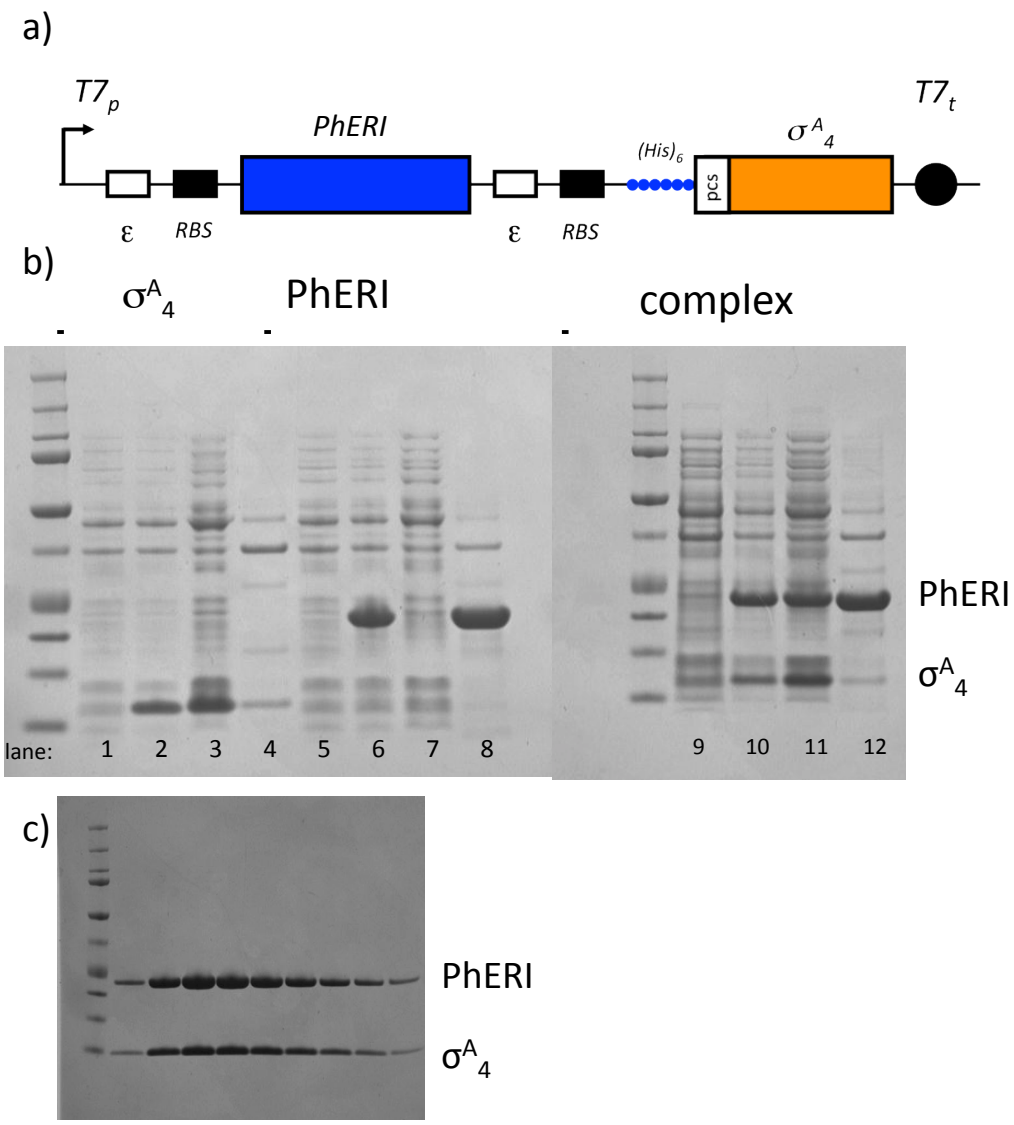


Figure 2.2: Cloning, expression and purification of the PhERI/σ^{A}_{4} complex. a) Co-expression cassette for PhERI/ σ^{A}_{4} expression in Eco cells. PhERI and σ^{A}_{4} are cloned as a single operon, both genes contain a ribosome binding site (RBS). σ^{A}_{4} is cloned with a 6(his) tag and cleavable precision protease (ppx) site for subsequent purification steps. b) Expression of PhERI/ σ^{A}_{4} complex in Eco cells. Lanes 1, 5 and 9: Preinduction. Lanes 2, 6 and 10: Post induction. Lanes 3, 7 and 11: Soluble protein. Lanes 4, 8 and 12: Insoluble protein. c) Purified PhERI/sigAd4 complex. 4-12% SDS-PAGE gel of PhERI/sigAd4 complex after purification on the sephadex 75 column.

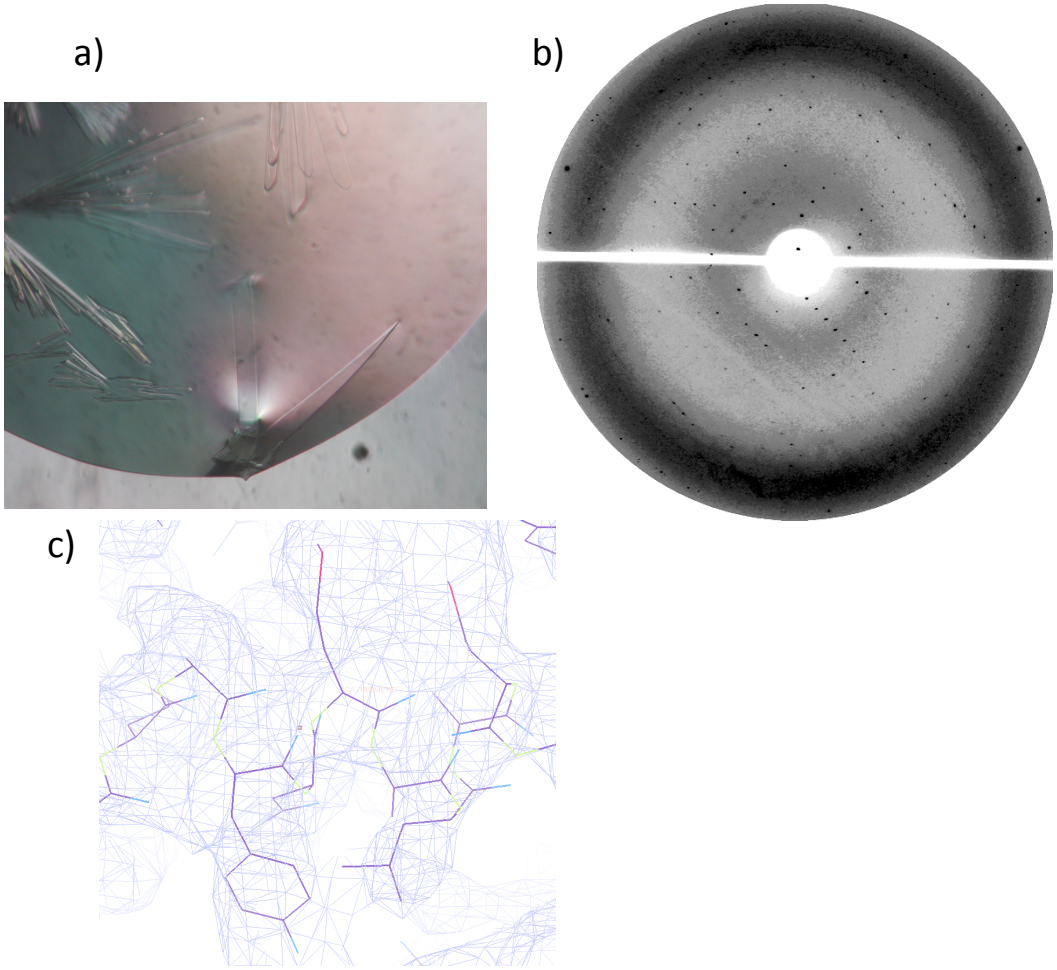


Figure 2.3: Crystallization, data collections and experimental electron density maps of the JCSG59 PhERI/ σ^{A}_{4} crystals. a) PhERI/ σ^{A}_{4} crystals. b) Diffraction pattern. Data were collected at X3A at NSLS at the Brookhaven National Laboratory. Native crystals diffracted to 3.0Å. c) Experimental electron density map showing the C-terminal helix of PhERI. Electron density is shown in blue and is contoured to a sigma of 1.

with SHARP. The crystallographic asymmetric unit contained two PhERI / σ^A_4 complexes were in the asymmetric unit. Initial models were built and refined against the native dataset.

σ^A_4 was not conformationally altered in either crystallographically independent complex PhERI (Fig 2.4). PhERI appeared as a largely helical protein with strong electron density for its C-terminal domain (Fig 2.4). However, the N-terminal region of PhERI was not well defined in the electron density. I could build some backbone in this region, and it appears that the secondary structure is largely characterized by β-sheets, as predicted (Fig 2.1a), but the density was not strong enough to build side chains or determine the sequence register. PhERI clearly does not reorganize the conformation of σ^A_4, but the details of the interaction between the two proteins remained unclear due to the relatively low resolution and the poor quality of the electron density map in PhERI's N-terminal region.

2.0 Å Structure of PhERI /σ^A_4 complex

A second crystal condition, ProteinComplex 38 (12% 1-propanol, 0.1M MES, pH 6.5, 10% PEG 5000 MME) produced crystals (Fig 2.5a) that diffracted to significantly higher resolution. To determine initial phases, I also crystallized seleno-methionine substituted protein complex. I collected native data on this crystal form to 2.0Å (Fig. 2.5b) and seleno-methionine substituted data to 2.2Å. Phases were solved with SHARP and density modification was performed using Solomon. A model was build into the electron density (Fig 2.5c) and refined against the native dataset. This crystal form had only one PhERI / σ^A_4 complex in the asymmetric unit.

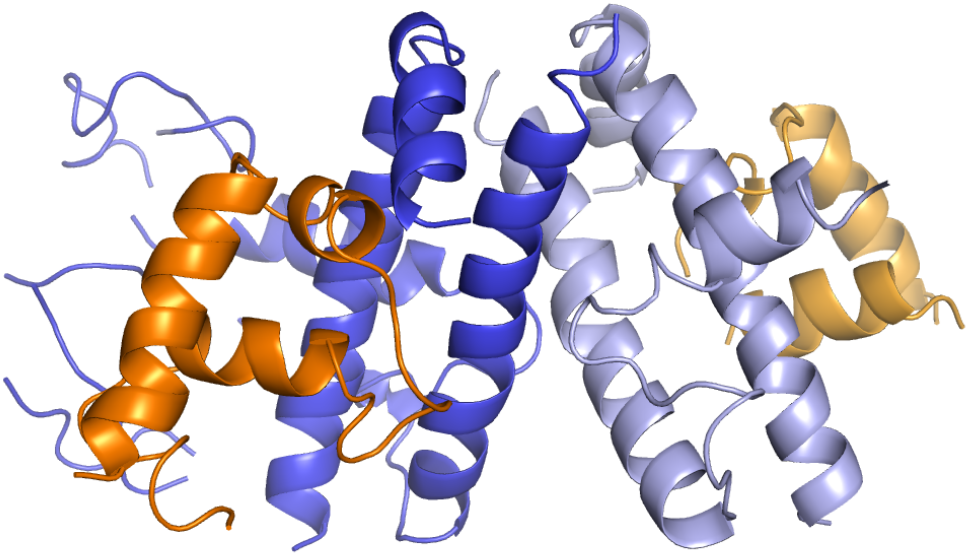


Figure 2.4: Initial model of the PhERI / σ^{A}_{4} co-crystal structure. There are two PhERI / σ^{A}_{4} complexes in the asymmetric unit. PhERI is shown in blue and σ^{A}_{4} in orange. The PhERI C-terminal domain, which is largely helical, is well defined in the density. The N-terminal domain, which is predicted to be β-sheet, is poorly defined in the density. σ^{A}_{4} is not structurally rearranged though its interaction with PhERI.

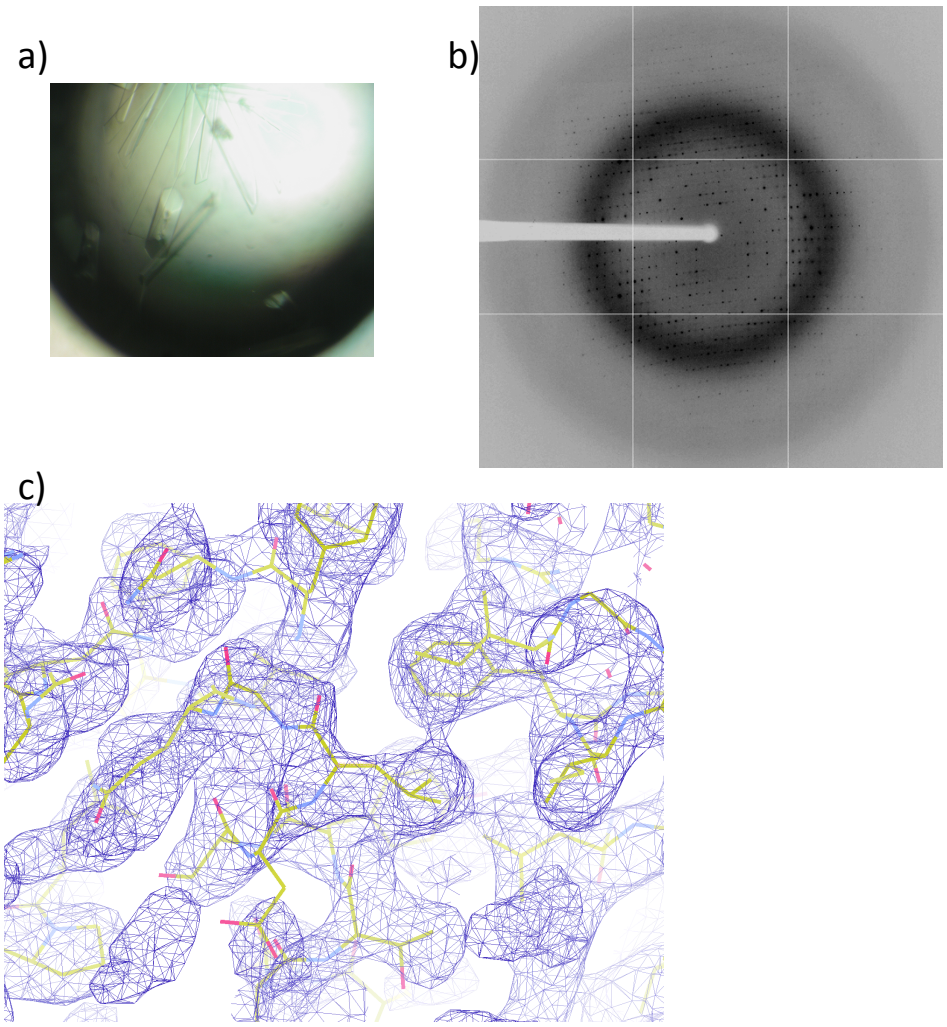


Figure 2.5: ProteinComplex 38 PhERI / σ^A_4 crystals. a) PhERI / σ^A_4 crystals. b) Diffraction of the PC38 PhERI/ σ^A_4 crystals. Data were collected at the Advanced Photon Source. Crystals diffract to 2.0Å. c) Experimental electron density maps showing the final PhERI / σ^A_4 model in yellow. Maps are countered to 1.0 sigma.

In this crystal form, I could build and refine both the N-terminal and the C-terminal domain of PhERI (Fig. 2.6a). Using the model from the PC38 structure, I was able to determine the sequence register and subsequently build much of the N-terminal domain in one protein complex in the JCSG59 structure (Fig 2.7). The second complex in the asymmetric unit of the JCSG59 structure has clear density for the N-terminal domain, however I was unable to build any side chains, even with the sequence register provided by the high-resolution model (Fig. 2.6a).

A comparison of the 3 crystallographically distinct PhERI / σ^A_4 complexes, two from the JCSG+59 crystal form and one from PC38 crystal form, shows that the PhERI / σ^A_4 complex is highly similar (Fig. 2.8)**.** Regions that differ between the 3 structures are also regions with elevated B-factors in the 2.0Å structure, indicating conformational flexibility within the crystal (Fig 2.8).

Overall PhERI /σ^A_4 structure

As noted above, PhERI does not alter the conformation of σ^A_4 (Fig. 2.6b, rmsd of 1.02 between Sau σ^A_4 and the 1.8Å structure of Tth σ^A_4). PhERI does not share any structural homology to any previously solved crystal structures and does not contain a previously described fold.

PhERI forms an extensive molecular interface with σ^A_4 (Figure 2.6a) through both its N-terminal beta-sheet rich domain (purple, 1031.8Å^2 of buried surface area) and its C-terminal helical domain (blue, 1757.3Å^2 of buried surface area), burying a total of 2789.1Å^2. The two proteins form extensive hydrogen bonds and salt bridges (Fig 2.9a) and hydrophobic interactions (Fig. 2.9b).

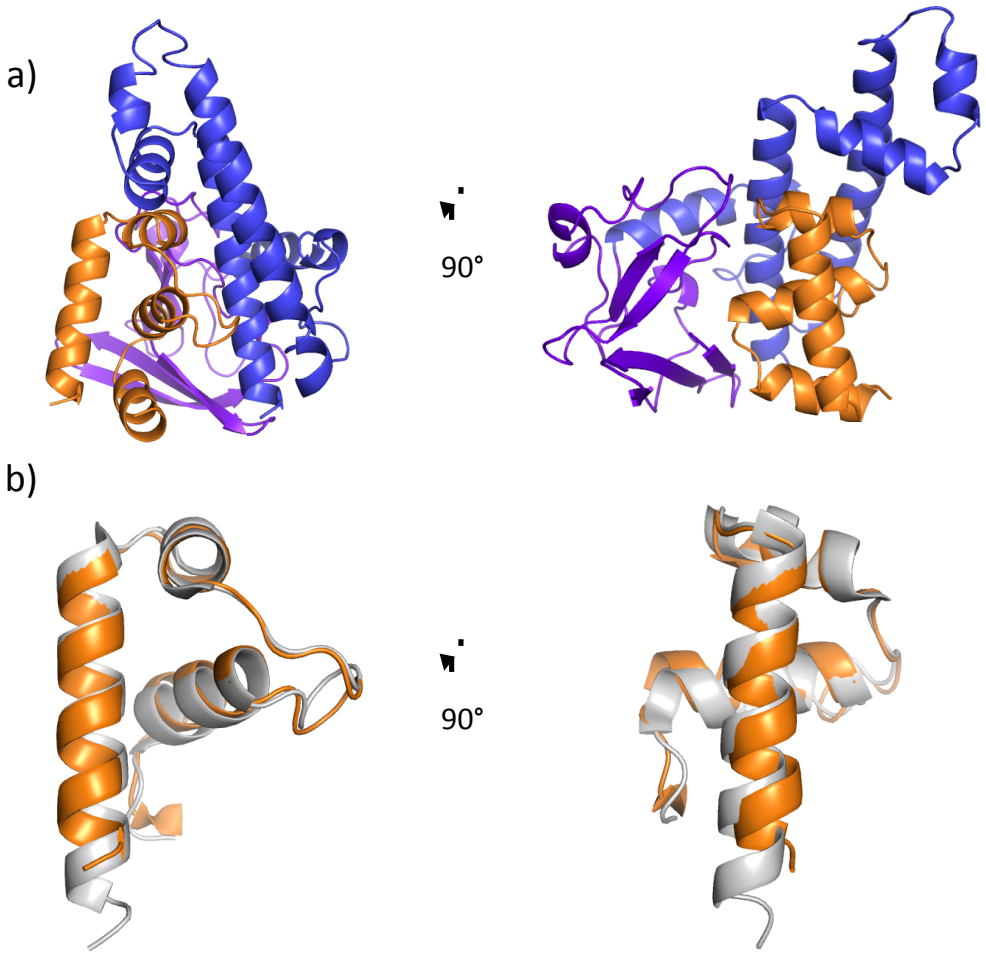
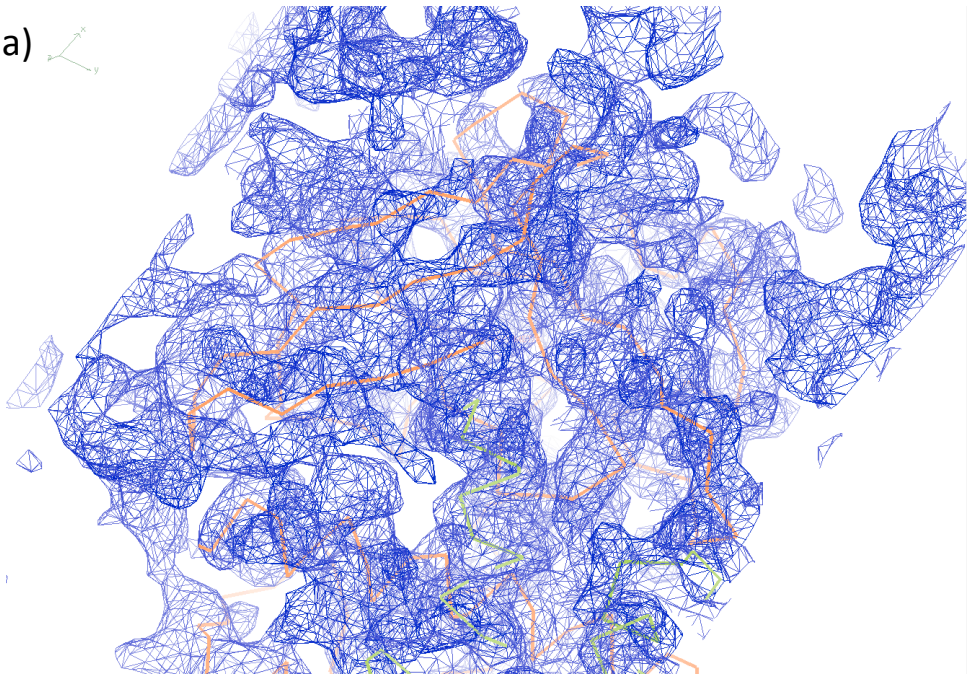


Figure 2.6: 2.0Å Structure of PhERI / σ^{A}_{4} complex. a)The PhERI / σ^{A}_{4} complex. σ^{A}_{4} is shown in orange, the PhERI-NTD in purple and the PhERI-CTD in blue. b) σ^{A}_{4} is not reorganized through its interaction with PhERI. Structure of Sau σ^{A}_{4} from the PhERI / σ^{A}_{4} complex (orange) aligned to the high-resolution (1.8Å) structure of Taq σ^{A}_{4} (orange).

Figure 2.7 (adjacent page): 3.0 Å structure of PhERI / σ^A_4. a) Placement of the poorly defined PhERI-NTD in the 3.0 Å electron density map using the 2.0Å model. 2Fo-Fc map contoured to 1.0 sigma after initial refinement with σ^A_4 and the PhERI CTD. Clear density for the β-sheets is seen, but side chains are not visible. b) Final 3.0Å structure of PhERI/ σ^A_4 showing 2 complexes in the asymmetric unit. PhERI is shown in blue (CTD) and purple (NTD) and σ^A_4 is shown in orange.

a)



b)

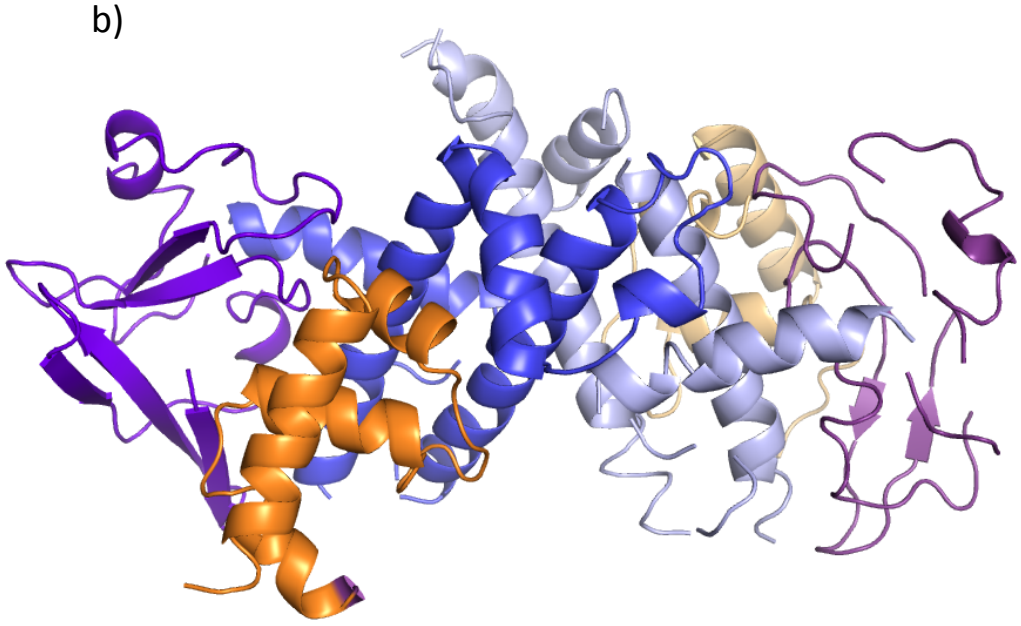


Table 2.1. Crystallographic statistics for PhERI/σ^{A}_{4} co-complex crystals (JCSG59).

	Native	Se1
Data collection		
Space group	C2	C2
Cell dimensions		
a, *b*, *c* (Å)	219.13, 57.03, 56.95	219.63, 56.83, 57.35
α, β, γ (°)	90, 104.79, 90	90, 104.49, 90
		Peak
Wavelength	1.03839	0.9790
Resolution (Å)	40-3.0	20-3.4
R_{sym} or R_{merge}	0.056 (.62)	0.124 (0.61)
$I / \sigma I$	36 (2.6)	(2.1)
Completeness (%)	100 (98.9)	99.8 (99.3)
Redundancy	6.1 (6.2)	2.7 (2.7)
Refinement		
Resolution (Å)	37-3.0	
No. reflections	13812	
R_{work} / R_{free}	0.2765/0.3317	
No. atoms		
Protein	3332	
Ligand/ion	0	
Water	0	
B-factors		
Protein	40.79	
Water		
R.m.s deviations		
Bond lengths (Å)	0.004	
Bond angles (°)	0.824	

*Values in parentheses are for highest-resolution shell.

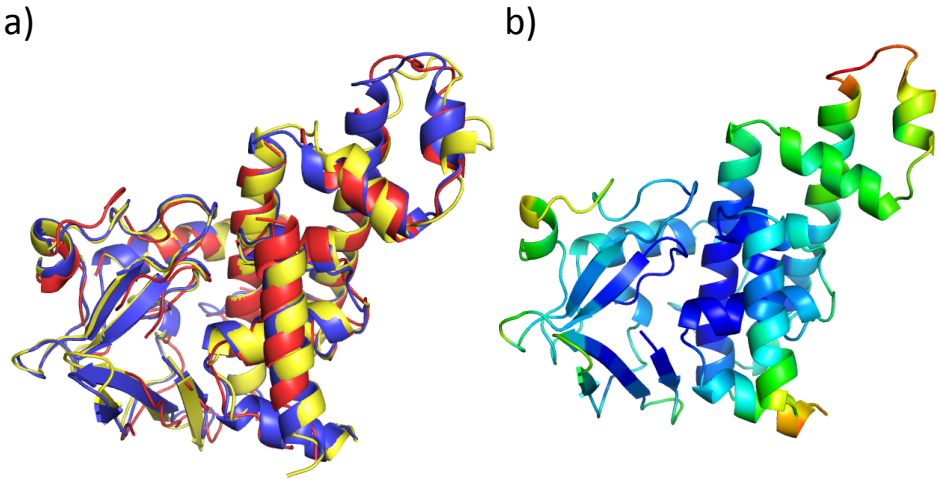


Figure 2.8: Comparison of the 3 PhERI/ σ^{A}_{4} structures. a) The 2.0Å structure (blue) and 3.0Å structures (Yellow and red) are aligned. The three independent structures are highly similar (RMSD=). b) 2.0Å structure of PhERI/σ^{A}_{4} colored by crystallographic B-factor. Regions that are flexible in the 2.0Å structure, as indicated by relatively elevated B-factor (red) show flexibility between the 3 crystal forms, as shown in a.

Table 2.2. Crystallographic statistics for PhERI/σ^{A}_{4} co-complex crystals (PC38).

	Native	Se1
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
a, *b*, *c* (Å)	38.38, 64.72, 108.72	38.35, 64.87, 109.74
α, β, γ (°)	90, 90, 90	90, 90, 90
		Peak
Wavelength	0.97949	0.97918
Resolution (Å)	30-2.0	30.0-2.20
R_{sym} or R_{merge}	0.047 (0.879)	0.057 (0.734)
$I / \sigma I$	43 (2.75)	22 (2.6)
Completeness (%)	98.3 (98.3)	98 (97.6)
Redundancy	5.5 (5.2)	3.1 (3.1)
Refinement		
Resolution (Å)	26.5-1.996	
No. reflections	18679	
R_{work} / R_{free}	0.2192/0.2506	
No. atoms		
Protein	2085	
Ligand/ion	0	
Water	56	
B-factors		
Protein	40.79	
Water		
R.m.s deviations		
Bond lengths (Å)	0.004	
Bond angles (°)	0.775	

*Values in parentheses are for highest-resolution shell.

The structure of σ^A_4 has been determined in many contexts, including bound to DNA (Campbell et al., 2002), RNAP (Murakami et al., 2002b; Vassylyev et al., 2002), RNAP and DNA (Murakami et al., 2002b), and anti-σ factors (Lambert et al., 2004; Westblade et al., 2004). Although PhERI interacts with a large portion of σ^A_4, the highly conserved helix-turn-helix (HTH) DNA-binding motif required for recognition of the -35 element (Campbell et al., 2002), and the region responsible for binding the RNAP β-flap (Figure 4c) (Murakami et al., 2002b) are exposed in the presence of PhERI, and PhERI does not appear that it would sterically clash with either of these σ^A_4 ligands (see Figure 2.10).

Model of PhERI bound to the RNAP holoenzyme

To understand how PhERI may function in the context of the RNAP holoenzyme, I docked PhERI onto models of the RNAP-holoenzyme open promoter complex (RPo) by superimposing the structural core of σ^A_4 (Fig. 2.10a) (Jain et al., 2005; Murakami et al., 2002b). PhERI binds to an external surface of holoenzyme and appears poised to interact with promoter DNA both upstream and downstream of the -35 element, flanking the σ^A_4 HTH-recognition helix positioned in the major groove of the -35 element (Fig. 2.10). PhERI has a stretch of positively charged residues in its N-terminal region, which are in close proximity to DNA between the -35 to -40 positions (Fig 2.10c). PhERI also does not block the protein-protein interaction between σ^A_4 and the RNAP α-CTD. PhERI is however in close proximity to promoter DNA in the region upstream of the -35 element, near the promoter binding site of the α-CTD (Fig 2.10).

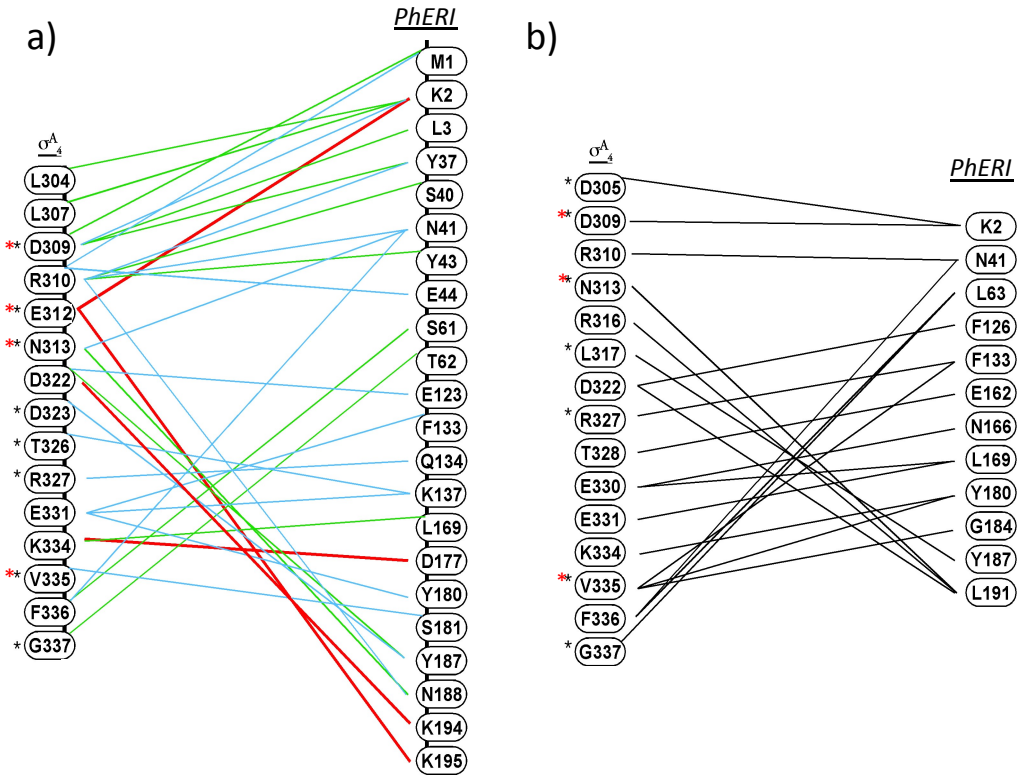


Figure 2.9: Interaction map between PhERI and σ^{A}_{4}. a) Hydrogen bond and ionic interaction map. Residues are listed and interactions are shown by lines connecting two residues. Hydrogen bonds are shown in green, hydrogen bonds that are mediated by a water molecule are shown in blue and ionic interactions are shown in red. Interactions with the protein backbone are signified by a line touching the residue number while interactions mediated by the protein side chain are shown by lines starting next to the residue number. Residues that differ between Sau and Eco are highlighted with an asterisk. Residues found to be binding determinants by the subsequent 2-hybrid analysis are shown with a red asterisk. b) Hydrophobic interaction map. Residues are listed as in part a and lines between residues denote Van der Waals interactions between the two proteins.

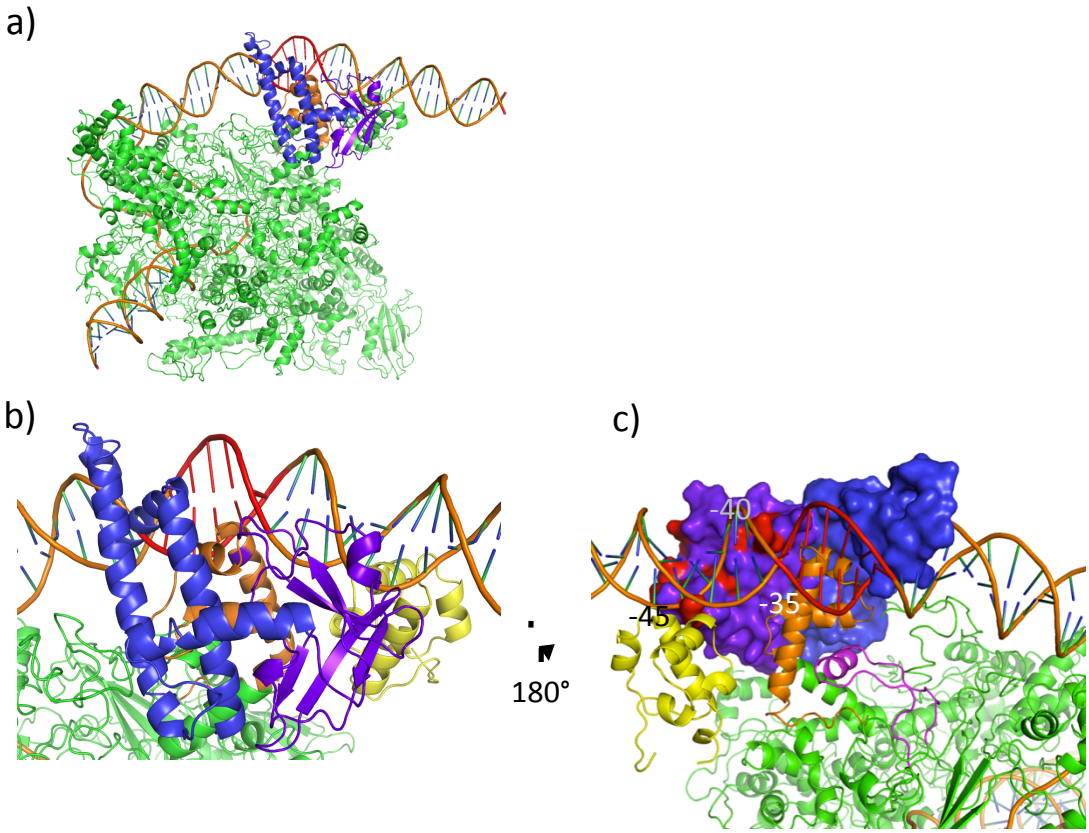


Figure 2.10: Model of PhERI bound to the RNAP open promoter complex (Rpo). a) PhERI was modeled onto the RNAP (Rpo) complex by aligning the conserved structure of σ^{A}_{4} from the 2.0Å PhERI co-complex and σ^{A}_{4} from the Rpo model. RNAP is shown in green, promoter DNA in orange (with the -35 element colored red), PhERI is colored blue/purple as in Figure 2.6. b) Zoomed view of the PhERI /RNAP interaction(s). PhERI interacts with an external surface of Rpo. The RNAP α-CTD is shown (yellow) interacting with an UP-element. c) PhERI does not appear to block the DNA or core binding surfaces of σ^{A}_{4}. PhERI is shown as a surface representation colored as above. The -35 element is colored red, and bases upstream of the -35 are labeled. The β-flap is shown in pink to highlight the β-flap / σ^{A}_{4} interaction. Positively charged residues in PhERI well poised to interact with promoter DNA are highlighted in red.

Identification of a σ^A mutant that does not bind PhERI

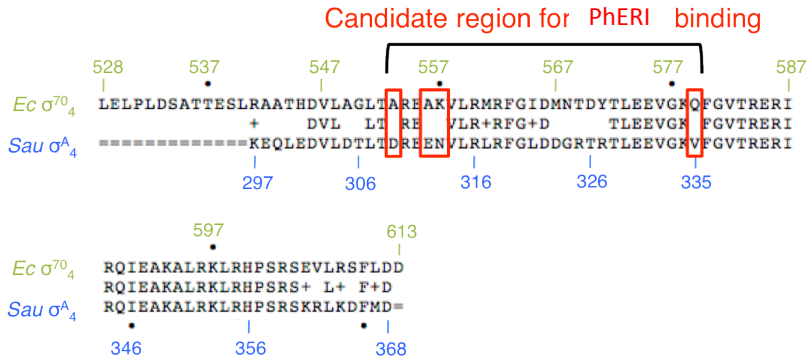
To validate the structure, and provide a useful biochemical tool to study PhERI's function, Cristina Montero-Diaz in Ann Hochschild's lab performed a screen for a $\sigma^A{}_4$ mutant that would abolish the interaction with PhERI. PhERI does not bind to Eco $\sigma^{70}{}_4$, while it does bind to Sau $\sigma^A{}_4$. These domains are highly conserved between the two organisms (Fig 2.11a) and therefore we could construct hybrid σ_4 domains and test their interaction with PhERI by bacterial two-hybrid. Briefly, one of the proteins of interest is fused to the DNA-binding protein λcl and the other replaces the α-CTD of RNAP. In the presence of a suitably strong protein-protein interaction, the λcl/RNAP complex will drive the transcription of a test promoter.

Two-hybrid experiments mapped the PhERI /$\sigma^A{}_4$ interaction to between Sau residues 309 and 335 (Fig 2.11). The interaction was further probed by swapping individual residues from the Sau to the Eco sequence. Four point mutations (D309A, E312A, N313K, V335Q) were required in combination to completely disrupt the PhERI / $\sigma^A{}_4$ interaction (Fig 2.11b and c). Mutations at these residues are not predicted to alter the interaction between σ and RNAP or σ and DNA.

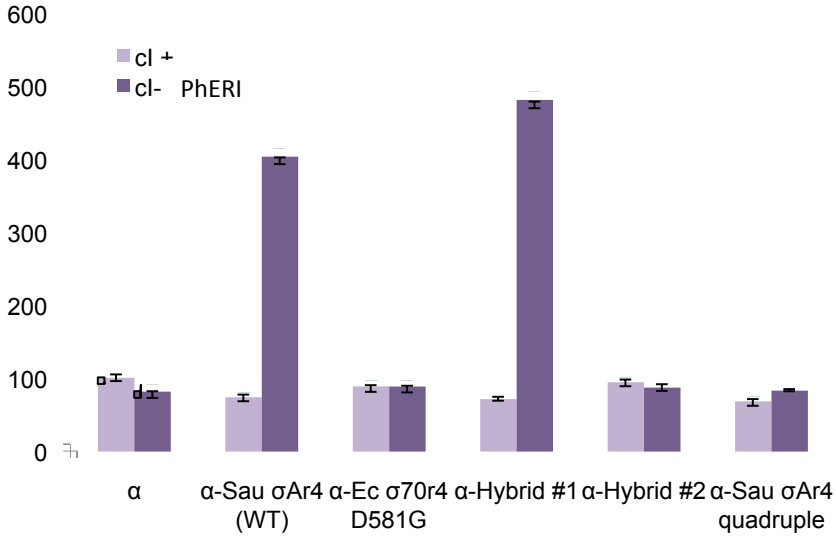
The σ residues identified in the 2-hybrid analysis sit at the interface between σ and PhERI in the co-crystal structure. Sau σ^A E312 and N313 form both hydrogen bonding and ionic interactions with two positively charged residues in PhERI (K2 and K195). The mutations E312A and N313K would not only disrupt these interactions but it would also introduce another positively charged residue in close proximity to the positively charged patch of PhERI (Fig 2.12a). V335 sits in a hydrophobic pocket in PhERI; the introduction of the polar glutamine residue at this location is likely

Figure 2.11 (adjacent page): Identification of a Sau σ^A mutant deficient in PhERI binding. a) Alignment of Eco σ^{70}, which cannot bind PhERI and Sau σ^A, which can. b) Bacterial 2-hybrid analysis to identify the Sau residues that confer PhERI binding specificity. PhERI was fused to lambda cl and $\sigma^A{}_4$ replaced the RNAP α-CTD. Binding of the 2 proteins was monitored by the expression of a test gene, LacZ. The analysis of hybrid 1 (containing Sau residues 309-335) and hybrid 2 (lacking these residues) identified the region of σ^A interacting with PhERI. The mutation of 4 individual residues (Sau $\sigma^A{}_4$ quadruple: D301A, E312A, N313K, V335Q) eliminated the interaction between SigAd4 and PhERI in the 2-hybrid experiment. Adapted from Diaz and Hochschild (personal communication).

a)


Candidate region for PhERI binding
528
537
547
557
567
577
587
Ec σ70 4
LELPLDSATTESLRAATHDVLAGLTARE AK VLRMRFGIDMNTDYTLEEVGK Q FGVTRERI
+ DVL LT RE VLR+RFG+D TLEEVGK FGVTRERI
Sau σA 4
KEQLEDVLDTLTDRE EN VLRLRFGLDDGRTRTLEEVGK V FGVTRERI
297
306
316
326
335
597
613
Ec σ70 4
RQIEAKALRKLRHPSRSEVLRSFLDD
RQIEAKALRKLRHPSRS+ L+ F+D
Sau σA 4
RQIEAKALRKLRHPSRSKRLKDFMD
346
356
368

b)


600
500
400
300
200
100
0
cl +
cl- PhERI
α
α-Sau σAr4 (WT)
α-Ec σ70r4 D581G
α-Hybrid #1
α-Hybrid #2
α-Sau σAr4 quadruple

c)


Hybrid #1
528
552
309
335
580
613
Hybrid #2
297
308
553
579
336
368
Sau σA 4 quadruple
297
D309A
E312A
N313K
V335Q
368

unfavorable (Fig. 2.12b) . The mutations selected in the screen for their importance to the PhERI / σ^A_4 interaction are in agreement with the structural information provided by the X-ray crystal structures.

Biochemical analysis of σ^A mutant

To show that the σ^A mutant identified in the bacterial 2-hybrid screen is deficient in PhERI binding but not RNAP binding, I performed a native gel shift analysis. PhERI, Sau RNAP, σ^A, and the σ^A quadruple mutant (4M, D309A, E312A, N313K, V335Q) were purified. On a native gel there is clear evidence for an interaction between PhERI and Sau holoenzyme, as previously described (Fig 2.12c lanes 2 and 3). While the σ^A mutant clearly interacts with RNAP core to form holoenzyme (Fig. 2.12c lane 4), no shift is evident upon the addition of PhERI (Fig. 2.12c lane 5). Additionally, the interaction between σ^A and PhERI is clearly visualized on the same gel (Fig. 2.12c lanes 2 and 3), but is absent when the σ^A 4M is used. Because the interaction surface between the two proteins is essentially the same in 3 independent structures and is validated by an *in vivo* bacterial 2-hybrid mutational analysis and *in vitro* gel shift assay, I am confident that I have identified the functionally relevant binding determinants of PhERI for the RNAP holoenzyme.

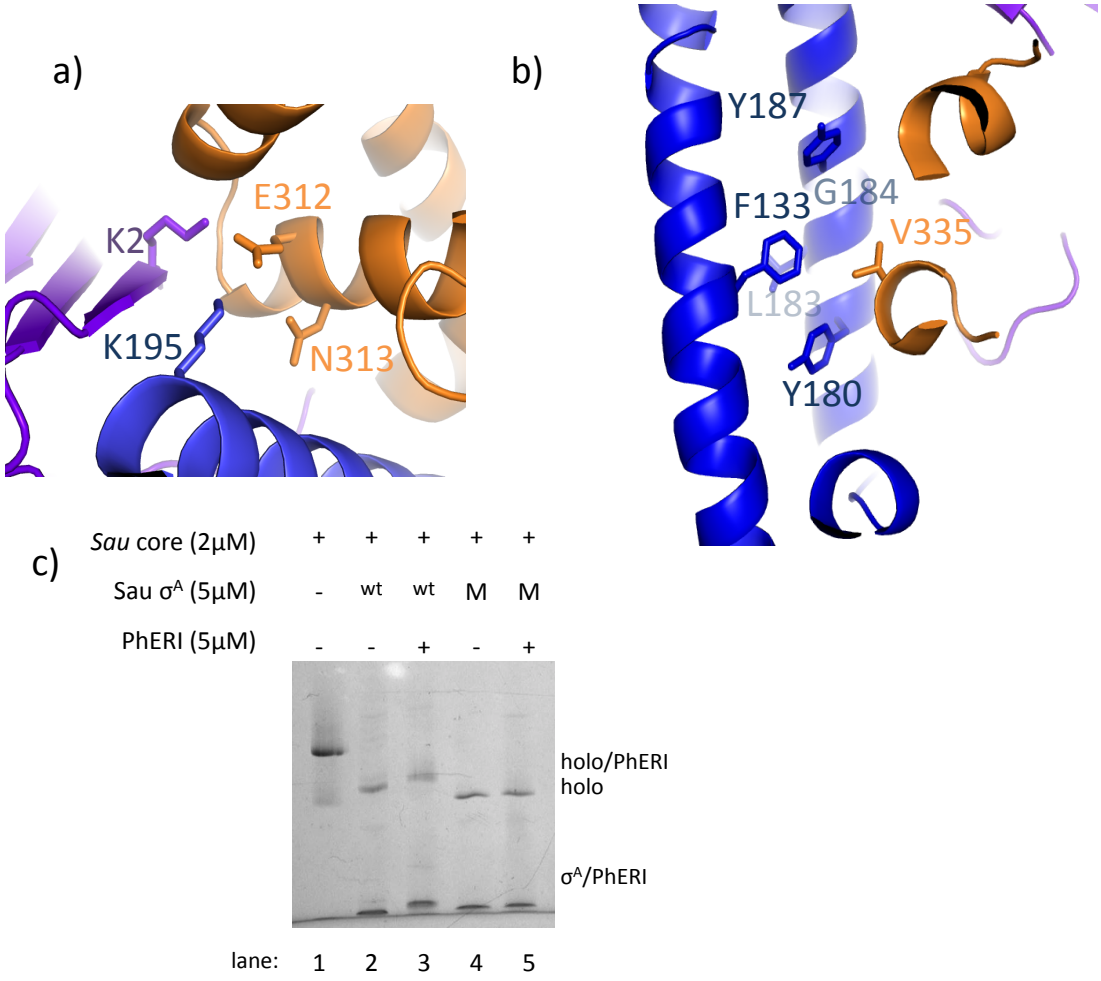


Figure 2.12: Structural analysis of the σ^A_4 quadruple mutant. a) The co-crystal structure of PhERI (purple: NTD, blue: CTD) interacting with σ^A_4. Mutations at σ^A position 312 and 313 would disrupt the interaction between E312 and PhERI K2 and K195 and add a positively charged residue likely to be repelled. b) σ^A V335 is situated in a hydrophobic pocket in PhERI, where the polar Q would be unfavorable. c) PhERI does not interact with the σ^A mutant *in vitro*. Purified Sau core, σ^A, σ^A quadruple mutant (M) and PhERI were incubated on ice at the indicated concentrations and run on a 4-12% native PAGE PhAST gel.

PhERI structural homology

BLAST searches for proteins with sequence homology to PhERI provided several homologous proteins from sequenced phage genomes. One homolog, was encoded in the Sau phage Twort; four others were found in phages that infect other fermicutes but not Sau. The relatively high levels of conservation between PhERI homologs argue that the proteins are functionally conserved (Fig 2.13).

The program Consurf can be used to map the conservation of residues onto protein structures. This analysis shows clearly that nearly all of the universally conserved residues in all the PhERI homologs are in the hydrophobic core of the protein, indicating that the overall fold of the homologs is similar. The most highly conserved region of PhERI maps to the interaction surface with σ^{A}_{4} (Fig 2.14), while the rest of PhERI's surface shows relatively low levels of sequence conservation. Because the region responsible for the direct interaction with σ^{A}_{4} is highly conserved, these homologous proteins are likely to be functionally related as well as structurally related.

A PhERI homolog was identified in an additional Sau specific phage, Twort. To test whether Twort ORF65 bound to Sau σ^{A} I cloned and purified the protein. Twort ORF65 also forms an interaction, as visualized on a native gel, with Sau σ^{A} (Fig. 2.14c). Interestingly, using the mutant σ^{A} that was identified as deficient in PhERI binding also eliminated the interaction with Twort ORF65 (Fig 2.14c). This shows that not only does Twort ORF65 also bind to σ^{A} but that it likely interacts by the same molecular determinants as the G1 protein PhERI. The two proteins are also sufficiently conserved to produce a structural model of ORF65 based on the structure of PhERI. These data

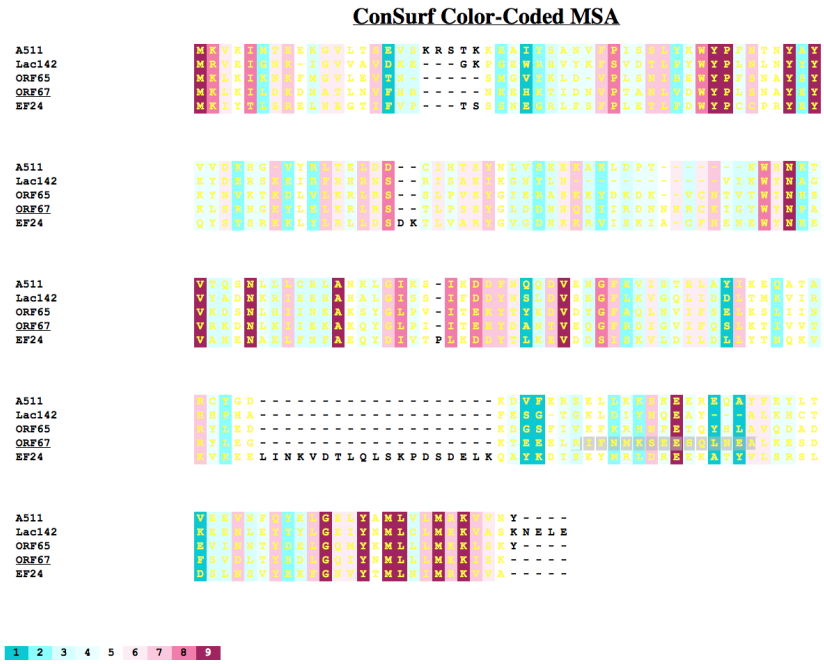


Figure 2.13: Alignment of PhERI homologs. The program ConSurf was used to produce a structure-based alignment. Residues are colored by conservation (blue = non-conserved ; red = highly conserved, see the color code in the inset). The PhERI homologs are labeled by the phage genome in which they were identified:

ORF67: Sau phage G1 PhERI

ORF65: Sau phage Twort PhERI

Lac142: Lactobacillus phage Lb338 PhERI

EF24: Enterococcus phage phiEF24C PhERI

A511: Listeria phage A511 PhERI

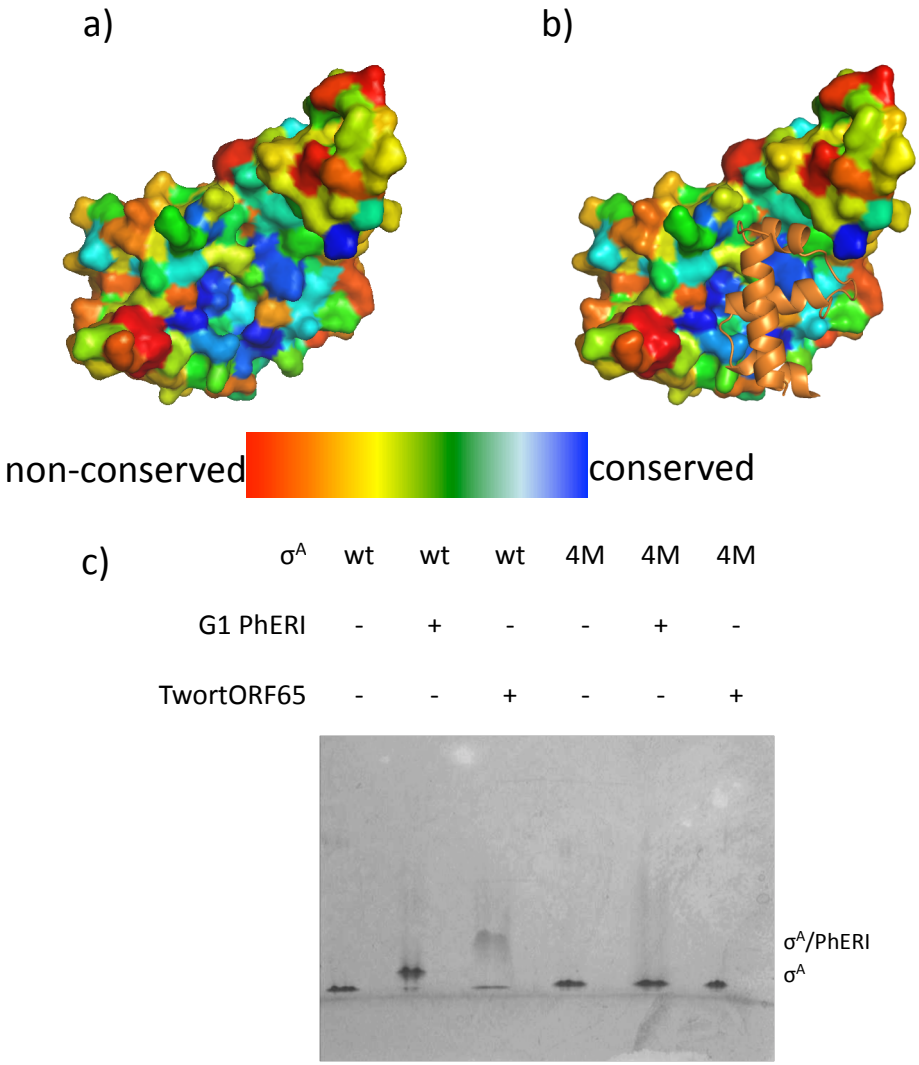


Figure 2.14: Structural conversation of PhERI. a) The structure of PhERI from the 2.0Å co-crystal structure with σ^A_4 is shown as a surface map colored by conservation. The structural conservation map was made by the program ConSurf. Highly conserved residues are shown in blue and poorly conserved residues in red. b) Highly conserved PhERI residues map to the binding site of σ^A_4. σ^A_4 is shown in orange as a cartoon model. Residues that are highly conserved (blue) cluster around the σ^A_4 binding site. c) The PhERI homolog from the Sau phage Twort interacts with σ^A, but not the σ^A mutant deficient in PhERI binding. Purified proteins were incubated at the indicated concentrations and run on a 4-12% native PhAST gel.

strongly suggest that the two proteins are not only structural but also functional homologs.

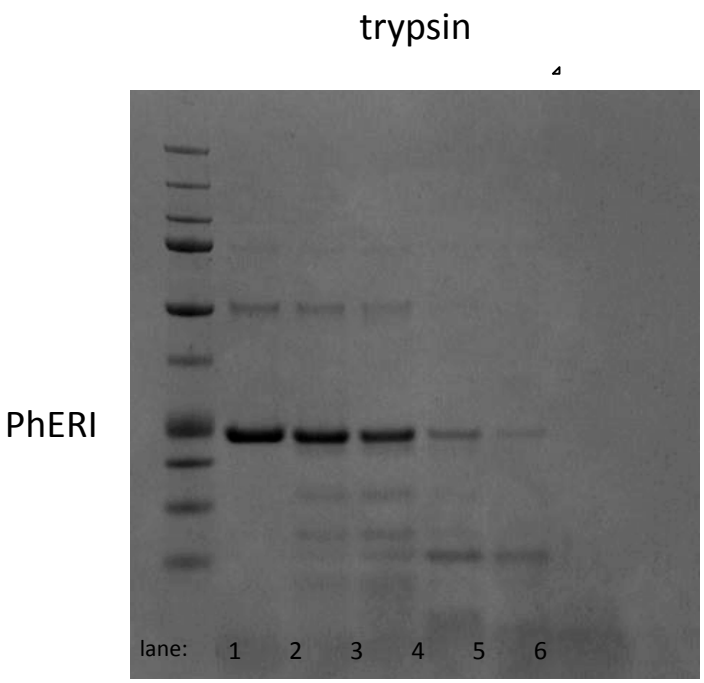
PhERI is conformationally stabilized through its interaction with σ^{A}_{4}

I solved the structure of PhERI complexed with Sau σ^{A}_{4}. To test whether PhERI undergoes any conformational changes upon binding σ, I attempted to crystallize PhERI alone. PhERI was purified and extensively screened for crystallization conditions. However, I found no evidence for PhERI crystallization under any conditions.

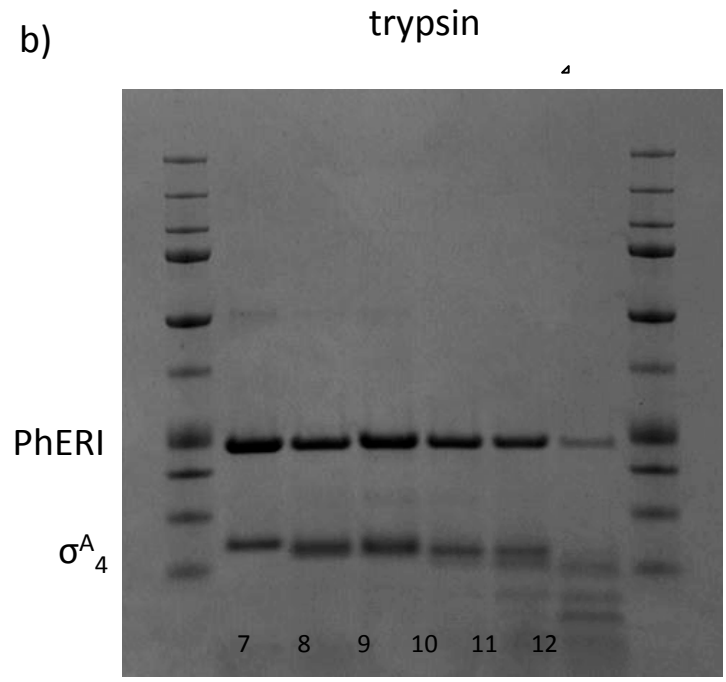
Proteins with conformational flexibility are difficult to study by X-ray crystallography because they are less likely to form the normally repeating units required for crystal packing. We used limited proteolysis as a tool to examine the conformational stability of PhERI alone and with its binding partner. Stably folded proteins are more resistant to digestion by proteases as many of their cleavage sites are buried within the core of the protein. PhERI alone is readily degraded by trypsin, indicative of a poorly folded protein (Fig 2.15a). However, when incubated with σ^{A} or σ^{A}_{4}, PhERI becomes significantly more resistant to degradation (compare Fig 2.15a lane 4 and 2.15b lane 10). Because we solved the crystal structure of PhERI bound to σ^{A}_{4} we know that a stable and well-folded complex is formed between the two proteins. However, it appears that PhERI may be largely unstructured or exhibit high levels of conformational flexibility when not bound to σ.

Figure 2.15 (adjacent page): PhERI is structurally stabilized by its interaction with σ^{A}_{4}. Limited proteolysis of PhERI alone (a) and PhERI in complex with σ^{A}_{4} (b). The complex was formed on ice in 1x proteolysis buffer at 5uM and incubated at 30°C with protein:protease concentrations of 1:0, 1000:1, 100:1, 50:1, 10:1, 5:1. After 20 minutes, reactions were stopped by the addition of 1mM PMSF and run on a 4-12% SDS-PAGE gel.

a)



b)



Conclusions

Structural techniques have been incredibly insightful in determining the molecular mechanisms of anti-σ factors (Fig 1.12). The group 1 σ factors are well studied both structurally and biochemically (Figs 1.3, 1.4, 1.7 and 1.12). The residues that interact with DNA and RNAP are known (Fig 1.4), and anti-σ factors typically block or disrupt these required σ functions.

The co-crystal structures I solved of PhERI bound to σ^{A}_{4} provide the molecular details of the interaction between the two proteins (Figs 2.6 and 2.9). The three structures are similar (Fig 2.8), and are in agreement with 2-hybrid and biochemical data on the interaction between the two proteins (Fig 2.11 and 2.12). I believe the structural analysis of PhERI binding to σ^{A}_{4} accurately describes the nature of the protein-protein interaction.

However, the mechanism through which PhERI inhibits Sau RNAP is not immediately apparent from the co-crystal structure. PhERI does not alter the conformation of σ^{A}_{4} nor does it appear to sterically block the residues required for DNA recognition or core RNAP binding (Fig 2.6 and 2.10). Previous work, suggesting that PhERI disrupts -35 element recognition by RNAP (Dehbi et al., 2008), is not supported by our structural data.

PhERI also does not appear to have any structural homology to previously studied anti-σ factors or any previously described protein fold. Because PhERI joins RNAP through an interaction with σ^{A}_{4}, and mutants in this region ablate PhERI binding to RNAP, PhERI is located well upstream of the transcription bubble, and more than 60Å from the RNAP active center. σ^{A}_{4} recognition of the -35 element occurs at an early stage in transcription initiation, when promoter DNA is fully double stranded. PhERI, because

it interacts with σ_4 was logically hypothesized to affect transcription by blocking σ^A_4 activity. Our structural data do not preclude PhERI acting by blocking DNA binding, or indeed being in a different conformation in the context of the full holoenzyme or open promoter complex than it is when bound to σ^A_4 alone as visualized in the structure. However, the structure does not support the previous hypothesis that PhERI blocks the -35 element DNA recognition (Dehbi et al., 2008). Biochemical and transcriptome data that I will present in chapters 3-5 conclusively show that PhERI indeed does not block -35 element recognition by RNAP holoenzyme.

The structural analysis described in this chapter provided valuable information detailing the interaction of PhERI with Sau RNAP. However, in this case, the molecular details of the interaction between PhERI and RNAP is not sufficient to propose a mechanism of inhibition for this protein factor. Subsequent chapters will describe the biochemical, genetic and genomic tools we used to characterize the activity of PhERI at Sau promoters.

Chapter 3:

PhERI Inhibits rRNA Transcription *in vitro*

The crystal structure of the PhERI / Sau σ^{A}_{4} complex, described in the previous chapter, provides a structural model for the interaction between PhERI and Sau RNAP. However, unlike previously studied anti-σ factors, PhERI does not appear to block the functional surfaces of σ^{A}_{4} (Figs 2.6 and 2.10). The previous hypothesis that PhERI inhibited -35 element recognition (Dehbi et al., 2008) is not supported by the co-crystal structure, but no alternative hypotheses are immediate apparent from the structural data alone.

RNAP activity has been studied using biochemical and genetic tools in model organisms for decades (Craig et al., 1998; deHaseth et al., 1998; Hinkle and Chamberlin, 1972a, b; Severinov and Darst, 1997; Siegele et al., 1988). Eco RNAP has been extensively studied at various host and phage promoters (Kumar et al., 1993; Lemke et al., 2011; Nechaev and Severinov, 1999; Pemberton et al., 2000; Saecker et al., 2002). The genetic tools available in Eco have made mutant polymerases readily available; many of these mutants are characterized for their ability to perform the various kinetic steps in the transcription cycle (Artsimovitch et al., 2003; Bartlett et al., 1998; Gardella et al., 1989; Severinov and Darst, 1997; Sun et al., 2004). The ability to use Eco RNAP to study PhERI function would allow us to rapidly test which kinetic steps (i.e. -35 recognition, open promoter complex stability, UP-element binding, closed to open complex isomerization) and which promoters (-35 element dependent versus -35 element

independent) are affected by this phage protein. The initial biochemical studies on PhERI used a hybrid RNAP holoenzyme comprising Eco core RNAP and Sau σ^A (Dehbi et al., 2008).

Transcription in Sau has not been extensively studied *in vitro* (Deora and Misra, 1996; Rao et al., 1995; Reynolds and Wigneshweraraj, 2011). While the Sau RNAP and σ^A have been identified and purified (Deora and Misra, 1996), few promoters have been tested and none have been studied in detail at various kinetic steps by footprinting. The most well studied promoters are involved in the switch from non-pathogenic to pathogenic growth, are characterized by weak -10/-35 elements, have non-optimal spacer length, and are known to be controlled by additional transcriptional activators and repressors (Agr P1 and P2, Fig 3.1a) (Deora and Misra, 1996; Rao et al., 1995; Reynolds and Wigneshweraraj, 2011).

In this chapter, I show that PhERI does not bind to or inhibit the previously studied Eco/Sau hybrid RNA holoenzyme. To study the mechanism of PhERI, I developed a fully native Sau *in vitro* transcription system and identified -10/-35 promoters in the Sau genome. I show that PhERI inhibits RNAP activity in a promoter-specific manner: PhERI is a potent inhibitor of transcription from the Sau rRNA promoters (rrnA and rrnB), but not all -10/-35 promoters *in vitro*, which is incompatible with a mechanism wherein PhERI blocks -35 recognition. Because the primary functional regions of $\sigma^A{}_4$ are not sterically blocked in the crystal structure of the PhERI / $\sigma^A{}_4$ complex and not all -10/-35 promoters are inhibited *in vitro*, PhERI must modulate a step other than -35 recognition or alter the interaction of RNAP with promoter DNA without directly affecting the activity of the σ factor. rRNA transcription is known to

affect the rate of cell division in prokaryotes; we hypothesize that PhERI inhibits Sau growth directly through inhibiting RNAP activity at the rrn promoters.

PhERI does not inhibit the hybrid Eco/Sau RNAP holoenzyme

I initially attempted to replicate the results of Dehbi *et al.* (2008) by testing PhERI's activity using Sau σ^A / Eco core RNAP on the well characterized -10/-35 λp_L promoter. Previous results showed that PhERI, but not a control phage protein, inhibits the hybrid RNAP at this promoter (Dehbi et al., 2008). In contrast, I show that PhERI, even at 100-fold molar excess, does not inhibit the hybrid RNAP holoenzyme at this same promoter (Fig 3.1b). Dehbi *et al.* only tested one PhERI concentration (2μM, with RNAP at 25nM), and showed only one time point after the initiation of transcription. Additionally, Dehbi *et al*. titrate Sau σ^A (100-500nM) above Eco core RNAP levels (25nM); while RNAP activity should saturate as 100% of core RNAP is bound by σ^A, transcription levels do not (Dehbi et al., 2008).

Due to these inconsistent results, I decided to compare the ability of PhERI to bind to Sau holoenzyme and the hybrid Eco/Sau holoenzyme by native gel shift (Fig 3.1c). While PhERI clearly shifts Sau holoenzyme (Fig 3.1c lanes 2 and 3), I see no evidence for a shift when the hybrid Eco / Sau enzyme is used (Fig 3.1c lane 6 and lane 4). Because I show that PhERI does not inhibit the hybrid Eco / Sau RNAP holoenzmye (Fig 3.1b), or even appear to interact with the hybrid holo (Fig. 3.1c), and because the biochemical results reported by Dehbi *et al.* were experimentally flawed, I decided to test the function of PhERI in a native Sau *in vitro* transcription system.

Figure 3.1 (adjacent page): PhERI does not inhibit the Sau σ^A/Eco core hybrid holoenzyme on the λpL promoter. a) Sequence of the λpL promoter. The -10 and -35 element are highlighted in red. b) PhERI does not inhibit the hybrid Eco core/Sau σ^A holoenzyme at the λpL promoter. PhERI, at 1μM or 5μM, was bound to Sau σ^A (100nM) and Eco core RNAP (50nM). Promoter DNA (50nM) was added and the reaction was incubated in 1x transcription buffer at 37°C. Reactions were initiated with NTPs (200μM GTP,CTP, and UTP, 50μM ATP and 0.1μl P32 ATP), stopped with 2x stop buffer at the times indicated and visualized on a 12% Urea-PAGE gel by autoradiography. c) PhERI forms a stable interaction with the Sau holoenzyme (lane 2 and lane 3), but not with the Eco core/Sau σ^A hybrid holoenzyme (lane 5 and lane 6). The proteins, at the indicated concentration, were incubated on ice for 10 minutes and run on a 4-12% native Phast gel.

a)

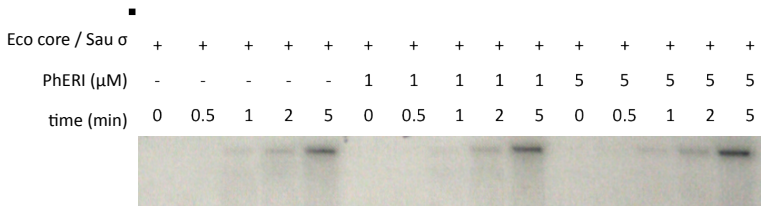
λ pL taaattatctctggcggtgTTGACAtaaataccactggcggtGATACTgagcacAtca

Agr P3 acgactagttaagaaaaaTTGGAAaataaatgcttttagcatgtTTTAATataactAg

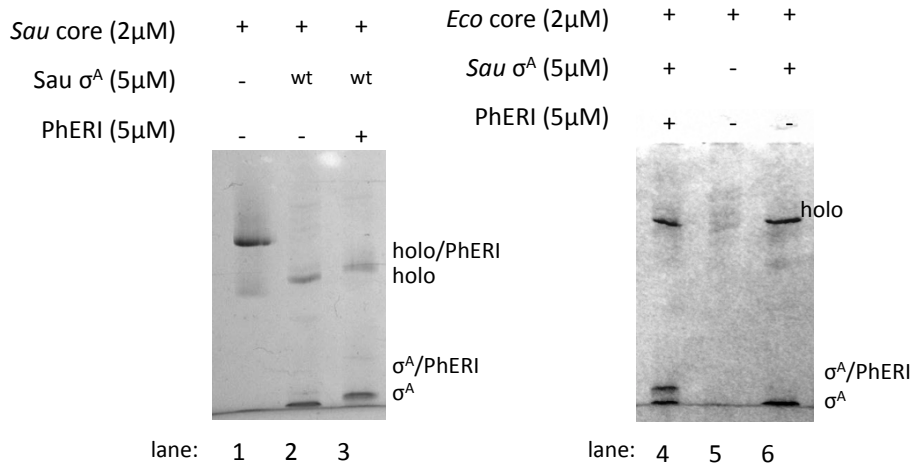
Agr P2 acattaggcaatataATGATAaaagattgtactaaatcgTATAATgacagtgAg

λpL promoter

b)



c)



Sau RNAP has been purified for biochemical analysis by native protein purification (Deora and Misra, 1996; Rao et al., 1995), overexpression of tagged subunits (Dehbi et al., 2008) and affinity purification using RNAP antibiodies (Reynolds and Wigneshweraraj, 2011). Using overexpressed subunits leads to an excess of the tagged domain of RNAP, and non-stoichiometric RNAPs have been shown to produce erroneous biochemical results (Vrentas et al., 2008). I therefore purifed native, untagged Sau RNAP from NCTC8325 cells, essentially as described previously (Fig 3.2; see Materials and Methods) (Deora and Misra, 1996).

I initially sought to replicate the *in vivo* result that PhERI inhibits RNAP activity (Dehbi et al., 2008). To this end, I used Sau genomic DNA as a template for transcription *in vitro* to test PhERI at all Sau promoters, similar to the experiment showing inhibitory activity *in vivo*. Standard transcription assays were performed, radiolabeling the RNA product with P^{32}, with genomic DNA (gDNA) as the template. To show the validity of this approach, I tested the inhibitory activity of the well-characterized anti-σ factor AsiA using Eco RNAP holoenzyme and Eco gDNA as the template. As expected, AsiA was a potent inhibitor of RNAP activity in this assay (Fig 3.3a). Because Sau RNAP was active in only in low concentrations of NaCl, I tested different buffer and salt conditions to optimize the transcription assay for Sau RNAP (Fig 3.3b). For subsequent experiments I used 0.1M NaCl, which was well tolerated by Sau RNAP, and proteins were stored in 0.2M sodium glutamate to avoid adding additional NaCl to the reactions with the protein buffers.

PhERI inhibits promoter-specific Sau RNAP-holoenzyme transcription from Sau gDNA (Fig. 3.3c, red bars), but not Eco RNAP-holoenzyme (Fig. 3.3c, yellow bars) nor

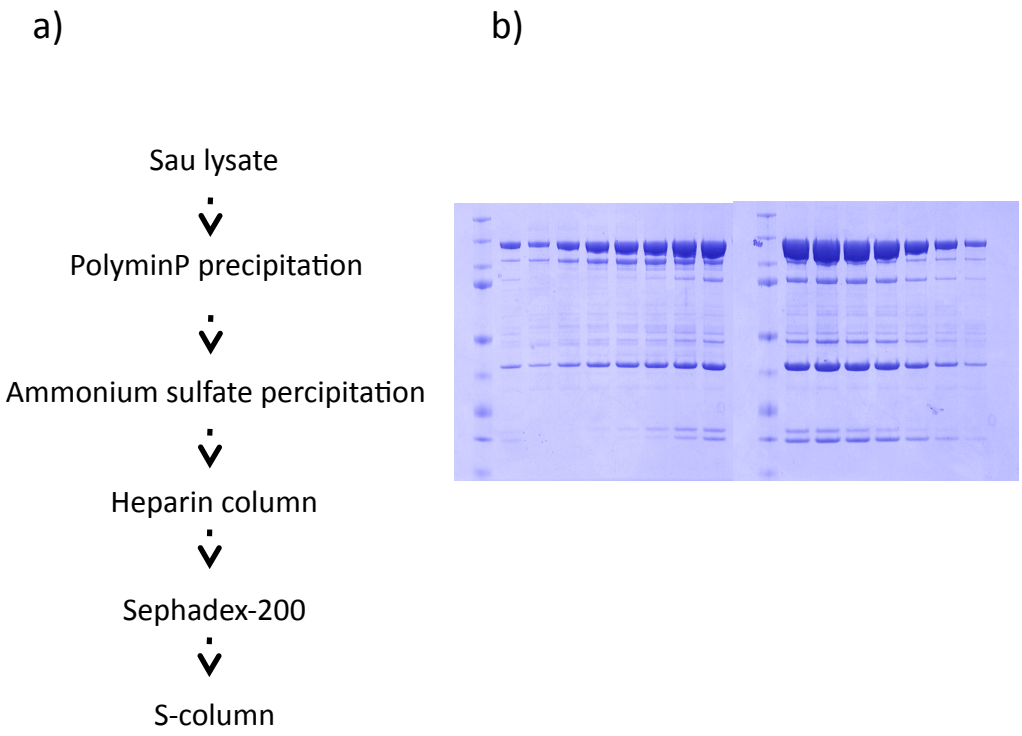
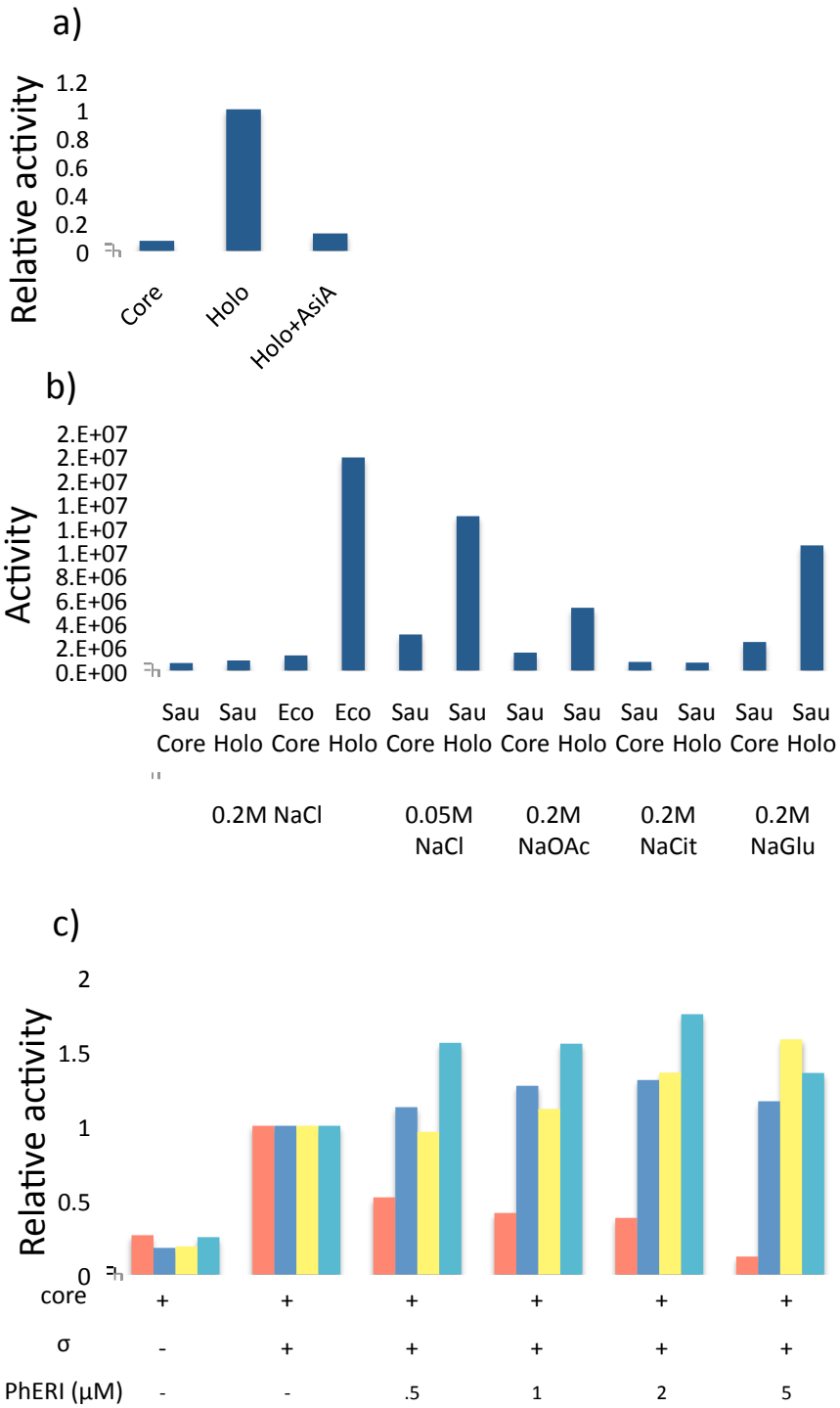


Figure 3.2: Purification of Sau core RNAP. a) Schematic of native, untagged Sau RNAP purification. b) 4-12% SDS-PAGE gel of Sau RNAP after the final purification step. Individual subunits are labeled.

Figure 3.3 (adjacent page): PhERI inhibits transcription from Sau RNAP but not Eco RNAP. a) The T4 anti-phage protein AsiA inhibits Eco RNAP holoenzyme activity using Eco gDNA as a template. b) High concentrations of NaCl inhibit Sau RNAP. c) PhERI inhibits transcription from Sau genomic DNA. Reactions contained σ^A (100nM), PhERI at the indicated concentrations and core RNAP (50 nM), 50ng gDNA in 1x Sau transcription buffer. Reactions were initiated with 200μM GTP,CTP, and UTP, 50μM ATP and 0.1μl P32 ATP, stopped after 5 minutes, pipetted onto DE81 paper, washed and quantified. Reactions were performed with Sau holoenzyme (red bars), Eco RNAP core complexed with Sau σ^A (blue bars), Eco holoenzyme (yellow bars) and Sau holoenzyme containing the σ^A mutant resistant to PhERI binding (green bars).


a)
Relative activity
1.2
1
0.8
0.6
0.4
0.2
0
Core
Holo
Holo+AsiA
b)
Activity
2.E+07
2.E+07
2.E+07
1.E+07
1.E+07
1.E+07
8.E+06
6.E+06
4.E+06
2.E+06
0.E+00
Sau Core
Sau Holo
Eco Core
Eco Holo
Sau Core
Sau Holo
Sau Core
Sau Holo
Sau Core
Sau Holo
Sau Core
Sau Holo
0.2M NaCl
0.05M NaCl
0.2M NaOAc
0.2M NaCit
0.2M NaGlu
c)
Relative activity
2
1.5
1
0.5
0
core
σ
PhERI (μM)
+ + + + + +
- + + + + +
- - .5 1 2 5

the hybrid holoenzyme formed by Eco RNAP/ Sau σ^A (Fig. 3.3c, blue bars), consistent with our in vitro results (Fig. 3.1b). Inhibition of Sau RNAP-holoenzyme by PhERI was dependent on the Sau σ^A / PhERI interaction, since RNAP containing the mutant Sau σ^A that abolished the PhERI interaction (but was normal in other respects, Fig. 2.11 and 2.12) was not inhibited (Fig. 3.3c, green bars). PhERI inhibition therefore requires a native Sau in vitro transcription system.

PhERI does not inhibit Sau -10/-35 promoters

While it is reassuring that PhERI has the same effect *in vitro* as *in vivo*, to study its mechanism we must identify individual promoters that are inhibited by this protein. Transcription from gDNA will not allow mechanistic assays as this experiment tests an ensemble of promoters with different binding affinities for RNAP and different kinetics of promoter recognition and transcriptional initiation.

Many early mechanistic studies of RNAP from Eco used phage promoters as template DNA (Letalaer and Jeanteur, 1971; Roberts, 1969; Rosenberg et al., 1982; Severinov and Goldfarb, 1994; Siebenlist, 1979; Siebenlist and Gilbert, 1980). In many phage, early phage promoters bind RNAP with high affinity and show robust activity to compete with host promoters for RNAP occupancy early in the phage infection. The phage T7 promoter T7A1 and the λ phage promoters pL and pR have been particularly useful for *in vitro* studies because of their stable intermediates and high levels of RNAP activity (Kadesch et al., 1982; Rosenberg et al., 1982).

Recent work sequenced the genome of many Sau phages (Kwan et al., 2005). In particular, I examined the genome of the Sau phage G1, which encodes PhERI. Due to

the short intergenic regions and the similarity of the phage promoters to the consensus sequence, I easily identified many strong phage promoters (Fig 3.4). The G1 promoters identified are all near to consensus and have ideal spacer length (17bp); most additionally have an extended -10 element and an A/T rich sequence upstream of the -35 element. These strong promoters are likely to be selected to compete for Sau RNAP upon the initial injection of the phage genome into the host cell. The coding region for PhERI is located downstream of one of these strong promoters (G1-pPhERI, Fig 3.5a). Therefore PhERI is almost certainly an early phage gene, transcribed by Sau RNAP upon initial phage infection.

To examine PhERI's function, I tested its activity at the phage promoter that drives its expression, G1pPhERI (Fig 3.5b). While RNAP has robust activity at this promoter, transcription is not affected by the presence of PhERI (Fig 3.5b). If PhERI is altering the interaction between σ^{A}_{4} and the -35 element, the phage promoters may not be informative as the majority have an extended -10 element and may therefore not require σ recognition of the -35 element.

To find -10/-35 promoters that may be affected by PhERI, I searched the Sau genome. While Sau virulence related promoters have been studied previously (Reynolds and Wigneshweraraj, 2011), I decided to search for promoters with more characteristic promoter elements that may be constitutively active in rapidly dividing cells. Early phage gene products often inhibit the host DNA replication (Datta et al., 2005; Yano and Rothman-Denes, 2011); inhibiting replication not only stops the cells from dividing but also allows the resources normally required for host replication to be allocated to the

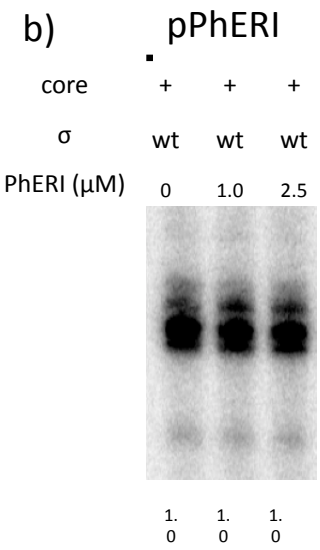
Figure 3.4: Promoters identified in the G1 Phage genome. ORF number is listed on the right and the putative promoter is listed. -35, -10 and extended -10 elements are highlighted in red.

Figure 3.5 (adjacent page): PhERI does not inhibit Sau RNAP at -10/-35 promoters for DNA replication and repair factors. a) Sequence of promoters used in subsequent experiments. -35 and -10 elements are highlighted in red. b) PhERI does not inhibit transcription from its own promoter. PhERI was incubated with σ^A (100nm), Sau core RNAP (50nM) and linear promoter DNA (50μM) in 1x Sau transcription buffer. Reactions were initiated by the addition of NTPs as described above, stopped with 2x formamide buffer, and visualized on a 12% Urea-PAGE gel by autoradiography. Bands were quantified in ImageQuant and the percent activity, relative to holoenzyme, is listed below each lane. c) PhERI does not inhibit RNAP activity at the dnaA, aag, and polIII promoters. Transcription assays were performed as in b.

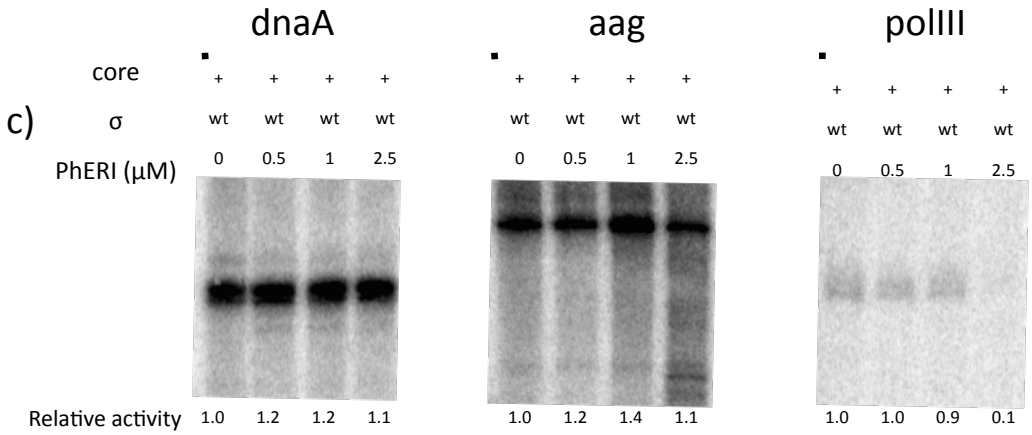
a)

dnaA	tttttagcaacatattcacaggtatTTGACAtatagagaactgaaaaagTATAATtgtgtgg
aag	agtacacatctatatggagactcatTTGAAAgtcaacgcttcgttaacTATACTaaaaatAt
polIII	gaacatttttattaattgttcaaTTAAGAagtaaaggtattatcaTGcTATAATgagaggt
G1-pPhERI	agtttaatttttaaaatataccacTTGACAttttatatgttaggTGgTATAATtattttAt

b)



c)



transcription and replication of the phage genome. I therefore looked for promoters upstream of genes required for DNA replication.

Identifying promoters in sequenced genomes is not a straightforward computational problem. Many promoters are highly divergent from the -10/-35 consensus sequences. Furthermore, Sau is A/T-rich, which creates many false positives for the -10 element (sequence: TATAAT). To facilitate the search for promoter sequences, we built a perl-based script to identify putative promoters. The program uses the likelihood of base identity at promoter positions in previously described promoters to score every 6 bases in a particular sequence (Fig 3.6) (short sequences are quickly searched, but the program can handle an entire prokaryotic genome in a matter of minutes). The initial search can be for either the -10 or the -35 sequence, and the threshold can be set by the user. Once a 6-base sequence has a score above the threshold, the program moves 15 bases downstream of a -35 sequence or 15 bases upstream of a -10 sequence and searches for the other promoter element. Spacers between 15 and 20 bases are allowed, and potential promoter sequences are output to a logfile. This program was built by a talented programmer, Michael Mosley, in collaboration with the author. While the author defined the parameters of the search algorithm, the details of the programming are out of the area of his expertise.

Using a combination of the Perl program described above and manual searches in the genome, I identified putative promoters upstream of three components of the DNA replication/repair machinery: dnaA, DNA polymerase III, and aag, a DNA repair factor (fig 3.5a). While Sau RNAP shows robust activity at the dnaA and aag promoters, PhERI

a)

A	0	0.10344828	0.03448276	**0.4568966**	0.36206897	**0.4396552**
C	0.15517241	0.12068966	0.12068966	0.27586207	**0.3706897**	0.0862069
G	0.18965517	0.04310345	**0.5172414**	0.14655172	0	0.28448276
T	**0.6551724**	**0.7327586**	0.32758621	0.12068966	0.26724138	0.18965517
	T	T	G	A	C	A
A	0.00862069	**1**	0.30172414	**0.4482759**	**0.3793103**	0.06034483
C	0.12931034	0	0.11206897	0.20689655	0.24137931	0.02586207
G	0.02586207	0	0.14655172	0.10344828	0.22413793	0
T	**0.8362069**	0	**0.4396552**	0.24137931	0.15517241	**0.9137931**
	T	A	T	A	A	T

b)

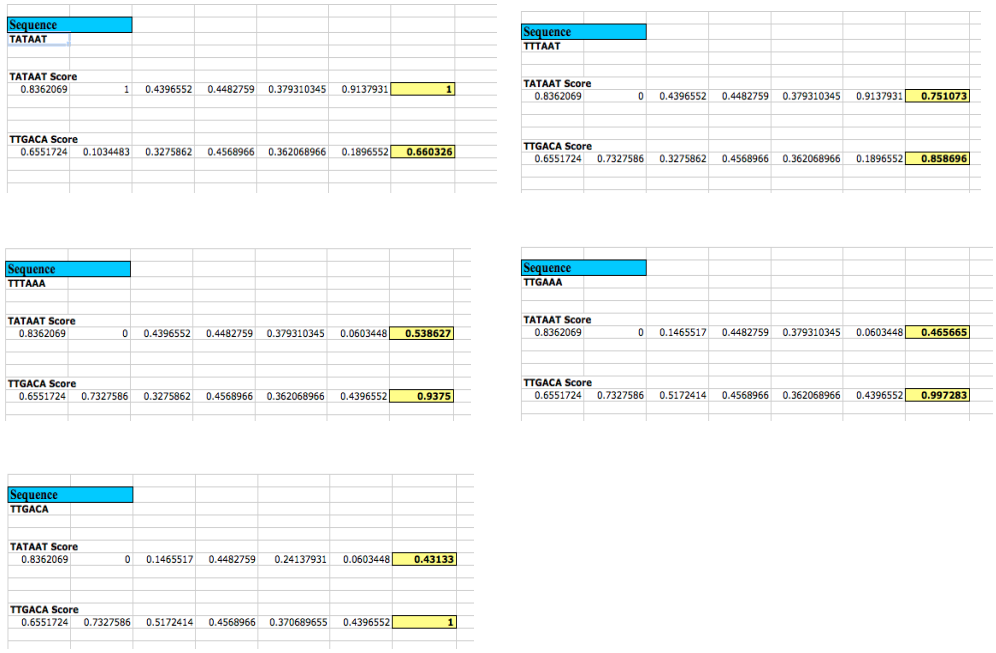


Figure 3.6: Bioinformatic tool used to identify putative promoters. a) Promoters were scored using an algorithm that used the likelihood of nucleotides at each promoter position. b) Example of promoter scoring. A consensus -10 element (TATAAT) sequence (1) was changed, base by base (2-4) to a consensus -35 sequence (TTGACA).

does not affect transcription (Fig 3.5c). PhERI has a small effect at very high concentrations on the already weak polIII promoter.

PhERI does not inhibit the G1 phage promoter pPhERI or the dnaA, aag or polIII promoters. Because these promoters, particularly the dnaA and aag promoters, almost certainly require an interaction with the -35 element (i.e. they have no extended -10 element), PhERI is unlikely to block the interaction between σ^{A}_{4} and the -35 element. This is in agreement with the structural analysis presented in the previous chapter.

PhERI inhibits RNAP at rRNA promoters

While PhERI appears to have no effect on the individual promoters I tested, it does inhibit RNAP activity *in vitro* when genomic DNA is used as the template (Fig 3.3c). The majority of transcription in log-growing prokaryotic cells is from the rRNA promoters (Gourse et al., 1996). We therefore hypothesized the PhERI may be inhibiting transcription from these promoters. The rRNA promoters in Sau have not been identified or studied previously, but the sequenced genome contains 5 rRNA operons, each containing the genes for the 16s, 23s and 5s rRNAs (Fig 3.7a). I searched for the promoters that would drive the expression of these operons. Two rRNA operons contained at least one easily identifiable promoter sequence within 200 bases of the sequence of the first structured RNA (Fig 3.7b).

Sau RNAP is active at the putative rrn promoters *in vitro* (Fig 3.7c), and furthermore PhERI inhibits this activity (Fig 3.7 c). While PhERI has no activity, even at high molar excess, at previously studied promoters, at the rrn promoters roughly

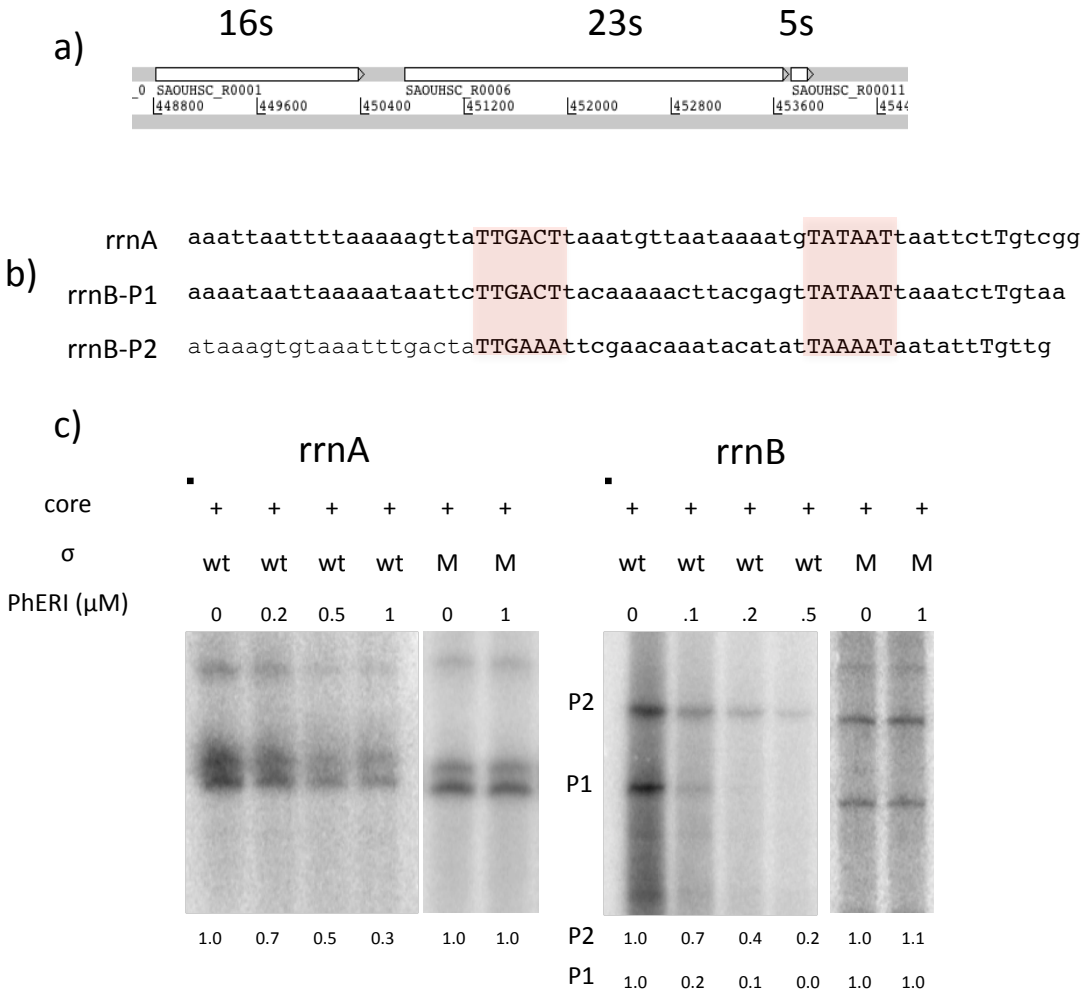


Figure 3.7: PhERI inhibits Sau RNAP at the rrn promoters. a) Schematic of the rrnA operon highlighting the 16s, 23s and 5s rRNA genes. b) Sequence of the rrn promoters tested in subsequent experiments. -35 and -10 elements are highlighted in red. c). PhERI inhibits Sau RNAP at the rrn promoters. PhERI (at the listed concentration) was incubated with σ^A (100nm), Sau core RNAP (50nM) and linear promoter DNA (50nM) in 1x Sau transcription buffer. Reactions were initiated by the addition of NTPs as described above, stopped with 2x formamide buffer, and visualized on a 12% Urea-PAGE gel by autoradiography. Bands were quantified in ImageQuant and the percent activity, relative to holoenzyme, is listed below each lane.

equimolar concentration of PhERI is sufficient to produce strong inhibition (0.1μM PhERI and 0.1μM σ^A).

These results demonstrate that PhERI is not a general inhibitor of RNAP activity but rather is a specific inhibitor only at certain promoters. Because PhERI does not inhibit RNAP activity at several -10/-35 promoters, it is unlikely to block the activities of σ^A_4 (DNA and core binding) required for RNAP activity most promoters. Therefore, PhERI must be modulating some other parameter specific to rrn and perhaps other Sau promoters. In subsequent chapters I will show that these results are reproducible *in vivo* and search for additional Sau promoters inhibited by PhERI. These studies will lead to a detailed model of PhERI's function at the molecular level.

Chapter 4:

RNA-seq Reveals PhERI Sensitive Promoters *in vivo*

The structural and biochemical studies of PhERI, and its activity at various promoters, have demonstrated that it is unlikely to act as a general anti-σ factor. While PhERI interacts with σ^{A}_{4} and forms a stable complex with the Sau RNAP holoenzyme (Fig 2.12), it fails to inhibit most promoters *in vitro* (Fig 3.5). Nonetheless, PhERI is a potent inhibitor of RNAP activity at the rrn promoters (Fig 3.7), which transcribe the structural rRNAs required for ribosomal biogenesis. How PhERI is able to target specific promoters, while having no apparent effect on others, is unclear.

To address PhERI's mechanism of inhibition at rrn promoters I used a genomic approach to simultaneously test the activity of PhERI on a large number and variety of Sau promoters. While *in vitro* analysis allows us to probe a small number of promoters in great detail, having a large number of PhERI-sensitive promoters may allow us to compare sequence features common to these promoters and make hypotheses about how PhERI inhibits RNAP activity. Studies examining the differential expression of genes by transcription factors and small molecules in Sau have successfully used gene-chip technologies to find differentially regulated genes (Hubscher et al., 2007; O'Neill et al., 2009; Truong-Bolduc et al., 2011).

In this chapter, I will describe my *in vivo* studies of the differential expression of genes by PhERI. I designed a PhERI expression vector that allows the inducible

expression of PhERI in Sau cells. I first show that PhERI inhibits rRNA synthesis *in vivo*, confirming our *in vitro* results. I will then describe a high-throughput RNA-seq analysis of cells expressing PhERI. These studies confirm that PhERI is not a general inhibitor of RNAP activity, and provide a library of PhERI-sensitive promoters that will be used in the next chapter to elucidate the mechanism of PhERI inhibition. I will also use the RNA-seq data to evaluate previous *in vitro* transcription experiments in Sau, determine robustly expressed mRNAs *in vivo* whose promoters may be useful for *in vitro* analyses, and compare gene expression profiles between Sau strains.

Construction of an Sau PhERI inducible vector

Sau has a well-developed genetic system including many vectors for expressing exogenous proteins (Corrigan and Foster, 2009; Dehbi et al., 2008; Truong-Bolduc et al., 2011). There is an electrocompetent Sau strain that is readily transformable. This strain, RN4220, is quite similar to the standard, non-pathogenic lab strain NCTC8325. The genomes of both strains have been sequenced and differ only by 121 single nucleotide polymorphisms (SNPs) and several small-scale insertions and deletions (indels) (Nair et al., 2011). The indels are clustered around phage-based transposable elements. Although some of the SNPs are in potentially important coding regions, including virulence and DNA repair factors, functional differences between the two strains are largely unreported (Nair et al., 2011).

PhERI was discovered in a high-throughput screen for proteins that inhibited Sau cell growth (Dehbi et al., 2008; Kwan et al., 2005; Liu et al., 2004). Because PhERI expression is known to be toxic in Sau cells (Liu et al., 2004), I opted to use an inducible

expression vector that has low levels of leaky expression (Corrigan and Foster, 2009). Many Sau expression plasmids drive healthy expression of the protein of interest but have constant low-level expression in the absence of inducer, which would greatly complicate our analysis. A recently described vector showed essentially no leaky expression of the gene of interest; further specificity was added by inducing with anhydrotetracycline, which has a reduced affinity for the bacterial ribosome and an increased ability to drive transcription from tet-inducible vectors (Fig 4.1) (Corrigan and Foster, 2009). I cloned both the phage G1 PhERI and its homolog from phage Twort, ORF65, into pRMC2 with Sau optimized Shine-dalgarno sequences upstream of the start codon. pRMC2, pRMC2-PhERI and pRMC2-TORF65 were electroporated into RN4220 cells and grown on plates containing chloramphenicol. In the absence of inducer, all strains grew with a doubling time comparable to that of the wild type (wt) RN4220 and NCTC8325 strains (Fig 4.2b).

When grown in the presence of even low levels of inducer, pRMC2-PhERI and pRMC2-TwortORF65 cells exhibit no evidence of cell growth over the course of several hours (Fig 4b (blue curve)). The phage Twort homolog, ORF65, clearly has the same function *in vitro* as phage G1 PhERI, as would be expected from its high level of sequence conservation and predicted structural similarity (Fig 4.2a). When cells were allowed to begin the logarithmic growth phase before the addition of inducer, pRMC2-PhERI and pRMC2-TwortORF65 containing cells exhibited an incomplete arrest of cell growth (Fig 4.2 a (orange and purple curves respectively) and b (orange curve)). While doubling times were significantly longer, the cells continued to divide, albeit slowly, and I saw no evidence for cell lysis or death.

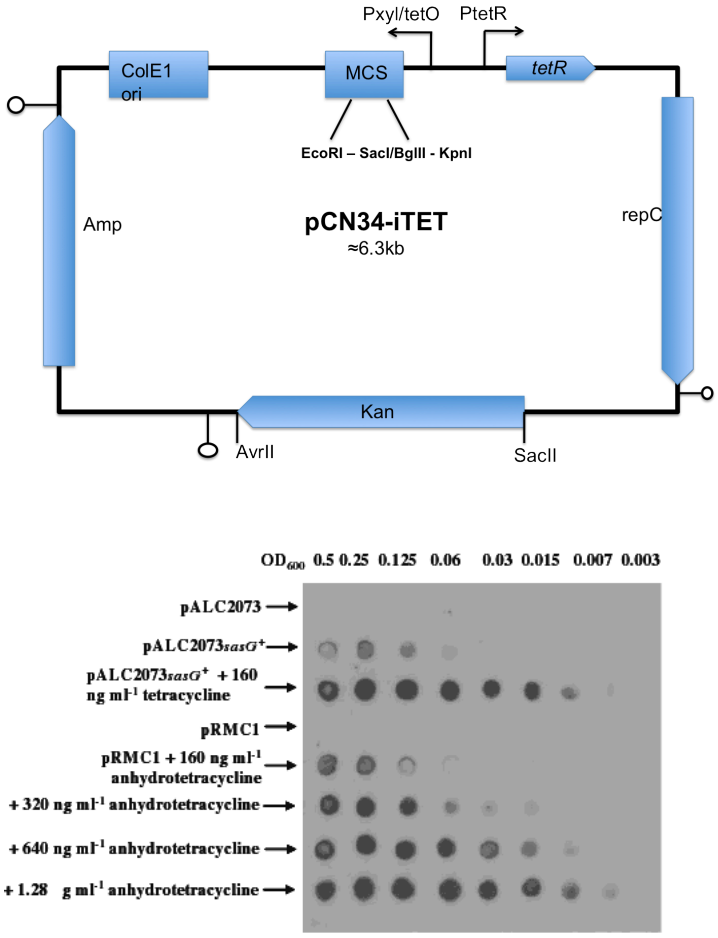
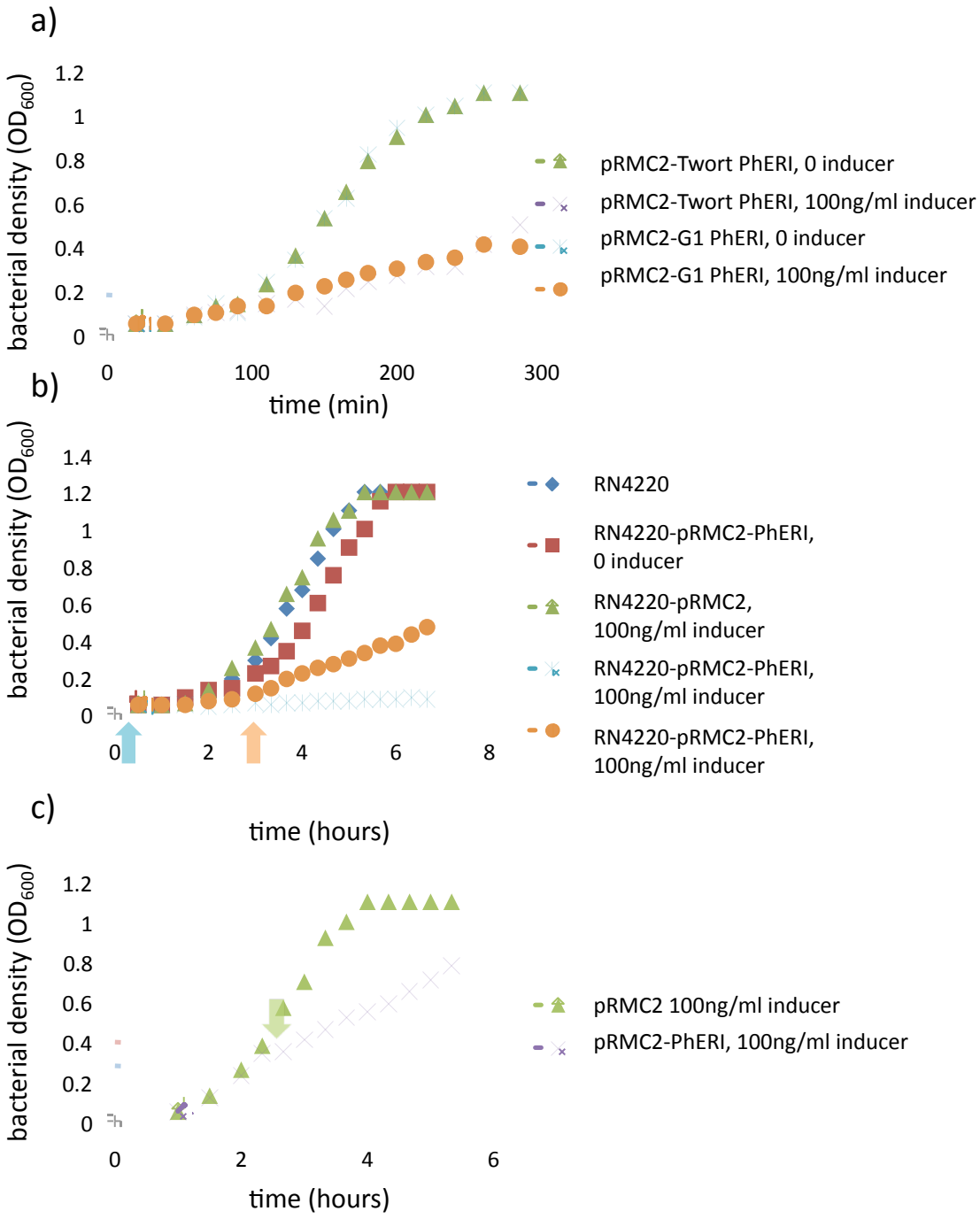


Figure 4.1: pRMC2 vector used for in vivo expression of PhERI.

a) Schematic view of pRMC2 showing the genes and cloning sites on the plasmid. b) pRMC2 has reduced leaky expression of the cloned gene. A previous Sau expression vector pALC2073, shows expression of the cloned protein. Lysates were applied to a membrane and an antibody specific to the cloned protein was used to detect protein expression. pRMC1, of which pRMC2 is a derivative, shows no evidence of leaky expression of the protein. Adapted from Corrigan and Foster, 2009.

Figure 4.2 (adjacent page). PhERI expression in Sau cells inhibits cell growth. a) PhERI and Twort ORF65 were cloned into the Sau expression vector pRMC2 and transformed into Sau RN4220 cells by electroporation. Cells containg pRMC2-PhERI or empty pRMC2 were grown in TS media containing chloramphenicol (25μg/ml). PhERI/ORF65 expression was induced by adding 100ng/ml anhydrotetracycline at the time indicated by the arrow and bacterial growth was monitored by the OD600. b) PhERI does not fully inhibit cell growth. When PhERI expression is induced at the beginning of the culture (as indicated by the blue arrow), no cell growth is evident. However, inducing PhERI expression after cells have entered the exponential growth phase (as indicated by the orange arrow and orange curve), cell growth is inhibited compared to normally growing cells (green, blue and red curves) but not completely. c) Cells as grown for the subsequent RNA-seq experiments. To purify cellular RNA for genomic analysis, RN4220 cells containing empty pRMC2 or pRMC2-PhERI were grown and inducer was added at an OD600 of 0.2 (red arrow). Cells were collected at OD600 of 0.4 for RNA purification (green and purple arrows).


a)
bacterial density (OD600)
1.2
1
0.8
0.6
0.4
0.2
0
0
100
200
300
time (min)
pRMC2-Twort PhERI, 0 inducer
pRMC2-Twort PhERI, 100ng/ml inducer
pRMC2-G1 PhERI, 0 inducer
pRMC2-G1 PhERI, 100ng/ml inducer
b)
bacterial density (OD600)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
8
time (hours)
RN4220
RN4220-pRMC2-PhERI, 0 inducer
RN4220-pRMC2, 100ng/ml inducer
RN4220-pRMC2-PhERI, 100ng/ml inducer
RN4220-pRMC2-PhERI, 100ng/ml inducer
c)
bacterial density (OD600)
1.2
1
0.8
0.6
0.4
0.2
0
0
2
4
6
time (hours)
pRMC2 100ng/ml inducer
pRMC2-PhERI, 100ng/ml inducer

PhERI inhibits rRNA synthesis *in vivo*

I first wanted to test whether PhERI is a direct inhibitor of RNAP activity at the rrn promoters *in vivo*. rRNAs are the most abundantly expressed RNAs in prokaryotic cells and are therefore excluded from both microarray and RNA-seq analyses. Because of their abundance, the 16s and 23s rRNAs can be easily visualized on a standard agarose or polyacrylamide gel stained with ethidium bromide (Fig 4.3a).

To test the impact of PhERI expression on rRNA abundance in cells, I added inducer to pRMC2-PhERI containing cells in early-growth phase (OD_{600} = 0.2) and took aliquots of cells at different times after induction. Cellular RNA was purified from $2x10^8$ cells using the RNA-easy kit (Qiagen) adapted for high-yield purification from Sau cells.

rRNA levels in Sau cells expressing PhERI do not significantly decrease even hours 4 hours after the addition of inducer (Fig 4.3a). While mRNAs are generally short-lived in prokaryotic cells, with a half-lives on the order of minutes, structured RNAs such as rRNAs are quite stable (Deutscher, 2003). The lack of depletion of rRNAs in pRMC2-PhERI cells could be because PhERI is not inhibiting rRNA synthesis *in vivo* or because rRNAs that have been produced prior to the expression of PhERI remain stable in Sau cells.

To differentiate between these possibilities, I used metabolic labeling to specifically visualize nascently transcribed RNA molecules. Adding radiolabeled inorganic phosphate to the growth medium causes radioactivity to be incorporated into all cellular nucleic acids within minutes. Because labeling is performed with P^{32}, radiolabeled transcripts can be separated by electrophoresis and visualized by standard techniques on a phosphoimaging screen (Wade et al., 1964).

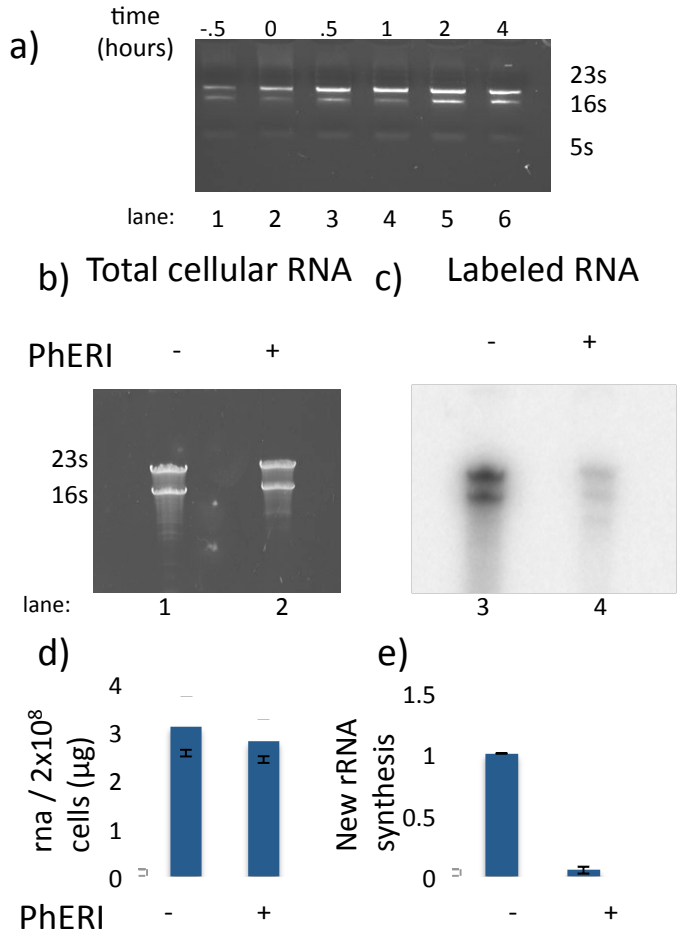


Figure 4.3: PhERI expression inhibits rRNA synthesis in vivo. a) PhERI expression does not reduce rRNA concentrations in Sau cells. Cells were collected at different time points before and after the induction of PhERI expression. b) PhERI inhibits new rRNA synthesis. Inducer was added to Sau RN4220 cells containing pRMC2 or pRMC2-PhERI, and new RNA synthesis labeled by subsequent the addition of P^{32} to the growth media. RNA was purified from $2x10^8$ pRMC2 and pRMC2-PhERI cells, run on a 6% Urea-PAGE gel, stained with GelRed to visualize all RNAs b) and exposed to a phosphoimaging cassette to visualize newly synthesized RNA c). d) Bulk RNA yield was quantified using a NanoDrop and newly synthesized RNA by quantifying the 16s and 23s rRNA bands on autoradiography images using ImageQuant.

I added inducer to pRMC2 and pRMC2-PhERI cells at early log-phase (OD = 0.2), allowed 20 minutes for PhERI to be expressed, and labeled nascently transcribed RNAs by addition of inorganic P^{32} to the growth media for 20 minutes. 2 x 10^8 cells were collected from the pRMC2 and pRMC2-PhERI strains and RNA was purified as described above. Total purified RNA was run on a 6% Urea-PAGE gel and visualized both by ethidium bromide staining (Fig 4.3b) and by autoradiography (Fig 4.3c). While the levels of RNA were unchanged between the cells expressing PhERI and those containing empty vector (Fig 4.3b), as visualized by the ethedium bromide staining and nano-drop quantification of total purified RNA (Fig 4.3d), nascent rRNA synthesis was significantly attenuated in the cells expressing PhERI (Fig 4.3c and e). This indicates that while rRNAs are not significantly depleted in cells expressing PhERI, PhERI does inhibit the synthesis of new rRNAs by Sau RNAP. This is consistent with the direct inhibition I showed at rrn promoters *in vitro.*

RNA-seq analysis in Sau

Gene chip analyses have been used extensively in Sau to test the differential expression of all genes in the genome under various conditions (Hubscher et al., 2007; O'Neill et al., 2009; Truong-Bolduc et al., 2011). While microarrays have become standard in Sau, I decided to use a relatively new technology: RNA-seq. The first RNA-seq analysis in Sau was published only after we began our analysis (Felden et al., 2011).

In both RNA-seq and genechip analyses, RNA is purified from cells and a cDNA library is prepared. Genechip technology tests the relative expression between the samples by annealing the library prepared from cells to DNA fragments immobilized on a

microchip. RNA-seq also requires a cDNA library. Once the library has been prepared, the samples are sequenced directly using Illumina technology. The output from RNA-seq technology is short sequences that can be mapped to unique locations in the genome. The number of sequences mapping to an individual location give the relative expression levels of that RNA in the original sample.

In addition to the relative expression levels between samples, which are provided in genechip analyses, RNA-seq provides additional layers of information. Regions with sufficient read coverage actually provide information about the genome sequence. SNPs can be mapped, both between samples and between the samples and a reference genome. As opposed to traditional genechip technologies, RNA-seq gives information for the expression levels over the entire length of the mRNA, including non-coding 5' and 3' UTRs. RNA-seq has been used, with additional preparation steps, to map promoter start sites (Mendoza-Vargas et al., 2009) and to locate RNAP binding sites on a genome-wide scale (Churchman and Weissman, 2011). RNA-seq technologies provide not only the ability to compare the expression of the same mRNA between two samples, but also information about the relative levels of expression of different mRNAs within samples. The added data provided by RNA-seq proved useful when identifying promoter sequences in the Sau genome and evaluating which promoters may be well suited for further study *in vitro*.

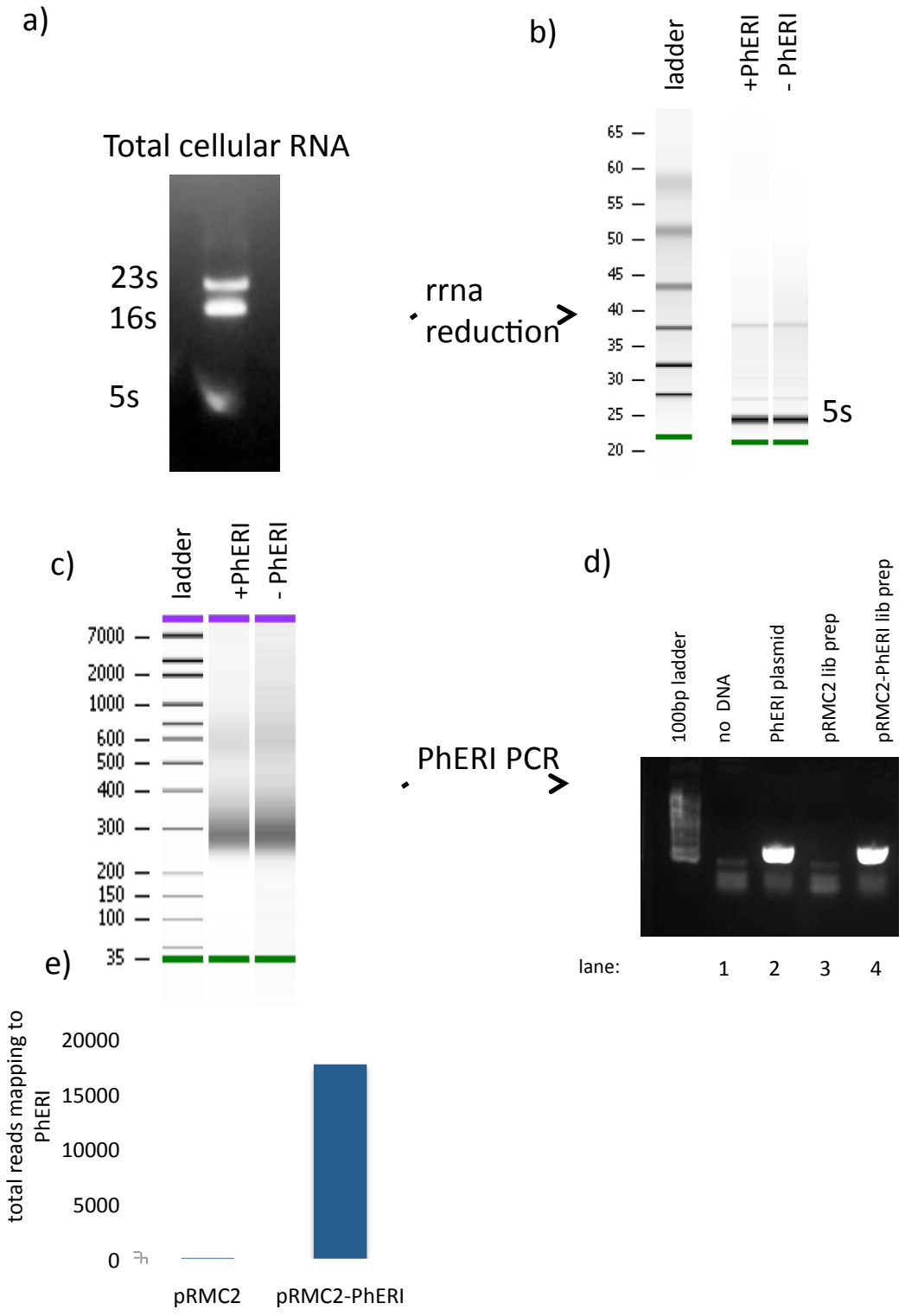
Sample preparation and validation for RNA-seq analysis

To prepare the samples for RNA-seq, I purified RNA from Sau cells, as described above. Cells were grown to early log phase (OD_{600} = 0.2) and inducer was added. Cells

were then collected at OD_{600} = 0.4 and RNAs were immediately stabilized by adding BioStabilize (Fig 4.2c) (Qiagen). Cells were then lysed as described above and RNAs purified by the RNeasy kit. Cellular RNA was checked for degradation by electrophoresis on an agarose gel, where the intact 23s and 16s rRNA bands can be visualized (Fig 4.4a). To prepare the samples for RNA-seq, rRNAs must be removed as they constitute the vast majority of the sample and can overwhelm the sequencing results. This is easily achieved in eukaryotic samples through an amplification step selecting for poly-A containing mRNAs. Prokaryotic mRNAs do not contain poly-A tails; therefore rRNAs must be physically removed from the sample. A recently developed kit uses immobilized oligos complimentary to the conserved 23s and 16s rRNA sequences to anneal to the large rRNAs in the sample and allow all other cellular RNAs, including the mRNAs, to flow through (Fig4.4b).

After this step, cellular RNA was examined on a bioanalyser chip (Fig 4.4b), which separates nucleic acids by size similar to standard gel electrophoresis, but allows the visualization of very small amounts of DNA or RNA. While it appeared that our rRNA depleted sample contained very little nucleic acid by nanodrop, there is clear evidence for a band on the bioanalyser chip corresponding to RNAs of roughly 200 bases (Fig 4.4b). The rRNA removal kit does not remove the 5s rRNA, tRNAs or other abundant, but small, cellular RNA species. This band at 200 bases likely corresponds to these cellular RNA species, while the mRNA in the sample correspond to the less abundant smear above this band. At this point, we decided to prepare a cDNA library using the standard random-primed PCR technique used for mRNA enriched eukaryotic samples (Fig 4.4c). The sample preparation, sequencing, sequence alignment and

Figure 4.4 (adjacent page): RNA purification and sequencing strategy. a) Total cellular RNA was purified from Sau cells expressing PhERI, and control cells containing empty vector. RNA quality was assessed on an agarose gel by visualizing the intact 16s and 23s rRNA bands. b) rRNA reduction was performed using standard procedure to remove the highly abundant 16s and 23s rRNAs prior to sequencing. RNA quality was assessed by running the samples on a BioAnalyzer. c) RNA-seq cDNA libraries were made by standard procedures and RNA quality was assessed by running the samples on a BioAnalyzer. d) cDNA libraries were analyzed for the presence of PhERI mRNA, which should only be present in cells containing pRMC2-PhERI. A PCR was performed using primers to amplify a 100bp fragment of PhERI from the RNA-seq library prepared from cells containing pRMC2 (lane 3) and pRMC2-PhERI (lane 4). e) Total PhERI sequencing reads that map to the gene for PhERI.


a)
Total cellular RNA
23s
16s
5s
rrna reduction
b)
ladder
+PhERI
- PhERI
65
60
55
50
45
40
35
30
25
20
5s
c)
ladder
+PhERI
- PhERI
7000
2000
1000
600
500
400
300
200
150
100
35
PhERI PCR
d)
100bp ladder
no DNA
PhERI plasmid
pRMC2 lib prep
pRMC2-PhERI lib prep
lane:
1
2
3
4
e)
total reads mapping to PhERI
20000
15000
10000
5000
0
pRMC2
pRMC2-PhERI

analysis were done in close collaboration with the Genomics Resource Center at The Rockefeller University. Connie Zhao and Scott Dewell were both essential at every stage of the RNA-seq sample preparation and data analysis.

Because standard cDNA library preparations are not generally performed with prokaryotic samples, I decided to check the mRNA levels in our cDNA library by non-quantitative PCR. The mRNA encoding for PhERI should only be present in the cells containing pRMC2-PhERI and absent in the cells containing the empty pRMC2 vector. I amplified an approximately 100bp fragment of PhERI from each prepared library, as well as from a prokaryotic expression plasmid containing the gene for PhERI. A strong band was amplified only from the library prepared from cells containing pRMC2-PhERI (Fig 4.4d lane 4), and not from control cells containing pRMC2 only (Fig 4.4d lane 3), confirming that the cDNA libraries do indeed contain cDNA corresponding to cellular mRNA and the samples differ predictably.

The two cDNA libraries were then sequenced directly by RNA-seq on an Illumina HiSeq2000 for 51 cycles using the standard protocols. Raw data was processed using SCS/RTA software to yield 51bp reads and subsequently aligned to the Sau NCTC 8325 genome using TopHat. We aligned to the NCTC 8325 genome, as opposed to using the genome for the strain we used, RN4220, because the quality of genome sequence is much higher, it differs from RN4220 in coding regions only by 121 SNPs, it is more fully annotated and is available in a downloadable format suitable for subsequent analyses. We also performed a search for reads corresponding to PhERI mRNA. Confirming the result from the PCR analysis of the cDNA libraries, only 47 total reads mapped to the PhERI mRNA in the cells containing pRMC2, while 17403 reads mapped to PhERI from

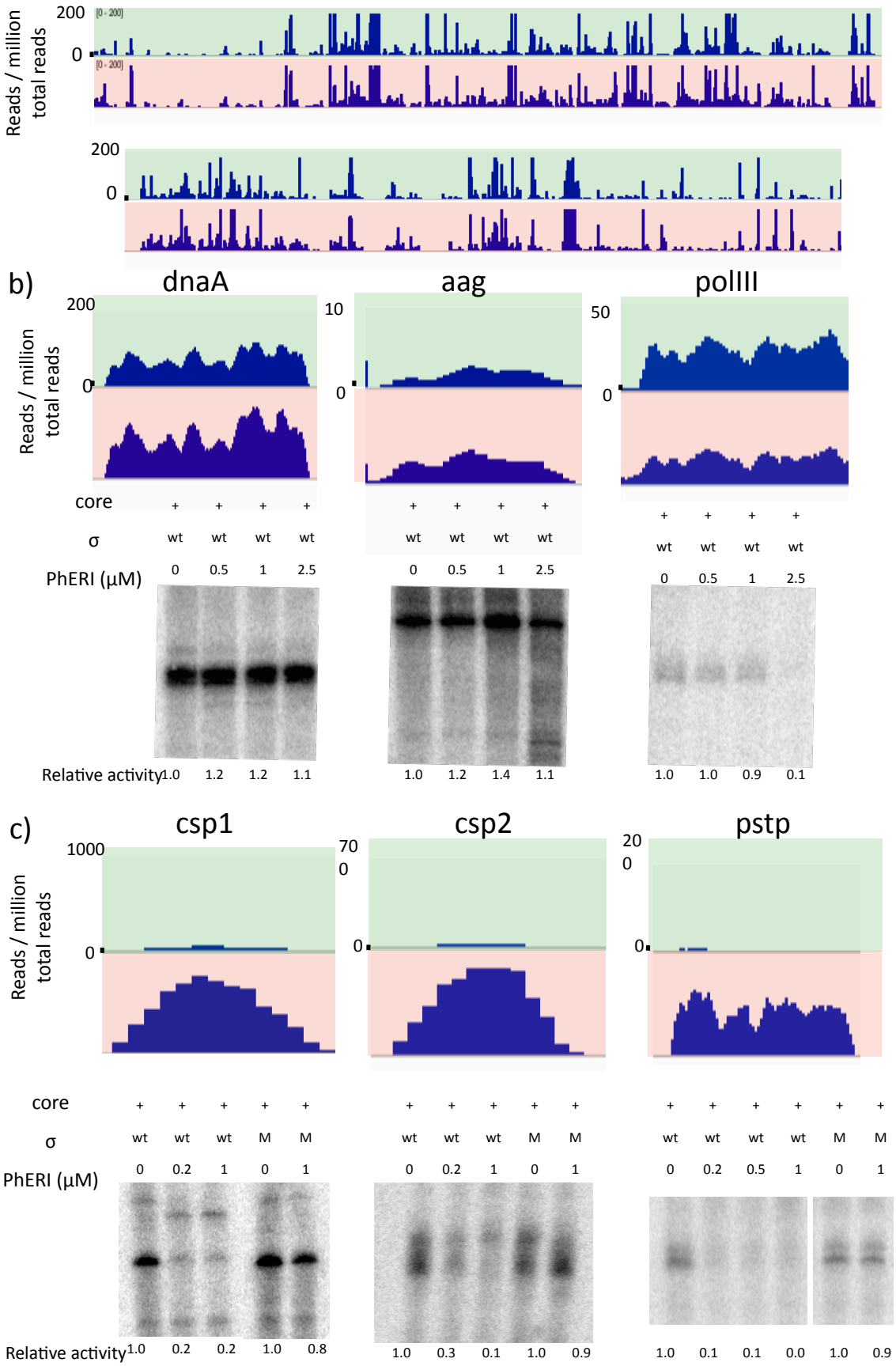
the pRMC2-PhERI library (Fig 4.5e). The RNA-seq data provided extremely good coverage, in part due to the relatively small size of the Sau genome compared to eukaryotic samples. Alignments reported from TopHat were processed by the Cufflinks software package (Trapnell et al., 2010) to determine differential expression of genes and transcripts between conditions.

RNA-seq analysis of differential gene expression by PhERI

An analysis of the differences in expression between control cells and cells expressing PhERI confirms that very few genes are expressed at significantly different levels between the two samples (Fig 4.5a). Importantly, the promoters I previously tested for PhERI activity *in vitro* (dnaA, aag, polIII) showed no significant differences between the two samples *in vivo* (Fig4.5b). The promoters that have been studied *in vitro* in previous work (Deora and Misra, 1996; Rao et al., 1995; Reynolds and Wigneshweraraj, 2011) also showed no difference in expression levels between cells expressing PhERI and control cells (Fig 4.6). Furthermore, these transcripts exhibited low levels of expression in log-growing cells, illustrating that while they may be important for the transcriptional switch to pathogenic growth, they are not useful tools to determine the mechanisms of transcriptional regulation under general growth conditions.

The Cufflinks software gives an output of all genes differentially expressed between samples. The statistical analysis evaluates expression levels and data quality at all loci to evaluate the probability of significant differential expression. In the analysis performed by Cufflinks, genes differentially regulated by 3 or more fold were found to be significant, when the data quality and number of reads were sufficient. By this analysis,

Figure 4.5 (adjacent page): RNA-seq reveals promoters sensitive to PhERI inhibition in vivo. a) PhERI expression in Sau RN4220 cells does not inhibit RNA levels from 95% of promoters. RNAs were sequenced directly and visualized by Integrated Genomics Viewer (IGV). Position on the genome is shown on the horizontal axis and the number of RNA reads per million total reads is shown on the vertical axis. Upper panel (green) represents RNA-seq data from pRMC2-PhERI cells and the lower panel (red) represents RNA-seq data from control pRMC2 cells. b) PhERI expression does not inhibit transcription from DNA replication promoters in vivo. RNA-seq data visualized as above from the dnaA, aag, and polIII loci. c) Genes negative regulated by PhERI *in vivo*. RNA-seq data visualized as above from the cold shock protein 1 (csp1), cold shock protein 2 (csp2), and pstp loci. PhERI inhibits transcription *in vitro* from promoters that are susceptible *in vivo*. In vitro transcription assays were performed as described above from purified, linear promoter DNA.


Reads / million total reads
200
0
200
0
b)
dnaA
aag
polIII
Reads / million total reads
200
0
10
0
50
0
core
σ
wt wt wt wt
PhERI (μM)
0 0.5 1 2.5
Relative activity
1.0 1.2 1.2 1.1
1.0 1.2 1.4 1.1
1.0 1.0 0.9 0.1
c)
csp1
csp2
pstp
Reads / million total reads
1000
0
70
0
20
0
core
σ
wt wt wt M M
PhERI (μM)
0 0.2 1 0 1
0 0.2 0.5 1 0 1
Relative activity
1.0 0.2 0.2 1.0 0.8
1.0 0.3 0.1 1.0 0.9
1.0 0.1 0.1 0.0 1.0 0.9

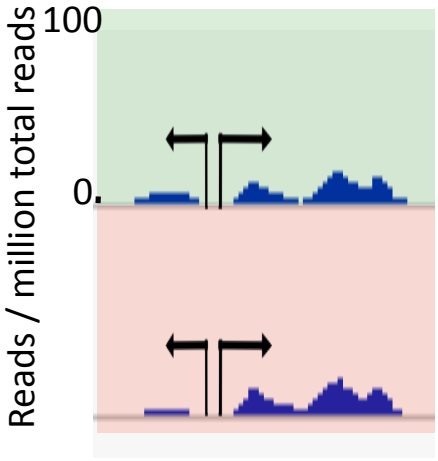


Figure 4.6: PhERI does not affect RNA levels at the agr promoters. RNA-seq data showing RNA levels at the agr operon in the absence (lower panel, red) or presence (upper panel, green) of PhERI.

fewer than 4% of all genes were repressed by PhERI expression (Table A.1a), while a further 5% were upregulated by PhERI (Table A.1b). More than 90% of transcripts were not differentially regulated by PhERI, confirming that most group 1 σ dependent promoters are not affected by PhERI. Again, this is consistent with a model in which PhERI does not disrupt essential functions of σ^{A}_{4}. In particular, this data is inconsistent with a model in which PhERI blocks -35 element recognition by σ^{A}_{4} because this would lead to repression of a large fraction of the -10/-35 promoters that do not contain other sequence elements (such as an extended -10 element).

The RNA-seq analysis only reveals RNA transcript levels, it does not differentiate between direct and indirect regulation of gene expression, nor does it reveal whether differential gene expression is due to changes in promoter binding and initiation or mRNA stability. To evaluate whether PhERI directly affects transcription at sensitive promoters identified using RNA-seq, I tested hits using the Sau in vitro transcription system.

mRNA processing enzymes can remove 5' and 3' UTRs from mRNAs in cells. Mapping promoter start sites requires enriching for primary transcripts that have not undergone processing in vivo (Mendoza-Vargas et al., 2009). For our analysis, I sequenced mRNAs from cells without subsequent enrichment for primary transcripts. Much of our RNA-seq data shows evidence of processing, with RNA-seq reads mapping to just upstream of the start codon for many predicted genes (Figure 4.7a). Generally, there is no putative promoter element immediately upstream of these transcripts (Figure 4.7a), arguing that the transcription start site is further upstream and the mRNA has been processed in vivo. However, some mRNAs in our data show clear evidence for a long 5'

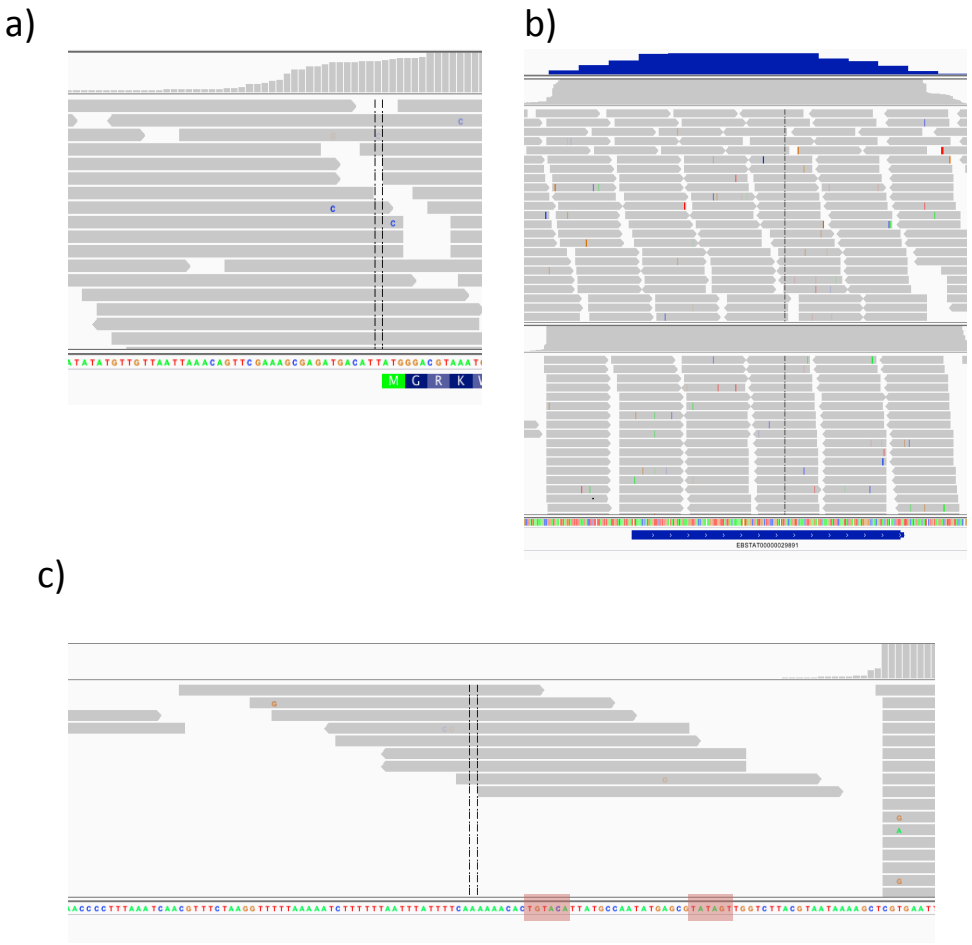


Figure 4.7: RNA-seq data aided in the identification of Sau promoters. While many genes had mRNAs with clear evidence of processing by cellular enzymes (RNA-seq reads beginning at or near the start codon, with no evidence of a promoter sequence upstream of the start codon), a subset of genes showed clear RNA-seq reads in a 5' UTR. Several of these genes contain a promoter sequence (-35 and -10 elements highlighted in red) upstream of the location where RNA reads begin mapping to the Sau genome.

UTR (Figure 4.7b). Additionally, many of these transcripts have strong putative promoters just upstream of the mapped 5' end of the mRNA (Figure 4.7c). This information is not generally provided by microarray analysis that contains only information about RNA expression in the coding region.

Likely candidate genes for PhERI inhibition assays *in vitro* had three characteristics: 1) High expression in the absence of PhERI (therefore likely downstream of a strong promoter); 2) High level of repression upon PhERI expression and 3) RNA-seq guided promoter identification, as described in the previous paragraph (Fig. 4.7).

Csp1, Csp2 and PSTP were identified as repressed *in vivo* by the expression of PhERI by the above analysis. When I tested these three promoters *in vitro*, there was clear evidence for inhibition by PhERI (Fig 4.5c). This inhibition was dependent on the interaction between PhERI and σ^{A}_{4} as holoenzyme containing the σ^{A} mutant deficient in PhERI binding was not affected. These data demonstrate that, at least at these 3 promoters, the depletion of cellular mRNA identified through the RNA-seq analysis is due to a direct inhibition of RNAP activity at the promoter. In addition to the rrn promoters shown in the previous chapter to be inhibited by PhERI, these promoters provide additional information to make hypotheses about the promoter elements that drive PhERI specificity.

PhERI inhibition of Sau cell growth

The RNA-seq data provide a comprehensive analysis of all transcripts regulated by the expression of PhERI. I purified RNA from cells expressing PhERI after the growth inhibition had begun and therefore the factor(s) leading to cell growth arrest

should be evident in our data. I show that PhERI inhibits rRNA synthesis *in vivo* (Fig 4.3) through direct regulation of RNAP activity at the rrn promoters (Fig 3.6). Inhibition of rRNA synthesis is a key regulatory step in the switch from logarithmic growth phase to stationary phase. rRNA synthesis therefore can directly regulate the rate at which prokaryotic cells divide (Gourse et al., 1996). Although PhERI's growth inhibition in cells is likely due to its activity at rrn promoters, it could be acting through the repression of gene expression of another required factor for cell division.

Recently published work used a combination of screening and genetic validation to identify a comprehensive list of required genes in Sau (Xu et al., 2010). The analysis was focused on identifying novel targets for small molecule inhibitors, and known drugs indeed do target many of the ORFs identified.

rRNA genes and other protein and RNA factors required for translation were found to be essential for normal cell growth in Sau (Xu et al., 2010), highlighting the important role of robust translation for logarithmic cell growth. 81 of the 308 (27%) genes found to be required for growth were related to translation.

I cross referenced all the genes shown to be downregulated by the expression of PhERI with the comprehensive list of all genes known to be required for cell growth in Sau. Only seven mRNAs downregulated by PhERI are essential, three of which (serS and pheS, tRNA synthetases, and rpsE, a ribosomal protein) are required for translation. Another recent paper showed that the regulation of RNAP activity at the promoters for ribosomal proteins is similar to that of the rrn promoters (Lemke et al., 2011) indicating that PhERI may act directly on these promoters by the same mechanism as at the rrn promoters.

Three additional downregulated genes are involved in metabolism and biosynthesis: glucose 6-phosphate isomerase (pgi), required for glycolysis; accB, required for fatty acid biosynthesis; and glutamine synthetase (glnA). The final gene is annotated as a putative mRNA degredation and processing factor, RNaseJ1/J2. Given that rRNA synthesis is known to directly regulate cell division in other organisms, that the inhibition of rRNA synthesis is known to be a direct effect of PhERI, and that only 4 required genes are downregulated by PhERI *in vivo* that do not directly affect translation, I conclude that PhERI blocks Sau cell division by inhibiting the rrn promoters directly.

Genes Upregulated by PhERI

I generally focused my analysis on genes found to be downregulated in PhERI expressing cells. PhERI was identified as a transcriptional inhibitor and my work focuses on defining the mechanism of that inhibition. However, in the high throughput *in vivo* analysis, more genes were stimulated by PhERI's presence than repressed. The upregulation of mRNAs by PhERI could be a direct effect on RNAP activity at promoters or an indirect effect of the repression of transcription at other promoters. In Eco, the downregulation of RNAP activity at rrn promoters frees a large concentration of RNAP; promoters, in particular the amino acid biosynthesis promoters, are directly sensitive to the concentration of free RNAP. Therefore, when the cellular concentration of RNAP increases due to decreased expression of rRNAs, transcription at these promoters is stimulated (Barker et al., 2001a).

Among the genes upregulated in the presence of PhERI are three phosphate transporters. My data from this chapter showing that PhERI is a direct repressor of rRNA

transcription *in vivo* relies on metabolic labeling. Radiolabled inorganic phosphate is added to the media, which cells import and incorporate into their nucleic acids. A decrease in phosphate transport, therefore, could explain the decrease in radioactive signal incorporated into the rRNAs in the cells expressing PhERI. However, the opposite is the case: phosphate transporters are upregulated in cells expressing PhERI and yet the radioactive signal indicating nascent rRNA transcription has decreased. This data supports the hypothesis that the decrease in signal for nascently transcribed rRNAs is due to a genuine and direct inhibition of RNAP by PhERI at the rrn promoters.

Inhibition of transcription at rRNA promoters is an important step in the stringent response as cells progress to stationary phase. Work in Eco shows that promoters regulating amino acid biosynthesis are directly regulated by the increased concentration of RNAP due to inhibition of rRNA synthesis. Among the genes upregulated by the expression of PhERI are several amino acid biosynthesis genes, which are indirectly upregulated by rRNA repression in Eco (Table A.1b). Recent work examined the global transcription changes in Sau induced by the small molecule Mupirocin, which induces the stringent response. In addition to the upregulation of amino acid synthesis related genes, genes for cellular transport processes were also upregulated (Reiss et al., 2012). I find evidence for upregulation of transport processes upon the induction of PhERI expression (Table A.1b). This indicates that PhERI inhibits rRNA synthesis *in vivo*, and therefore indirectly stimulates transcription at other promoters, and that regulation of the switch between log-growing cells and cells at stationary phase may have some conserved elements between Eco and Sau, including the indirect upregulation of amino acid biosynthesis genes.

Comparison of relative gene expression between Sau strains

Unlike microarray transcriptome analysis, RNA-seq provides information about relative expression of different genes throughout the genome. I used the RNA-seq data from all genes in the Sau genome to evaluate which genes are most highly expressed in log-growing cells. Recent work has examined the genomic differences between the commonly used, electroporatable Sau strain RN4220 (Nair et al., 2011) and its parent strain NCTC8325-4 (O'Neill, 2010). NCTC8325-4 differs from the fully sequenced NCTC8325 by the curing of 3 phage infections (O'Neill, 2010). To evaluate the transcriptional differences between NCTC8325-4 and RN4220, and to ensure that RN4220 carrying an empty expression vector was not misrepresentative of baseline transcription in NCTC8325-4, we sequenced RNA purified from NCT8325-4 cells containing no expression vector.

I evaluated the levels of gene expression in NCTC8325-4 and RN4220. Among the 100 most highly expressed genes, none differed significantly in expression levels between these two strains (Table A.2). Similarly, among the genes with no evidence for RNA-seq reads, none differed between the two.

Among the 100 most highly expressed genes in RN4220 and NCTC8325-4, the majority (62) were ribosomal proteins or proteins otherwise involved in translation (the 30 most abundant mRNAs in RN4220 and NCTC8325-4 are shown in Table A.2). This is in good agreement with the observation that log-growing prokaryotic cells expend most of their transcriptional resources on maintaining the translational machinery. Other highly expressed mRNAs corresponded to genes for gylcolysis and sugar metabolism

(12), fatty acid biosynthesis (6), chaperones (3), transcription/transcription regulation (3), and redox regulation (3). 9 of the 100 most abundant mRNAs were for conserved proteins of unknown function. The remaining 2 genes were a GTPase required for cell division and protein translocase.

An additional 212 genes, mostly of unknown function, had no evidence for RNA-seq reads in either NCTC8325-4 or RN4220. Whether any of these genes are upregulated as cells enter stationary phase, or respond to cellular stress signals, is unknown but could potentially be evaluated by sequencing RNA from cells under various conditions.

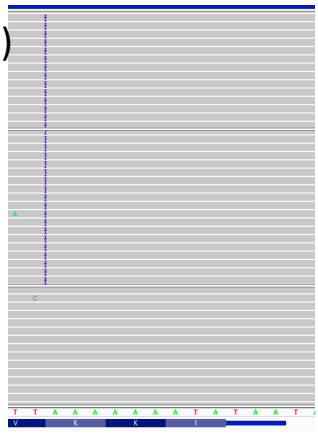
Analysis of Single Nucleotide Polymorphisms between RN4220 and NCTC8325-4

The genome of RN4220 was recently sequenced (Nair et al., 2011). In the genome sequence of RN4220, SNPs were identified that differ from NCTC8325 and NCTC8325-4. The authors suggested that RN4220 may be deficient in factors required for normal cellular responses to stress and virulence regulation (Nair et al., 2011). Additional work characterized SNPs in NCTC8325-4 relative to NCTC8325 (O'Neill et al., 2009). Through our RNA-seq analysis, we can identify SNPs in both the NCTC8325-4 and the RN4220 transcriptome, and map these SNPs to the NCTC8325 genome.

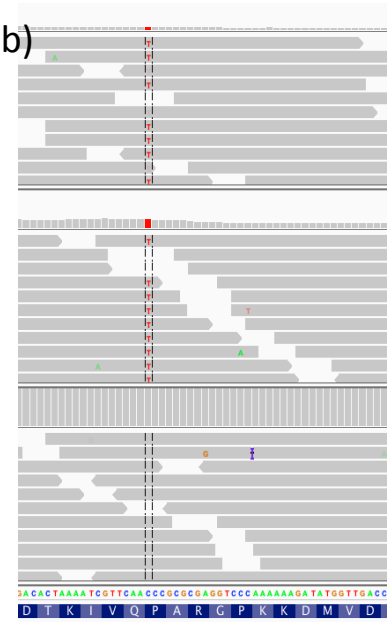
NTCT8325-4, as analyzed by O'Neill (2010), was found to differ from the NCTC8325 genome at 12 locations, and RN4220 had 121 SNPs, including those previously identified in NTCT8325-4. However, there is clear evidence for SNPs identified as unique to RN4220 in NCTC8325-4 (Figure 4.8 c-e). Importantly, SNPs known to be unique to RN4220 and to cause functional differences between these two strains, such as the frame shift in the virulence transcriptional regulator AgrA (Figure

Figure 4.8 (adjacent page): Single nucleotide polymorphisms (SNPs) between RN4220 and NCTC8325-4. a) A previously identified single nucleotide insertion at the c-terminus of the AgrA gene causes a frame shift mutation. The blue bars indicate the single nucleotide deletion, which is present only in RN4220 cells. RNA-seq data was visualized by the Integrated Genomics Viewer (IGV). b) The previously identified non-synonmyous SNP in the UvrC gene is found only in RN4220 cells and not NCTC8325-4 cells. c) A SNP in the GroEL gene that was previously identified in RN4220, but not NCTC8325-4, is identified in both strains. d) A SNP in the EzrA gene that was previously identified in RN4220, but not NCTC8325-4, is identified in both strains.

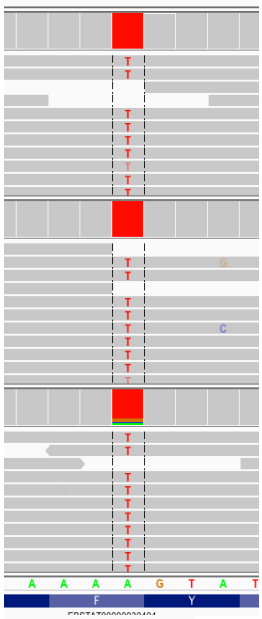
a)



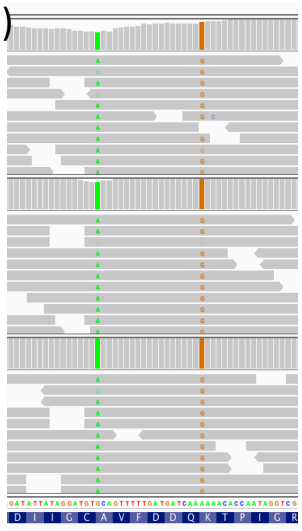
b)



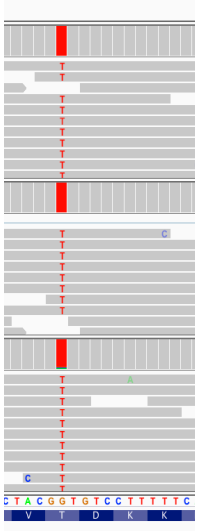
c)



d)



e)



4.8a) and the DNA repair factor UvrC (Figure 4.8b), are found only in RN4220. However, the SNPs in GroEL (Figure 4.8c), RimM (Figure 4.8d), and EzrA (Figure 4.8e), which the authors of the genome sequence of RN4220 argued may effect the fitness of this strain (Nair et al., 2011), were found in the ancestral NCTC8325 genome as well as in RN4220. Roughly half of the additional SNPs identified as unique to RN4220 were similarly found in NCTC8325 in our analysis, but not in the analysis done by O'Neill (Table A.3a). Subsequent to our analysis, these results were confirmed by resequencing the NCTC8325 genome (Berscheid et al., 2012). These mutations (RimM, EzrA, MurA, and GroEL) are therefore present in the parental strain NCTC8325 and are not unique to either NCTC8325-4 or RN4220. Berscheid *et al.* further show that RN4220 and the parental strain NTCT8325 have the same fitness levels in laboratory tests. I identified five novel SNPs in NCTC8325-4 (Table A.3b) that are previously unreported.

RNA-seq can only identify SNPs in genomic regions present in the transcriptome. However, many of the non-coding SNPs identified in the RN4220 genome (Nair et al., 2011) were located in 5' or 3' UTRs that were covered in our transcriptome data. I could therefore also identify these SNPs in our data (Figure 4.9a) and show that while some were unique to RN4220, others were also found in NCTC8325 (listed in table 4.3). Other SNPs were located in regions that were not covered in our analysis, and I therefore cannot verify their presence in RN4220.

An additional advantage of the RNA-seq method is that sequencing reads correspond to a single molecule of RNA purified from cells. Therefore, not only can SNPs be mapped to individual locations in the genome, but heterogeneity at chromosomal locations is also apparent (Figure 4.9b). At some locations where a SNP

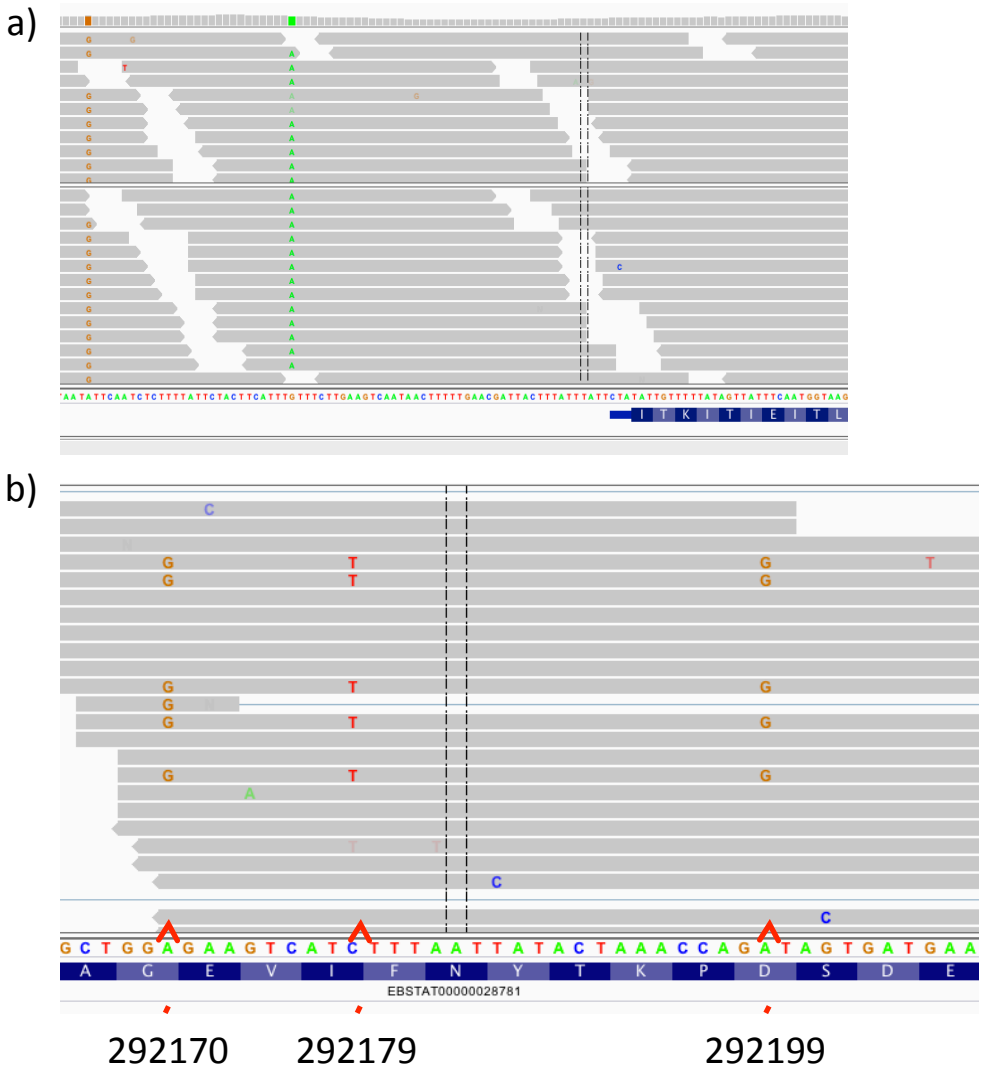


Figure 4.9: RNA-seq reveals SNPs in 5' and 3' UTRs as well as heterogeneity within the genomic sample. a) RNA-seq data provided strong coverage of many 5' and 3' UTRs, allowing us to unambiguously map SNPs in these regions. SNPs A-2244467-G, G-2244495-A are shown in the NCTC8325-4 RNA-seq data. b) Heterogeneity in the RNA-seq data. Individual sequencing reads correspond to a single molecule of cellular RNA, allowing the visualization of heterogeneity at genetic loci. Here we show that previously identified SNPs (292179 and 292199) in RN4220 are not only present in NCTC8325-4, but are a heterogeneous mix of wild type and mutant at these loci.

was reported in RN4220 but not NCTC8325-4, there is genetic variation in NCTC8325-4; most reads contain the wild type nucleotide, but the mutation in RN4220 is also present. This argues that RN4220 gained mutations that were present in NCTC8325-4 at low frequency either due to selective forces or, more likely, due to bottleneck effects.

RNA-seq reveals differential gene expression between two Sau strains

The authors of the RN4220 genome sequence argue that the SNPs that differ between the strains may cause functional differences in cellular responses to stress and to the switch to virulent growth (Nair et al., 2011). Subsequent work showed that the RN4220 and the parental strain NCTC8325 have similar fitness levels in laboratory conditions (Berscheid et al., 2012). Because RN4220 is electrocompetent and capable of being transformed by expression plasmids, it is well suited to genetic analysis and laboratory studies (Schenk and Laddaga, 1992). To examine the differences in gene expression between the two strains, we sequenced the transcriptome of NCTC8325-4 cells growing in the absence of plasmid or antibiotic used for selection. I then compared gene expression between NCTC8325-4 cells and RN4220 cells containing pRMC2 and selected by addition of chloremphencol to the growth media. While the genomes of all these strains have been sequenced and examined for genomic variations (Berscheid et al., 2012; Iandolo et al., 2002; Nair et al., 2011; O'Neill, 2010; Ohta et al., 2004), transcriptional differences have not been examined to our knowledge.

35 genes were found to be differentially expressed between these two Sau strains, representing 1.5% of all genes (Table A.4) by the Cufflinks software described above. To validate the technique for transcriptome comparisons between the two strains, I

examined a previously described transcriptional difference. RN4220 has a mutation in the AgrA gene that causes a frame shift near the C-terminus of the protein (Figure 4.8a). This mutation is known to cause disregulation of RNAIII; cells containing the AgrA mutation show a delayed upregulation of RNAIII, which is a key regulator in the switch to virulent growth (Traber and Novick, 2006). Only four genes are significantly downregulated in RN4220 compared to NCTC8325-4 cells (Figure 4.10) (Table A.4a). RNAIII is one of these genes, in agreement with the previous data on the mutation in AgrA (Traber and Novick, 2006). These data show the power of RNA-seq compared to other methods for transcriptome analysis: in one set of data I can identify both the SNP in AgrA that alters its function and the downregulation of RNAIII that is a direct result of this mutation. RNAIII is the most highly repressed gene in RN4220 compared to NCTC8325-4, arguing for the importance of the mutation in AgrA for regulation at this locus.

The other three downregulated genes in RN4220 are an acetoactate synthase, which catalyzes the formation of 2-acetolactate from pyruvate during stationary phase and an alpha-acetolactate decarboxylase from the same operon. The final downregulated gene encodes a protein of unknown function. Interestingly, four SNPs identified in the RN4220 genome (A-2244467-G, G-2244495-A, and deletions of C-2244932 and T-2244933) all cluster around this gene (2244539-2244724). These mutations were identified in the RN4220 genome sequence, but I see clear evidence for their presence in NCTC8325 genome (Figure 4.9b). The function of this gene, and of these mutations, are all unknown.

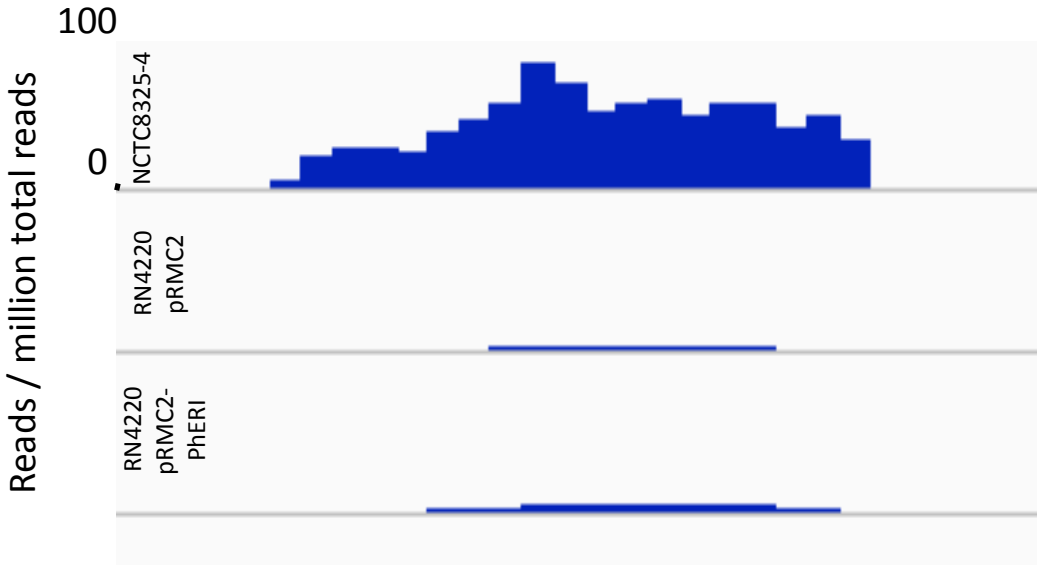


Figure 4.10:RNAIII is downregulated in RN4220 compared to NCTC8325-4. RNA-seq reads mapping to the gene for RNAIII from NCTC8325-4 (upper panel), RN4220-pRMC2 cells (middle panel) and RN4220 cells expressing PhERI (lower panel). The previously described frameshift mutation in AgrA has been shown to delay the expression of RNAIII in RN4220 cells. In one data set, we are able to visualize the SNP in AgrA (Figure 4a) and the subsequent downregulation of the non-coding RNAIII.

31 genes are upregulated in RN4220 carrying an expression cassette and under antibiotic selection compared to NCTC8325-4 cells (Table A.4b). Among these upregulated mRNAs, nine encode putative or confirmed ABC transporters. This may be due to the addition of chloramphenicol to select for RN4220 cells containing pRMC2; sequencing of RNA from RN4220 cells not containing an expression vector would clarify if this difference is inherent to the strains or rather is a response to the addition of antibiotic to the growth media. ClfB, a clumping factor, is also upregulated in RN4220. This could potentially compensate for the ClfA mutation previously identified in RN4220.

Identification of a putative orphan CRISPR element in Sau

Clustered regularly interspaced short paleidromic repeats (CRISPRs) are bacterial RNA elements that provide an adaptive immune response to phage infection (Marraffini and Sontheimer, 2010). CRISPRs are organized in bacterial genomes with many interspaced repeats that create a long RNA followed by the Cas genes encoding the protein machinery required to process the RNA into functional units. After processing, crRNAs can interact with phage or invasive DNA with sequence specificity and induce cleavage (Marraffini and Sontheimer, 2010). Sau is not thought to have a functional CRISPR system; no genes in the Sau genome have any homology to previously identified Cas proteins. Genomic searches for putative CRISPR elements in the Sau NCTC8325 genome reveal only three weak hits (Grissa et al., 2007).

I used our RNA-seq data to determine whether RNA is being expressed at any of the putative CRISPR loci in Sau. While two of the three putative CRISPR elements were

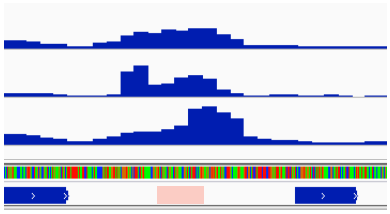
located in annotated ORFs, and contained no signal for an additional RNA element in our RNA-seq data, the third putative CRISPR was located in an intergenic region and showed clear evidence for RNA-seq reads (Figure 4.11). The putative CRISPR has only one repeating unit and no downstream Cas genes that would be required for active crRNA function (Marraffini and Sontheimer, 2010). BLAST searches for the CRISPR element revealed that the spacers map to several locations in the Sau genome including both coding and non-coding regions (Grissa et al., 2007). This element may be an orphan CRISPR, and that reintroduction of Cas genes into Sau may activate this putative RNA element.

RNA-seq to screen for PhERI suppressor mutations

RNA-seq allows gene expression and SNPs to be rapidly quantified in one dataset as cells respond to various stimuli. The expression of a toxic protein, or the addition of drugs or antibiotics to the growth media, provides a strong selective pressure on cells to evade the effect of the protein or small molecule. Recent work has illustrated how RNA-seq can be used to identify small-molecule binding targets in eukaryotic cells (Wacker et al., 2012).

To evaluate whether there are PhERI resistance mutations in our RNA-seq data, I searched for SNPs that were present only in RN4220 cells expressing PhERI. In our data, PhERI was only induced for sufficient time to allow its expression and repression of Sau cell growth. Longer expression times may allow more resistance mutations to accumulate. Even with the short expression time, I identified two mutations present in cells expressing PhERI, whereas I identified no mutations unique to cells containing only

a)



b)

Figure 4.11: Identification of a putative CRISPR element in Sau. a) RNA-seq reads mapping to an intergenic region containing a putative CRISPR element. The putative CRISPR sequence is denoted by a red box. b) View of the putative CRISPR element showing relative expression from NCTC8325-4 and RN4220 cells. The putative CRISPR repeats are highlighted in red.

empty vector. The two mutations were both in coding regions (Figure 4.12). One was a glycine to alanine mutation in the transcription elongation factor GreA (Laptenko et al., 2003; Stebbins et al., 1995) and the other was a valine to alanine mutation in a putative RNaseJ protein (Figure 4.12) (Even et al., 2005; Newman et al., 2011).

Gre factors are known to bind to RNAP and modulate its activity during elongation phase (Borukhov et al., 1992). PhERI also interacts with RNAP but through an interaction with the Sau σ factor. It is unclear how a mutation in GreA, which has not been studied in Sau, may alleviate PhERI activity, but finding mutations in another RNAP binding protein is potentially physiologically relevant (Figure 4.13a). PhERI does not significantly affect the expression levels of GreA; although GreA mRNA is roughly 1.5 fold more abundant in cells expressing PhERI, it is not statistically significant ($p = 0.13$).

The mutation in RNAseJ may also be functionally important. RNAseJ is a putative member of a family of proteins (RNAse J1/J2) required for mRNA processing and degradation in gram-positive organisms (Even et al., 2005). Both putative RNaseJ proteins in Sau are significantly downregulated by PhERI expression. One paralog was found to be required for normal cell growth in Sau (Xu et al., 2010); the other paralog is mutated in PhERI expressing cells. Whether this mutation (V29A) is functionally relevant is under investigation (Figure 4.13). A mutation increasing the activity of this required protein, which is downregulated by PhERI, may be particularly advantageous for cells growing in the presence of PhERI.

I searched for mutations in the known binding partner of PhERI, the group 1 Sau σ factor. There is no evidence for accumulation of mutations in RpoD in the RNA-seq

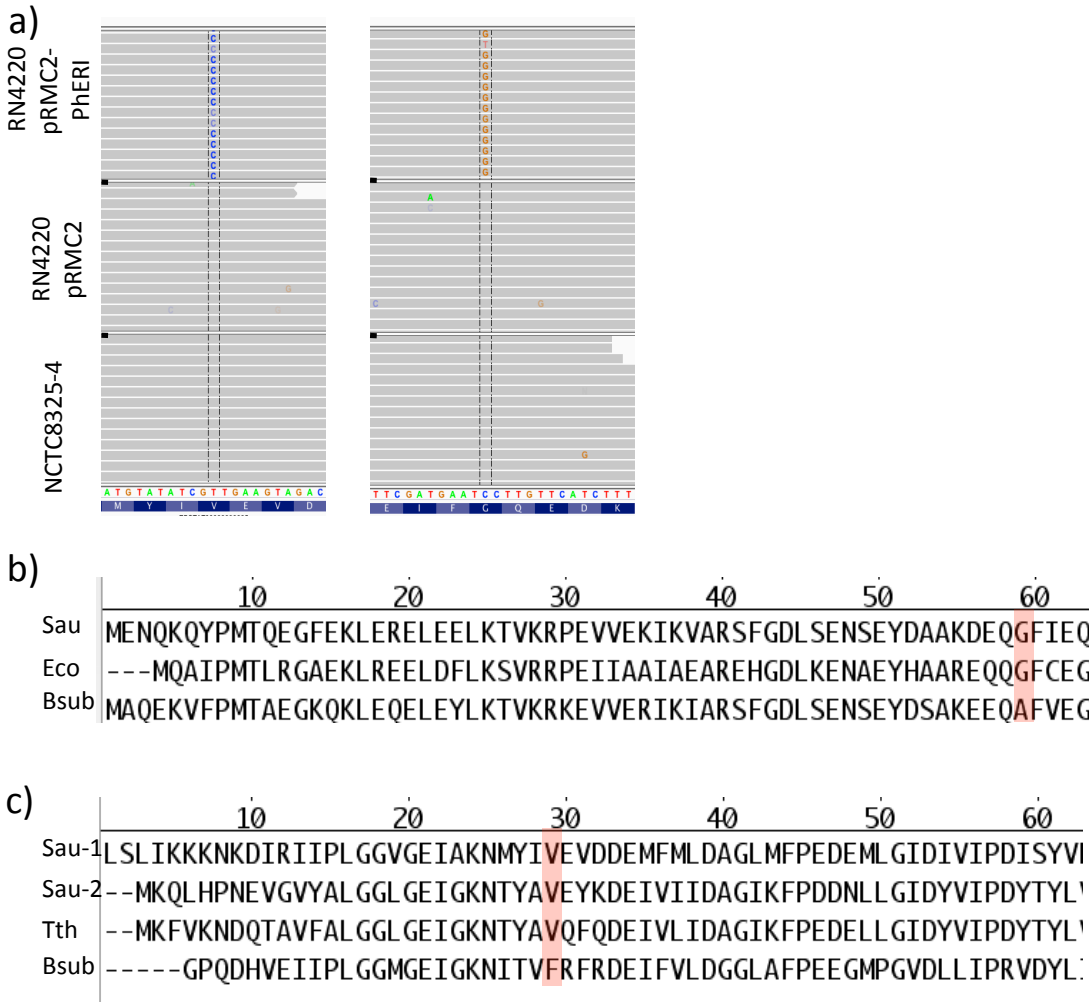
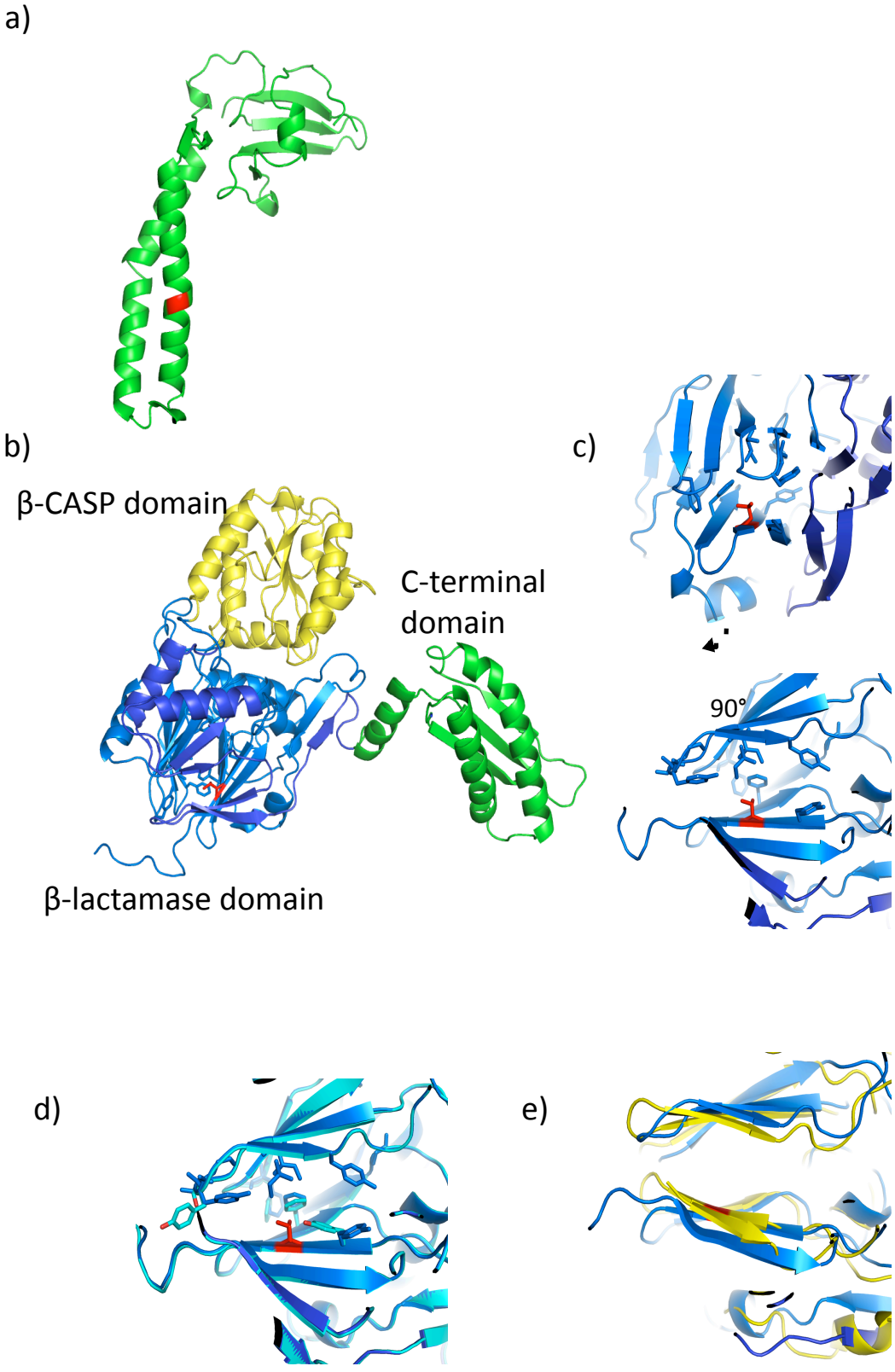


Figure 4.12: Identification of SNPs from cells expressing PhERI. a) Identification of 2 SNPs only present in RN4220 cells expressing PhERI (upper panel) but absent in RN4220 cells containing empty vector (middle panel) or NCTC8325-4 cells (bottom panel). GreA (left panel) and a putative RNAseJ1/J2 protein (right panel) both contain coding changes in pRMC2-PhERI cells only. b) Alignment of the coding change in GreA (G59A) compared to with sequences for the well studied GreA proteins from Eco and Bsub. c) Alignment of the coding change in the putative RNAseJ1/J2 protein (V29A), the other RNAseJ1/J2 paralog in Sau, and the well studied homologous proteins from Tth and Bsub.

Figure 4.13 (adjacent page): Structural models of the coding changes from PhERI expression Sau cells. a) Crystal structure of Eco GreA. Eco G56 (Sau59) is highlighted in red. b) Model of the putative Sau RNAseJ1/J2 protein. Model was made using modeller-v9.10 with the high-resolution Tth structure (pdb:3BK1) as a model. The domains are labeled (β-CASP domain: yellow; β-lactamase domain: blue; C-terminal domain: green). V29 is highlighted in red. c) Zoomed view on V29, the location of the mutation in PhERI expressing cells. V29 is located in a hydrophobic pocket and is contacting multiple aromatic and hydrophobic residues. d) Model of V29A mutation (cyan) compared to the Sau wt model (dark blue). Multiple hydrophobic residues are predicted to be in alternate conformations in the V29A model. e) Structural changes in the β-lactamase domain upon binding to substrate. V29 is highlighted in red, yellow indicates the structure bound to RNA substrate. The distance between the two beta-strands in 12.5 in the apo enzyme, but 10.2 in the open form that is capable of binding to substrate. All figures were generated in PyMol using the indicated pdb files.


a)
b)
β-CASP domain
C-terminal domain
β-lactamase domain
c)
90°
d)
e)

data. Our structural and biochemical information demonstrate that the two proteins form an extensive molecular interaction. 2-hybrid and biochemical data show that four individual mutations in combination are required in σ^{A}_{4} to eliminate PhERI binding. Mutations eliminating the interaction between PhERI and RNAP would be likely to completely alleviate PhERI repression of cell growth. However, the requirement for multiple point mutations to eliminate the interaction and the short time that cells were exposed to PhERI expression possibly prevent the accumulation of mutations in rpoD. RNA-seq could be performed at additional times long after the expression of PhERI to identify additional suppressor mutations, potentially in σ^{A}.

We demonstrate that RNA-seq is a valuable tool to examine gene expression in Sau. RNA-seq provides the information that was previously accessible by microarray analyses in this organism plus much additional data. I was able to identify promoters specifically and directly inhibited by PhERI through differential gene expression analysis by RNA-seq (Figure 4.5)

High throughput sequencing provides additional information that was only previously accessible through using multiple, complimentary techniques. Because prokaryotic genomes are generally small, and contain relatively short intergenic distances with limited non-coding regions, we sequence the majority of the Sau genome through RNA-seq analysis of the transcriptome. We were able to identify almost all of the SNPs previously found in RN4220, including 23/26 that were in non-coding regions. RNA-seq has become increasingly cost effective and the sample preparation has been standardized for eukaryotic cells. I believe similar standardization of RNA-seq for prokaryotic

samples, and routine transcriptome analysis using high-throughput sequencing provides a significant advantage over the previously used microarray based techniques.

Chapter 5:

PhERI Inhibits RNAP at rrn Promoters by Blocking UP-element Binding

The structural, biochemical and genomic analyses all show that most -10/-35 promoters are not affected, positively or negatively, by PhERI expression. These complimentary methods provide data demonstrating that PhERI does not inhibit the normal functions of σ^{A}_{4}. How, then, does PhERI inhibit Sau RNAP at certain promoters, and how does PhERI discriminate between promoters that will be inhibited and promoters that will not be affected?

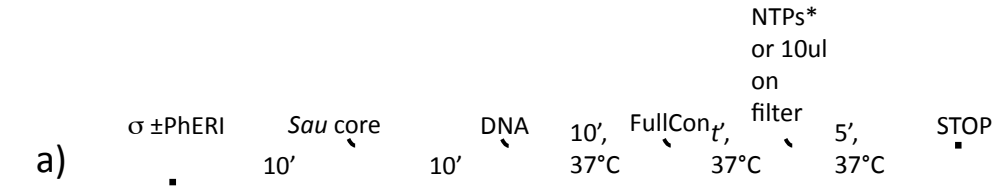
In this chapter, I will detail our mechanistic studies to understand the function of PhERI. Using the DNA sequences identified in chapters three and four, I will show that PhERI specifically targets Sau promoters containing an A/T-rich UP-element upstream of their -35 element. While PhERI binds to RNAP holoenzyme at all promoters, it blocks α-CTD binding to the UP-element and therefore only inhibits promoters dependent on this interaction for robust transcription, including the rrn promoters. PhERI is not an anti-σ factor in the classical sense. PhERI is a phage encoded transcription factor that inhibits host RNAP in a targeted manner through a novel mechanism. While it joins the RNAP holoenzyme through its interaction with σ^{A}_{4}, it modulates RNAP activity through an interaction with promoter DNA and another RNAP domain. These results will be discussed in the context of their importance in the phage life cycle, Sau transcriptional regulation, and rRNA regulation more generally.

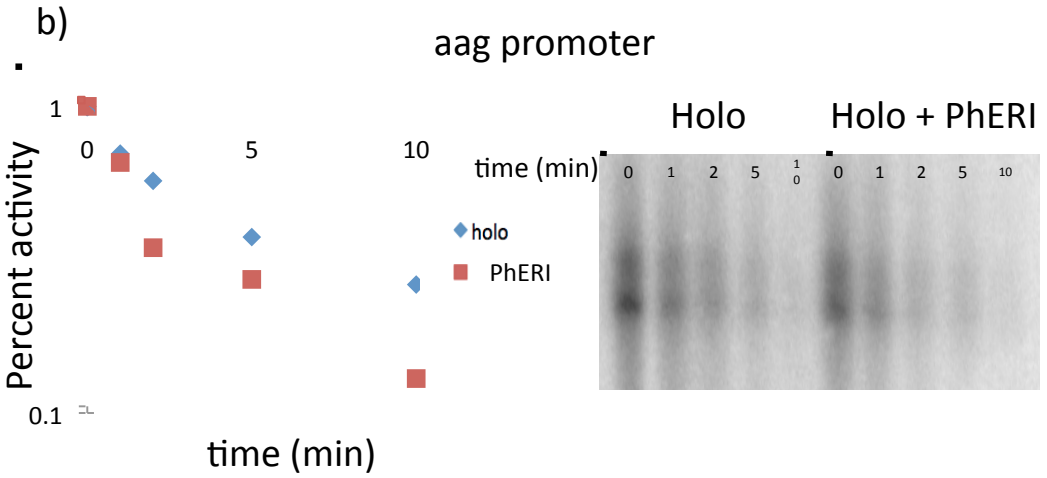
PhERI does not modulate the stability of Open Promoter Complexes

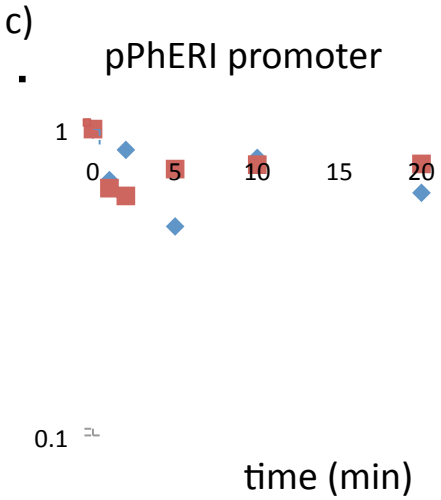
The rrn promoters have been extensively studied in Eco. Because of their important role in the switch from logarithmic growth to stationary phase, the rrn promoters are extensively regulated by extrinsic factors as described in Chapter 1. While most promoters form stable OPCs, rrn promoters form characteristically unstable OPCs. The stability of OPCs is rate-limiting for these promoters, and therefore proteins and small molecules (DksA and ppGpp respectively) that modulate the stability of the OPC can dramatically affect output from rrn promoters but not other -10/-35 promoters. DksA and ppGpp show a reproducible effect on OPC stability at all promoters but specifically affect transcription at rrn promoters due to their special properties (Barker et al., 2001b; Paul et al., 2004). I hypothesized that PhERI could inhibit the rrn promoters but not most -10/-35 promoters by modulating the stability of the OPC; I therefore tested the OPC stability in the presence and absence of PhERI at the promoters that I described in Chapter 3.

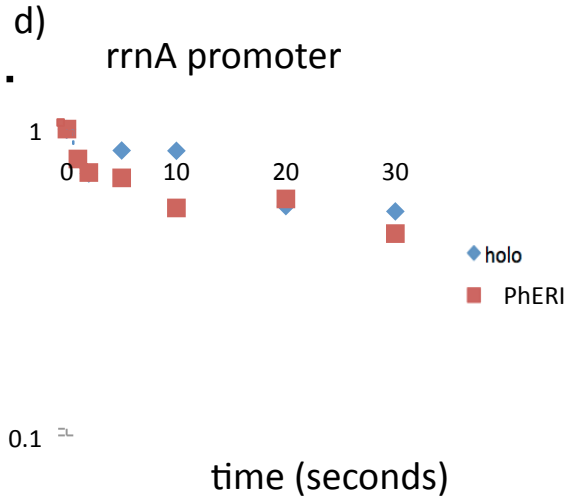
PhERI has no significant effect on OPC stability on any promoter tested (Fig 5.1). Importantly, I find that rrn promoters in Sau are quite unstable (Fig 5.1d) compared to other promoters tested (Fig 5.1 b and c). The half-lives ($t_{1/2}$) of the rrnA promoter was only 28 and 23 seconds in the presence and absence of PhERI, respectively. The aag promoter had a half-life of minutes (4.5 minutes and 3.1 minutes in the presence and absence of PhERI, respectively) while the phage G1-pPhERI promoter had an extremely stable half-life (24 minutes in the absence of PhERI and 23 minutes in the presence of PhERI). If PhERI were decreasing the stability of the OPC, one could explain its narrow

Figure 5.1 (adjacent page): Figure 5.2: PhERI does not modulate the stability of Open Promoter Complexes (OPCs). a) Schematic of the OPC stability assay. Briefly, OPCs were formed, after which they were challenged with the addition of excess the dsFullCon promoter fragment. At different times after the addition of challenge, reactions were started by the addition of NTPs or pipetted onto a filter paper. In either case, the signal represents the amount of RNAP bound to promoter after the challenge. b) PhERI does not alter the stability of RNAP OPCs on the aag promoter. OPCs were monitored by transcriptional output and visualized on a 12% Urea-PAGE gel (right panel). Bands were quantitated, normalized to time 0 and plotted (left panel). c) PhERI does not modulate the OPC stability of the G1 phage pPhERI promoter. OPCs were monitored using P^{32} labeled linear DNA in a filter binding assay. d) PhERI does not modulate the stability of the OPC at the rrnA promoter as monitored by transcription output.


a)
σ ±PhERI
Sau core
10'
DNA
10'
10', 37°C
FullCon
t', 37°C
NTPs* or 10ul on filter
5', 37°C
STOP


b)
aag promoter
Holo
Holo + PhERI
time (min)
Percent activity
1
0.1
0
5
10
holo
PhERI
time (min)


c)
pPhERI promoter
1
0.1
0
5
10
15
20
time (min)


d)
rrnA promoter
1
0.1
0
10
20
30
holo
PhERI
time (seconds)

spectrum of inhibition and the apparent absence of a DNA sequence unique to PhERI sensitive promoters. However, it does not appear that PhERI functions by modulating OPC stability at Sau promoters.

PhERI inhibits transcription from promoters containing an A/T-rich sequence upstream of the -35 element

Because PhERI is unlikely to block -35 element recognition, we hypothesized that other DNA elements may confer susceptibility to PhERI inhibition. I aligned all promoters tested in vitro for direct PhERI activity by their -10 and -35 elements (Fig 5.2). The sensitive promoters do not share an obvious common sequence, but contain an A/T-rich region upstream of their -35 elements (80 – 100% A/T-rich in the region expected for an UP-element), whereas this feature was less prominent in promoters resistant to PhERI inhibition (65% A/T-rich, similar to the A/T content of the entire Sau genome, which is ~67%; Figure 5.2).

To test whether the region upstream of the -35 element was important for PhERI function, I constructed hybrid promoters that swapped the DNA immediately upstream of the -35 element between a PhERI-sensitive promoter, rrnA, and a PhERI-resistant promoter, dnaA. As observed previously, PhERI inhibited transcription from rrnA (Fig. 5.3a, lanes 1, 2). In the absence of PhERI, the rrnA(dnaA-UP) promoter showed decreased activity, similar to the PhERI-inhibited rrnA (Fig. 53a lane 3), and PhERI had no additional effect on this hybrid promoter (Fig. 5.3b, lane 4).

PhERI did not affect transcription from the dnaA promoter (Fig. 5.3b, lanes 5, 6). However, when its upstream sequence was replaced by the rrnA A/T rich sequence, the

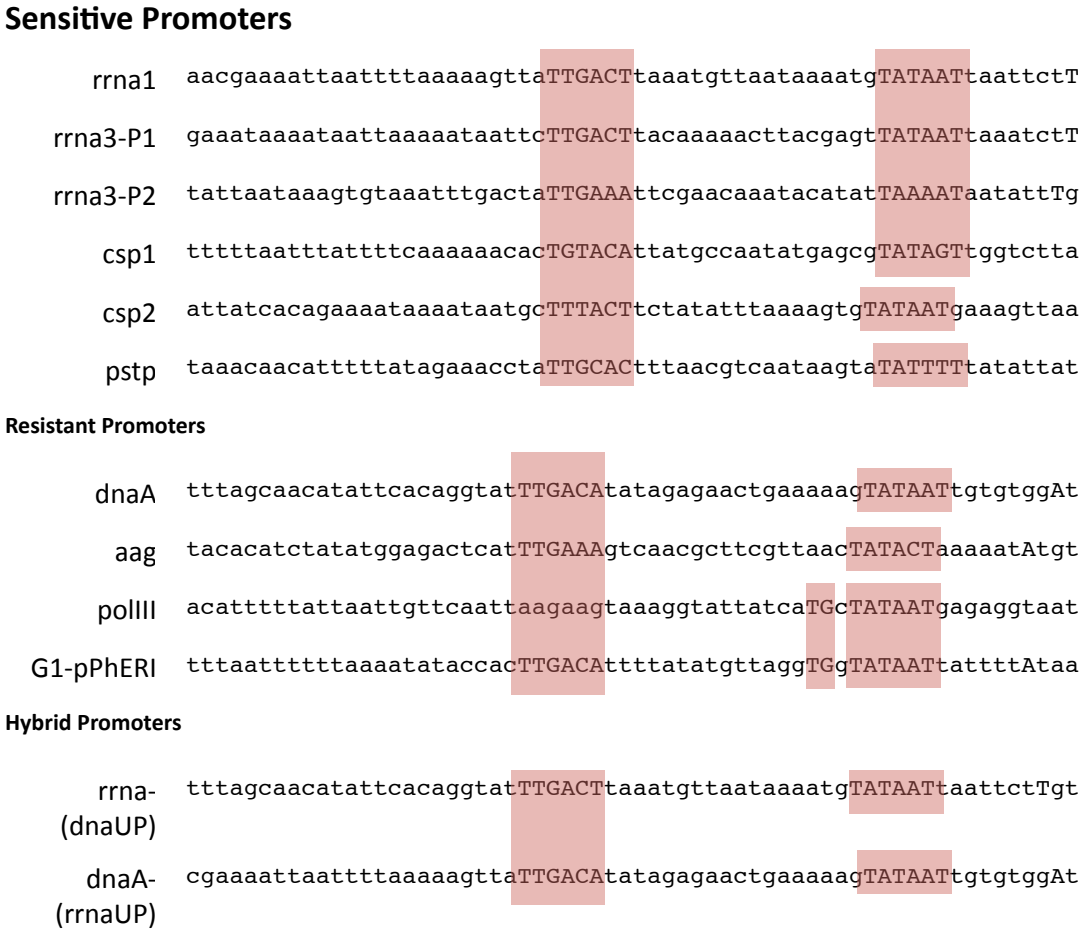


Figure 5.2: Sequences of Sau promoters sensitive and resistant to inhibition by PhERI. -35 and -10 and extended -10 elements are highlighted in red. Putative UP-elements are highlighted in green. Hybrid promoters, which will be used in subsequent experiments, swap the region immediately upstream of the -35 element between a sensitive (rrnA) and resistant (dnaA) promoter.

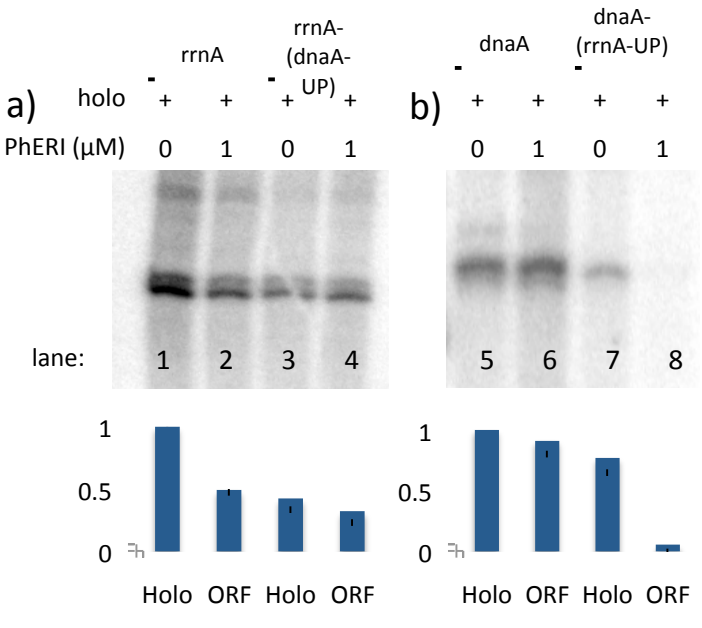
resulting dnaA(rrnA-UP) promoter became highly sensitive to PhERI (Fig. 5.3b). The loss of activity between rrnA and rrnA(dnaA-UP) (compare Fig. 5.3, lanes 1 and 3) demonstrates that the A/T-rich sequence just upstream of the rrnA -35 element contributes significantly to rrnA activity, as with the Eco rrn promoters (Rao et al., 1994). Moreover, PhERI sensitivity of a promoter appears to be determined almost totally by the sequence just upstream of the -35 element. These results argue that PhERI's activity is mediated by a promoter feature upstream of the -35 element and implicate an A/T-rich sequence in this inhibitory effect.

PhERI alters RNAP interactions with promoter DNA upstream of the -35 element

To test whether PhERI directly modulates RNAP binding to promoters with A/T rich elements, I used DNAseI footprinting. DNAseI cleaves at exposed minor grooves of the DNA double-helix with some sequence dependence. Cleavage is particularly sensitive to deformations or bends in the DNA double helix that widen the minor groove (Fox, 1997). Because RNAP extensively bends promoter DNA, DNAseI has characteristic cleavage patterns on DNA in the OPC, including cleavage between the -10 and -35 elements, and cleavage upstream of the -35 element (Severinov and Darst, 1997).

Because Sau rrn promoters have never been tested biochemically, and UP-element binding has not been shown in this organism, I first tested the DNAseI cleavage pattern of the rrnA promoter using the well characterized Eco RNAP. I observed the typical DNAse cleavage patterns on this promoter (Fig 5.4b) with a hypersensitive band immediately upstream of the -35 element and protection upstream of the -35. To test whether protection in the putative UP-element on rrnA is due to RNAP α-CTD binding, I

Figure 5.3 (adjacent page): An element upstream of the -35 is required for PhERI inhibition. *In vitro* transcription assays were performed as in chapter 3 on hybrid promoter DNA that swapped the region upstream of the -35 element from a PhERI sensitive promoter (rrnA) and a resistant promoter (dnaA). a) PhERI inhibits transcription from the rrnA promoter (lane 1 and 2). Replacing the upstream A/T rich region decreases the output at this promoter (lane 3) similar to the effect of PhERI (lane 2). PhERI has no additional effect on this promoter (lane 4). b) PhERI has no effect on transcription at the dnaA promoter (lanes 5 and 6). When an A/T-rich region is placed upstream of the -35 at this promoter (lane 7), it becomes highly sensitive to PhERI inhibition (lane 8). Lower panels show the average of 3 independent normalized experiments with the error bars showing one standard deviation above and below the mean.



c)

Sensitive Promoters

rrna1 aacgaaaattaattttaaaaagttaTTGACTtaaatgttaataaaatgTATAATtaattctT

Resistant Promoters

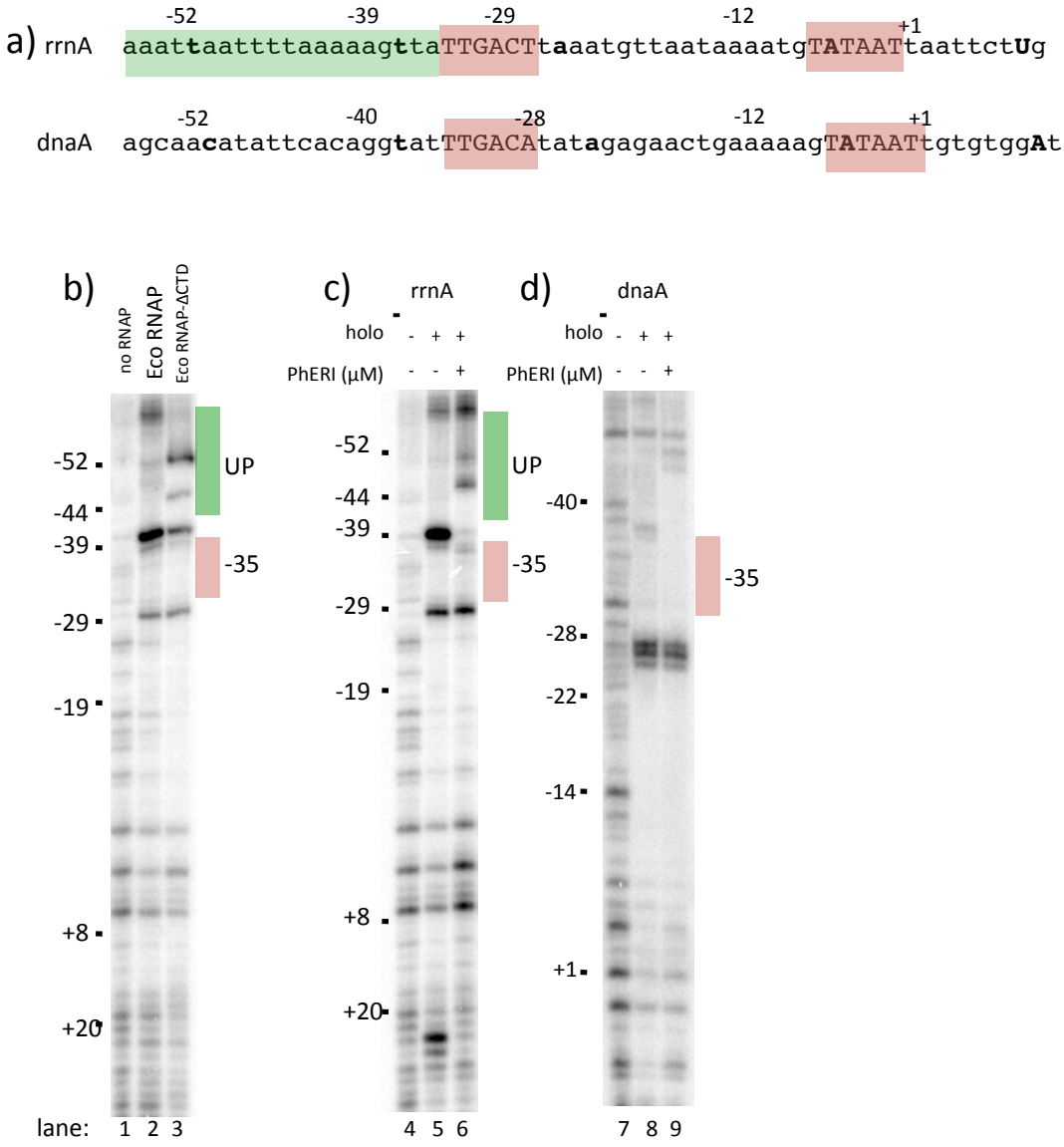
dnaA tttttagcaacatattcacaggtatTTGACAtatagagaactgaaaaagTATAATtgtgtgg

Hybrid Promoters

rrna-(dnaUP) tttttagcaacatattcacaggtatTTGACTtaaatgttaataaaatgTATAATtaattctT

dnaA-(rrnaUP) aacgaaaattaattttaaaaagttaTTGACAtatagagaactgaaaaagTATAATtgtgtgg

Figure 5.4 (adjacent page): DNAse footprinting reveals that PhERI blocks protection of upstream A/T-rich DNA elements. a) Sequence of promoters tested below. -35 and -10 elements are highlighted in red, UP-element in green. Position relative to the first transcribed base (+1) are listed at locations that are identified experimentally in the G/A ladder. b) Eco holoenzyme and Eco holoenzyme with the α-CTD truncated were incubated with promoter DNA end labeled on the template strand in 1x DNase buffer. DNAse (0.1μg/ml) was added for 1 minute before reactions were stopped with 0.1M EDTA, boiled and visualized on a 6% Urea-PAGE gel. c) Sau holoenzyme (1μM) was incubated with PhERI (2μM) or buffer in 1x DNAse buffer and assayed as above. DNA regions corresponding to the -10 and -35 elements are highlighted in red, and the UP-element in green.


a)
rrnA
-52
-39
-29
-12
+1
aaattaattttaaaaagttaTTGACTtaaatgttaataaaatgTATAATtaattctUg
dnaA
-52
-40
-28
-12
+1
agcaacatattcacaggtatTTGACAtatagagaactgaaaaagTATAATtgtgtggAt
b)
no RNAP
Eco RNAP
Eco RNAP-ΔCTD
-52
-44
-39
-29
-19
+8
+20
UP
-35
lane: 1 2 3
c)
rrnA
holo - + +
PhERI (μM) - - +
-52
-44
-39
-29
-19
+8
+20
UP
-35
4 5 6
d)
dnaA
holo - + +
PhERI (μM) - - +
-40
-28
-22
-14
+1
-35
7 8 9

used Eco RNAP lacking the α-CTD. In the absence of the α-CTD, the hypersensitiviy upstream of the -35 element is dramatically decreased and there are prominent cleavage events in the A/T-rich UP-element (Fig5.4 lane 3). Thus, two DNaseI cleavage properties appear to be diagnostic of an α-CTD/UP-element interaction i) Hypersensitivity at a site just upstream of the -35 element, and ii) absence of cleavage events within the UP-element itself.

Sau RNAP also confers the characteristic protection from DNAseI indicating OPC formation (Fig 5.4c lane 5). At the rrnA promoter, OPC formation induces strong hypersensitivity just upstream of the -35 element, and the UP-element itself is protected from cleavage (Fig 5.4c, lane 5), indicative of α-CTD/UP-element interaction. Addition of PhERI results in a complete loss of the sensitivty immediately upstream of the -35 element and a loss of protection within the UP-element (Fig 5.4c, compare lanes 5 and 6), qualitatively similar to the effect of deleting the Eco α-CTD entirely. On the dnaA promoter, there is no significant protection by RNAP in the region upstream of the -35 element (Fig 5.4d, lane 8). PhERI addition leads to protection immediately upstream of the -35 element (Fig 5.4d, lane 9) indicating that while PhERI does not affect transcription from this promoter *in vitro* or *in vivo*, it is bound to RNAP and positioned near the DNA in this region (Fig 5.4d).

To show that the A/T rich element is responsible for the protection upstream of the -35, I tested the DNAseI digestion pattern on the hybrid promoters described above (Fig 5.5). Removal of the A/T-rich sequence from the rrnA promoter leads to DNA cleavage events between the -40 and -52 positions and a decrease in the hypersensitivity of the cleavage at the -40 position, indicating a loss of α-CTD binding (Fig 5.5a, compare

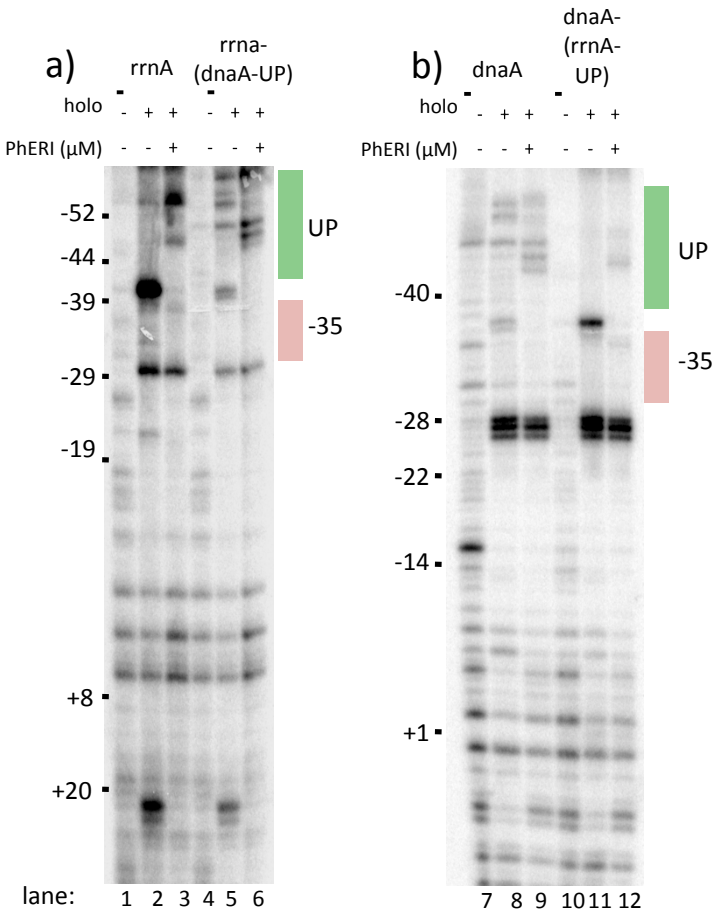


Figure 5.5: DNAseI footprinting assays performed on the hybrid promoters described above. DNAseI digestions were performed as in Figure 5.4.

lanes 2 and 5). Moving the rrnA A/T-rich sequence upstream of the -35 position in the dnaA promoter confers resistance to DNAse cleavage throughout the upstream region and increased hypersensitivity of the cleavage event immediately upstream of the -35 (Fig 5.5b, compare lanes 8 and 11), indicative of α-CTD binding. Addition of PhERI to the dnaA hybrid promoter cause a complete loss of the hypersensitivity, then protects the DNA immediately upstream of the -35 and induces cleavage in the A/T-rich element (Fig 5.5b, compare lanes 11 and 12), indicating loss of α-CTD binding.

Together, these results show that PhERI is closely associated with DNA immediately upstream of the -35 element on all promoters regardless of its activity, in agreement with the structural model, and prevents functional binding of the RNAP α-CTD to an upstream A/T-rich promoter element. Based on the footprinting, it appears that PhERI protects cleavage from roughly the -35 element to base -45 relative to the promoter start site. By blocking the UP-element interaction, PhERI inhibits only transcription from the subset of promoters requiring α-CTD binding for full activity, including the rrn promoters.

Role of PhERI in the phage life-cycle

Sau phages have been studied as possible therapeutic agents against highly resistant infections. Much of this research has focused on the ability of phages, or their lytic enzymes, to clear resistant infections (Fischetti, 2010). While the genomes of many Sau phages have been sequenced (Kwan et al., 2005), very little is understood about the basic molecular mechanisms through which Sau phages initially inhibit cell growth and co-opt the molecular machinery to favor completion of the phage life-cycle.

In the previous chapters, I showed that the G1 phage protein PhERI inhibits cell growth in Sau by binding to the host RNAP and targeting UP-element dependent promoters, including the rrn promoters. The G1 phage is nearly identical to the Sau phage K (Kwan et al., 2005) and PhERI shares 100% sequence identity between these phages. PhERI has homologs in five firmicute specific phages, including in the Sau phage Twort. Recent publications have tested the ability of Phage K and Twort lytic enzymes to lyse Sau strains (Paul et al., 2011), potentially preparing for their therapeutic application (Kelly et al., 2011). The work I present is the first understanding of the early stages of infection from any Sau lytic phage.

Because PhERI is expressed from a -10/-35 promoter that is highly transcribed by Sau RNAP in vitro (Fig. 5.6 and 3.5b), and is located downstream of strong -10/-35 promoters in all phages that encode a PhERI homolog (Fig. 5.6), it is likely one of the early proteins transcribed after initial injection of the double stranded phage genome into the host cell. Work on the T4 phage protein AsiA shows that phage promoters escape inhibition by containing an extended -10 element (TGn immediately upstream of the -10 element) or an UP-element. To understand how the PhERI homologs are transcriptionally regulated, I searched for the promoter sequences that drive PhERI homolog expression in the phage genomes.

Interestingly, three of the five promoters (Staphylococcus phages G1 and Twort and Lactobacillus phage Lb338-1) that express PhERI homologs contain an extended -10. Three of these promoters do contain A/T-rich regions upstream of their -35 elements (100%, 93% and 89% A/T, compared to ≈67% in the Sau genome), but these promoters also contain an extended -10 element. The two additional phage promoters (Listeria

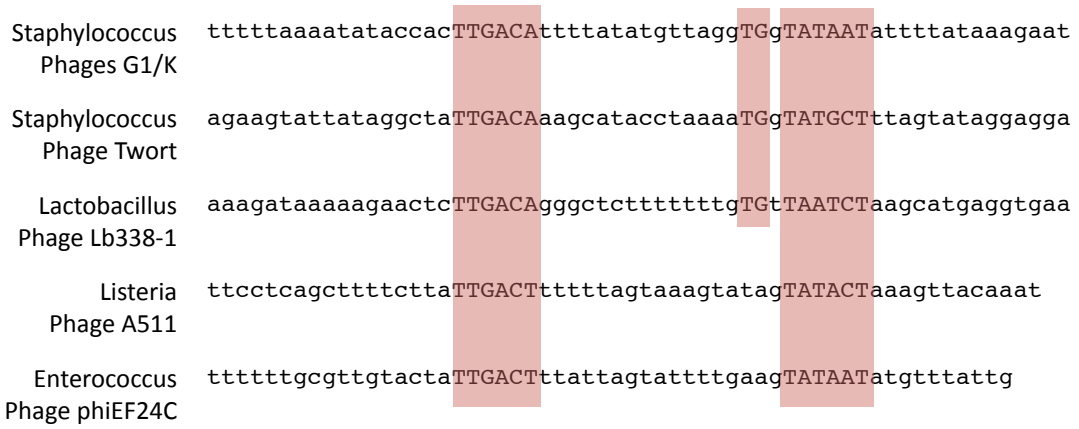


Figure 5.6: Promoters that drive the expression of PhERI and PhERI homologs in phage genomes. The fully sequenced phage is listed on the left along with the species of bacteria it infects. The -35, -10 and extended -10 elements are highlighted in red.

Phage A511 and Enterococcus Phage phiEF24C) that do not contain an extended -10 element do not have an A/T-rich region upstream of their -35 elements (63% and 68% A/T-rich), and are therefore not likely to be targeted by PhERI inhibition. It appears that PhERI containing phages evolved a mechanism through which they could inhibit host rRNA transcription while phage promoters themselves remained unaffected through various mechanisms (extended -10 elements or UP-element independence).

Upon being transcribed and translated, PhERI binds to the host RNAP holoenzyme (Fig. 2.2) and suppresses production of rRNAs (Fig. 3.7 and 4.3) by selectively inhibiting UP-element-dependent promoters (Fig. 5.3) while allowing transcription from the majority of -10/-35 promoters (Fig. 3.5 and Fig 4.5a), including the phage G1 pPhERI promoter (Fig 3.5) and other potential phage early promoters. The phage ultimately will require the use of host ribosomes to translate middle and late phage gene products; however, because rRNA is quite stable in prokaryotic cells (Deutscher, 2003), previously formed ribosomes are abundant in Sau cells for hours after PhERI expression (Figure 4.3a). Our model for PhERI inhibition of Sau RNAP is summarized in Figure 5.7.

During log-phase growth, the majority of RNAP in prokaryotic cells is occupied in actively transcribing rRNA. Inhibition of rRNA transcription not only leads to arrest of cell division (Gourse et al., 1996), but would free a large pool of host RNAP (Barker et al., 2001a) that could then be recruited to the strong phage early promoters. The T4 phage anti-σ factor, AsiA, inhibits host RNAP by blocking -35 element recognition while an additional protein, MotA, binds and recruits the RNAP complex to phage promoters (Hinton et al., 2005). The T4 phage has additional protein factors that ADP-ribosylate the

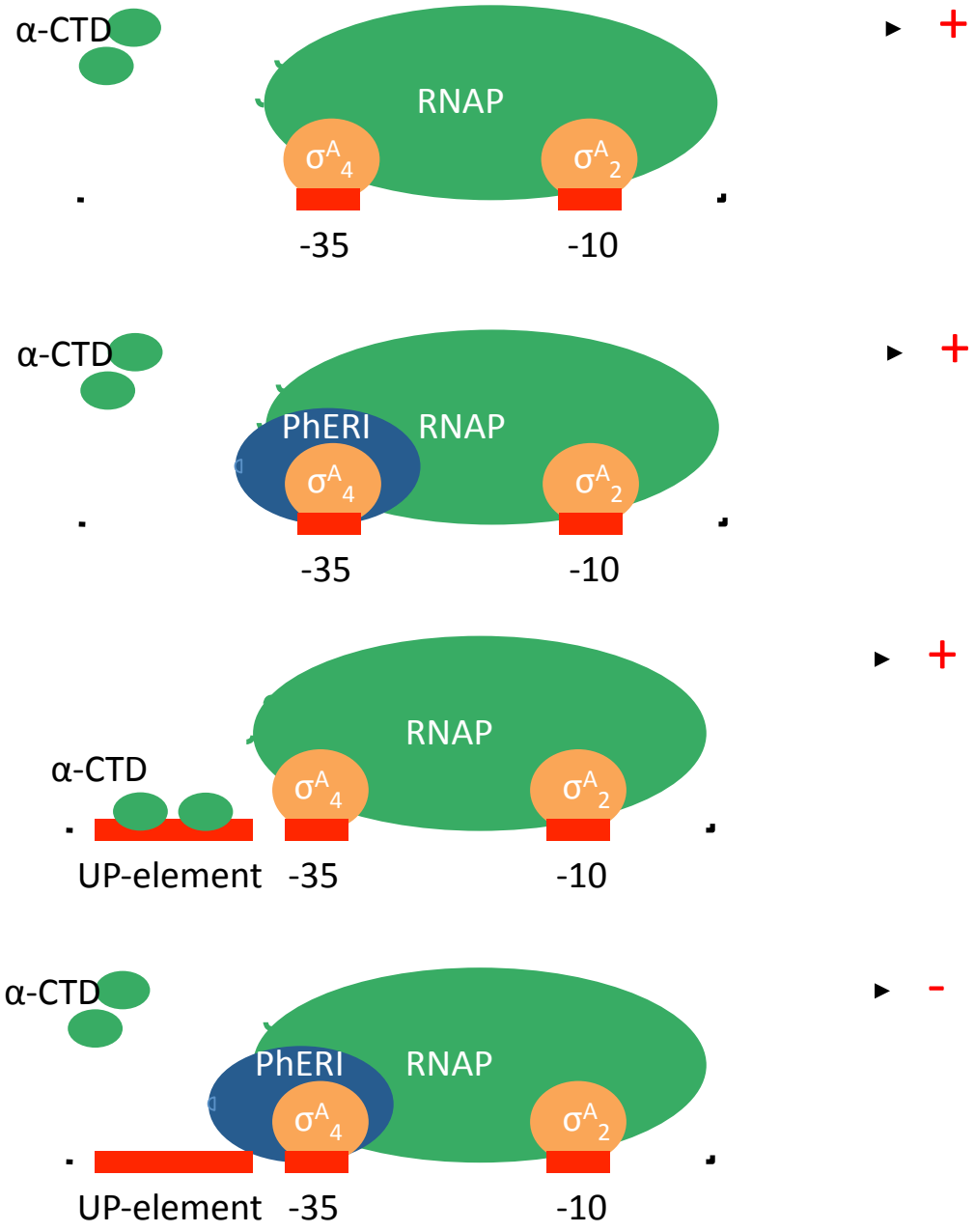


Figure 5.7: Model for PhERI function. At -10/-35 promoters, PhERI joins the Sau RNAP holoenzyme through its interaction with σ^{A}_{4}. However, PhERI does not block any of the functions of σ^{A}_{4} and does not inhibit RNAP. At promoters that require UP-element binding, PhERI joins RNAP holoenzyme through its interaction with σ^{A}_{4} and blocks the UP-element binding by the RNAP α-CTD.

RNAP α-CTD, leading to recruitment of RNAP to phage promoters (Tiemann et al., 2004). Whether phages K and G1 have additional proteins that actively recruit RNAP to phage promoters, or whether PhERI binding can directly stimulate RNAP activity at phage promoters, is under investigation. As demonstrated previously, PhERI expression alone is sufficient to inhibit cell growth (Figure 4.2) (Dehbi et al., 2008; Liu et al., 2004). I show that PhERI arrests cell growth by binding to RNAP and specifically inhibiting only a small subset of Sau promoters, including rRNA promoters, highlighting ribosomal synthesis as a strong target for novel therapeutics.

Sau UP-element binding and Sau promoters

Previous *in vitro* transcription work in Sau only studied highly regulated virulence promoters. Here, I identify and study RNAP activity and regulation at two classes of promoters, -10/-35 promoters (Fig. 3.5) and UP-element dependent promoters (Fig. 3.7). rRNA transcription has not been previously studied in Sau. My work identified the first rRNA promoters in this organism and demonstrated their activity *in vitro* (Fig. 3.7).

rRNA promoters in other organisms are dependent on UP-element binding for robust transcriptional activation. Similarly, in Sau, I show rRNA promoters require UP-element / α-CTD interaction for full transcriptional activity (Fig 5.3). Replacing UP-element DNA sequences with more G/C-rich sequences decreases the activity of these promoters *in vitro* (Fig 5.3a, compare lanes 1 and 3). PhERI, which blocks UP-element binding by the RNAP α-CTD, inhibits rRNA transcription and a minority of other promoters in the Sau genome (Fig 4.5, table 4.2). This is the first demonstration of UP-

element activation in Sau, and further shows that as in Bsub and Eco, rRNA transcription in Sau is activated by the UP-element/α-CTD interaction.

The identification of Sau and Sau phage promoters was not trivial. The lack of well-characterized Sau promoters for *in vitro* studies was a significant problem in the early stages of my work. PhERI did not appear to inhibit Sau RNAP at the Eco and Eco phage promoters commonly used for *in vitro* analysis. Continued work on *in vitro* transcription systems from organisms other than Eco is critical to our understanding of transcription throughout the prokaryotic kingdom, including in pathogenic organisms. The work presented in this and previous chapters demonstrates the importance of studying protein factors in a fully native system *in vitro*, using appropriate test promoters and RNAP. The promoter sequences I identified here, in particular the rrnA and dnaA promoters, appear to be valid promoters *in vitro*: I can swap promoter sequences between them with predictable results and observe RNA products of the expected size (Fig 5.3). These promoters show expected patterns of DNaseI digestion, indicating that they form normal OPCs, and again I can manipulate the promoter with predictable changes in their digestion patterns (Fig 5.4 and 5.5). These promoters should facilitate subsequent studies on Sau RNAP activity at -10/-35 promoters and should expedite further work on protein factors and small molecules that modulate RNAP activity in Sau.

Chapter 6:

Structural Studies of Sau RNAP

The work I presented in the previous chapters provides the most extensive *in vitro* analysis of transcriptional regulation in Sau to date. I identified and studied -10/-35 promoters, including the Sau rrn promoters for the first time, and show UP-element activation in Sau. These studies were significantly hindered by the paucity of biochemical and structural information on RNAP and promoter sequences from this organism.

RNAP is a known target of small molecule inhibitors. Rifampicin, which is currently used to treat *Mycobacterium tuberculosis* infections, binds to prokaryotic RNAP and inhibits RNAP synthesis (Campbell et al., 2001; Ezekiel and Hutchins, 1968). While the structure of Rif bound to Taq RNAP has been solved (Campbell et al., 2001), it and other small molecules do not actually inhibit the thermophilic enzymes, making structural studies difficult. Crystallizing Eco RNAP, or RNAP from another species inhibited by small molecules, has been a priority in the prokaryotic transcription field for decades (Twist et al., 2011a).

RNAP from different species also have varying abilities to bind and be regulated by protein and small molecule factors (Aiyar et al., 2002; Krasny and Gourse, 2004). We show that PhERI inhibition requires a fully native Sau RNAP *in vitro* transcription system (Figs 3.1 and 3.3). Work done in collaboration with a graduate student in Ann Hochschild's lab, Cristina Montero-Diaz at Harvard, identified a σ^{70} mutant that interacts

with PhERI. Even using this Eco RNAP that binds to PhERI, it was not inhibited by PhERI at the Sau rrn promoters. Only when studied at the Eco promoter rrnB promoter, known to be dependent on UP-element activation in Eco, did PhERI have an inhibitory effect (Cristina Montero-Diaz and Ann Hochschild, personal communication). I, and others, showed that Sau RNAP does not recognize all Eco promoters (Deora and Misra, 1996; Rao et al., 1995). Additional biochemical and structural data will help explain the differences in RNAP activity and regulation between species and open avenues for studies of proteins and small molecules that bind and modulate RNAP activity.

Sau is an important human pathogen and the increasing antibiotic resistance in this species is creating an urgent need for novel therapeutics (Klein et al., 2007b; Lowy, 1998; Pastagia et al., 2011). RNAP is a validated drug target and high-resolution data from Sau would provide invaluable information to design potential therapeutics.

This chapter will describe my efforts to study the structure of Sau RNAP. We initially attempted to crystallize a three-protein complex of PhERI, σ^{A}_{4} and the Sau β-flap to understand how PhERI may interact not only with σ^{A}_{4} but how it may potentially alter the interaction between σ and RNAP core. We grew crystals containing all 3 proteins that diffracted to near 4.0Å-resolution; however, due to issues with data quality and crystal twinning, we have yet to determine the orientation of the three proteins in the crystal lattice. As part of this work, we solved a 2.5Å crystal structure of the Sau β-flap by molecular replacement. While the structure is highly similar to previously solved the β-flap structures, the flap-tip helix, which interacts with σ^{A}_{4}, is in a novel conformation, revealing the extent of the flexibility of this region. Finally, I grew crystals that are likely to contain full-length Sau core RNAP. While these crystals are small, they diffracted to

better than 4.0Å-resolution and appeared to contain a unit cell volume compatible with RNAP. I have attempted to reproduce these crystals to collect data of high enough quality to solve the structure, but have been unable to regrow the crystals in the same condition.

Structural studies of a 3-protein complex containing PhERI/σ^{A}_{4}/RNAP β-flap

This work was done in close collaboration with Ann Hochschild's lab at Harvard University. The 2-hybrid studies I describe were performed by a graduate student in the Hochschild lab, Cristina Montero-Diaz. The crystallography was done in collaboration with a summer undergraduate student in the Darst lab, Lizzy Hubin.

To validate the original PhERI/σ^{A}_{4} structure, and specifically the hypothesis that PhERI does not block the interaction between σ^{A}_{4} and the RNAP β-flap (Fig 2.x), we performed bridging two-hybrid experiments. We identified a 2-hybrid interaction between Sau σ^{A}_{4} and the Sau RNAP β-flap. In the bacterial 2-hybrid system, expression of an unfused protein that disrupts this interaction can be visualized by a decrease in the transcription of the reporter gene upon its expression. As we expected, PhERI did not disrupt the interaction between σ^{A}_{4} and RNAP. However, we were surprised to find that unfused PhERI appeared to increase the expression of the reporter gene upon its exogenous expression (Fig. 6.1a). This result could indicate a direct interaction between PhERI and the RNAP β-flap or by an indirect effect in which PhERI stabilizes a conformation of σ^{A}_{4} that binds to the flap domain. A direct interaction between PhERI and the flap would not be predicted based upon our previous structural work (Fig 6.1b), but could help explain why PhERI is unable to interact with the hybrid RNAP-holoenzmye comprising Sau σ^{A} with Eco core RNAP (Figure 3.1c). We reasoned that a

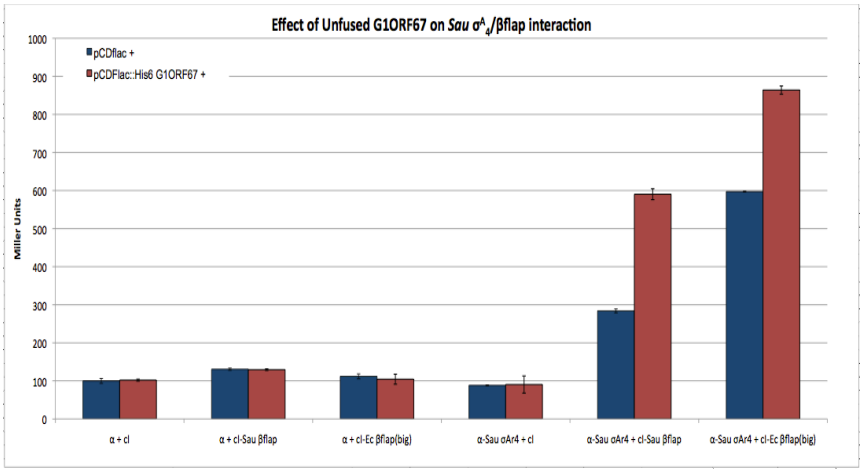
3-protein crystal structure could determine if PhERI can interact with the β-flap and therefore whether large-scale conformational shifts of PhERI, σ^A and RNAP may occur after the initial interaction between PhERI and σ^A.

We cloned and purified the Sau β-flap, as well as the complex of PhERI and $\sigma^A{}_4$. The 3-protein complex was then screened for crystallization conditions. Because the β-flap from many organisms is known to crystallize readily, we also set up trays with the Sau β-flap alone. We had previously screened for conditions in which the PhERI/$\sigma^A{}_4$ complex crystallizes (see Chapter 2). We therefore searched for conditions that gave crystals with the 3-proteins together but that produced neither crystals of the β-flap or the PhERI/$\sigma^A{}_4$ complex alone.

Microcrystals, needles and urchin-like crystals were produced in several conditions containing $MgCl_2$ and PEG4000 that were unique to drops containing all three proteins (Fig. 6.2a). 2-dimensional screens consistently produced microcrystals under these conditions, and seeding was required to produce large single plates (Fig 6.2b). To determine the protein composition of the crystals, we collected several crystals, washed extensively in mother liquor, and redissolved the crystals in 2x SDS loading buffer. After extensive heating at 95°C, the crystal solution was run on a 4-12% SDS-PAGE Phast gel. We clearly see evidence for all 3 proteins in the crystals (Fig. 6.2c).

Crystals were cryoprotected in mother liquor plus 20% glycerol and flash frozen in liquid nitrogen. The 3-protein complex crystals consistently diffracted to ~6Å-resolution (Fig 6.2d). One crystal diffracted to beyond 4.0Å-resolution (Fig 6.2e). I indexed this crystal and collected a full dataset (P222 a = 76.95 b = 84.50 c = 120.31; α = β = γ = 90°). The data were highly anisotropic, for which we corrected using the

a)



b)

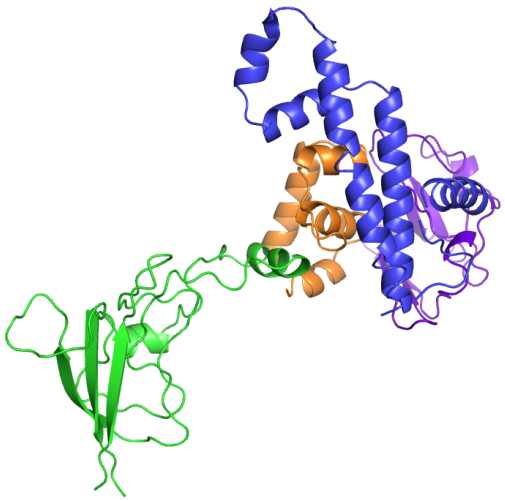


Figure 6.1: ORF67 stabilizes the interaction between σ^{A}_{4} and the RNAP β-flap. a) 2-hybrid data showing that the interaction between σ^{A}_{4} and the β-flap (blue bars) is increased upon the expression of unfused ORF67 (red bars). b) Structural model of ORF67, σ^{A}_{4} and the RNAP β-flap. The co-crystal structure of ORF67 and σ^{A}_{4} was aligned with the structure of RNAP holoenzyme (Vassylyev et al., 2002) to model the beta-flap domain bound to sigAd4. No contacts between ORF67 and the β-flap would be predicted based on this structural model. Part a) adapted from Diaz and Hochschild (personal communication).

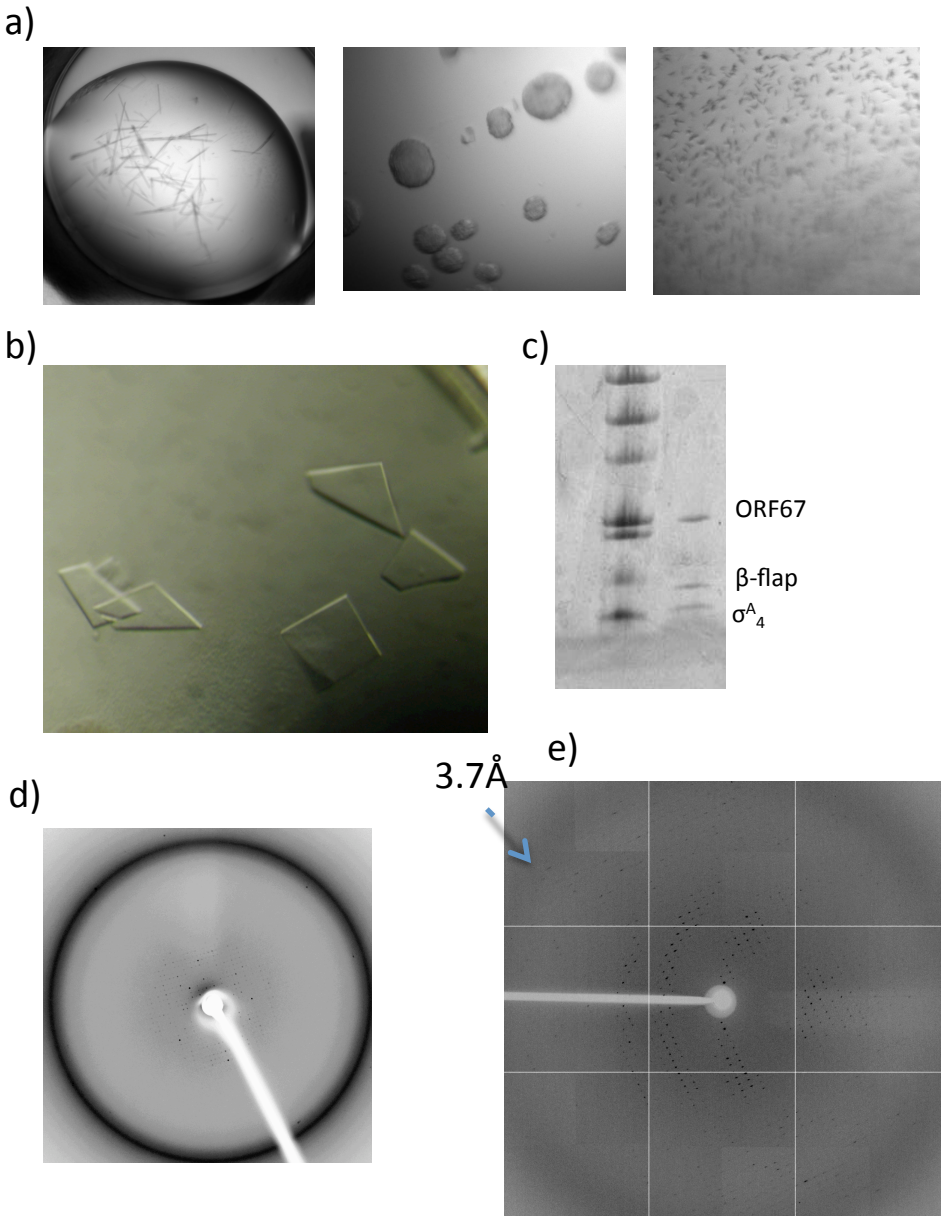


Figure 6.2: Crystallization of the 3-protein (ORF67, σ^A_4 and the Sau β-flap) complex. a) Initial hits showed needles and microcrystals. b) Final crystals after refinement of the condition and seeding. c) SDS-PAGE PhAST gel of resuspended crystals showing clear evidence of all three proteins. d) Diffraction of a typical crystal on the R-AXIS source. e) Diffraction of the one crystal that gave spots to under 4.0Å resolution. Diffraction was evaluated on the X29 beamline at Brookhaven National Laboratories (BNL) as described in the materials and methods.

diffraction anisotropy server (Strong et al., 2006) (Fig. 6.3a). We searched for a molecular replacement solution in all possible point groups using both the PhERI/σ^A_4 complex and PhERI and σ^A_4 separately. Phaser found initial solutions (TFZ = 8.8) in point group $P22_12_1$. However, initial rigid body refinement failed to reduce the R factors under 0.50 and the density was difficult to evaluate.

Due to our difficulty refining the molecular replacement model, I tested the initial dataset for quality. The data shows clear evidence for twinning (Fig 6.3b), and therefore the data were overmerged in the initial indexing. The true space group for the data appears to be P2 (a = 72.32, b = 83.23, c = 116.5; $\alpha = \gamma = 90°$ $\beta = 91.5°$). The completeness of the data was significantly reduced due to the reindexing in P2 and the resolution was cut at 4.0 Å to maintain an overall completeness near 80%. Phaser still found molecular replacement solutions with TFZ scores above 8.0. When molecular replacement searches were performed with the PhERI/σ^A_4 complex or with the proteins alone, PhERI and σ^A_4 were consistently placed in the same location in the unit cell (Fig. 6.4a). The solutions when the proteins were searched separately were consistent with the co-crystal structure of the two proteins. Density is clearly seen for the β-flap tip helix bound to the proper surface of σ^A_4 (Fig. 6.4b). Phaser placed the β-flap at multiple, but restricted, locations in the unit cell (Fig 6.4a). Low-resolution DEN refinements were further used to place the β-flap in the unit cell.

While there are clear blobs of density for the β-flap, the low resolution and poor quality of the data (anisotropic, twinned and low completeness) made placing the structure of the β-flap in the proper orientation extremely difficult. To properly determine the relative position of the 3 proteins in the unit cell, higher quality diffraction

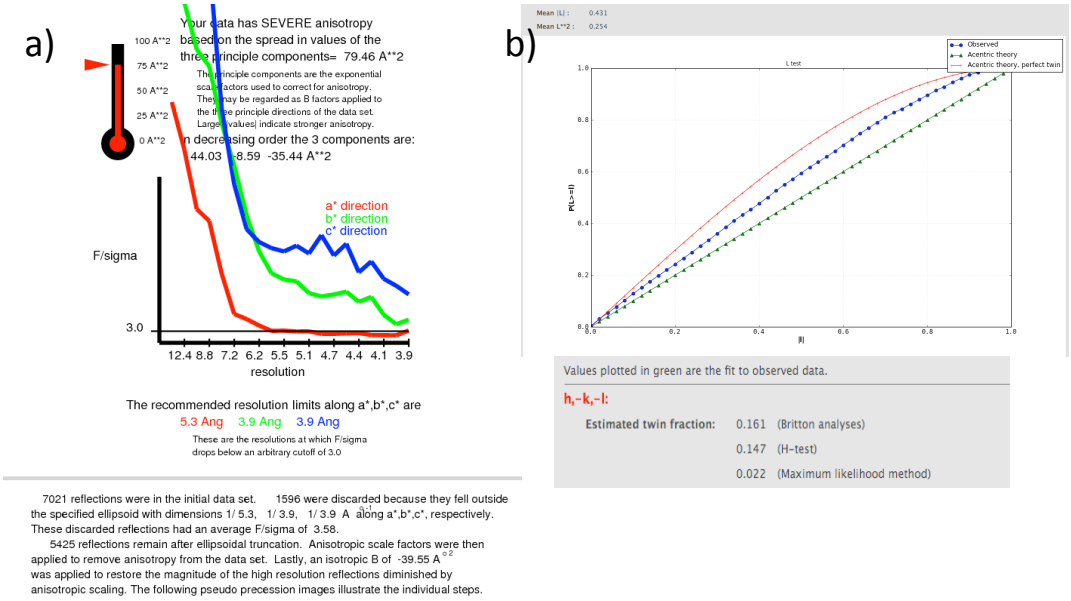
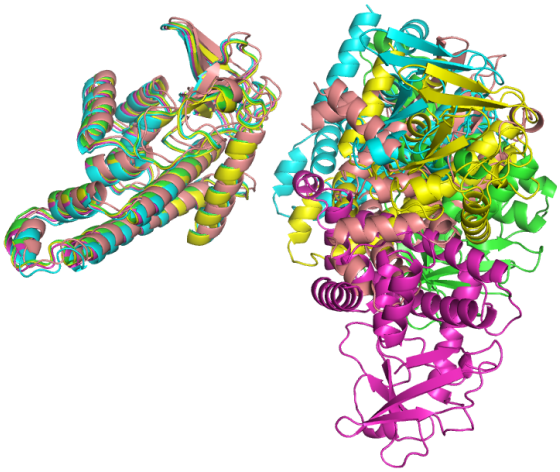


Figure 6.3: 3-protein complex crystals show high anisotropy and twinning. a) Anisotropy analysis of the diffraction data. The anisotropy served was used to truncate data not present due to anisotropy. b). Twinning analysis using Phenix (Adams et al., 2010). The crystals show clear evidence of twinning in the P2 space group.

a)



b)

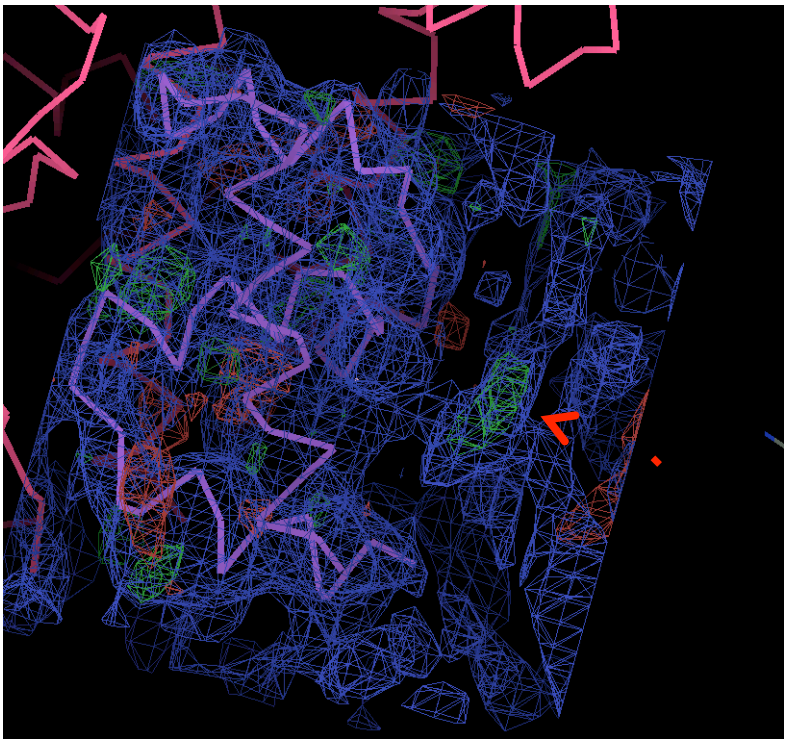


Figure 6.4: Molecular replacement solutions and initial density maps of the 3-protein complex. a) Molecular replacement solutions showing the complex of ORF67 and σ^{A}_{4} is placed consistently within the unit cell, but the β-flap is placed in multiple, but constrained locations. b) 2Fo-Fc (blue) and Fo-Fc (red/green) maps showing clear density for the β-flap tip helix (indicated by the red arrow) at its binding site to σ^{A}_{4}.

data is likely required. Data collection in the proper space group should facilitate the collection of more complete data, and correcting for the twinning of the crystals appears to allow for proper refinement of the initial models. Only one crystal out of dozens screened diffracted to better than 6.0 Å-resolution. Increasing the resolution further may require additional screening of many crystals.

Structure of the Sau β-flap

For the structural studies described above, we purified large quantities of the Sau β-flap. As a control, we also screened for crystallization conditions of the β-flap alone. The Sau RNAP, and indeed RNAP from gram-positive organisms, has not been extensively studied from a structural perspective. No structures of regions of Sau RNAP have been published. Our structure of the Sau σ^{A}_{4} (Fig 2.8) provides the first crystal structure of a region of Sau RNAP. This domain is highly conserved in all organisms and is structurally conserved in Sau. The Sau β-flap readily crystallized in many conditions, one of which (Fig 6.5a) produced crystals that diffract to 2.5Å (Fig. 6.5b) (0.1 M Sodium acetate pH 4.6, 8% (w/v) PEG 4000). Crystals were cryoprotected by a quick-dip in mother liquor plus 20% glycerol. Data were indexed collected and indexed ($P3_121$, a = 92.651 b= 92.651, c= 129.971; $\alpha = \beta = 90°$, $\gamma = 120°$) at X29A at Brookhaven National Laboratories. We searched for molecular replacement solutions in both Phaser and MolRep using the Taq β-flap domain with all sidechains and the flexible flap-tip helix removed. Likely solutions were found in both programs (TFZ > 8.0). Evaluation of the initial electron density showed clear evidence for the β-flap tip helix that was not present in the search model (Fig 6.5c).

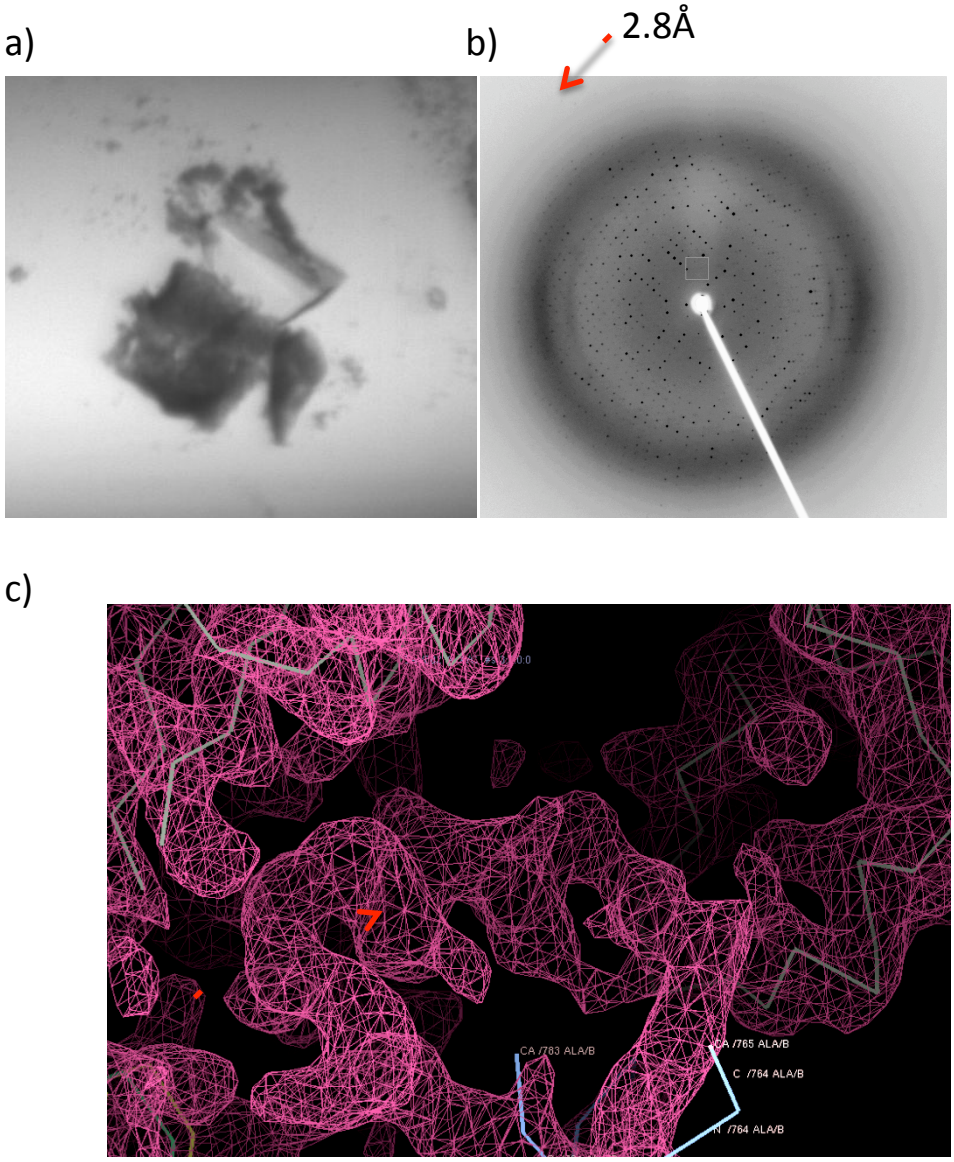


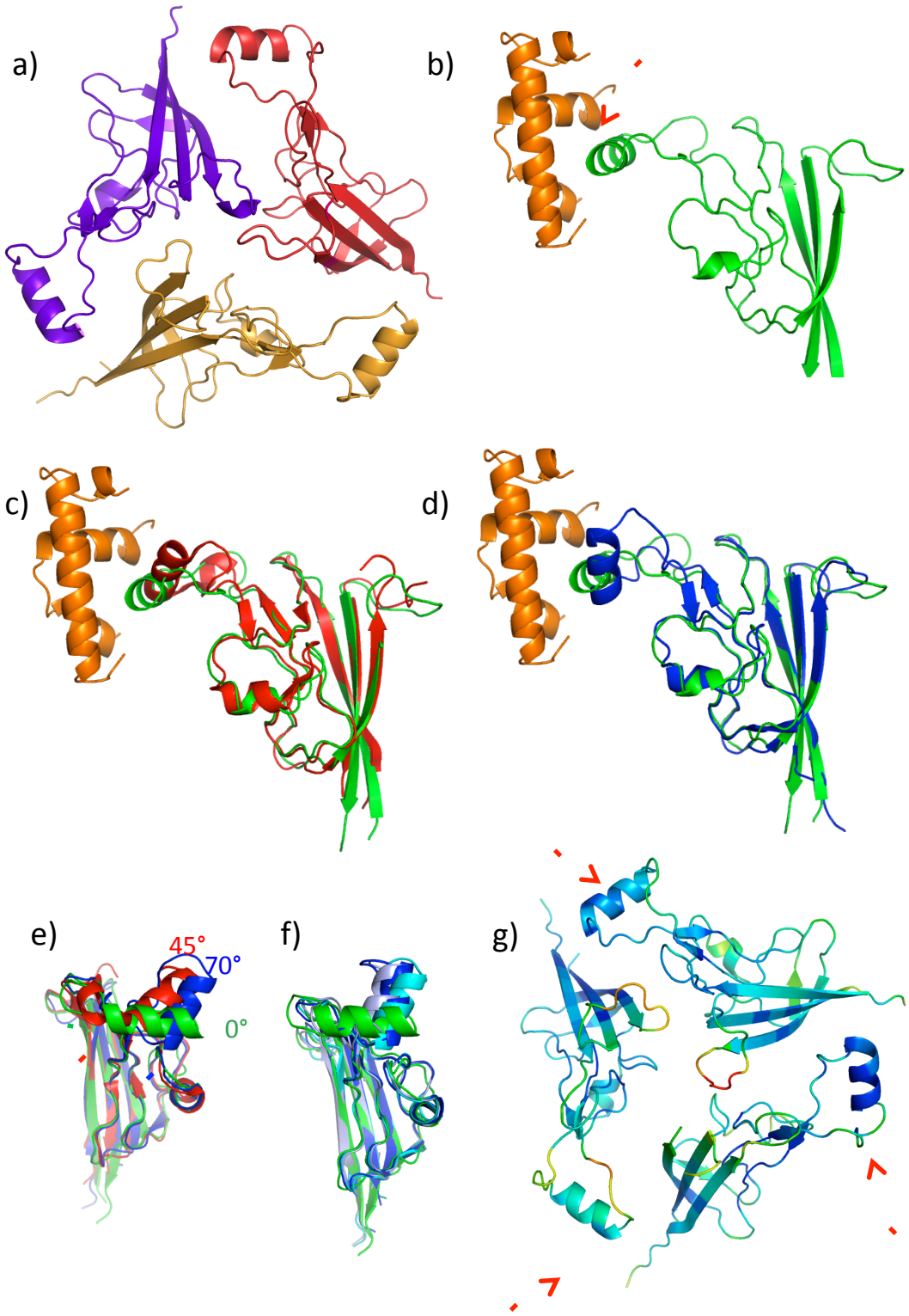
Figure 6.5: Crystallization of the Sau β-flap. a) β-flap protein crystals. b) Diffraction pattern collected on the R-AXIS source. Diffraction is evident to 2.8Å. Diffraction was initially evaluated on an R-AXIS Xray source. Data for the final structure were collected on beamline X29 at Brookhaven Natinoal Laboratories. c) 2Fo-Fc maps from the molecular replacement solution showing clear evidence for the β flap-tip helix (red arrow), not present in the search model.

The β-flap crystals contained three monomers in the asymmetric unit (Fig 6.6a). The three individual structures were quite similar to one another (Rmsd = 1.1 and 1.2) and also highly similar to previously solved structures of the β-flap from other organisms (Rmsd = 1.3 with the flap-tip helix excluded from the Eco β-flap). The β-flap tip helix is known to be flexible (Twist et al., 2011b). This helix is responsible for interacting with, and organizing σ^{A}_{4} such that it can properly recognize the major grove of the -35 promoter element (Murakami et al., 2002b; Vassylyev et al., 2002) (Fig 6.1b and 6.6b). This helix is the target for proteins that regulate RNAP activity, including the phage protein gp33 that binds to the flap-tip helix (Twist et al., 2011b).

The flap-tip helix in our structure is in a conformation incompatible with an interaction with σ^{A}_{4} (Fig 6.6d). This helix is maintained in its conformation due to extensive crystal packing contacts (Fig 6.7), both within one asymmetric unit and between asymmetric units. Compared to previous structures compatible with σ^{A}_{4} binding (Fig 6.6b), the flap-tip helix is rotated 70° (Fig 6.6 d and e). It is similarly distorted in a co-crystal structure with the Eco phage protein gp33 (Fig 6.7c and e). Our structure shows even more distortion than when the helix is stabilized by a phage transcriptional regulator. This is clear evidence of the extent of the flexibility of this helix in solution. While it can be maintained in a stable conformation by binding other proteins, such as σ or gp33, crystal packing alone can capture a functionally distorted conformation of the β-flap.

Our structure of the Sau β-flap provices evidence that the protein in solution samples a wide variety of conformations, one of which is captured and stabilized by interaction with σ^{A}_{4} or other protein factors. The interactions required to form the RNAP

Figure 6.6 (adjacent page): Structure of the Sau β-flap. a) Crystallographic trimer of Sau β-flap. b) interaction between the β-flap domain (green) and σ^{A}_{4} (orange) from the Tth holoenzyme structure (Vassylyev et al., 2002). c) Structure as shown in b overlayed with the structure of the Eco β-flap bound to the phage protein gp33 (red) (Twist et al., 2011). The β flap-tip helix bound gp33 is in a conformation incompatible with σ^{A}_{4} binding. c) Structure as in b aligned to the Sau β-flap (blue) showing the flap-tip helix in a further contorted conformation. d) Overlay of the Tth β-flap in the conformation required for σ^{A}_{4} (green), the Eco β-flap bound to gp33 (red) and the Sau β-flap (blue). e) Structure of the Tth β-flap in the conformation bound to σ^{A}_{4} (green) and the 3 structures of the Sau β-flap present in the asymmetric unit (blue). f) Structure of the crystallographic trimer of the Sau β-flap colored by B-factor. The flap-tip helices are identified with red arrows.


a)
b)
c)
d)
e)
45°
70°
0°
f)
g)

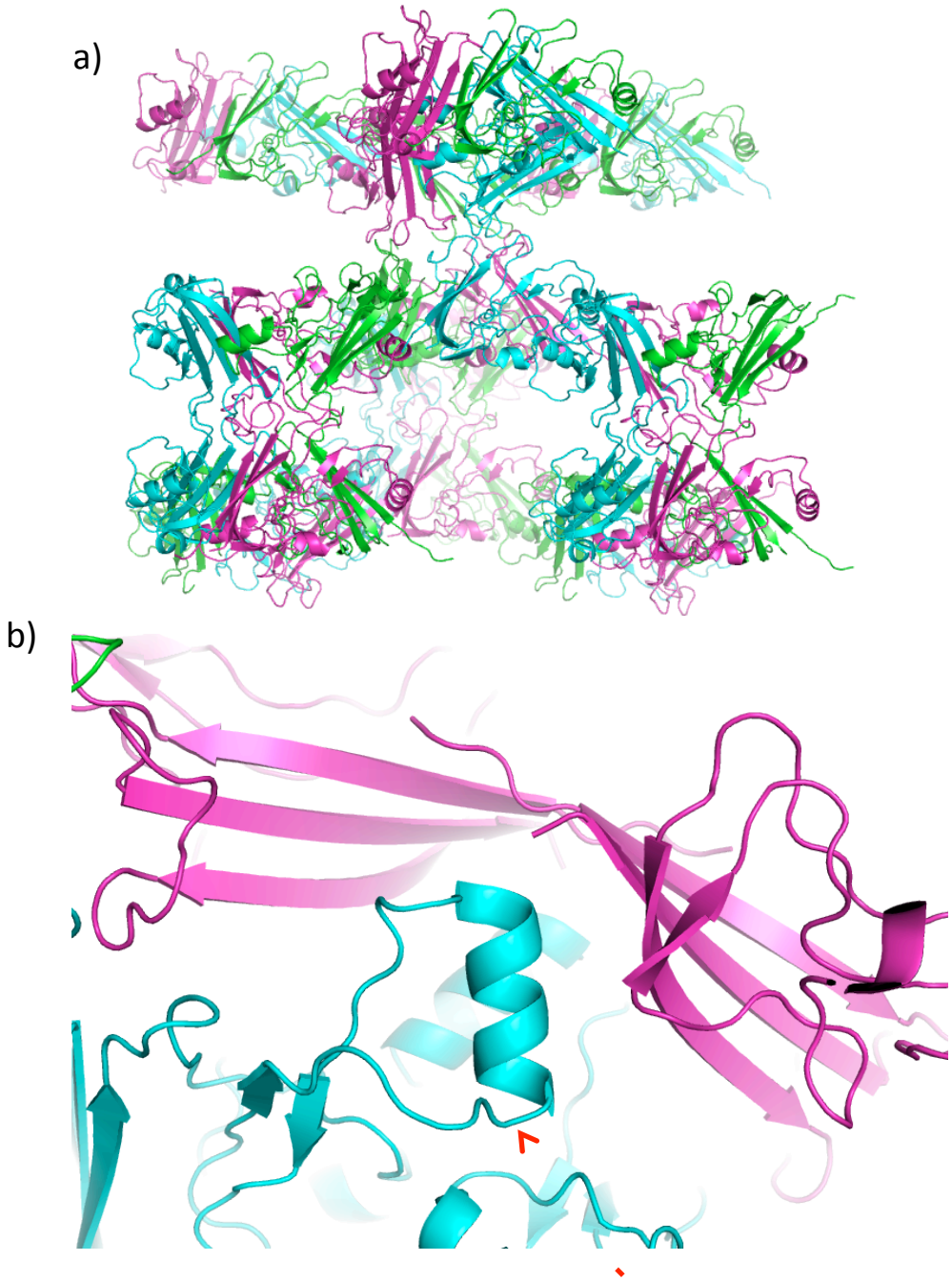


Figure 6.7: Packing of the Sau β-flap crystals. a) Crystal packing of the trimer (cyan/pink/green). b) The β flap-tip helix from one asymmetric unit (cyan, indicated with a red arrow) interacts with a crystal packing neighbor (pink).

holoenzyme, including the specific contacts between the β-flap and σ_4, occur as one of the first kinetic steps in the transcription cycle, prior to promoter specific DNA recognition and melting. Regulation of this step would have dramatic effects on RNAP activity at a wide variety of promoters in prokaryotic cells. This conformational flexibility of the β-flap is a site of regulation by proteins that not only sterically block σ binding but may also maintain the β-flap in a conformation incompatible with σ recognition (Berdygulova et al., 2012; Deighan et al., 2008; Twist et al., 2011b; Yuan et al., 2009).

Structural studies of the Sau RNAP core enzyme

The structures of proteins and biological molecules can be studied by a variety of techniques. Both electron microscopy (EM) and X-ray crystallography have been used to study the structures of prokaryotic RNAPs (Campbell et al., 2002; Murakami et al., 2002b; Polyakov et al., 1995). EM technologies produce structures of moderate resolution (Polyakov et al., 1995) and cannot be used to study RNAP at atomic resolution for small-molecule binding studies or the placing of individual amino acids in the density maps. X-ray crystallography, which is capable of producing electron density maps giving atomic level detail of protein structure and allowing for unambiguous placement of protein residues within the density maps, is limited by our ability to grow crystals of a protein or complex that diffract to sufficiently high resolution. Crystallization of a protein generally requires a large amount (10mgs or more) of pure protein. To date, the only structures of prokaryotic core and holo RNAPs that have been solved to atomic resolution come from thermophilic bacteria (Murakami et al., 2002b; Vassylyev et al.,

2002). X-ray crystal structures of RNAP from additional organisms would provide invaluable information about the regulation and function of this important class of enzymes.

Sau RNAP is inhibited by a small molecule inhibitor

One of the driving forces behind the work to crystallize non-thermophilic RNAPs is that the thermophilic enzymes are not effectively inhibited by many small molecule drugs that target RNAP function. Many small molecules are potent inhibitors of Eco RNAP and therefore crystals of Eco RNAP would allow for co-crystallization studies with the small molecules to determine their binding sites and mechanisms of inhibition. Small molecule inhibitors have not been tested *in vitro* using Sau RNAP. Before attempting to crystallize Sau RNAP for subsequent structural studies, I wanted to determine whether this enzyme is inhibited by small molecules that are active against Eco RNAP. I tested the activity of one Rifampicin derivative, Rifalazil, on Sau RNAP activity at three promoters (Fig 6.8). This small molecule does not effectively inhibit the thermophilic enzymes but is active against Eco RNAP. I show that it is an inhibitor of Sau RNAP at all three promoters (Fig 6.8), indicating that an RNAP core crystal structure would likely open up many avenues for soaks or co-crystalization with small molecule regulators and inhibitors of RNAP function.

Crystalization of Sau RNAP

For our biochemical studies, I prepared highly pure Sau RNAP (Fig. 3.2). An attempt to crystallize the Sau RNAP required only a scale-up of the small scale RNAP

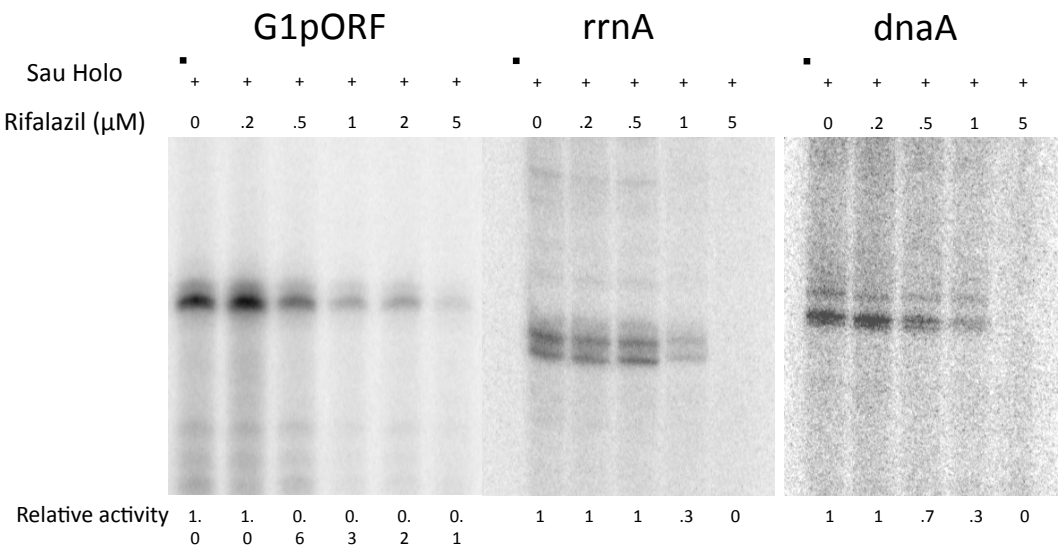


Figure 6.8: The Sau RNAP is inhibited by the antibiotic Rifalazil. Transcription assays were performed using Sau holoenzyme on 3 Sau specific promoters identified in chapters 3 and 4. RNAP core (50nM) was added to Sau σ^A (100nM); linear promoter DNA (50nM) was added to the reaction in 1x Sau transcription buffer. Rifalazil, added to the reactions prior to initiation at the indicated concentrations, inhibits RNAP activity.

purification. Because I purify untagged Sau RNAP natively from cells, a large-scale purification would require a large amount of biomass. Through a fermentor facility, I obtained 300 L cultures of Sau NCTC8325 to saturation, producing ~300g of solid biomass. I purified Sau RNAP from ~150g of biomass for subsequent structural studies. Following the same purification protocol (Fig 3.2) (polymin P and ammonium sulfate precipitations followed by purification on heparin, sephadex-200 and S columns), producing ~5mgs of Sau RNAP. The sample was concentrated to 12mg/ml and screened extensively for crystallization conditions.

One condition (0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$, 30% PEG400) produced small protein crystals after approximately 14 days (Fig 6.9a). I managed to reproduce these crystals with Sau RNAP from the same purification, but was limited by the quantity of sample that remained. Crystals were cryoprotected by a quick dip in the mother liquor with the addition of 5% glycerol and flash frozen in liquid nitrogen. Due to the small size of these crystals (5μM), we evaluated their diffraction on the microdiffractomoter at the Advanced Light Source at Argonne National Laboratories. Of the 6 total crystals we screened, only one showed evidence of diffraction. This crystal, however, clearly diffracted to beyond 4.0 Å-resolution (Fig 6.9b). While the quality of the spots was good, due to extensive ice formation on the loop, we attempted to anneal the crystal. After annealing, the spots appeared streaky (Fig 6.9b). Although the data quality after annealing was quite poor, we indexed the crystal and collected a dataset.

Crystals of protein purified natively from cells may not contain the protein of interest, but rather be composed of a crystallizable impurity. RNAP, due to its large size, produces crystals with characteristically large unit cells. Although a large unit cell is not

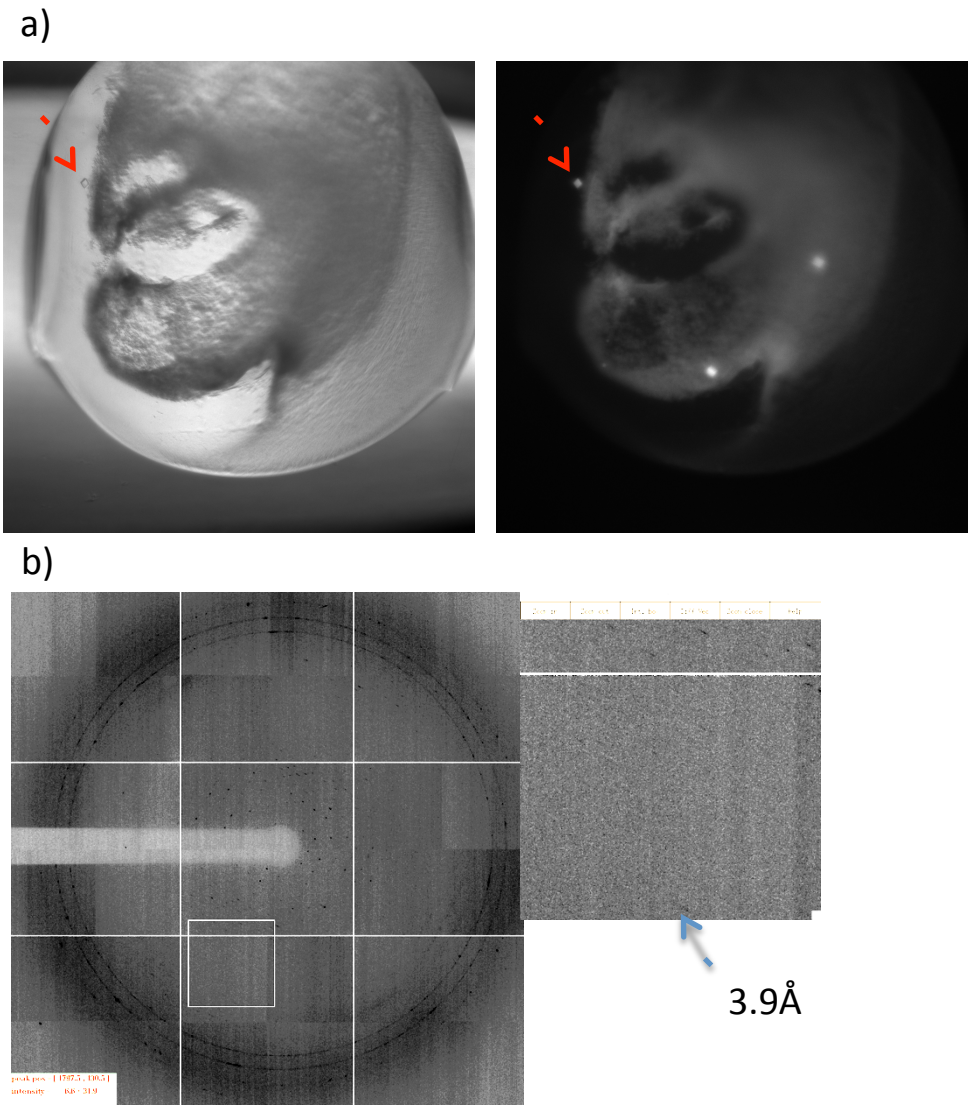
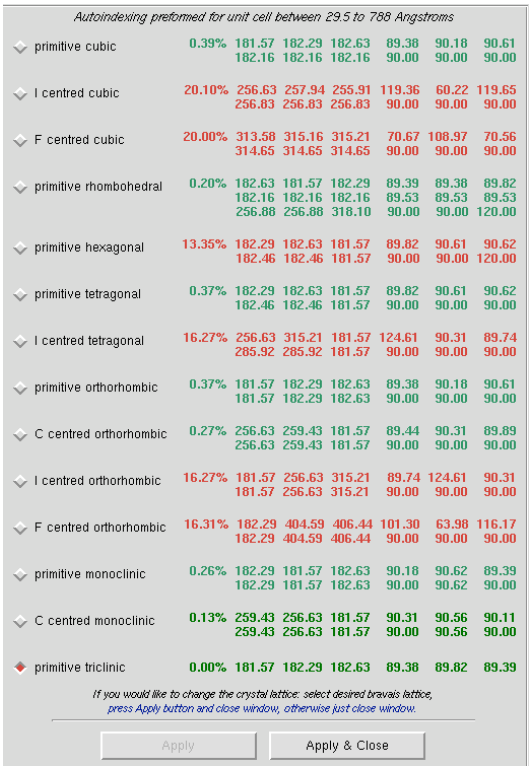


Figure 6.9: Crystallization of Sau RNAP. a) Putative Sau RNAP protein crystals. Left image is a brightfield, right is UV absorption. One crystal is indicated by a red arrow in both images. b) Diffraction pattern of putative Sau RNAP crystals. Diffraction was evaluated on the 24-ID beamline at the Advanced Photon Source (APS) as described in the materials and methods.

fully indicative of a crystal containing RNAP, it is a good indication that the crystal is at least compatible with RNAP crystal packing in the unit cell. I was able to index the Sau RNAP crystals to evaluate their unit cell. While the data is compatible with many different space groups (Fig 6.10), many of which consist of unit cells large enough to contain RNAP, an initial analysis of the statistics of indexing of the data in all the possible space groups identified P4 as the likely correct space group. The unit cell in P4 ($a = 182.46$, $b = 182.46$, $c = 181.57$; $\alpha = \beta = \gamma = 90°$) could accommodate one Sau RNAP molecule (MW ~ 350kD) with a solvent content of 70.4% (Fig 6.10). Because of the poor quality of the data, attempts to scale the dataset failed and molecular replacement to solve the phases could not be attempted.

Given the promising initial diffraction data, I attempted to repeat the large-scale Sau RNAP purification and produce additional crystals under this same condition. I also tried to improve the crystals, in particular their size, by screening additives to the crystallization. Subsequent studies could use micro or macro seeding to improve crystal size and quality. However, I was never able to reproduce crystals under the same condition from subsequent large-scale RNAP purifications. RNAP purified from subsequent fermetor batches proved not only to be uncrystallizeable, but also completely inactive biochemically. Although the purification appeared to proceed normally, either the RNAP was inactived by small molecule(s) in the sample or conformationally unable to undergo transcription. We hope to resolve the issues in the fermentor growth and/or RNAP purification to again produce large amounts of highly active RNAP suitable for crystallization. Given the promising initial diffraction by the putative Sau RNAP crystal,



P4
1 Sau RNAP (~350kD)
Solvent content: 70.4%

Figure 6.10: Indexing of putative Sau RNAP crystals. Indexing was performed using HKL2000, and the statistics of indexing were used to determine the likely space group. χ^2 values of space groups higher than P4 were elevated.

we believe that Sau may be a suitable organism for high-resolution studies of mesophilic RNAP.

Chapter 7:

Structure and Function of Tth CarD

Mycobacterium tuberculosis (Mtb) is a pathogenic bacterial species and is the primary causative agent of tuberculosis, a serious infection of the lungs. More than 30% of the world's population is currently infected with latent Mtb and Mtb reactivation in this subset of the population is responsible for 1.3 million deaths a year (WHO, 2009). Mtb infects tissues in the lung where it invades the host macrophages where it evades the immune response to continue to divide. As with Sau, multi-drug resistance in Mtb makes infections difficult and costly to treat (Heep et al., 2000; Ramaswamy and Musser, 1998). Mtb cells divide only once every 15-20 hours, as opposed to Eco and Sau cells which divide every 20-30 minutes (Stallings and Glickman, 2011). This slow growth rate has made Mtb difficult to study in the laboratory, and it makes Mtb difficult to fully eliminate from patients.

Mtb's ability to persist in patients for years and even decades depends on a complex cellular response to the anaerobic conditions and various stressors inherent to the host immune response. The stringent response, outlined in the introduction, adjusts the prokaryotic growth regime from rapid growth and cell division when resources are abundant to the stress response required to survive when resources are scarce (Gourse et al., 1996; Traxler et al., 2008). The stringent response, including the downregulation of rRNA synthesis, is thought to be important to the ability of Mtb to persist in the stressful environment of the human host (Stallings and Glickman, 2011).

CarD was identified in a high throughput screen for genes upregulated by the DNA damage response in Mycobacterium smegmatis (Msm) (Stallings et al., 2009), a model organism closely related to Mtb but with a faster doubling time (and therefore more amenable to laboratory study). The original work on CarD showed that it is essential for cellular survival, critical for the response to oxidative stress and DNA damage, and is a negative regulator of the transcription of ribosomal components (Stallings et al., 2009). CarD contains a conserved RNAP Interacting Domain (RID), similar to the RID that mediates the interaction between the transcription repair coupling factor (TRCF or Mfd) and RNAP (Stallings et al., 2009). While the N-terminal region of CarD contains the RID, the C-terminal region has no apparent sequence homology to any other protein family.

CarD pulls down RNAP subunits in Msm and interacts with the RNAP β subunit, the site of TRCF-RID interaction, by 2-hybrid assays. Because CarD binds to RNAP and appears to regulate rRNA transcription, it was hypothesized to be a regulator of the stringent response, similar to DksA in Eco (Stallings et al., 2009) (see Chapter 1). CarD is widely conserved in prokaryotic species, with homologs in Bsub, the thermophilic bacterium Thermus Thermophilus (Tth) and Myxococcus Xanthus (Stallings and Glickman, 2011; Stallings et al., 2009). Unfortunately, no homolog exists in Eco, and therefore the rich genetic and biochemical tools that have been developed in Eco for studying RNAP regulation have not been accessible to examine CarD function.

CarD is essential in Mtb and Msm, which has complicated studies of its function (Stallings and Glickman, 2011; Stallings et al., 2009). DksA, which regulates rRNA synthesis at the transition into stationary phase, is not essential in Eco (Paul et al., 2004),

and therefore it has been hypothesized that while CarD binds to RNAP, it may have some other required function relating to DNA replication and DNA damage (Stallings and Glickman, 2011). Specifically, because rRNA operons are so highly expressed in log-growing prokaryotic cells, collisions between RNAP and the DNA replisome are likely to occur at these loci. Such collisions between RNAP and the replisome are generally bypassed by the DNA replication machinery; however, if RNAP is not removed and the replisome cannot clear the collision, DNA replication is disrupted and the DNA damage response is activated. Because collisions of RNAP and the replisome traveling in the same direction are more easily bypassed that head-on collisions, it is generally assumed that rRNA operons are oriented in the direction of DNA replication to alleviate the probability of deleterious RNAP/replisome collisions (Stallings and Glickman, 2011). RNAP binding proteins, including TRCF and DksA, have been shown to help remove RNAP from stalled complexes, eliminating replisome roadblocks (Pomerantz and O'donnell, 2008, 2010; Tehranchi et al., 2010).

Mtb only has one rRNA operon from which all ribosomal RNAs must be transcribed (Stallings and Glickman, 2011). In contrast, Eco contains 7, Bsub 10 (Stallings and Glickman, 2011) and Sau 5 (Baba et al., 2008; Wada et al., 1993). Whether CarD is essential or not appears to be correlated with the number of rRNA operons in an organism: organisms with fewer rRNA operons, such as Mtb and Msm, require CarD while Bsub does not (Stallings and Glickman, 2011). Whether CarD is directly modulating RNAP activity at promoters, and whether modulation of RNAP activity is its required function, remain unclear.

Subsequent work showed that while CarD depletion leads to an upregulation of rRNA and transcription related proteins, mutating the CarD-RID to disrupt its interaction with RNAP leads to a decrease in RNAP activity at rRNA promoters (Stallings, CL, personal communication). These seemingly disparate results may point to multiple important roles of CarD in cells, some relating to its interaction with RNAP and others perhaps independent of this interaction.

In this chapter, I will present our work to understand the structure and function of CarD directly on RNAP at promoter sequences. The work presented here is a collaboration between our group and Christina Stallings at Washington University, and the *in vivo* data that I will briefly discuss come from Christina's lab. The structure of CarD, as well as the modeling of CarD onto RNAP, was performed by a post-doc in the Darst lab, Devendra Srivastava, and Elizabeth Campbell, who was essential to the project throughout. I will largely focus on my own work to describe the direct function of CarD on RNAP activity.

Structure of Tth CarD

Thermophilic organisms can produce proteins with unique characteristics. Thermophilic enzymes tend to be optimally active at temperatures at which many proteins from non-thermophilic organisms denature, can be uniquely stable (Kelch and Agard, 2007), and are often well suited for crystallographic studies. Because CarD has a homolog in Tth, Devendra initially attempted to solve the crystal structure of both Tth and Mtb CarD. He also attempted to co-purify CarD and the RNAP β-1 domain to study the structure of the complex.

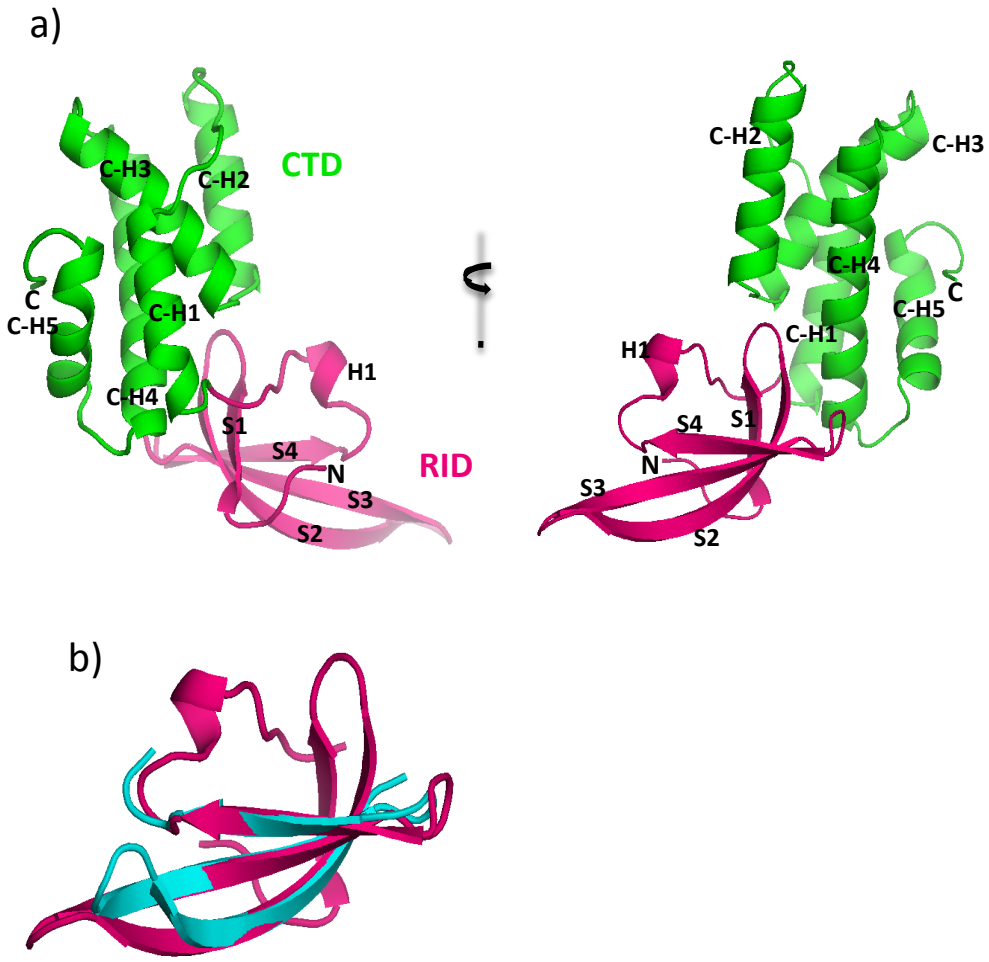


Figure 7.1: Structure of Tth CarD. a) CarD contains two domains, an N-terminal RNA polymerase Interacting Domain (RID, pink) and a C-terminal domain (green). The helices and sheets are numbered as in the sequence alignment in Fig. 7.3. b) CarD RID (pink) is structurally homologous to the previously solved Transcription Repair Coupling Factor (TRCF) RID (cyan). RMSD = 1.45 (C(alpha)).

The Tth CarD crystal structure (Fig. 7.1) was solved by Devendra Srivastava and Elizabeth Campbell. The structure reveals that the CarD-RID is similar to the Mfd-RID (Fig. 7.1b), arguing that the two proteins likely interact with RNAP in a conserved fashion (RMSD = 1.45). While the CarD-CTD is connected to the RID by a linker, and therefore could be flexible with respect to the RID and the RNAP interaction, extensive inter-domain interactions between the RID and the CTD (Fig. 7.2) are likely to stabilize CarD in the structure we visualize. Because of the relatively high conservation between Tth and Mtb CarD, we were able to make a structural model for the Mtb CarD based on the Tth data (Fig. 7.3).

The structure of the TRCF-RID has been solved in complex with the RNAP β1 domain. CarD has a structurally homologous RID and also interacts specifically with the β-1 domain of RNAP (Stallings et al., 2009). Therefore, we can use the previous structures of a RID bound to RNAP to model CarD onto RNAP. Because the RID interacts directly with the RNAP β-subunit, part of the catalytic core of the enzyme that is present throughout the transcription cycle, it is unclear whether CarD should be modeled onto structures of RNAP holoenzyme at promoter DNA or RNAP core bound to elongation substrates. CarD can be accommodated by both the RNAP holoenzyme open promoter complex (OPC) (Fig. 7.4a) and the RNAP elongation complex (Fig. 7.4b).

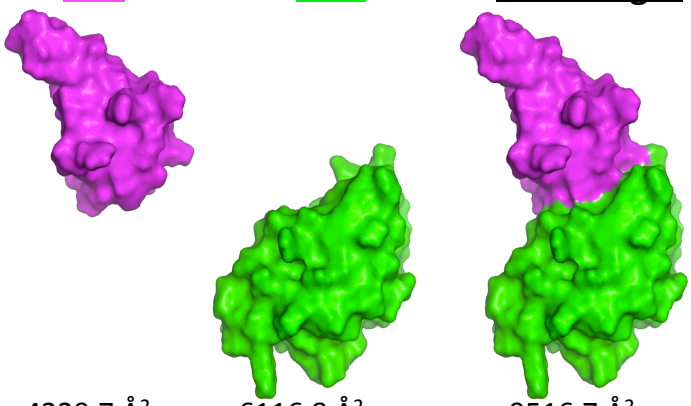
To determine at which point in the transcription cycle CarD interacts with RNAP, and dissect which kinetic steps to test for modulation by CarD, work in Christina Stalling's lab used ChIP-Seq. CarD was HA-tagged in Msm, growing cells were crosslinked, and lysates were collected and passed over a column that binds to the HA tag or contains an RNAP specific antibody. After elution from the column, crosslinks were

Figure 7.2 (Adjacent page): CarD RID and CTD form extensive interdomain contacts. a) Interaction map showing specific interactions between residues on the RID and the CTD. Ionic interactions are shown in red, hydrogen bonds in blue and hydrophobic interactions in black. b) Buried surface area analysis. The surface area of the CarD-RID, CarD-CTD and full length CarD (upper row) were used to calculate the buried surface area between the two domains (829.8 $Å^2$).

a)

RID	CTD
K9	
V11	D99
P13	G100
P14	P102
G16	R131
V17	E138
G18	E139
V19	Q142
R60	S143
A63	E145

b)



Total surface area: 4229.7 Å^2 6116.8 Å^2 9516.7 Å^2

Buried surface area: 829.8 Å^2

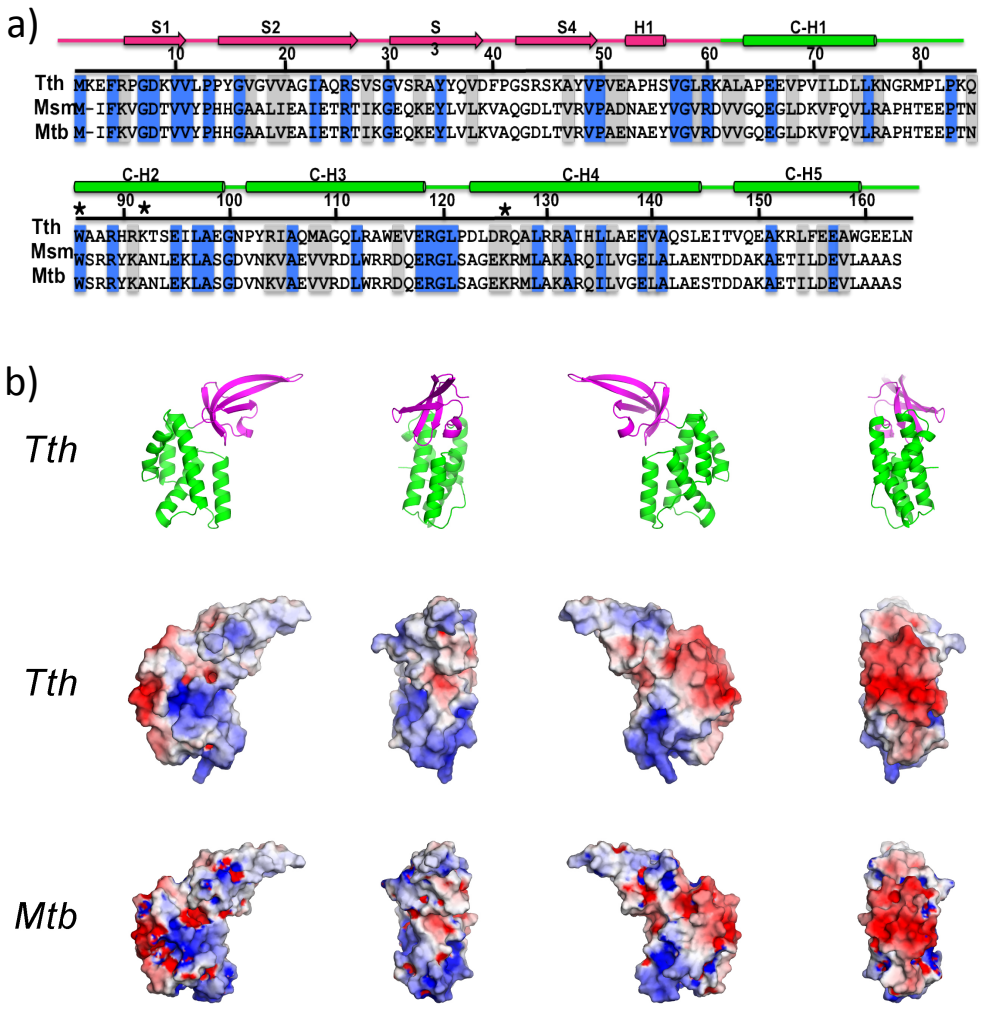


Figure 7.3: Structural modeling of Mtb CarD. a) Alignment of Tth, Msm, and Mtb CarD with the strands and helices numbered corresponding to Fig. 7.1. b) Structural model of Mtb CarD based on the structure of the Tth protein. Top panel is the structure of Tth CarD shown as a cartoon and colored as in Fig 7.1. Middle panel is the experimental structure of CarD shown as a surface representation colored by electrostatic potential (negative charge = red; positive charge = blue). Bottom panel is the Mtb structural model based shown as a surface representation and colored by electrostatic potential as above. Both the Tth and the Mtb CarDs contain a positively charged path of amino acids at the tip of the CTD.

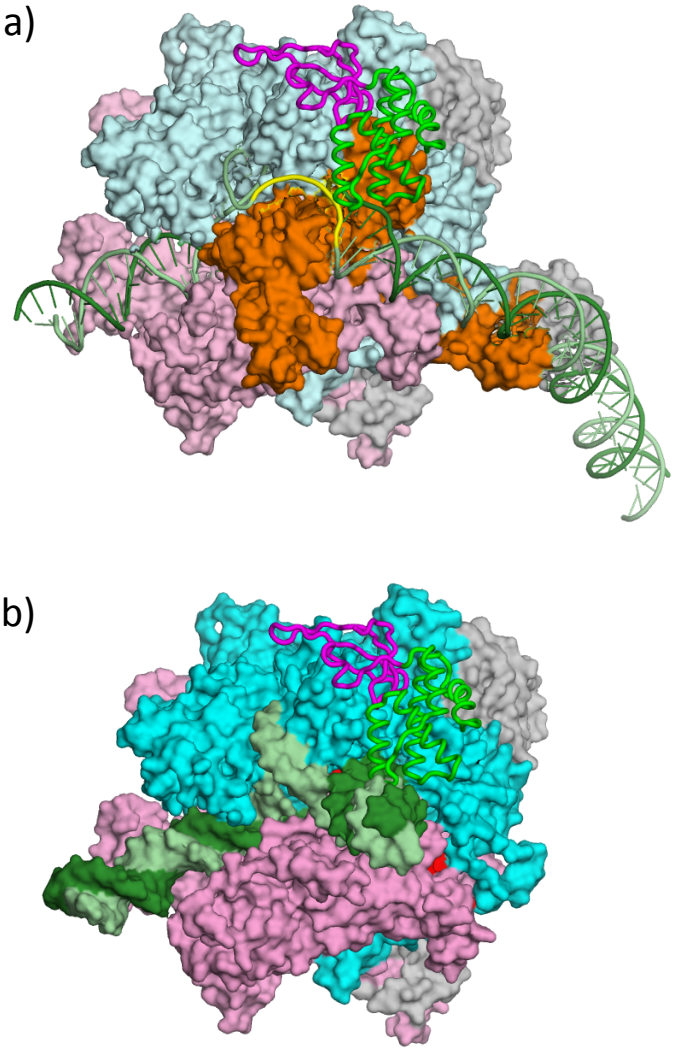


Figure 7.4: Model of CarD interacting with RNAP. a) Structural model of CarD bound to the RNAP open promoter complex. CarD was modeled onto RNAP by alignment with the structure of the TRCF RID bound to the β-1 lobe (Westblade et al., 2010). σ^A is shown in orange, β in cyan and β′ in pink. CarD is represented as a ribbon structure and colored as in Figure 7.1. b) Structure of CarD modeled onto a transcription elongation complex (Vassylyev et al., 2007) by aligning the CarD RID as above.

reversed and bound DNA sequences were determined by high-throughput direct sequencing. While RNAP is bound to DNA throughout the length of genes and operons (Fig 7.5, red lines) as expected, CarD appears to interact with DNA at promoter sequences (Fig. 7.5 blue lines) and to correlate to regions of DNA where σ is bound (Fig 7.5, green lines). Based on these results, CarD appears to associate with RNAP at promoter sequences where σ is present.

CarD stimulates RNAP activity at Tth rrn promoters

The results of experiments examining CarD function(s) *in vivo* have produced seemingly contradictory results. While it is clear that CarD interacts with RNAP (Stallings et al., 2009), the direct effects on RNAP activity are unclear. Because CarD may have multiple functions *in vivo*, only some of which involve RNAP binding, we decided to test its function on RNAP in a purified, *in vitro* transcription system. The work from the previous chapters illustrates the importance of studying transcription factors in their native systems. I tested Tth CarD using Tth core RNAP, Tth σ^A and Tth promoters. CarD appears to alter rRNA transcription in Mtb and Msm *in vivo*; I therefore assayed CarD's activity on the Tth rRNA promoters *in vitro*.

CarD has a robust stimulatory effect on rRNA transcription *in vitro* (Fig. 7.6). Two Tth rRNA promoters have been previously studied, the 16s and 23s rRNA promoters (Fig. 7.6a). CarD stimulates RNAP activity at both promoters (Fig. 7.6b). RNAP activity at low temperatures (42°C) required the use of a truncated version of Tth σ^A with region 1.1 deleted (σ^A-Δ1.1). Region 1.1 is auto-inhibitory; in the absence of core RNAP, it interacts with σ_4 to prevent DNA binding. We also tested full-length σ^A at

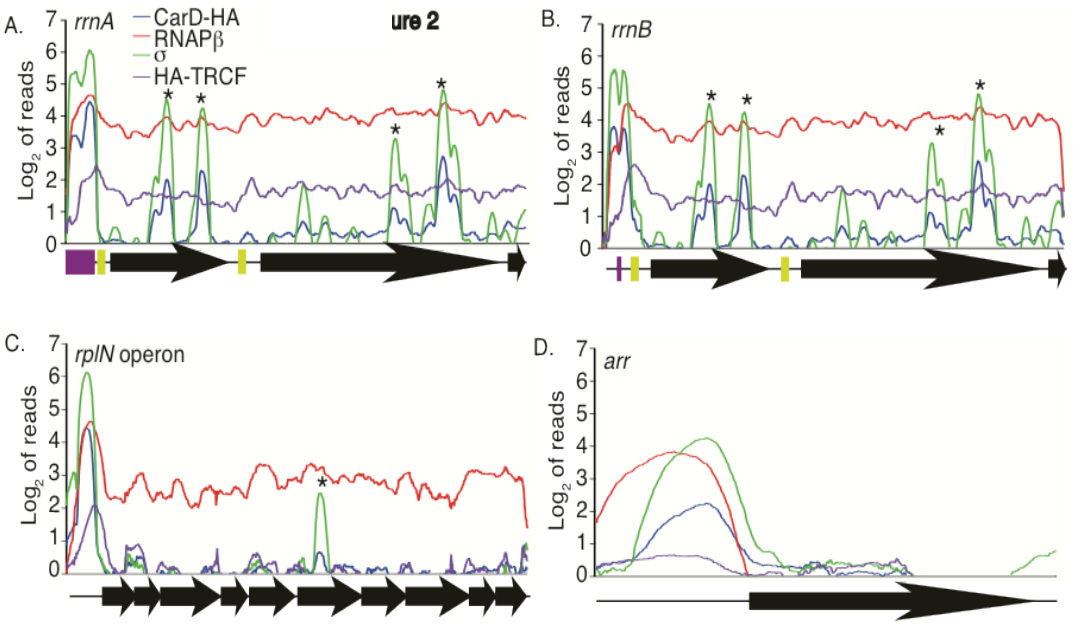


Figure 7.5: CarD binds to transcription initiation complexes. Protein nucleic acid complexes containing CarD-HA, HA-TRCF, and HA were immunoprecipitated from Msm with a monoclonal antibody specific for HA. RNAP and σ were immunoprecipitated with monoclonal antibodies specific for these subunits. Co-precipitated DNA was sequenced and the number of sequence reads for every base pair was normalized to the total DNA co-precipitated and expressed as a log2 value. The number of reads co-precipitated with HA alone served as the background and was subtracted from the other samples. The legend in panel A is the same for all the panels, genes are designated with black arrows, and annotated promoters and antitermination boxes are purple and yellow boxes respectively.

a)

Tth 16s promoter CTCGCAAGCCTTGACAAAAAGGAGGGGGATTGATAGCATGGCTTTTCTGCG

Tth 23s promoter GGGGGCCCTTGACAAAGGCCATGCCTCCTTGGTATCTTCCCTTTTGCGCTGC

T7A1 promoter CAAAAAGAGTTGACTTAAAGTCAACCTATAGGATACTTACAGCCATCGAGAG

b)

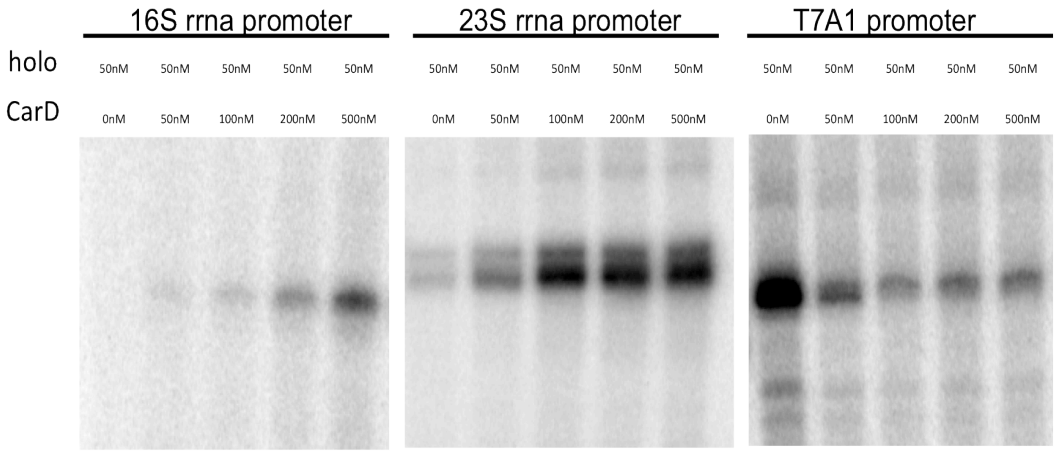


Figure 7.6: CarD stimulates RNAP activity at rRNA promoters but inhibits transcription at the T7A1 promoter. a) Sequences of promoters used in subsequent experiments. -10 and -35 elements are highlighted with red boxes. b) Open promoter complexes were formed by first incubating CarD (200nM) with RNAP holoenzyme (50nM, containing $\sigma^{A}\Delta$N1.1), adding linear template DNA (50nM) and incubating at 65°C as described in the methods. Varying concentrations of CarD (0, 50, 100, 200 and 500nM) were added to RNAP core, and reactions initiated by the addition of NTPs. After five minutes, reactions were stopped and run on a Urea-PAGE gel.

high temperature (65°C) and confirm that CarD stimulates RNAP activity at the two rRNA promoters regardless of the presence of region 1.1 (Fig. 7.7). These results are the first evidence that CarD has a direct effect on RNAP activity at promoters.

CarD inhibits RNAP at the T7A1 promoter

To determine whether the effect of CarD on rRNA promoters is specific to this class of promoter, we also tested CarD's activity on the -10/-35 Eco phage promoter T7A1, which has been extensively studied *in vitro*. Interestingly, while CarD stimulates transcription from rRNA promoters, it inhibits RNAP activity at the T7A1 promoter (Fig. 7.6b).

CarD stabilizes Open Promoter Complexes

To evaluate the mechanism of CarD's effects on RNAP activity at these promoters, we tested CarD's ability to modulate the production of abortive transcripts. We formed OPCs and initiated transcription with only two NTPs for the 23s rRNA promoter to give a 4-base product and with the dinucleotide primter ApU and labeled CTP for T7A1 to produce a 3-base product (Fig 7.8a). CarD stimulates the production of abortive products at both the 23s rRNA and the T7A1 promoter (Fig 7.8b).

While CarD has the same effect on the production of abortive products from the 23s rRNA promoter, it increases abortive transcription and decreases run-off transcription at the T7A1 promoter. The T7A1 promoter is a phage promoter known to bind to RNAP with high affinity and form stable OPCs (Kadesch et al., 1982; Rosenberg et al., 1982) while rRNA promoters from various organisms are characterized by their unstable OPCs

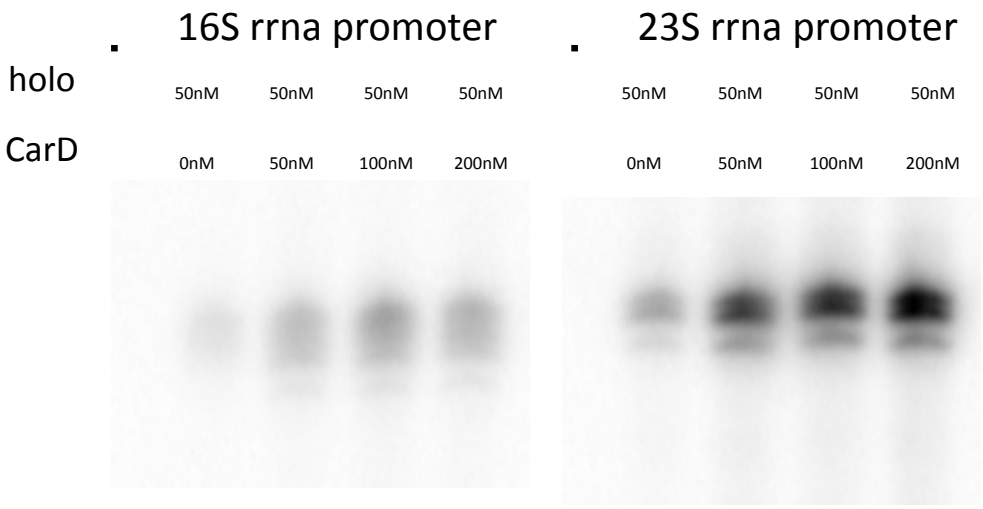


Figure 7.7: CarD stimulates RNAP activity on rRNA promoters with Full Length Tth σ^A. Transcription assays were performed as described in Figure 7.6.

a)

Tth 23s promoter GGGGGCCCTTGACAAAGGCCATGCCTCCTTGGTATCTTCCCTTTTGCGCTGC

T7A1 promoter CAAAAAGAGTTGACTTAAAGTCAACCTATAGGATACTTACAGCCATCGAGAG

b)

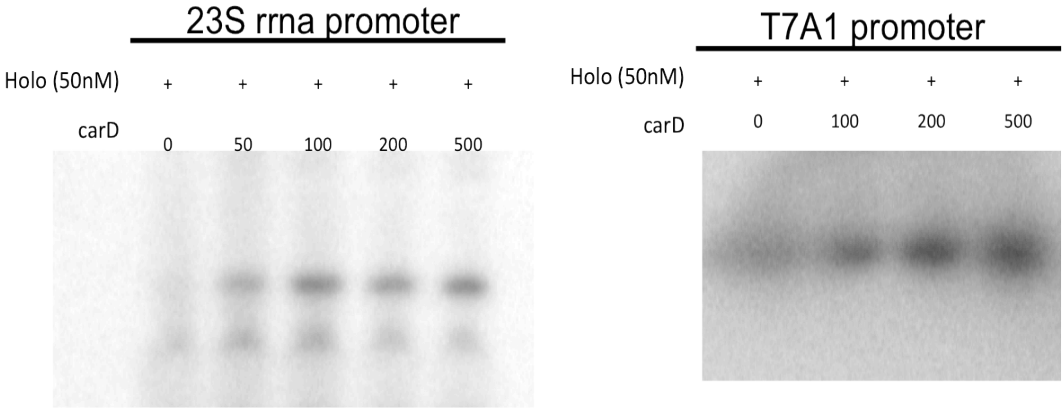


Figure 7.8: CarD stabilizes RNAP open promoter complexes (OPCs). a) Sequences of promoters used in subsequent experiments. Abortive RNA products produced by the nucleotides added to the reactions are colored in red. b) CarD stimulates abortive initiation at the 23s rRNA and the T7A1 promoter. Open promoter complexes were formed by first incubating various concentrations of CarD with RNAP holoenzyme (50nM), adding linear template DNA (50nM) followed by incubation at 65°C as described in the methods. Reactions were initiated by the addition of NTPs to produce short RNA products. After five minutes, reactions were stopped and run on a 23% Urea-PAGE gel.

(Barker et al., 2001b; Krasny and Gourse, 2004; Paul et al., 2004). CarD could act upon the same kinetic step in these two promoters and have different effects on the output due to the differences in the relative stabilities of the intermediates of transcription initiation.

We tested the stability of open promoter complexes in the presence and absence of CarD on the 23s and T7A1 promoters. We found that heparin actively inhibited Tth RNAP activity, but that the double stranded DNA fragment, FullCon, at 20 fold molar excess to the promoter DNA fragment (Paul et al., 2004), was able to fully compete for RNAP binding. The FullCon promoter was developed by *in vitro* selection for high affinity binding to RNAP holoenzyme (Gaal et al., 2001). To test the stability of OPCs, I first formed OPCs at 65°C. For the 23s rRNA promoter, we measured OPC stability using transcription as an output. Briefly, after the formation of OPCs, we added FullCon DNA to challenge the complexes, after which we initiated with NTPs and allowed the reaction to proceed. The transcriptional output is a measure of how much RPo was present at the time of initiation. The half-life of the 23s rRNA promoter was quite short (t=10s). CarD stabilized the half-life significantly (Fig 7.9a).

For the T7A1 promoter, we measured OPC stability by filter binding. The T7A1 promoter fragment was end-labeled with P^{32} and used to form OPCs at 65°C. Once OPCs were formed, we challenged with unlabeled FullCon DNA. At time points after the addition of the FullCon fragment, we aliquoted 10µl of the reaction onto a prewashed filterpaper that binds specifically to protein and not nucleic acid. Therefore, radiolabel present on the filterpaper records T7A1 DNA bound to RNAP. The half-life of the T7A1 promoter was remarkably stable (t = 71 min) and CarD further stabilized this promoter (Fig. 7.8c).

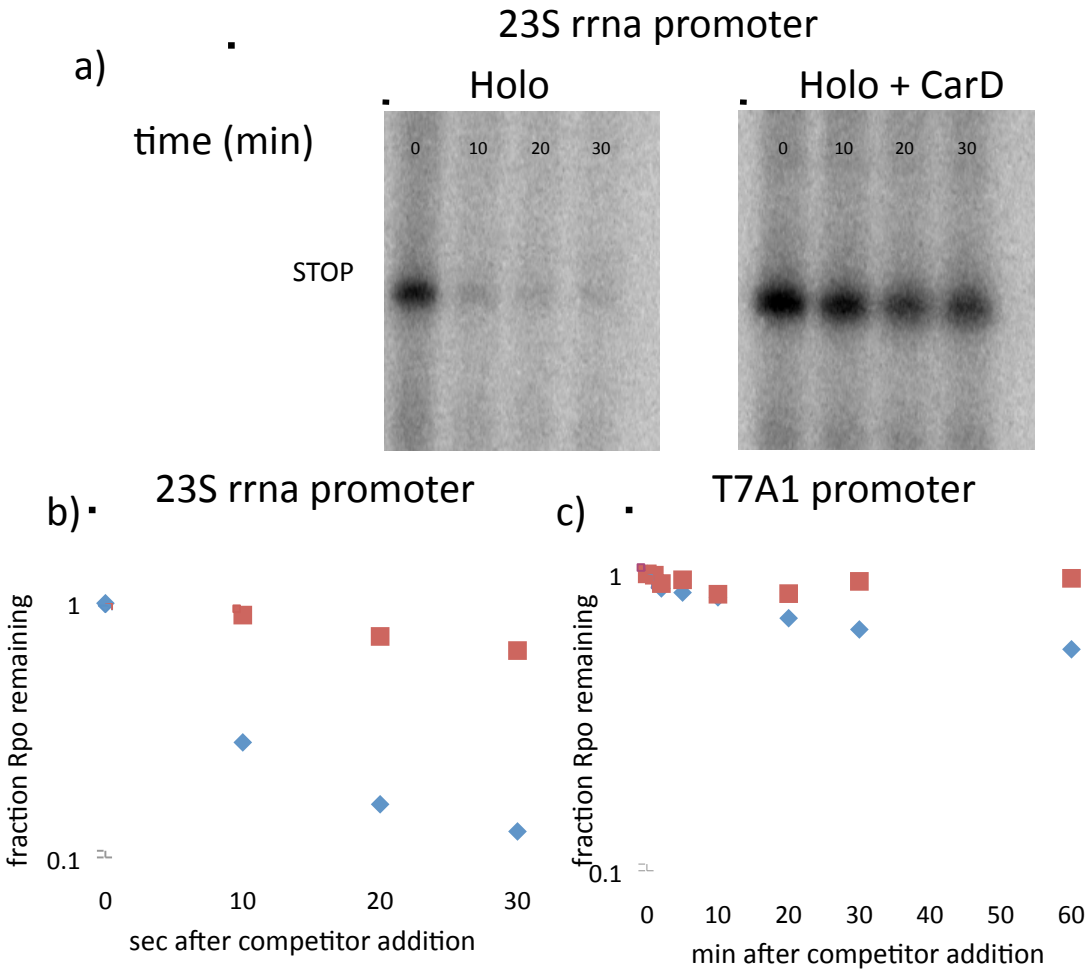


Figure 7.9: CarD stabilizes OPCs. Open promoter complexes were formed by first incubating CarD (200nM) with RNAP holoenzyme (50nM), adding linear template DNA (50nM) followed by incubation at 65°C as described in detail in methods. At t=0, 1μM FullCon DNA was added as a competitor, and open complex stability was monitored over time by run-off transcription assays for the 23S rrna promoter and filter binding for T7A1. The OPC stability at the rRNA promoter was visualized by Urea-PAGE gel (a), bands were quantified and normalized to t=0 (b). c) T7A1 OPCs were measured by filter binding using radiolabeled linear promoter fragment. Signal was normalized to t=0.

rRNA promoters from multiple organisms have short OPC half-lifes (Krasny and Gourse, 2004; Paul et al., 2004). Proteins such as DksA, and small molecules, such as ppGpp, that modulate the half-life of OPCs at all promoters affect the RNAP output from rRNA promoters primarily due to their unique kinetic parameters (Paul et al., 2004). CarD also appears to modulate the stability of OPCs, but unlike DksA and ppGpp, CarD stabilizes OPCs. The stabilization of RPo stimulates transcription from the rRNA promoters, where RPo is unstable (half life of 10s versus 71 minutes at the T7A1 promoter). However, on the already stable T7A1 promoter, further stabilization of RPo may prevent RNAP promoter clearance and subsequently inhibit run-off transcription.

CarD mutants stimulate rRNA transcription

CarD modeled onto the RNAP OPC places the CarD-CTD near promoter DNA at the non-template strand -10 element and downstream where the template and non-template strand reform duplex DNA (Fig. 7.4a). CarD contains a number of positively charged amino acids in this region that would be well placed to make contacts with the DNA backbone and potentially alter RNAPs interactions with promoter DNA. Work in Christina's lab showed that Mtb CarD alone in solution has a weak and non-specific interaction with DNA. Mutation of the positively charged residues at the CTD tip, which potentially interact with DNA in the model, disrupted the DNA binding activity of CarD by gel shift (Stallings, CL, personal communication).

We sought to determine the ability of Tth CarD mutants in this region to activate transcription from the rRNA promoters. The mutagenesis and protein purification for this project was performed by a technician in the Darst lab, Katherine Leon. None of the

mutants predicted to interact with DNA, which block DNA binding of CarD in solution, have any effect on CarD activity (Fig 7.10a and b).

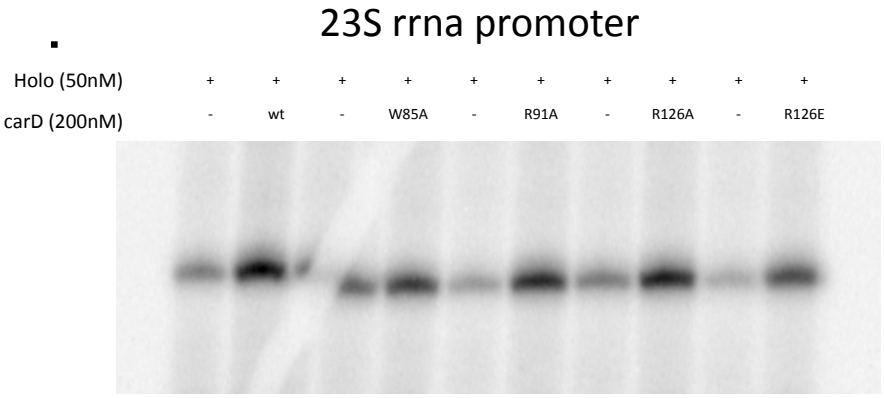
Because there are several positively charged residues in this region, and because RNAP binding recruits CarD to the site of its potential interaction with DNA, single mutations that affect binding to DNA in solution may have a reduced effect on CarD activity in the context of RNAP. To show that the CarD-CTD is required for CarD's activity, we tested the CarD-RID alone. The CarD-RID is unable to stimulate RNAP activity at the rRNA promoters, even at large molar excess (Fig 7.10c), arguing that the CarD-CTD is critical for its activity at promoters.

However, the CarD-RID alone may have a reduced affinity for binding to RNAP. The complex of CarD bound to the RNAP β1 domain was relatively unstable during purification. To determine whether the defect in transcriptional activation of the CarD-RID was due to an inability to interact with RNAP, I probed the interaction *in vitro*. Tth RNAP is (10)his tagged, but Tth CarD and σ^A both have their affinity tags cleaved through the purification process. I used Ni-bead pull downs to test the ability of CarD and the CarD-RID to bind to immobilized Tth RNAP holoenzyme. While CarD clearly interacts with RNAP holoenzyme (Fig. 7.11a, lanes 7 and 8), we see no evidence of the CarD-RID bound to RNAP (Fig 7.11b, lanes 7 and 8).

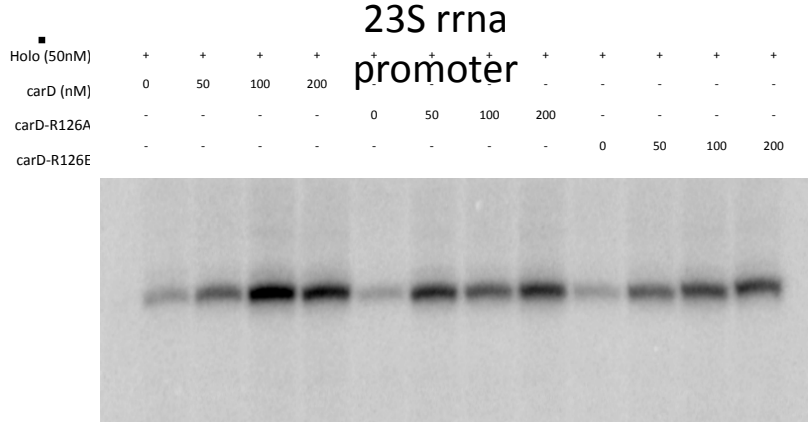
The Ni-bead pulldowns were performed at relatively low protein concentrations. Based on previous reports and the data I present here, we believe that CarD interacts with RNAP through its RID, which therefore must be capable of binding to β-1 lobe of RNAP both *in vivo* (Stallings et al., 2009) and *in vitro*. I therefore performed native gel shifts at both low and high protein concentration to determine whether the CarD-RID/RNAP

Figure 7.10 (Adjacent page): CarD C-terminal mutants do not affect CarD stimulation of RNAP activity at the 23s rRNA promoter. a) Transcription assays were performed as above with 200nM wild type (wt) CarD and the CarD CTD mutants listed. b) Transcription assays were performed as above with wt CarD and the CarD mutants listed at increasing concentrations. c) The CarD-RID alone does not stimulate RNAP activity at the 23s rRNA promoter. Wt CarD and the CarD RID were added at the indicated concentrations to transcription assays performed as above.

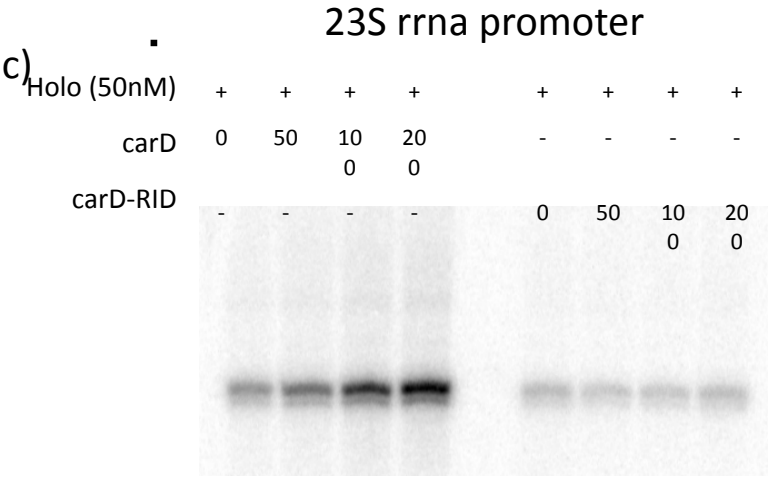
a)


23S rrna promoter
Holo (50nM) + + + + + + + + + +
carD (200nM) - wt - W85A - R91A - R126A - R126E

b)


23S rrna promoter
Holo (50nM) + + + + + + + + + + + +
carD (nM) 0 50 100 200 - - - - - - - -
carD-R126A - - - - 0 50 100 200 - - - -
carD-R126E - - - - - - - - 0 50 100 200

c)


23S rrna promoter
Holo (50nM) + + + + + + + +
carD 0 50 100 200 - - - -
carD-RID - - - - 0 50 100 200

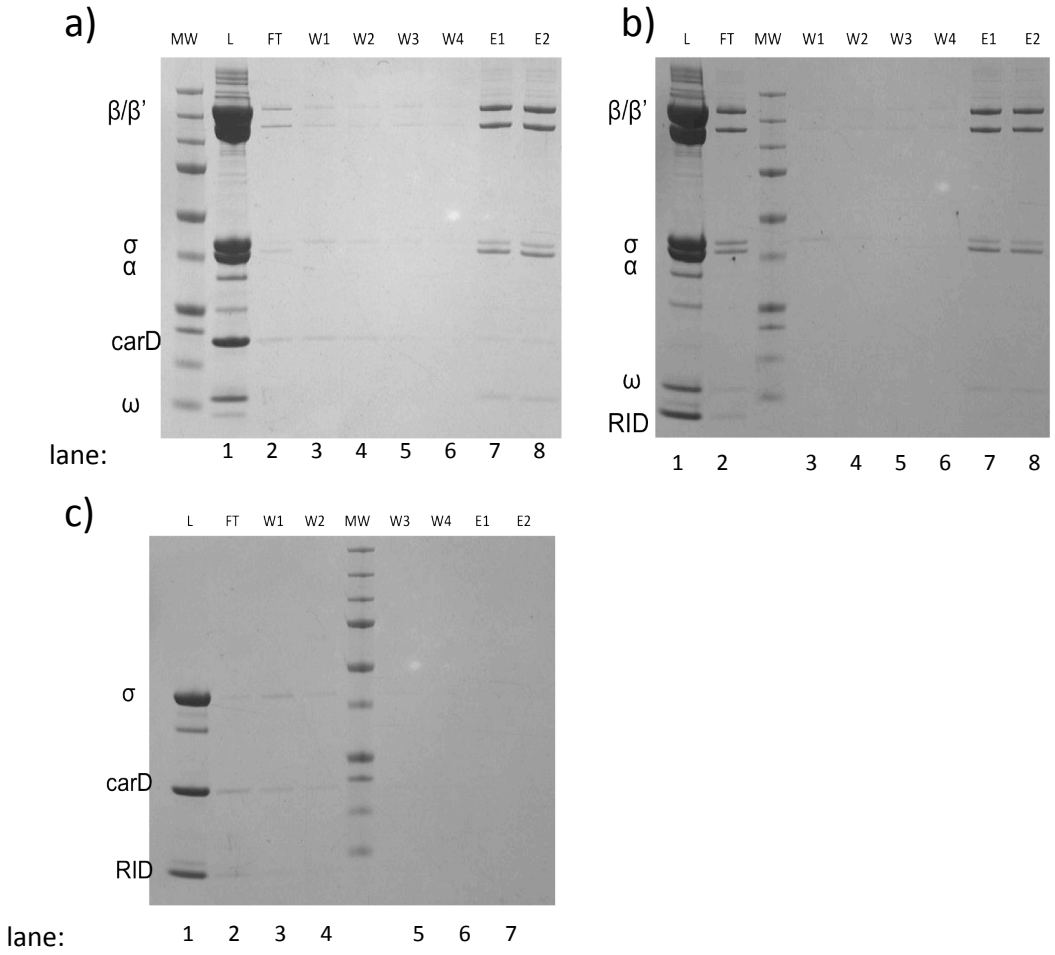


Figure 7.11: The CarD-RID does not interact with Tth RNAP by Ni-bead pulldowns. His-tagged Tth RNAP (2uM) was incubated with untagged $\sigma^A\Delta$N1.1 (5uM) and untagged CarD or CarD-RID (10uM). The mixture was then bound to Ni-agarose resin, washed extensively with wash buffer (10mM Tris HCl pH 8.0, 250mM Na Cl, 25mM Imidazole) and eluted with 500mM Imidazole. Aliquots from the load, wash and elution steps were run on a 4-12% SDS-PAGE gel.

interaction is low affinity and therefore requires high protein concentrations. While the gel shift performed at low concentration confirms the Ni-bead pulldown experiments (i.e. no apparent binding of the CarD-RID to RNAP; Fig. 7.12a, compare lanes 4 and 5), at high concentration we see clear evidence for a CarD-RID/RNAP interaction (Fig 7.12b, lane 3, red arrow). These data argue that the CarD-RID interacts with RNAP, but with a lower affinity than full length CarD. Therefore, the absence of apparent stimulation of RNAP activity by the CarD-RID at the rRNA promoters could be explained by either poor binding or the mechanistic importance of the CarD-CTD. We are currently planning experiments to test these possibilities.

CarD and PhERI target unique characteristics of rRNA promoters

rRNA promoters are uniquely tuned to be both abundantly expressed in log-growing cells and to be regulated by a variety of small molecules and protein factors as cells progress into stationary phase. Early work on rRNA regulation in prokaryotes focused on the importance of small molecules (Barker et al., 2001a; Barker et al., 2001b; Gourse et al., 1998). ppGpp, a modified nucleotide formed under stress conditions by the ribosome associated protein RelA, directly modulates the activity of RNAP at rRNA promoters (Barker et al., 2001b). These promoters are also uniquely sensitive to the concentration of initiating nucleotide (Krasny and Gourse, 2004). As cellular pools of NTPs decrease during starvation, rRNA promoters are inhibited without the aid of any additional signals or protein factors. The kinetics of rRNA promoters, in particular the characteristic instability of their open promoter complexes, allows them to respond to such cellular signals (Barker et al., 2001a; Barker et al., 2001b; Krasny and Gourse,

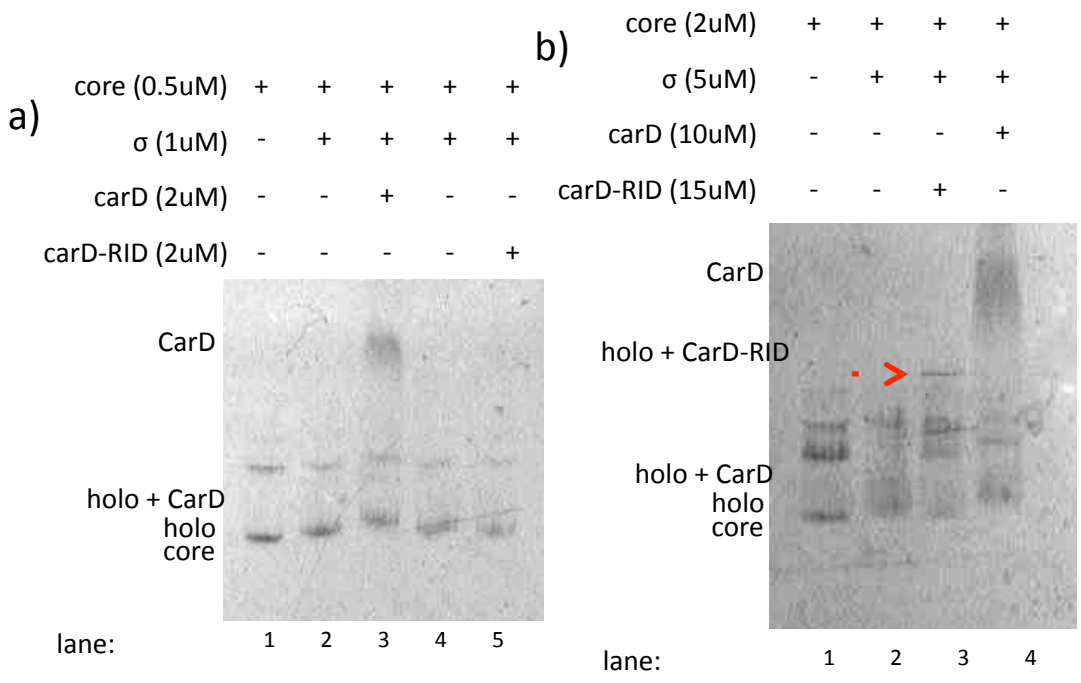


Figure 7.12: The CarD-RID only interacts with Tth RNAP at high protein concentration. a) Native-PhastGel analysis of CarD and CarD-RID binding to Tth RNAP at low concentration. Proteins were incubated at the listed concentration and electrophoresed on a 4-12% native Phast gel. While a shift is evidence for the CarD/RNAP holoenzyme complex, no shift is apparent upon binding of the CarD-RID. b) CarD and CarD-RID interaction with Tth RNAP holoeznyme at high protein concentration visualized on a native Phast gel. Proteins were incubated at the listed concentration and visualized as above. A band is visible indicating an interaction between the CarD-RID and RNAP holoenzyme (red arrow).

2004). rRNA promoters in various organisms also require UP-element binding and in Eco are further regulated by an additional activator, Fis (Krasny and Gourse, 2004; McLeod et al., 2002). In Bsub, the signals that regulate rRNA transcription are different than in Eco. Small molecules, in particular NTP concentration, feature prominently in the regulation (Krasny and Gourse, 2004).

The hypothesis that rRNA regulation is driven exclusively by small molecules was disproved in Eco by the discovery of the stringent response regulator DksA and activator Fis (Paul et al., 2004; Perederina et al., 2004). In the absence of DksA, Eco cells are unable to properly downregulate the expression of rRNAs upon entering stationary phase (Paul et al., 2004). DksA, which is structurally related to the Gre factors (Perederina et al., 2004), binds to the RNAP secondary channel, and decreases the stability of OPCs, thereby decreasing transcriptional output from rRNA promoters, which are rate-limited by this step. ppGpp also decreases the stability of OPCs, and ppGpp and DksA have a synergistic effect on the inhibition of RNAP activity at rRNA promoters in Eco (Paul et al., 2004). rRNA regulation in Eco therefore depends on both protein factors and small molecule signals to ensure the appropriate response to stress or starvation. No DksA homologs have been found in many organisms, including the gram-positive organisms Bsub (Krasny and Gourse, 2004) and Sau. There is also no evidence for a DksA homolog in thermophilic bacteria, hindering structural studies of DksA bound to RNAP holoenzyme and RPo.

The G1 phage protein PhERI and Tth/Mtb CarD both exploit the unique characteristics of rRNA promoters to have a profound impact on rRNA transcription. PhERI is the first protein factor that joins RNAP through an interaction with the global

transcription regulator σ but regulates RNAP activity though modulation of another RNAP domain, the α-CTD. PhERI, by blocking productive α-CTD binding to UP-elements, inhibits rRNA transcription, and thereby halts normal logarithmic cell growth (Chapters 1 – 5). The majority of Sau promoters, having no dependence on UP-element binding, are not inhibited by PhERI. PhERI thus exploits one unique characteristic of rRNA transcription previously described in other organisms: UP-element activation.

CarD, like DksA, modulates the stability of RPo, but rather than destabilizing RPo to decrease output from rRNA promoters, CarD increases the stability of RPo and stimulates RNAP activity at rRNA promoters (Fig. 7.8). Whether this is CarD's only, or even primary, function *in vivo* remains the subject of active research. However, CarD's effect on rRNA transcription *in vitro* demonstrates that it is a direct modulator of RNAP at rRNA promoters.

Neither Tth, Mtb or Sau have DksA homologs. Originally it appeared that CarD may be a functional homolog of DksA because it can compliment a DksA knockout in Eco cells (Stallings et al., 2009). However, we show that CarD acts by stabilizing RPo, while DksA destabilizes RPo.

rRNA expression is one of the most important transcriptional switches in prokaryotic cells. RNAP activity at rRNA promoters is regulated by small molecules that modulate RNAP activity (Gourse et al., 1998; Krasny and Gourse, 2004). These promoters are finely tuned to be able to both have robust RNAP activity in log-growing cells but respond rapidly to changing cellular conditions. While regulation may come from various small molecule signals of cell stress or starvation, proteins have co-opted the unique kinetic parameters of rRNA promoters to also facilitate the regulation of

RNAP activity. We describe the prokaryotic protein CarD and the phage protein PhERI that both target rRNA transcription through unique parameters of rrn promoters: RPo stability and UP-element binding respectively. While only three proteins in all prokaryotic organisms have been shown to directly target rRNA transcription specifically, two of which are described in this thesis, we believe the uniqueness of rRNA promoters can be a prime target for regulation not only be small molecules but also by protein effectors.

Chapter 8:

Identification of the plastid RNA polymerase in *Plasmodium falciparum*

Malaria, an infection by parasites from the family *Plasmdoium*, affects 300 million and causes over one million deaths per year (2010; Kar and Kar, 2010; Murray et al., 2012). *Plasmodium falciparum* (Pf) causes the majority of malaria cases throughout sub-Saharan Africa (2010; Kar and Kar, 2010). Pf has a complex life-cycle that involves infection both of human and mosquito hosts; humans are infected when bitten by a Pf carrying mosquito (Cox, 2010). In the initial stages of the infection, Pf parasites divide in the liver before being released into the blood where they infect red blood cells (RBCs), causing high fever, nausea, headaches, and other symptoms of infection (Trampuz et al., 2003).

Malaria is difficult to treat due to the lack of diagnostic tools and high levels of resistance to small-molecule therapeutics (Wernsdorfer and Noedl, 2003). Malaria is traditionally diagnosed by visualizing blood smears on a microscope, technologies that are not widely available in developing countries (Tangpukdee et al., 2009; Wongsrichanalai et al., 2007). Small-molecule interventions to malaria infection have suffered from widespread resistance (Wellems, 2002). Because malaria generally affects those in developing countries, research and drug development on this important disease have lacked funding until recently (Murray et al., 2012). Many traditional antimicrobials

are not effective against *Plasmodium* parasites due to their eukaryotic nature. Treatment is further complicated by the lack of accessibility to regular and quality healthcare in regions with endemic infections (Murray et al., 2012).

Quinine, a small molecule isolated from the bark of Cinchona tree, and chloroquine were the first active anti-malarial agents. However resistance rapidly arose to these treatments (Chaturvedi et al., 2010; Wernsdorfer, 1994). Army research (a response, largely, to the malarial burden of US soldiers in the war in Vietnam) in the United States developed the quinine derivative mefloquine (Miller and Su, 2011) as a next-generation malarial drug. Work in China identified artemisinin, derived from a medicinal plant (Miller and Su, 2011). The use of mefloquine has been generally low due to its side effects and the emergence of resistance (Wernsdorfer, 1994). Currently, artemisinin, and its derivatives, are the drug of choice for severe malaria infections (Miller and Su, 2011), although resistance to artemisinin has been described (Dondorp et al., 2009).

Pf and related parasites harbor an organelle, termed the apiocoplast or plastid, descended from an endosymbiotic event with a cyanobacterium (Walter and McFadden, 2005) (Fig. 8.1a). The plastid has been shown to be required for Pf viability by chemical and genetic means (Nair and Striepen, 2011), but its essential function is not fully established (Roos et al., 1999). It has been hypothesized that the plastid is required for biosynthetic pathway(s), and isolates reactions that produce reactive byproducts away from cellular proteins and nucleic acids. Recent work shows that the inactivation of the plastid can be fully rescued by the addition of isopentenyl pyrophosphate (IPP). Pf parasites can be grown indefinitely in the presence of antibiotics that block plastid

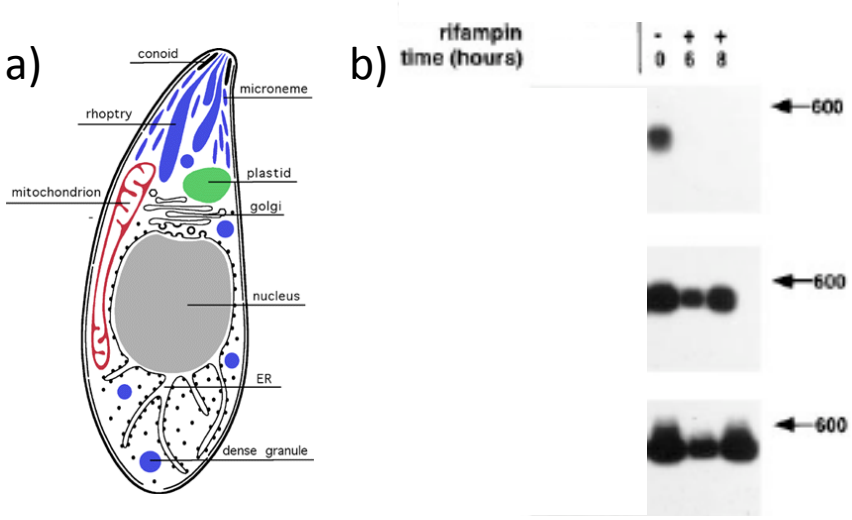


Figure 8.1: Schematic of apicoplast structure. a) Schematic view of *Toxoplasma gondii*, a parasite closely related to Pf. Cellular organelles are labeled; the plastid is shown in green. b) Rif inhibits the plastidic RNAP. Rifampicin was added to parasites in culture and RNAs visualized by northern blot. rRNA and a nuclear gene, MSA, are not affected by Rif addition, but the plastid encoded rpoB/C mRNA is repressed by Rif addition. Adapted from (McConkey et al., 1997).

function if they are supplemented with IPP (Yeh and DeRisi, 2011). This work argues that the only required function of the plastid in blood stage parasite growth is the biosynthesis of IPP. However, it is unknown if the plastid has any other required roles at different stages of the Pf life cycle, either in the human or the mosquito host.

Like the mitochondrion or chloroplast, the plastid harbors its own genome (Saxena et al., 2010; Wilson et al., 1996). Indeed, the identification of the plastid began with the isolation of its circular genome (Gardner et al., 1991). Very few genes are encoded on the plastid, arguing that plastidic genes have been shuttled to the nuclear genome (Wilson et al., 1996). Import into the plastid itself is controlled by a bipartitate N-terminal import signal (Fig 8.2) (Waller et al., 2000; Zuegge et al., 2001). Hundreds of nuclear genes are annotated as likely plastidic proteins, and many have been shown to localize to the plastid by co-localization studies (Nair and Striepen, 2011). The plastidic genome contains genes for translational and transcriptional machinery, including rRNAs and tRNAs. The large (β and β') subunits of a prokaryotic like RNAP are also encoded on the plastidic plasmid (Wilson et al., 1996).

RNA polymerase is a validated drug target being inhibited by the small molecule Rifampicin (Rif) (Hartmann et al., 1967; Hinkle et al., 1972b). Rif binds near the RNAP active site and blocks the path of nascently transcribed RNA (Campbell et al., 2001). Rif and Rif derivatives inhibit RNAP Sau (Fig 6.8), Eco and Mtb (Zenkin et al., 2005). Rif has been tested as an anti-malarial agent in patients, alone and in combination with other antibiotics. Rif showed moderate antimalarial activity in patients but is less effective than other therapies and therefore is not used clinically (Pukrittayakamee et al., 1994).

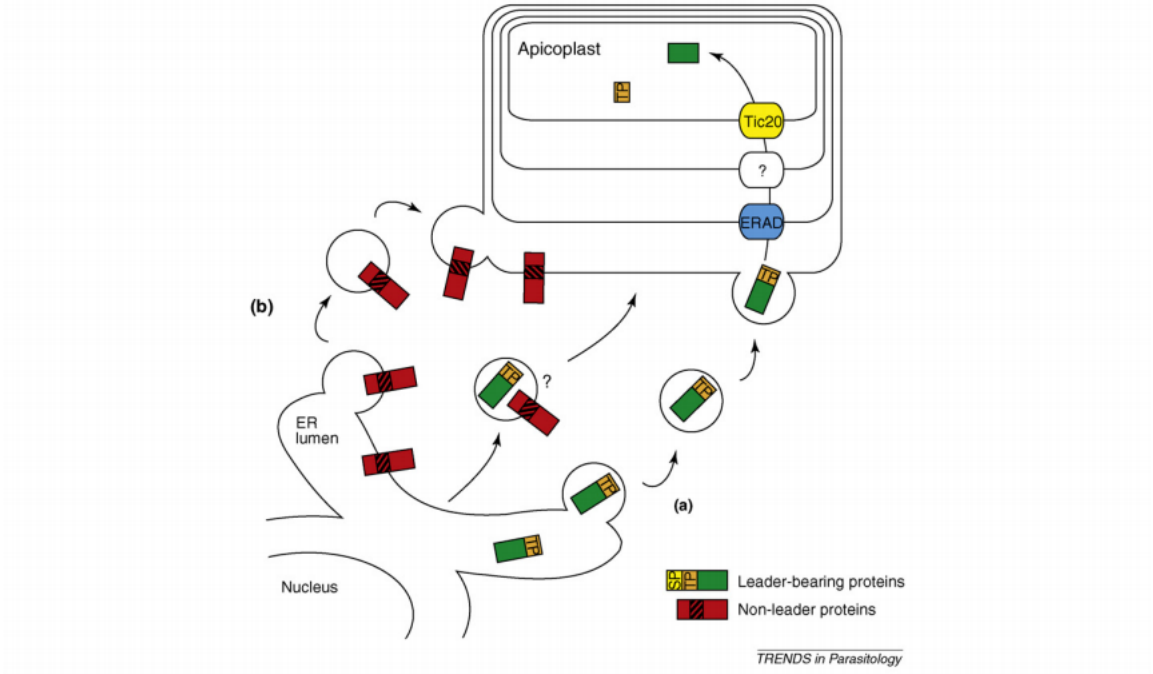


Figure 8.2: Schematic view of plastid membranes and protein import into the plastid in Pf. Proteins containing the bipartite N-terminal signal sequence (green protein; yellow/orange targeting sequence) are imported across the four membranes of the apicoplast. An additional pathway through which proteins localize to the apicoplast is shown (red proteins). Adapted from (Lim et al., 2009).

The RNAP subunits encoded in the plastid are most closely related to cyanobacterial RNAP, which has not been previously studied *in vitro* or *in vivo*. The addition of Rif to Pf parasites in culture inhibits parasitic growth, and also inhibits the synthesis of plastidic mRNA as assessed by northern blot (Fig. 8.1b) (McConkey et al., 1997). Whether cyanobacterial RNAP is directly inhibited by Rif, or any Rif derivatives, has not been established.

The plastid is required for viability of malaria parasites and therefore may provide novel drug targets (Nair and Striepen, 2011; Yeh and DeRisi, 2011). RNAP activity in the plastid appears to be essential for viability of the parasite (Dahl and Rosenthal, 2007; McConkey et al., 1997; Pukrittayakamee et al., 1994), and is a known target for small molecule inhibitors (Campbell et al., 2001; Ezekiel and Hutchins, 1968; Hinkle et al., 1972a). This chapter will detail our efforts to identify, validate and clone the plastidic RNAP from Pf. To evaluate the potential for small molecule development against the prokaryotic-like RNAP in malaria, I identified the subunits of RNAP holoenzmye (α, β, β', and σ) in the plastidic and nuclear Pf genome sequences. To show that our putative RNAP subunits assemble *in vivo*, I attempted to tag the nuclear-encoded α subunit with affinity and fluorescent tags. These studies examine an interesting target for drug development but also attempted to study a cyanbacterial-like RNAP for the first time.

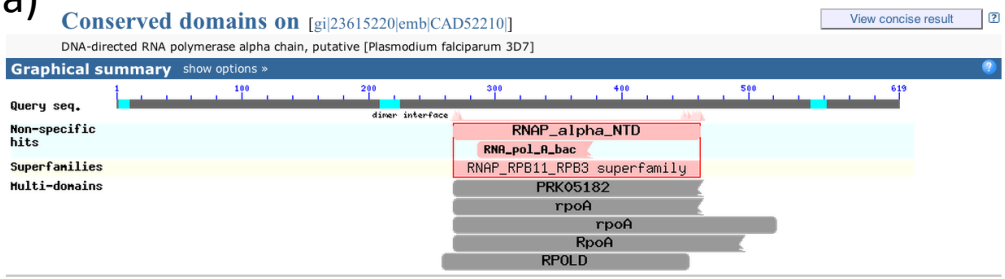
This project was done in close collaboration with Kirk Deitsch at Weill Cornell Medical College. Kirk allowed me, and several of my summer students to culture and transfect Pf parasites in his lab. The pull-downs and mass spectrometry experiments were planned in collaboration with the laboratory or Brian Chait here at The Rockefeller University.

Identification of the RNAP α-subunit

The β and β' subunits of RNAP are encoded in the plastidic genome, but every known bacterial RNAP also requires two copies of the α subunit for basic function and a σ subunit for initiation. α and σ are not identifiable in the plastidic genome. Therefore, if a functional RNAP assembles in the plastid, the α and σ subunits must be imported from the nuclear genome. Bioinformatic tools are available to predict whether nuclear genes are likely to be imported into the plastid. Blast searches in the Pf genome using the Eco α RNAP protein sequence identify a putative α RNAP subunit in the nuclear genome. The gene (XP_001349803.1) is annotated as an RNAP subunit (Fig 8.3). When evaluated for import into the plastid by the PATS server (Waller et al., 2000; Zuegge et al., 2001), this protein has a score of 0.989 (with a score of 1 being the most likely to be imported into the plastid), indicating a likely plastid import signal. The predicted secondary structure of the Pf α and Eco α are listed in Fig 8.4a.

The α subunit of RNAP from other organisms has been amenable to crystallographic studies. To attempt to crystallize this protein, a summer student in the Darst Lab, Leigh Harris, attempted to clone, express and purify Pf α constructs (Fig 8.4b). All constructs, except a small C-terminal fragment, were insoluble when expressed in Eco, even with the addition of solubilizing tags. This C-terminal fragment was purified (Fig. 8.5b) and screened extensively for crystallization conditions, but no crystallization conditions were identified.

a)



b)

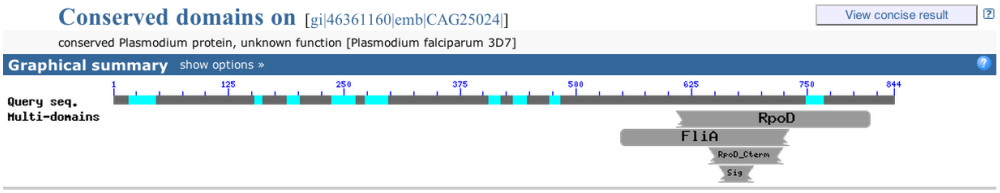
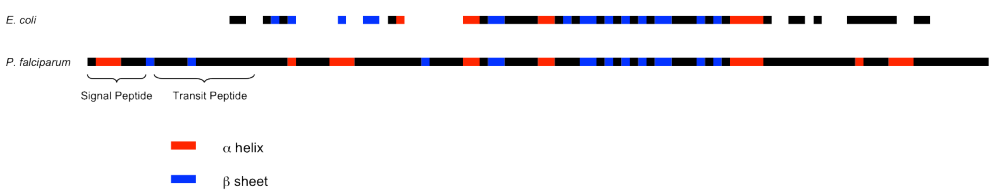


Figure 8.3: Domain conservation in the putative Pf α and Pf σ. a) Pf α was blasted against the Eco genome and the rpoA/RNAP α fold was identified. b)) Pf σ was blasted against the Eco genome and the rpoD/RNAP σ fold was identified. The N-terminus of this protein has no known fold or homology to any protein in the Eco genome.

a)



b)

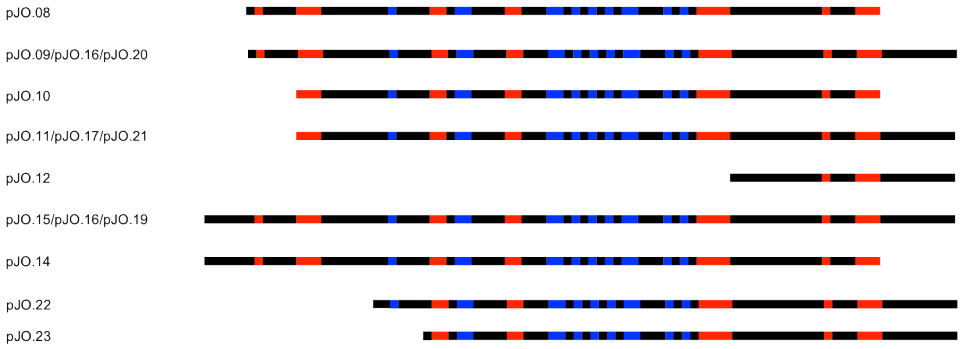
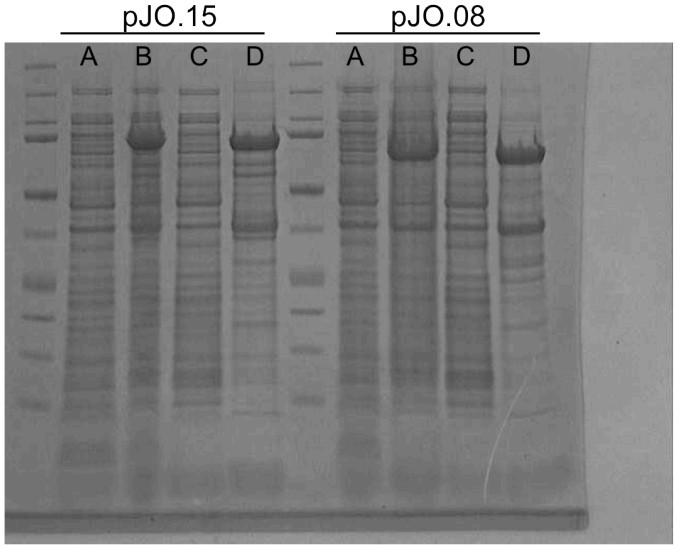


Figure 8.4: Schematic view of the putative Pf α RNAP subunit. a) Comparison of the secondary structure predictions (sheet: blue; helix: red) for Pf α to the known secondary structure of Eco α. b) Pf α constructs cloned into Eco expression vectors for protein expression and purification.

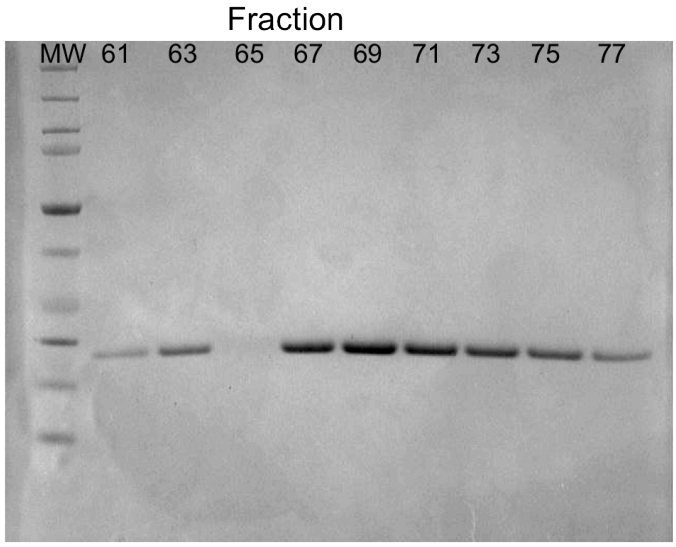
Figure 8.5 (adjacent page): Expression and purification of Pf α and Pf α fragments. a) Expression and solubility gel of full length Pf α (pJO.15) and an α truncation (pJO.08). Both proteins are present in the insoluble fraction (lane D). b) Purification of a C-terminal fragment of Pf α. Protein was visualized on a 4-12% SDS-PAGE gel after the gel purification step and prior to crystallization trials.

a)



A: Uninduced B: Induced C: Soluble D: Insoluble

b)



Identification of nuclear σ subunit

BLAST searches with Eco σ^{70} identified no σ homologs in Pf. σ factors are required for promoter recognition and DNA melting, and it is unlikely that a functional cyanobacterial RNAP could exist without a σ subunit. While group 1 σ factors are well conserved throughout bacterial species, differences between Eco and cyanobacterial RNAP sequences could explain the apparent lack of BLAST search hits. When BLAST searches were performed with cyanobacterial σ factors, a Pf homolog was identified in the nuclear genome (Fig 8.3b). Like the putative α-subunit, the putative σ factor (gene: XP_966194.1) is likely to be imported into the plastid (PATS score: 0.991). This protein is annotated as being of unknown function.

BLAST searches for the RNAP ω subunit from various organisms, including Eco and several sequenced cyanobacteria, yielded no obvious homologous protein in Pf. Cyanobacteria have a clear ω homolog. This protein is small (<90 amino acids) and is not required for RNAP catalytic activity. Whether Pf contains an ω homolog that could not be identified in our analysis is unclear.

Validation of putative plastidic RNAP: pA pull-down

The β and β' subunits of the cyanobacterial-like RNAP, encoded in the plastid, are almost certainly incorporated into a cyanbacterial-like transcriptional apparatus. However, the α and σ subunits we identified in the nucleus may not be properly imported into the plastid, despite their high PATS score values, and may not be the functional RNAP subunits. We also failed to identify an ω subunit *in silico* although it may be present, but highly divergent in sequence, in the Pf genome. To show that the complex

we identified *in silico* forms a complex *in vivo*, we attempted to purify the putative plastid-RNAP complex (pRNAP) by affinity tagging subunits *in vivo* followed by purification and mass-spectrometry. Genetic tools are not available to make mutants in plastidic proteins in Pf. Genetically altering nuclear genes is not trivial in this organism, but has been done previously (Deitsch et al., 2001; Epp et al., 2008).

To grow Pf in the laboratory requires cell culture conditions wherein the parasite can infect human RBCs. Therefore, growing large-scale cultures requires using large volumes of RBCs and is technically difficult. *In vivo* protein-tagging for identification of complexes requires a large amount of biomass to perform the purification steps. Only one study has previously used mass spectrometry to identify *in vivo* assembled protein complexes in Pf, and large (2-3L) fermentor batches of RBCs were required for starting materials (Takebe et al., 2007).

For subsequent purification and mass spectrometry analyses, I decided to tag the putative nuclear encoded α subunit. The σ subunit binds to core RNAP only at promoters, and therefore may not be an ideal target for pull-down experiments, whereas α is incorporated into RNAP at all steps in the transcription cycle. I therefore cloned the nuclear α-subunit with a C-terminal protein-A (pA) tag. Previous pull-down experiments in Pf used a PTP tag (Takebe et al., 2007), which I also cloned onto the C-terminus of the α-subunit. While the endogenous, untagged α subunit was not removed from the Pf genome, we reasoned that it would not complicate our analyses.

These constructs were transfected into wild type Pf lab strain C3 by standard procedures (Epp et al., 2008). Both produced potentially transfected parasite lines after two months of selection for the plasmid. To ensure that the Pf lines were indeed

transfected and expressing our tagged RNAP α subunit constructs, I recovered plasmid DNA from the transfected lines, as well as purified mRNA and evaluated protein levels by western blot. Plasmid DNA was recovered from transfected lines matching the constructs we expect. RT PCR was used to identify mRNA corresponding to the gene of interest. I used primers annealing to the tag for this purpose: endogenous RNAP α would be amplified by primers annealing just to the gene. I see clear evidence for mRNA expressing out tagged constructs (Fig 8.6a and b). Finally, I performed western blots using an antibody specific to the pA tag. As a positive control, I tested a yeast strain previously shown to express pA tagged Nup130; the negative control was C3v86, the wild type Pf strain used for the initial transfection.

By western blot, we see clear evidence for pA tagged protein in the pA-tagged α transfected cell line (Fig 8.6b). However, there is no distinct band at the correct molecular weight, as is the case for Nup130. This is likely evidence of protein processing and/or degradation. Proteins are known to be processed as they are transported across the plastidic membranes (Fig. 8.2), however we would still expect to see one predominant band somewhat under the size of the theoretical mass (74 kDa). Whether the degradation apparent in the western blot is occurring in the cells or after we lyse the cells and prepare the sample is unclear.

At this point, satisfied that pA tagged protein is being expressed in our transfected Pf lines, we decided to proceed with pA pull-downs and subsequent analysis. Samples were sent to the Ben Mamoun lab at Yale for large-scale growth in their fermentor facility. However, we were never able to achieve enough sample to proceed beyond this

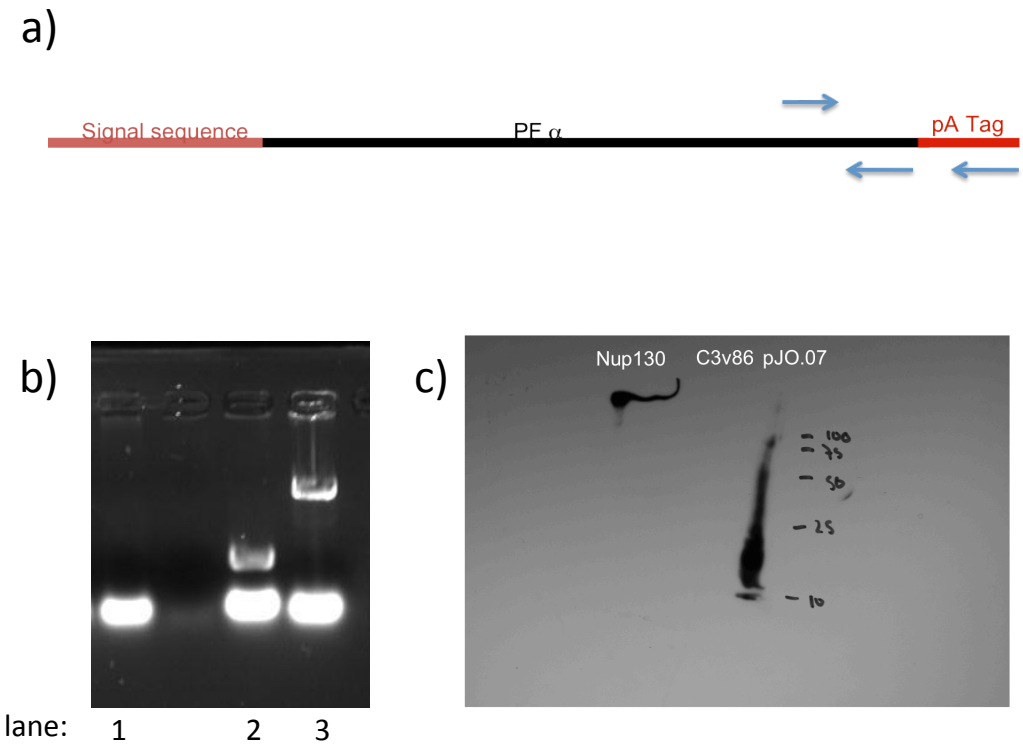


Figure 8.6: Expression of pA tagged Pf α in C3 Pf parasites. a) Schematic of PCR used to identify tagged Pf α. The Pf α gene is shown in black with the N-terminal signal sequence and the C-terminal pA tag highlighted in red. Primers used for PCR are shown as blue arrows. The second reverse primer anneals to a sequence in the pA tag and therefore can differentiate between endogenous and exogenously expressed Pf α. b) Non-quantitate RT-PCR of control cells (lane 1) and cells expressing pA tagged Pf α (lanes 2 and 3). c) Western blot with anti-pA antibody. pA tagged Nup130 was used as a positive control, C3v86 which does not express a pA tagged protein was used as a negative control. pJO.07 corresponds to C3 parasites containing the pA tagged Pf α subunit.

step. The Ben Mamoun lab was unable to grow the required sample for our pull-down experiments due to logistical issues.

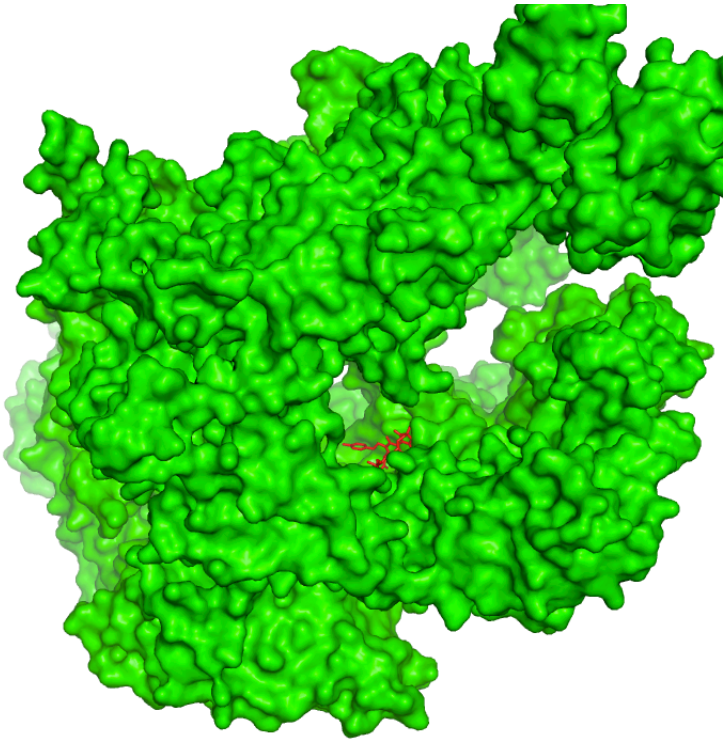
Validation of putative plastidic RNAP: Rif pull-outs

Due to the inherent difficulty of transfecting Pf parasites, the time required and the possibility that our exogenously expressed, tagged protein was being degraded *in vivo*, we decided to use chemical means to pull-out unusual RNAPs. Rifampicin binds to RNAP from many organism with high affinity and slow K_{off} rates (Feklistov et al., 2008). We reasoned that Rif, bound to a resin, could interact with RNAPs and facilitate the purification of RNAP from organisms that are not genetically tractable. Rif binds near the active site of RNAP, deep inside the cleft formed by the β and β' subunits (Fig. 8.7) (Campbell et al., 2001). We therefore reasoned that Rif would have to be attached to the resin via a long, flexible linker, allowing it to reach its RNAP binding site. This work was done in collaboration with a fantastic chemist, Arkady Mustaev, and a talented summer student in the Darst lab, Fatmata Bah.

Arkady Mustaev chemically synthesized agarose resins bound to Rif via linkers of increasing length (17 C-C bonds, 42 C-C bonds and 52 C-C bonds). The distance from the surface of RNAP to the Rif binding site is approximately 45 Å (Fig 8.7b). I initially used purified Eco RNAP to show that it bound specifically to Rif-conjugated resins. To remove protein from the resin, I used Guanidinium to denature all proteins; adding excess Rif was insufficient to remove bound RNAP from the column (likely due to the extremely slow K_{off} of Rif bound to RNAP). Gels were silver stained to allow visualization of small amounts of protein. RNAP binds specifically to resin conjugated to

Figure 8.7: Structure of Rif bound the RNAP. a) Surface representation of Taq RNAP (green) bound to Rif (red). Rif binds deep within the cleft of RNAP adjascent to the active site. b) Zoom of Rif (red) binding pocket showing shortest path (yellow) to the surface of RNAP. The distance was measured using PyMol.

a)



b)

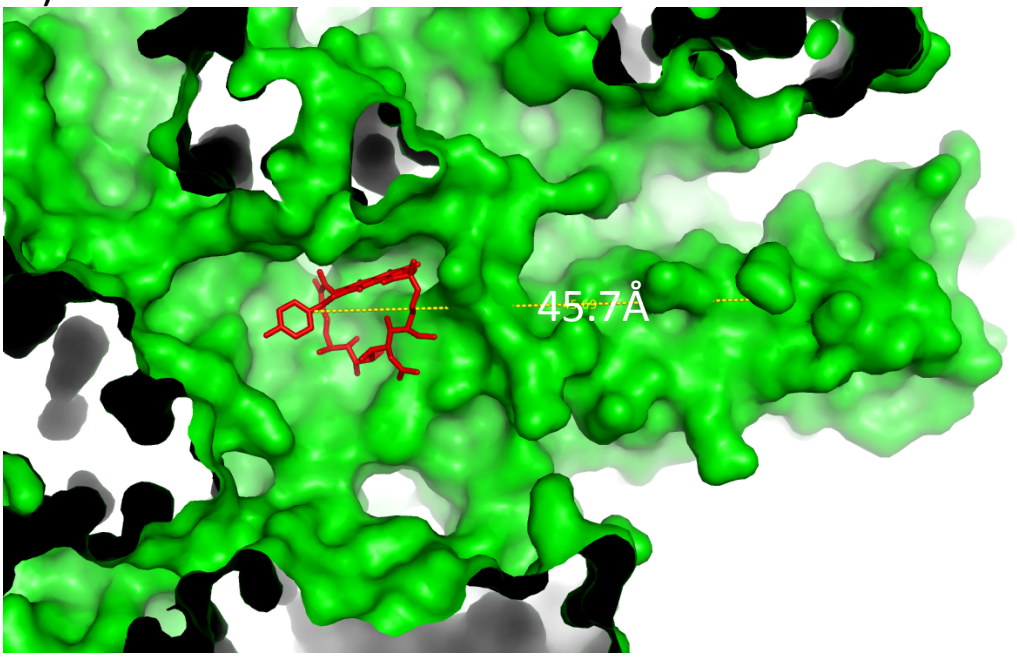
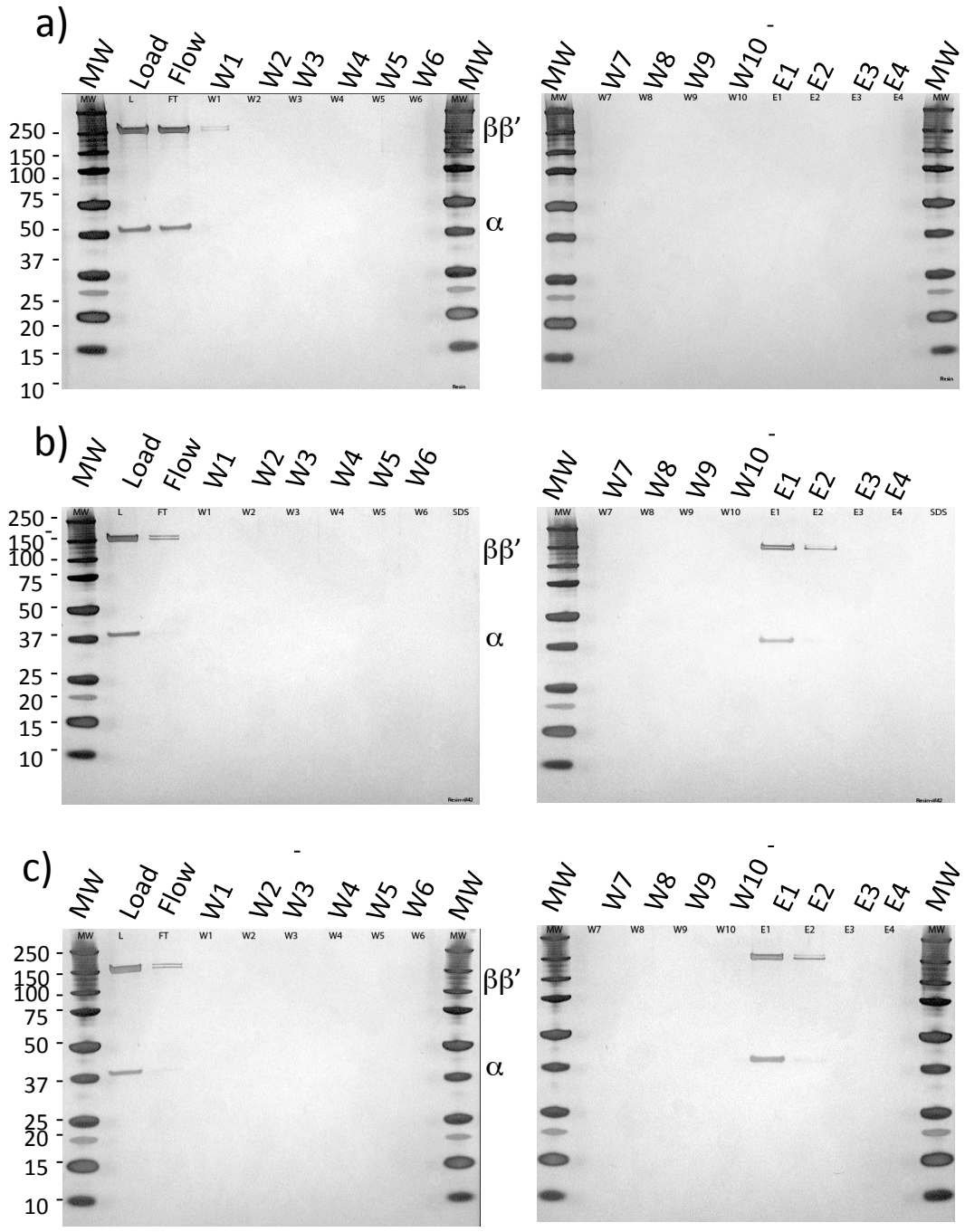


Figure 8.8: Rif pull-outs of Eco RNAP. a) Rif was conjugated to agarose beads with a 17 carbon-carbon bond containing linker. Eco RNAP was incubated with the beads, followed by thorough washing with 1x binding buffer. Bound protein was eluted using 6M guanidinium. Aliquots from each sample were visualized on a 4-12% SDS-PAGE gel with silver staining. b) Rif conjugated to agarose beads using a 42 carbon-carbon linker. Pull-outs were performed as in part a. c) Rif conjugated to agarose beads using a 52 carbon-carbon linker. Pull-outs were performed as in part a.


a)
MW
Load
Flow
W1
W2
W3
W4
W5
W6
MW
W7
W8
W9
W10
E1
E2
E3
E4
MW
250
150
100
75
50
37
25
20
15
10
ββ′
α
b)
SDS
c)

Rif with 42 or 52 C-C linkers, but not with the 17 C-C linker, which we predict would be too short to reach the Rif binding site on RNAP (Fig 8.8).

Next I attempted to specifically purify RNAP from Eco lysate using the same techniques. While purified RNAP appeared to bind to the Rif-conjugated resins, we see no RNAP enrichment in the elution fractions from Eco lysates (Fig 8.9). While some bands appear that are approximately the right size for RNAP in elution fractions from the 42 C-C and 52 C-C conjugated resins these results proved difficult to reproduce systematically.

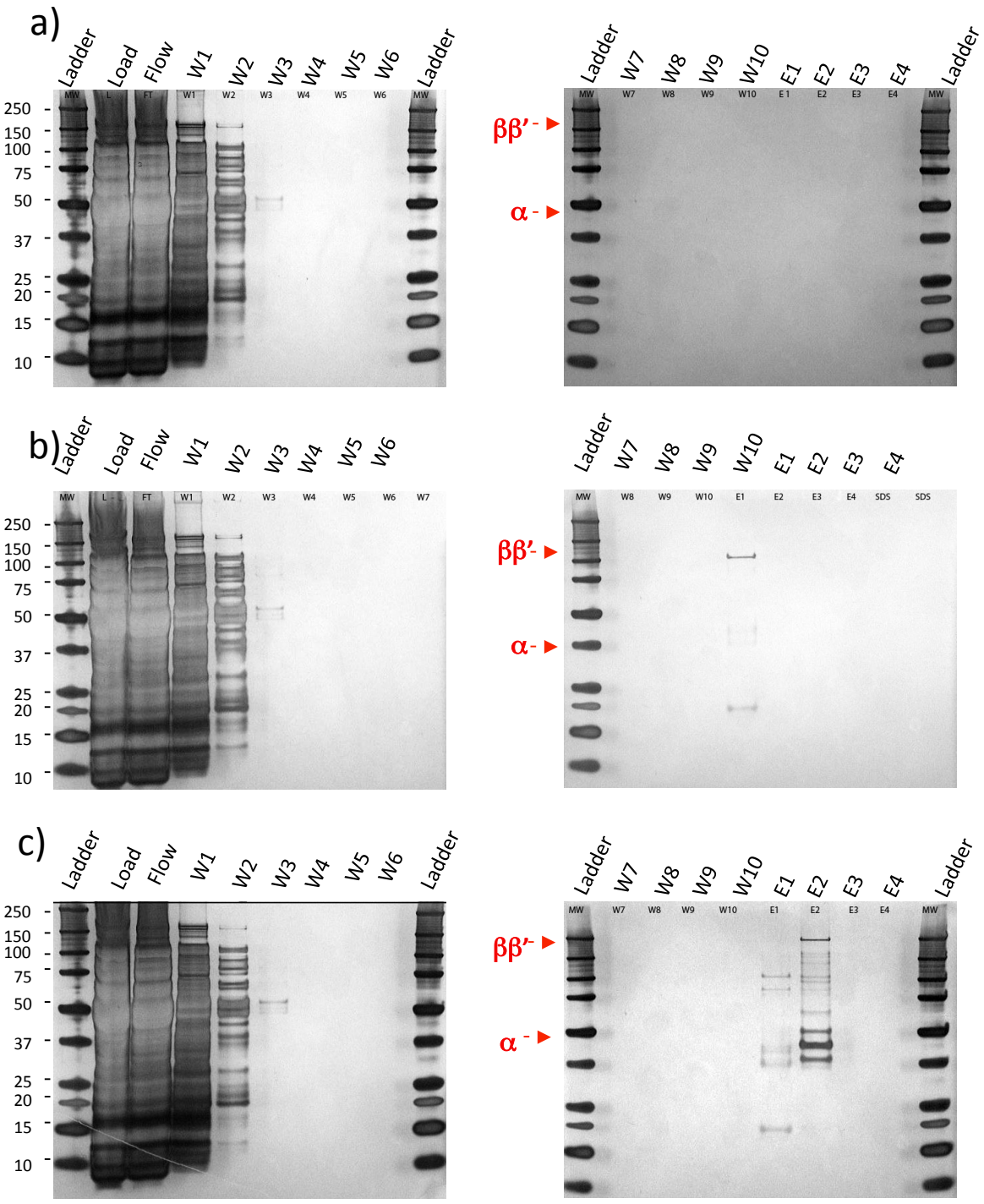
Validation of putative plastidic RNAP: mCherry tagged α

Because pull-downs using pA tagged RNAP and pull-outs using chemical ligands proved difficult, we decided to attempt to tag the α subunit with a fluorescent tag that would allow direct visualization by standard microscopic techniques. This work was done with the help of a talented undergraduate, Ronnie Almonte. We cloned the RNAP α subunit with mCherry, a monomeric RFP, fused to the C-terminus of the protein. We then attempted to transfect wild type Pf parasite lines and a parasite line with a plastid-localized protein fused to GFP for co-localization studies. However, after two attempts at transfection, we were never able to attain a single parasite line containing our gene of interest.

Conclusions and future directions

The identification of the plastidic RNAP in Pf proved technically challenging at almost every point. Even relatively simple techniques, such as molecular cloning, are

Figure 8.9: Purification of RNAP from Eco lysate. a) Eco lysate was loaded onto Rif conjugated beads with a 17 carbon-carbon linker. Purification was performed as in Figure 8. b) Eco lysate was loaded onto Rif conjugated beads with a 42 carbon-carbon linker. Purification was performed as in Figure 8. c) Eco lysate was loaded onto Rif conjugated beads with a 52 carbon-carbon linker. Purification was performed as in Figure 8.


a)
Ladder
Load
Flow
W1
W2
W3
W4
W5
W6
Ladder
250
150
100
75
50
37
25
20
15
10
Ladder
W7
W8
W9
W10
E1
E2
E3
E4
Ladder
ββ'-
α-
b)
Ladder
Load
Flow
W1
W2
W3
W4
W5
W6
W7
250
150
100
75
50
37
25
20
15
10
Ladder
W7
W8
W9
W10
E1
E2
E3
E4
SDS
SDS
ββ'-
α-
c)
Ladder
Load
Flow
W1
W2
W3
W4
W5
W6
Ladder
250
150
100
75
50
37
25
20
15
10
Ladder
W7
W8
W9
W10
E1
E2
E3
E4
Ladder
ββ'-
α-

difficult in Pf. The genome content in Pf is nearly 90% A/T rich and Pf proteins are often characterized by long, unstructured regions, making them notoriously difficult to study or purify using model organisms such as Eco. DNA synthesis is increasingly cost effective and may be particularly useful when attempting to clone Pf genes due to their wildly abnormal codon usage in Eco. Using codon optimized DNA could potentially make the expression and purification of the Pf plastidic RNAP domains and complexes feasible in Eco.

The Pf plastidic RNAP is related to the cyanobacterial enzymes. In these organisms, some of which have been fully sequenced, RNAP genes are annotated and likely to form functional complexes. Studying RNAP from cyanobacteria may provide insights into the Pf enzyme and, even if conclusions on the Pf enzyme would be limited, would certainly be interesting from an evolutionary standpoint. As in Sau, promoters and regulatory factors are completely uncharacterized in the cyanobacteria.

Recent work has characterized the putative replicative DNA polymerase in the Pf plastid. Work on this enzyme is facilitated by the fact that a single protein encodes both helicase and DNA polymerase activities and is clearly imported into the Pf plastid. Identification of the active RNAP required for transcription in the plastid would provide a rich biochemical system for evaluating gene expression in this organelle. Furthermore, given a purified *in vitro* system, testing or designing small-molecule inhibitors of RNAP activity would be relatively trivial. This research would provide a framework for understanding transcription in the Pf plastid, and allow for the development of small-molecule inhibitors active against this unique RNAP.

Chapter 9:

Materials and Methods

Protein Expression and Purification

The PhERI / σ^{A}_{4} (Sau σ^{A} residues 297-368) complex was cloned into a single operon as described (Patikoglou et al., 2007). Both proteins were cloned to be expressed from a single mRNA transcribed by T7 RNA polymerase using pET vectors or the modified pET vector pSKB2, and to have a ribosome binding site. σ^{A}_{4} was cloned with a cleaveable 6(his) tag whereas PhERI contains no affinity tag (see figure 2.2). The complex was expressed in Eco BL21(DE3) cells with 1mM IPTG at 37°C for 3h and purified using Ni-affinity chromatography using 1x protein purification buffer (20mM Tris HCl pH 8.0, 0.5M NaCl, 5% glycerol, 0.5mM β-ME). The his(6) tag was subsequently cleaved from σ^{A}_{4} by incubation overnight with precision protease and the complex was further purified by substractive Ni-affinity and size exclusion chromatography in 10mM Tris HCl, 0.5M NaCl, 1mM DTT. PhERI and σ^{A}_{4} formed a stable, stoichiometric complex throughout the purification. We dialyzed the complex into crystallization buffer (10mM Tris-HCl pH 8.0, 0.5M NaCl) and screened for crystallization conditions.

PhERI was cloned into pET29a using NdeI and XhoI, removing the native stop codon, to produce a C-terminally His-tagged protein. PhERI was expressed in BL21(DE3) cells with 0.5mM IPTG overnight at 18°C and purified using standard Ni-affinity chromatography in 1x protein buffer.

Sau σ^A was cloned into PSKB2, a modified pET vector, using NheI and HindIII. The His-tagged protein was expressed in BL21(DE3)plysS cells at 25°C for 5 hours. The protein was purified using Ni-affinity chromatography, dialysed into low salt buffer (20mM Tris HCl pH 8.0, 0.1M NaCl, 1mM DTT, 5% glycerol), and subsequently purified on a Q-sepharose column using a linear salt gradient (0.1M – 0.6M). Eco core was removed by a final purification step using size-exclusion chromatography in 20mM Tris HCl pH 8.0, 0.25M NaCl, 5% glycerol, 1mM DTT (SD75). The σ^A-mutant was made by megaprimed-PCR, cloned into the same expression vector and purified by the same protocol. σ^A and PhERI for subsequent use in biochemical experiments were stored in Sau protein storage buffer (10mM Tris HCl pH 8.0, 150mM NaGlutamate, 15% glycerol and 1mM DTT).

Crystallization of the PhERI / σ^A_4 4 complex

Crystals of the PhERI / σ^A_4 complex were grown under two different conditions at 22°C: JSCG59 (0.16M CaAcetate, 0.08M NaCacodylate, 15% (w/v) PEG8000 and 20% glycerol) and PC38 (0.1M MES pH 6.5, 10% (w/v) PEG5000 MME, and 20% 1-propanol). Sitting and hanging drops were used and crystals were formed using a 1:1 ratio of protein complex (10mg/ml) and reservoir solution. JCSG59 crystals were flash frozen in their mother liquid while PC38 crystals were cryoprotected briefly in 15% glycerol before being frozen. Selenomethionine substituted protein was purified and crystallized under the same conditions. Data were collected at the X3A beamline at the Brookhaven National Laboratories and 24-ID at the Advanced Photon Source. The native PC38 crystals diffracted to under 2.0Å. The data was processed using HKL2000

(Otwinowski and Minor, 1997), selenomethionine sites were found using Shake-and-Bake (Weeks et al., 2002), SAD phases calculated using SHARP (de La Fortelle et al., 1997) and initial density modification performed with Resolve (Terwilliger, 2000). Coot (Emsley and Cowtan, 2004) was used for model building and the structures were refined using Phenix (Adams et al., 2010). The high-resolution model was used late in the refinement process for the JCSG59 structure to aid in the placement of poorly defined density.

Calculation of buried surface area and protein contacts

Buried surface area was calculated using the CCP4 (Bailey, 1994; Potterton et al., 2003) based program AREAIMOL by measuring the total surface area of PhERI and σ^A_4 individually, and then measuring the total surface area of the complex.

To evaluate the contacts between the two proteins, we used the CCP4 (Bailey, 1994; Potterton et al., 2003) based program Contact. Putative contacts were then evaluated in the structural data using PyMol. Hydrogen bonds, or water mediated hydrogen bonds, were accepted if they were within 4Å.

Purification of Sau RNAP

Sau RNAP was purified natively from cells essentially as described (Deora and Misra, 1996). Briefly, Sau NCTC8325-4 cells were grown to an O.D. of 1.0, collected by centrifugation, washed in a high salt buffer, and resuspended in grinding buffer (TGED (10mM Tris HCl pH 8.0, 5% glycerol, 0.1mM EDTA and 1mM DTT) + 0.2M NaCl). Cells were lysed by French-Press and the cleared lysate was precipitated with

polethyleneimine at 0.6% (v/v). After centrifugation, the pellets were washed with TGED + 0.45 M NaCl, RNAP was eluted with TGED + 1.0M and precipitated with 35% (w/v) ammonium sulfate. The pelleted protein was resuspended in TGED and diluted to a conductivity equal to TGED + 0.1M NaCl. Protein was purified sequentially by a herparin column using a linear salt gradient and gel filtration column. An S-sepharose column was required as a final step to remove holoenzyme from core RNAP. RNAP was stored at -20°C in Sau protein storage buffer (10mM TrisHCl pH 8.0, 0.15M NaGlu, 30% glycerol, 1mM DTT). Protein purity was analyzed at each step in the purification by running protein samples on a 4-12% SDS-PAGE gel.

Native gel eletrophoresis

Proteins were incubated for 10 minutes on ice at the indicated concentrations in 1x protein buffer (10mM Tris HCl pH 8.0, 150mM Sodium Glutamate, 1mM DTT) supplemented with 10% glycerol to facilitate subsequent gel loading. After incubation, samples were loaded onto a 4-12% PhastGel using native buffer strips (GE Healthcare). Gel migration was monitored by adding 0.05% bromphenol blue to the loading buffer. Gels were stained with coomassie blue.

Limited proteolysis

PhERI or σ^{A}_{4}, or the reconstituted complex of the two proteins were incubated at 5μM in 10μl in 1x proteolysis buffer (100 mM Tris HCl, pH 8.0 and 20 mM $CaCl_2$) on ice for 10 minutes. Reactions were brought to 30°C before the addition of trypsin at the following molar ratios (protein:protease): 1:0, 1000:1, 100:1, 50:1, 10:1, 5:1. Reactions

were allowed to proceed for 30 minutes and stopped with the addition of 1µl of 100mM PMSF. 10µl SDS loading buffer was added and samples boiled for 5 minutes at 95°C before being run on a 4-12% SDS-PAGE gel (Invitrogen).

Preparation of DNA template for *in vitro* transcription assays

Promoter DNA fragments were made by PCR. Briefly, DNA oligos annealing to more than 100 bases upstream of the promoter and roughly 100 bases downstream were ordered. PCR was then used to amplify the promoter fragment directly from the Sau or Phage G1 genome. Promoter fragments were at least 250 bases in length; shorter promoter fragments, or DNA oligos corresponding to promoter fragments of 100-150 bases, produced no activity *in vitro*. PCR products were then purified on a 1.5% agarose gel and electroeluted into a 15kD dialysis tubing before subsequent phenol:chloroform extraction and ethanol precipitation. Promoter fragments were then resuspended in 10mM Tris pH 9.0 at 1µM final concentration.

***In vitro* transcription assays**

Proteins used in *in vitro* transcription assays were diluted into 1x protein storage buffer. σ^A (100nM) was preincubated with PhERI (or buffer) for 10 minutes on ice, followed by the addition of Sau core RNAP (50nM). After 10 minutes, DNA (50ng of genomic DNA or 50nM of purified promoter DNA) was added and the reaction brought to 20µl in 1x Sau transcription buffer (40mM Tris-acetate pH 7.9, 10mM MgCl2, 1mM EDTA, 50ug/ml BSA, 100mM NaCl and 1mM DTT). Reactions were incubated for 10 minutes at 37°C to form open promoter complexes and then initiated with NTPs (200uM

GTP, CTP, UTP, 50uM ATP, 0.1ul alpha-P32ATP). After 5 minutes reactions were stopped with 2x formamide buffer (98% formamide, 5mM EDTA) and run on a 12% Urea PAGE gel. Products were visualized on a phosphoimging screen and where applicable quantified using ImageQuant.

For the assays using genomic DNA as the transcription template, reactions were stopped with 20mM EDTA and 0.5% SDS, pipetted onto Whatman DE81 paper, washed 5 times with Sodium Phosphate (50g/L), rinsed with water, dried and visualized on a phosphoimaging screen.

Sau promoter identification

Sau promoters were identified through the use of a bioinformatic tool and manually by searching in the annotated Sau genome using the program Artemis (Rutherford et al., 2000). Briefly, the DNA sequence upstream of genes of known function (i.e. DNA polIII or dnaA) was manually searched for promoter-like sequences (TTGACA/TATAAT). When obvious promoters were not apparent, a program designed by Michael Mosely was used. The PromoterScore program is a Perl based script that searches genomic sequences for putative -10 or -35 elements. The user can select a score threshold above which a sequence is considered a putative -10 or -35 element. The scoring is based on the observed likelihood of bases at each position in the -10 and -35 elements. Once a sequence has reached the score threshold for the initial search, the program moves 15 bases downstream of a putative -35 or 15 bases upstream of a -10 and searches for the other sequence element, again using a threshold for putative sequence elements defined by the user. Putative promoters are output to a text file.

Strains and plasmids

RN4220 was obtained from Peter Moyle in Tom Muirs lab. pRMC2 and NCTC8325-4 were a generous gift from Ramesh Wigneshweraraj at Imperial College London.

PhERI expression in vivo

PhERI was cloned into the Sau expression vector pRMC2 (Corrigan and Foster, 2009) using primers containing a consensus Shine-Dalgarno sequence and BglII upstream of the start codon and a stop codon and EcoRI site downstream. pRMC2-PhERI and empty pRMC2 were then transformed into Sau strain RN4220 by standard electroporation (Schenk and Laddaga, 1992) and transformants selected on trypticase soy (TS) plates containing Chloramphenicol (10ug/ml). RN4220 containing empty pRMC2 and pRMC2-PhERI were grown in TS broth containing Chloramphenicol and transgene expression was induced with 100ng/ml anhydrotetracycline, which was the minimum required concentration for maximal cell growth inhibition by PhERI.

RNA purification

RNA was purified from cells at mid-log phase growth ($O.D._{600}$ 0.3-0.4) using the RNeasy kit from Qiagen. Briefly, $2x10^8$ cells were removed from growing cultures, immediately added to 2 volumes of BioStabilize solution and incubated for 5 minutes at room temperature. Cells were then collected by centrifugation, resuspended in TE buffer containing 1mg/ml lysostaphin and 200ug proteinase K and incubated for 15 minutes at

RT. 100µl zirconia beads (0.1mm) were added and the cells lysed for 3 x 2minutes, with a 1-minute rest on ice, in a bead-beater at top speed. The lysate was centrifuged briefly to remove the beads and the remaining procedure was carried out to the manufacturer's specifications. Purified RNA was quantified using a NanoDrop spectrometer.

Metabolic labeling of Sau rRNA

In vivo labeling of nascent RNAs was carried out as described (Wade et al., 1964). Briefly, cells containing pRMC2-PhERI or empty pRMC2 were induced with 100ng/ml anhydrotetracycline at O.D.600 0.2, allowed to grow for 40 minutes (1 normal doubling time), after which 200µCi of P^{32} labeled orthophosphoric acid was added directly to the growth media. $2x10^8$ cells were collected after 20 minutes and RNA purified as described above. To visualize rRNA, total RNA was run on a 6% polyacrylamide gel, which was stained with 1x gelred (Phenix Research) to visualize total RNA. The same gel was then exposed to a phosphoimaging cassette to visualize P^{32} incorporation into cellular RNA. P^{32} containing rRNA was quantified using ImageQuant. Three independent experiments were performed and relative P^{32} incorporation was averaged. Total RNA was quantified by Nanodrop and averaged over three independent experiments.

RNA-seq: Sample preparation and Sequencing

Total RNA was purified as described above from Sau RN4220 cells containing pRMC2 and pRMC2-PhERI at OD_{600} = 0.4 after the addition of anhydrotetracycline (100ng/ml) at OD_{600} = 0.2. For RNA-seq analysis, $2x10^8$ cells were immediately added

to 2x volumes of BioStabilize (Qiagen) and incubated at room temperature for 10 minutes before RNA purification.

For the RNA-seq analysis, mRNAs are generally enriched to exclude the majority of large, structured rRNAs. RiboZero rRNA removal kit for gram-positive organisms (Epicenter) was used to eliminate the 16s and 23s rRNA species prior to sequencing analysis. RNA quality was then checked on a BioAnalyzer (Agilent) chip prior to cDNA library synthesis. cDNA libraries were prepared by standard techniques for subsequent Illumina sequencing using the mRNA-seq Sample Prep kit (Illumina) eliminating the step for mRNA amplification. Briefly, after the rRNA reduction, RNA was fragmented and used as a template for a randomly primed PCR. After the amplification, ends are repaired and ligated to Illumina adapters. The cDNA library is then verified for appropriate fragment size (200-300bp) on a BioAnalyzer chip.

Samples were amplified onto flowcells using an Illumina cBot and sequenced on an Illumina HiSeq2000 for 51 cycles per manufacturer protocols. Raw sequencing data was processed using the onboard SCS/RTA software yielding 51bp reads.

RNA-seq: Data Analysis

Sequencing reads were processed using TopHat (Trapnell et al., 2009), an alignment package designed to align sequencing reads derived from transcribed RNA. Briefly, the program aligns reads to a reference genome, identifying regions of coverage that correspond to transcribed RNA. These regions are joined and queried for potential junctions by attempting alignment of reads that did not initially align. Reads aligning to multiple locations are kept (to a maximum of 20 potential positions) to assist constructing

gene models for genes with repetitive or low complexity features. When aligning reads, 2 mismatches to the reference (Ensembl S_aureus_nctc_8325.EB1.fa) were allowed.

Alignments reported from TopHat were processed by the Cufflinks software package (Trapnell et al., 2010) to determine differential expression of genes and transcripts between conditions.

Alignments were quantitated against the Ensembl annotation: (S_aureus_nctc_8325.EB1_s_aureus_nctc_8325.gtf).

Expression values are reported as fragments-per-kilobase-of-gene-per-million-mapped reads (FPKM). Data were visualized using the Integrated Genomics Viewer (Robinson et al., 2011).

Open Promoter Complex Stability Assay

Open promoter complexes were assayed as previously described. Briefly PhERI (1μM), or empty protein buffer, was incubated with σ^A (100nM) for 10 minutes before the addition of Sau core RNAP (50nM). The complex was incubated on ice for 10 minutes. For the aag and rrnA promoter, 50nM of linear DNA was added to the reaction in 1x Sau transcription buffer, as described. Open promoter complexes were allowed to form for 20 minutes at 37°C before being challenged by the addition of 1μM double stranded FullCon promoter fragment. At time points after the addition of the FullCon promoter fragment (Gaal et al., 2001), reactions were initiated by the addition of 200μM NTPs with α-P^{32} labeled ATP. Reactions were stopped after 5 minutes by the addition of 2x stop buffer and electrophoresed on a 12% Urea-PAGE gel as described above.

For the G1-PhERI promoter, we assayed open promoter complexes by filter binding. The linear promoter fragment was end-labeled with P^{32} using polynucleotide kinase (PNK, New England Biolabs) under standard conditions. Open complexes were formed by incubation of holo +/- PhERI at the above concentrations at 37°C and challenged by the addition of 1μM unlabeled FullCon promoter fragment. At time points after the addition of challenging DNA, 10μl aliquots were pipetted onto prewashed filter papers (MF-Membrane Filters, 0.45μM, Millipore), allowed to bind for 10 seconds, and washed with 1x wash buffer (10mM Tris pH 8.0, 0.2M NaCl). Filter papers were dried and quantified on a phosphoimaging screen. All samples were normalized to the signal at time = 0.

Construction of hybrid promoters

Hybrid promoters for transcription assays and DNAse footprinting were made by megaprimed PCR. Breifly, a DNA oligo containing the DNA upstream of the -35 from the rrnA promoter and the -35 element of the dnaA promoter was used, with an rrnA upstream primer, to amplify the upstream promoter region. This was then used as a megaprimer using a dnaA downstream primer to amplify the full hybrid promoter. The dnaA(up)-rrnA(down) promoter was made using the inverse strategy: an initial PCR using a primer for upstream dnaA promoter and the rrnA -35 element. The first PCR was then used to megaprime using the rrnA downstream primer. Sequences were verified by standard DNA sequencing.

DNAse I footprinting

DNA fragments were made by PCR and end-labeled on the template strand. The reverse PCR oligo (100pmol) was gel purified, P32 end labeled with polynucleotide kinase (PNK), and unincorporated nucleotide was removed on a sephadex G-50 spin column. The radiolabeled primer was then added to a standard PCR reaction using a blunt-end PCR enzyme and the PCR product was gel purified.

Reactions were performed by forming a complex between sigma (5μM) PhERI (5μM) and core (2μM) on ice. Labeled promoter DNA (0.1μl) and DNAseI reaction buffer (1x : 5mM Tris-acetate pH 7.9, 5mM KCl, 1mM MgCl2) were added and incubated at 37°C for 10 minutes. DNAseI (0.1μg/ml) was added for 1 minute and reactions stopped with 15mM EDTA. Samples were boiled for 5 minutes, separated on a 6% UREA PAGE gel and visualized on a phosphoimager screen.

β-flap : Protein Expression and Purification

The Sau β-flap (Sau β residues 789-916) was cloned into the modified pET vector pSKB2 using the restriction enzymes NdeI and BamHI to contain an N-terminal 6(his) tag upstream of a precision protease cleavage site. The Sau β-flap was then transformed into Eco BL21(DE3) cells and induced at 37°C for 3h using 1mM IPTG.

Lysate from 2L of BL21(DE3) cells was lysed by French Press, cleared by centrifugation and loaded onto a HiTrap IMAC column (GE Healthcare). The column was washed with 20mM, 40mM and 60mM Imdazole in purification buffer (20mM Tris HCl pH 8.0, 0.5M NaCl, 5% glycerol and 0.5mM β-ME). Protein was eluted from the column in 250mM Imdazole, dialyzed overnight in the presence of precision protease

(50:1 protein:protease) into purification buffer, and reloaded onto a HiTrap IMAC column to remove uncleaved protein and tag. The β-flap was finally purified by size exclusion chromatography on an SD75 gel filtration column (GE healthcare).

Crystallization of β-flap/PhERI/σ^A_4 complex

The PhERI/σ^A_4 complex was purified as described above, and the Sau β-flap was purified separately. The PhERI/σ^A_4 complex was concentrated to 10mg/ml and the β-flap to 5mg/ml; both proteins were buffer exchanged into crystallization buffer (10mM Tris HCl pH 8.0 and 0.5M NaCl). The two protein solutions were mixed (10mg/ml PhERI / σ^A_4 complex; 5mg/ml Sau β-flap) on ice and incubated for 10 minutes prior to setting up crystallization screens. Initial hits were further refined by 2-dimensional screening. Seeding from initial needles into two conditions (0.1 M MES pH 6.5, 2% (w/v) PEG 10,000, and 0.2M Magnesium chloride, and 0.1 M HEPES pH 7.5, 2% (w/v) PEG 4000) reproducibly produced large plates. We evaluated the protein content of these crystals by looping several large plates, washing them 8 times in fresh mother liquor and resuspending them in 2x SDS-loading buffer. After extensive boiling, the solution was run on an 8-25% PhastGel and stained with coomassie blue. For diffraction analysis, the crystals were cryoprotected in mother liquor plus 20% glyercol for both conditions before being flash frozen in liquid nitrogen.

Data Collection and Analysis

The crystals of the 3-protein complex were tested for diffraction at the X29 beamline at Brookhaven National Laboratories. One crystal diffracted to under 4.0Å

resolution. Data were indexed and scaled using HKL2000 (Otwinowski and Minor, 1997). We searched for molecular replacement solutions using the PhERI/σ^{A}_{4} complex using both Phaser (Mccoy et al., 2007) and MolRep. Data was analyzed for diffraction defects using Phenix (Adams et al., 2010) and low-resolution DEN refinements were attempted using CNS (Brunger et al., 1998; Schroder et al., 2010).

Crystallization of the β-flap

The Sau β-flap alone was purified as described, dialyzed into crystallization buffer (10mM Tris HCl pH 8.0 and 0.5M NaCl) and screen for crystallization conditions. Large, high quality crystals were produced in many conditions, including 0.1 M Sodium acetate pH 4.6, 8% (w/v) PEG 4000. These crystals were cryoprotected in 20% glycerol before being flash frozen in liquid nitrogen. Data were collected on the X29 beamline at Brookhaven National Laboratories. Data were indexed and scaled using HKL2000 (Otwinowski and Minor, 1997) and molecular replacement solutions identified using Phaser (Mccoy et al., 2007). The model building was performed using Coot (Emsley and Cowtan, 2004) and refined using Phenix (Adams et al., 2010).

Sau Transcription Assays using Rifalazil

Sau transcription assays were performed exactly as described above, with the addition of Rifalazil at the indicated concentrations. Briefly, open promoter complexes were formed by mixing Sau core with σ^{A}, adding linear promoter DNA fragments, and incubating at 37°C for 10 minutes. 2 minutes prior to the initiation of the reaction,

Rifalazil was added to the complex. Reactions were then initiated by the addition of NTPs exactly as previously above.

Crystallization of Sau RNAP

Sau RNAP was purified as described above. After the final step in the purification, protein was dialyzed into 1x crystallization buffer (10mM Tris HCl pH 8.0, 0.2M NaCl, 1mM DTT) and extensively screened for crystallization conditions. In one condition (0.1 M HEPES pH 7.5, 0.2 M $MgCl_2$, 30% PEG400), small crystals grew after approximately 14 days. We evaluated the protein content of these crystals by visualizing them on a UV microscope (JANSi UVEX). Sau RNAP crystals were cryoprotected by a quick dip in mother liquor supplemented with 5% glycerol and flash frozen in liquid nitrogen. We evaluated crystal diffraction on the microdiffractomoter (beamline 24-ID-E) at the Advanced Photon Source at Argonne National Laboratories. The one diffracting crystal was indexed using HKL2000 (Otwinowski and Minor, 1997).

CarD: *In vitro* transcription assays

Proteins used in *in vitro* transcription assays were diluted into 1x protein storage buffer (10mM TrisHCl pH 8.0, 0.15M NaGlu, 15% glycerol, 1mM DTT). CarD, at the indicated concentration, was preincubated with core RNAP (50nM) for 10 minutes on ice, followed by the addition of σ (Tth σ^A full length or $\sigma^A\Delta$N1.1, as indicated; 100nM). After 10 minutes, linear promoter DNA (50nM) was added and the reaction brought to 20µl in 1x transcription buffer (40mM Tris-acetate pH 7.9, 10mM $MgCl_2$, 1mM EDTA, 50ug/ml BSA, 100mM NaCl and 1mM DTT). Reactions were incubated for 10 minutes

at 65°C to form open promoter complexes and then initiated with NTPs (200uM GTP, CTP, UTP, 50uM ATP, 0.1ul α-P^{32}ATP). After 5 minutes reactions were stopped with 2x formamide buffer (98% formamide, 5mM EDTA) and run on a 12% Urea PAGE gel. Products were visualized on a phosphoimging screen and quantified using ImageQuant. Promoter fragments were prepared by PCR, purified on a 1.5% agarose gel, and electroeluted into dialysis tubing. Following phenol/chloroform extraction, DNA was ethanol precipitated and resuspended in 10mM TrisHCl pH 8.0 to 1μM final concentration.

CarD: Abortive Initiation Assays

Abortive transcription assays were performed as above, but initiated with NTPs suitable to give only short products. The 23s rRNA promoter was tested by forming open promoter complexes as above. Transcription was initiated by adding the initiating nucleotide, GTP (200nM) and labeled CTP (50μM CTP, 0.1μl α-P^{32}CTP) to form a 4-base RNA product. T7A1 was tested by forming open complexes as above and initiating with ApU dinucleotide primer (20μM) corresponding to the +1 and +2 nucleotides and radiolabeled CTP (50μM CTP, 0.1μl α-P^{32}CTP).

CarD: Open Complex Stability assays

Reactions were prepared as above for the *in vitro* transcription assays. Briefly, CarD or buffer, was preincubated with core RNAP (50nM) for 10 minutes before the addition of $\sigma^{A}\Delta$N1.1 (100nM). Linear promoter DNA (50nM) was added and the reactions were incubated at 65°C for 10 minutes to form open complexes. Once open

complexes were formed, they were challenged with the FullCon double stranded DNA fragment in large molar excess (1μM). For the 23s rRNA promoter, open complexes were quantified by initiating transcription with NTPs at the concentrations used above at various times after the addition of competing DNA. 5 minutes after the addition of NTPs, reactions were stopped with 2x formamide buffer, run on a 12% Urea PAGE gel, visualized and quantified as above.

On the T7A1 promoter, open complexes were quantified by filter binding, as described above. Briefly, linear promoter DNA fragments were end-labeled with P^{32}. Labeled promoter DNA (50nM) was added to RNAP holoenzyme +/- CarD. After incubation at 65°C to form open complexes, a large excess (1μM) of the double stranded FullCon promoter fragment was added. 10μl aliquots were taken at different time points after the addition of challenging DNA and bound to pre-washed filters (MF-membrane filters, Millipore) and immediately washed with 4ml 1x wash buffer (10mM Tris HCl pH 8.0, 200mM NaCl). Radioactive signal, corresponding to labeled promoter DNA fragment bound to RNAP, was quantified after exposure on a phosphoimaging cassette using ImageQuant.

CarD: Ni-bead pulldown

10(his)-tagged Tth RNAP (2μM) was first incubated with CarD (10μM) or the CarD-RID (10μM) and $\sigma^{A}\Delta N1.1$ (5μM) for 10 minutes on ice in 1x pull-down buffer (20mM Tris HCl pH 8.0, 0.25M NaCl, 1mM DTT). The proteins were then added to pre-washed and equilibrated Ni-NTA agarose beads (Qiangen) and incubated with gentle mixing for 1h at 22°C. After binding, the slurry was centrifuged for 3 minutes at 4000xG

to separate the beads from the buffer. Supernatent was carefully removed, and the beads were washed extensively (5 times 500µl, 10 minutes with gentle shaking per wash) with 1x pull-down buffer containing 25mM imidazole. After each wash, the slurry was centrifuged and supernatent removed. Protein was eluted from the Ni-NTA resin with pull-down buffer containing 500mM Imidazole (in 50µl aliquots). 10µl aliquots were removed from each step (loading, washes and elution), added to 10µl 2x SDS loading buffer, and visualized on a 4-12% SDS-PAGE gel. A negative control, using CarD (10µM), the CarD-RID (10µM) and $\sigma^A\Delta$N1.1 (5µM), none of which contain his-tags, was performed to evaluate non-specific binding to the Ni-NTA agarose resin.

Construction of Pf α expression cassettes

Full-length Pf α was amplified from Pf genomic DNA using a modified PCR protocol as described. Primers for cloning into a modified pET vector, pRMC2, were used with an NheI site upstream of the start codon and EcoRI site downstream of the stop codon. Sequences were verified by DNA sequencing

For expression of tagged α in Pf, I used the vector pHBIRH. Pf α was amplified by PCR with an upstream primer containing a SpeI site and a downstream primer containing an in-frame NotI site. The pA or PTP or mCherry tag was amplified by PCR with a primer containing a NotI site upstream of the first amino acid (containing an additional base to maintain the coding frame) and a SacI site downstream of the stop codon. Both PCRs were digested with NotI and ligated for 1 hour at room temperature under standard conditions. A second PCR was performed with the upstream α primer and the downstream tag primer to amplify a single linear DNA containing the α coding

region and the tag in a single reading frame. This linear DNA fragment was then cloned into pHBIRH (with SpeI and SacI) by standard molecular techniques. Sequences were verified by DNA sequencing.

Culture and transfection of Pf parasites

Pf C3 parasites were cultured and transfected by standard procedures (Deitsch et al., 2001; Epp et al., 2008). Briefly, parasites were cultures at 5% haematocrit in RMPI 1640 media with the addition of 0.5% Albumax II, 0.25% sodium bicarbonate and 0.1mg/ml gentamicin to prevent bacterial contamination. Cultures were grown at 37°C in 90% nitrogen.

Transfections were performed using loaded RBCs. 0.175ml of erythrocytes and 50-75µg of plasmid DNA were added to 0.2cm electroporation cuvettes in cytomix. The DNA/erythrocyte mix was elecetroporated with 0.31kV and 960mFD. Electroporation was performed once, followed by incubation for 24-48 hours and a second electroporation. After 2-4 days, WR99210 was added to 40ng/ml to select for the presence of the transgene.

RNA extraction and RT-PCR

RNA was extracted from Pf cells essentially as described. Briefly, 20mls of Pf parasites at 5% parasitimia was centrifuged to collect RBCs. After one wash in 1x PBS, cells were resuspended in 1ml PBS with the addition of 10% saponin to lyse RBCs. After centrifugation, parasites were washed two times in 1x PBS, after which 750µl of tri-reagent was added. The aqueous phase was further purified by the Qiagen RNeasy

RNA purification kit. RNA was eluted in 15µl and quantified by Nanodrop. RNA was treated with DNaseI to remove all traces of genomic and plasmid DNA and cDNA was synthesized by reverse transcriptase (Invitrogen). PCR from the cDNA was performed using Accuprime DNA polymerase and a 1:10 dilution of the cDNA.

Western Blot

20mls of Pf parasites at 5% parasatimia were collected, centrifuged to collect RBCs and washed with PBS. RBCs were lysed in 1ml PBS with the addition of 10% saponin. Parasites were subsequently washed in 1ml PBS, centrifuged and flash frozen in liquid nitrogen.

20µl of 2x SDS load buffer was directly to the parasite pellet, and the samples were boiled for 30 minutes at 95°C. Nup130, expressed in yeast, was used as a positive control and treated essentially as the parasite pellet. Samples were electrophoresed on a 4-12% SDS-PAGE gel and transferred to a membrane for blotting in 1x CAPS buffer in 10% methanol. Membranes were blotted using an anti-pA antibody (IgG; Cappel) and visualized with a peroxidase doupled secondary antibody (Amersham NA934).

Rifampicin pull-outs

Rifampicin coupled agarose beads were synthesized with increasingly long linkers by Arkady Mustaev. Purified Eco RNAP was used for initial binding tests. Briefly, Eco RNAP (50µl at 0.5mg/ml) was loaded onto the beads in 1x pull-out buffer (10mM Tris HCl pH 8.0, 0.2mM NaCl, 1mM DTT). After one hour with gentle mixing, the beads were collected by gentle centrifugation and the supernatant removed. Beads

were washed 10 times with 500µl pull-out buffer and bound protein eluted from the beads by denaturation with 6M guanidinium. Eluted fractions were precipitated with acetone and resuspended in 2x SDS load buffer. Load, flow-through and wash (10µl aliquots) and elution fractions were visualized on a 4-12% SDS-PAGE gel by silver staining.

RNAP was purified from Eco lysates as above. 50ml of Eco CC118 cells were collected at OD_{600} of 0.6 and collected by centrifugation. Cells were resuspended in 1x pull-out buffer and lysed by sonication. Lysate was cleared by centrifugation at high speed and 200µl of lyaste was added to Rif-conjugated agarose beads. Purification was performed as decribed above.

References:

Adams, P.D., Afonine, P.D., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica D Biological Crystallography *D66*, 213-221.

Bailey, S. (1994). The Ccp4 Suite - Programs for Protein Crystallography. Acta Crystallogr D *50*, 760-763.

Brunger, A.T., Adams, P.D., Clore, G.M., Delano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S.*, et al.* (1998). Crystallography & NMR system: a new software system for macromolecular structure determination. Acta Crystallographica D Biological Crystallography *D54*, 905-921.

de La Fortelle, E., Irwin, J.J., and Bricogne, G. (1997). SHARP: A maximum-likelihood heavy-atom parameter refinement and phasing program for the MIR and MAD methods. In Crystallographic Computing, P. Bourne, and K. Watenpaugh, eds. (Boston, Kluwer Academic Publishers), pp. 1-9.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallographica D Biological Crystallography *60*, 2126-2132.

Gaal, T., Ross, W., Estrem, S.T., Nguyen, L.H., Burgess, R.R., and Gourse, R.L. (2001). Promoter recognition and discrimination by EsigmaS RNAP. Molecular Microbiology *42*, 939-954.

Mccoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology *276*, 307-326.

Potterton, E., Briggs, P., Turkenburg, M., and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. Acta Crystallogr D *59*, 1131-1137.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., and Barrell, B. (2000). Artemis: sequence visualization and annotation. Bioinformatics *16*, 944-945.

Schroder, G.F., Levitt, M., and Brunger, A.T. (2010). Super-resolution biomolecular crystallography with low-resolution data. Nature *464*, 1218-1222.

Terwilliger, T.C. (2000). Maximum-likelihood density modification. Acta Crystallogr D *56*, 965-972.

Wade, H.E., Lovett, S., and Robinson, H.K. (1964). The autodegradation of 32-P-labelled ribosomes from Escherichia coli. Biochem J *93*, 121-128.

Weeks, C.M., Blessing, R.H., Miller, R., Mungee, R., Potter, S.A., Rappleye, J., Smith, G.D., Xu, H., and Furey, W. (2002). Towards automated protein structure determination: BnP, the SnB-PHASES interface. Z Kristallogr *217*, 686-693.

Appendix:

RNA-seq data tables

Table A.1a: Genes significantly downregulated by PhERI expression. Only genes with a P-value < 0.05 are listed.

Genome position	Gene function	NCTC 8325-4	RN 4220	2ln(fold change)	locus (SAUOHSC)
9754		118.291	21.109	-1.723	0007
12785	trna synthesis	338.936	69.359	-1.587	***serS
78526	Fe transport	320.21	51.904	-1.820	00074 (sirA)
99487		751.303	128.28	-1.768	sodM
164124	Pyrimidine metabolism	109.03	17.41	-1.835	00152 (187)
164748	indolpuyruvate decarboxylate	139.571	21.343	-1.878	00153 (0188)
166582		628.011	93.171	-1.908	ptsG
199582	isoprenyl cystein carboxyl methyltransferase	382.718	86.026	-1.493	00182 (0220)
222778		120.479	20.511	-1.770	0200
271578		431.454	60.837	-1.959	00253 (280)
297407		285.614	47.161	-1.801	00282 (306)
341745		148.352	33.927	-1.475	00328 (0346)
342463		129.186	24.217	-1.674	00329 (347)
342787		168.878	31.855	-1.668	00330 (348)
366132		187.023	41.175	-1.513	00357 (0373)
402507		44.8496	7.0150	-1.855	00399 (set15)
500137	NTP synthesis	742.144	89.702	-2.113	pdxS
501028		841.017	116.51	-1.977	pdxT (sav0520)
536141	2-amino-3-oxobutane from acetyl coa	242.677	39.254	-1.822	00532 (550)
547607		36.7493	7.0773	-1.647	00543.1
551104		507.965	83.640	-1.804	sdrD
690241		195.406	23.006	-2.139	0707
691928	frutcose permease	316.081	52.638	-1.793	00708 (FruA)
722747		1358.73	86.276	-2.757	00738 (SAV0727)
776118		3048	443.72	-1.927	00794

					(gapR)
801076	cold shock	5402.94	322.97	-2.817	cspC
804681		207.55	48.308	-1.458	00827 (0824)
806086	028697 acyl transferase	352.807	59.718	-1.776	00830 (0826)
864859	pgi glucose 6-phosphate (required)	1297.41	192.49	-1.908	***pgi
890664		999.459	146.08	-1.923	00919 (0982)
909471		411.913	67.071	-1.815	00937 (1000)
940007		185.156	31.379	-1.775	00963 (1028)
947204		111.982	25.901	-1.464	00973
990820		853.732	147.18	-1.758	01021 (1075)
993034		579.556	64.304	-2.199	01022 1077)
1004297	mrna degredation and processing (required)	966.567	134.26	-1.974	***01036(1089)
1006241		914.677	119.55	-2.035	01037 (def)
1028809		702.261	97.723	-1.972	01064 (pycA)
1033752	heme biosynthesis	527.373	69.373	-2.028	ctaB
1034688		473.681	67.918	-1.942	01067 (1118)
1051882	trna synthetase	281.582	53.940	-1.653	***pheS
1114346	uracil permease	372.335	32.925	-2.426	pyrP
1115681	pyrmadine biosynthesis	409.408	36.638	-2.414	pyrB
1116580	Pyrimidine biosynthesis	413.906	54.705	-2.024	01168 (pyrC)
1117856	Pyrimidine biosynthesis	430.515	85.961	-1.611	01169 (pyrAA)
1122232	Pyrimidine biosynthesis	461.721	87.618	-1.662	01172 (pyrF)
1148092		1135.87	185.42	-1.813	fapR (1228)
1148669	negative regulator of fatty acid biosynthesis	1206.48	210.10	-1.748	plsX
1204422		571.444	71.310	-2.081	01252 (1275)
1242420		1896.92	236.95	-2.080	01285 (glnR)
1242807	glutamine synthetase (required)	2175.49	290.89	-2.012	***01287 (glnA)
1249421		46.319	7.8498	-1.775	1297
1253698		69.4637	10.140	-1.924	01309 SAV1316
1268367	amino acid transporter	597.115	92.257	-1.868	1326
1272662	NTP biosynthesis	304.507	41.202	-2.000	guaC (sav1337)
1273814		458.992	76.713	-1.789	01332

					(1338)
1304208	Oxidative stress	874.423	115.66	-2.023	msrA1
1426309	tRNA synthetase	888.695	146.32	-1.804	asnS
1458084		643.869	125.43	-1.636	01505 (1485)
1545327	malonyl co-a fatty acid biosynthesis (required)	525.605	68.133	-2.043	***01624 (accB)
1659667	MreD	238.947	27.886	-2.148	01759 (1648)
1798610	riboflavin biosynthesis	673.018	142.48	-1.553	ribH
1799087	riboflavin biosynthesis	602.566	124.79	-1.575	01887 (ribA)
1800279	riboflavin biosynthesis	647.036	107.00	-1.800	01888 (ribB)
1800918	riboflavin biosynthesis	654.379	115.33	-1.736	01889 (ribD)
1811161	fructose 6 phosphate aldolase	817.344	170.19	-1.569	01901 (1781)
1883879		1006.38	193.93	-1.647	01979 (1847)
1991203	trna synthesitase	953.649	201.14	-1.556	gatC
2103845		593.509	125.44	-1.554	rex 2273 (xcription reg) (2046)
2152964		248.686	53.497	-1.537	02319 2084)
2189428		261.971	54.169	-1.576	02367 (2126)
2256513	Fe transport	477.757	96.856	-1.596	02430 (htsA)
2270735	prostoglandin dehydrogenase, putative	177.138	38.462	-1.527	02447 (2187)
2272019		83.2365	16.023	-1.648	02448 (2188)
2348592		452.331	43.248	-2.347	02554 (2284)
2357498	transcriptional regulators: GntR family	1032.03	200.06	-1.641	sarR
2370059	ribosomal protein E (required)	394.45	42.534	-2.227	***rpsE
2370857		200.504	20.811	-2.265	2580.1
2438501		146.33	33.948	-1.461	02652 (2370)
2453507		181.402	38.899	-1.540	02668 (2385)
2454270		825.607	75.899	-2.387	sarZ
2454932	protein oxidoreductase	533.963	113.27	-1.551	02670 (2387)
2477970		129.194	23.611	-1.700	02694 (2409)
2478588		86.8084	14.415	-1.795	02695 (2410)

2595325		94.3532	15.658	-1.796	02816 (2513)
2619737		92.976	15.579	-1.786	02845 (5326)
2620123	glycolosis	98.7279	11.306	-2.167	02846 ptsG
2691654	Na/SO4 symporter	2120.95	361.13	-1.770	fda (2606)
2801301	Cold shock protein	147.207	27.530	-1.677	03030 (2694)
2813019	trna synthesis	6199.66	634.56	-2.279	cspB

Table A.1b: Genes significantly upregulated by PhERI expression. Only genes with a P-value < 0.05 are listed.

Genome position	Gene function	NCTC 8325-4	RN 4220	2ln(fold change)	locus (SAUOHSC)
42680	trna modification	183.00	1346.6	2.00	39
110292		6.90	31.31	1.51	106
119491	polysaccharide binding prot	4.25	22.63	1.67	114
121648	nadp sugar binding	3.07	14.48	1.55	117
125615	cell wall biosynthesis	4.00	18.92	1.55	120
130861	cell wall biosynthesis	9.13	40.99	1.50	125
133575	cell wall biosynthesis	15.50	66.86	1.46	128
140421		20.74	147.88	1.96	134
141657	ABC transporter	37.82	159.99	1.44	136
190529		33.36	208.49	1.83	174
303758	pfkb carbohydrate kinase	2.41	16.74	1.94	291
308682	sialic acid synthethetase	3.14	17.61	1.72	nanA
330822	MepB (MFS family) transporter	34.68	152.72	1.48	316
337879		4.60	24.69	1.68	325
339970		7.63	38.28	1.61	327
349924	METHIONINE BIOSYNTHESIS	15.25	124.41	2.10	340
365422		26.05	432.25	2.81	356
366695		30.63	660.34	3.07	358
564127	PROLINE TRANSPORT	110.78	589.76	1.67	556
567449	acetyl co-a transferase	24.70	157.40	1.85	558
568931		33.70	187.21	1.71	560
610261		6.33	122.99	2.97	619
614349		9.42	64.79	1.93	624
614928	translation	17.25	86.63	1.61	mnhB2
619570	translation	15.67	95.92	1.81	mnhG2
628913	glycosyl transferase ie WecG / TagA	193.52	1213.98	1.84	640
657303	PHOSPHATE TRANSPORTER	562.03	4493.62	2.08	669
657936	PHOSPHATE TRANSPORTER	533.30	4734.35	2.18	670
688883		226.10	1372.71	1.80	704
716261	sulfate transporter	123.67	1232.61	2.30	731
718985	Histidine biosynthesis	28.07	298.82	2.36	hisC
720394		160.51	912.64	1.74	734
816025		30.42	884.46	3.37	845
855576		298.98	1401.18	1.54	892
889973		196.80	1643.28	2.12	918

901688	peptide transporter	4.00	17.65	1.49	929
908437		5.42	25.51	1.55	936
924132		45.07	435.87	2.27	950
957785	Cystein protease	28.61	132.72	1.53	sspC
960953	Aspartate amino transferase	179.25	1737.92	2.27	989
1025648		289.56	1585.69	1.70	1060
1025817		288.33	1646.66	1.74	1061
1081622	ARG BIOSYNTHESIS	7.08	70.45	2.30	1128
1082646	Carbamate kinase	10.42	73.49	1.95	arcC1
1083816	anion permease	16.52	71.89	1.47	1130
1086146	mallic acid transporter	165.48	891.37	1.68	1133
1087342		126.91	850.30	1.90	1134
1162044		43.21	256.51	1.78	1213
1186353	zinc metallo protease	346.89	2482.23	1.97	1239
1187557	trna synthetase	580.53	4381.59	2.02	proS
1276192		205.49	903.45	1.48	1336
1329326	phosphate transporter	4.43	19.67	1.49	1386
1332693		5.91	68.22	2.45	A01332
1349110		6.26	49.50	2.07	1410
1446046		674.41	5336.69	2.07	gpsA
1447061		593.55	4578.22	2.04	engA
1531152	ArgE, Arg biosynthesis	72.93	438.67	1.79	1606
1596249	ROS scavenging	73.59	361.01	1.59	1686
1635662		35.81	346.15	2.27	1729
1635848		33.56	337.70	2.31	1730
1692301	Co-A biosynthesis	202.54	957.01	1.55	1795
1705673		6.24	40.28	1.87	1804
1706497		10.18	61.67	1.80	1805
1723310		105.23	498.33	1.56	1817
1770560		76.57	813.58	2.36	1864
1795367		8.52	101.50	2.48	1880
1806781		1009.79	8086.02	2.08	1895
1813963		3.74	27.20	1.99	1905
1843315	Serine protease	7.15	32.86	1.53	splE
1844133	Serine protease	4.94	25.96	1.66	splD
1845029	Serine protease	4.11	44.18	2.37	splC
1845806	Serine protease	3.09	29.60	2.26	splB
1846653	Serine protease	2.76	19.67	1.96	splA
1889484		125.92	915.47	1.98	1984
1889998		8.95	45.97	1.64	1985
1978642		18.50	179.01	2.27	2104
1979280		221.48	1459.98	1.89	2106
1980013		202.69	1335.89	1.89	2107
1982200		16.68	295.79	2.88	2109
2031740		71.08	472.18	1.89	hlb
2202995		938.94	14904.3	2.76	2381
2208189	nadp binding prop	9.36	94.24	2.31	2387

2223263		14.04	68.34	1.58	2402
2264436		4.57	24.59	1.68	2437
2266096		44.58	1452.99	3.48	asp23
2266668		14.04	489.08	3.55	2442
2266920		19.98	513.40	3.25	2443
2267631		3.47	72.07	3.03	2444
2279891	Lactose metabolism	71.09	478.15	1.91	lacR
2289533		3.54	33.20	2.24	2466
2290322		33.92	142.67	1.44	2467
2389508		296.07	1437.99	1.58	2600
2391719		41.85	188.11	1.50	2603
2420343		52.11	235.33	1.51	2632
2425298		5.91	68.22	2.45	A02505
2503504	glycerate kinase	85.32	512.81	1.79	2723
2504706		201.33	1558.81	2.05	2724
2516693	NADP binding	176.09	1067.86	1.80	2737
2517700		52.52	797.26	2.72	2738
2519027	drug transporter	392.01	3899.53	2.30	2740
2525681	amino acid transporter	36.93	372.04	2.31	2750
2549036	2-deoxy-D-gluconate 3-dehydrogenase NADP	6.40	29.76	1.54	2772
2590644		28.17	208.43	2.00	2812
2591199		34.47	190.78	1.71	2814
2650766	Squalene synthase	7.71	34.11	1.49	crtN
2652286	Squalene desaturase	5.28	30.00	1.74	crtM
2653186		3.32	20.07	1.80	crtQ
2654319	Phytoene dehydrogenase	2.91	16.62	1.74	crtP
2659919	acetyl transferase	18.93	138.41	1.99	2886
2735411		46.69	299.58	1.86	2973
2775173	Polysaccharide polymerase (biofilm)	7.20	44.80	1.83	icaD
2797472		315.36	1869.31	1.78	3024
2798580	Pyroglutamyl peptidase	30.13	177.12	1.77	pcp
2800599		9.03	54.80	1.80	3028
2816354		210.35	1261.91	1.79	mnmG
2819017	translation	149.09	832.60	1.72	mnmE
2820535	translation	146.23	801.59	1.70	rnpA

Table A.2a: 30 Most Highly Expressed Genes in NCTC8325-4 and RN4220

Genome position	Gene function	NCTC 8325-4	RN 4220	2ln(fold change)	p value	locus SAUOHSC
1142973-1143162	translation	21944	19572	-0.17	0.871	rpmB
1345916-1346117	cold shock, putative	18639	18982	0.03	0.979	cspA
520793-521162	translation	14647	15704	0.10	0.925	rplL
1151544-1151778	fatty acid biosynthesis	13995	10009	-0.48	0.624	acpP
518656-519079	translation	13826	19981	0.53	0.622	rplK
529959-530373	translation	13589	15650	0.20	0.849	rpsL
1587819-1587996	translation	13307	8910	-0.58	0.550	rpsU
1201451-1201721	translation	12420	12584	0.02	0.985	rpsO
530438-530909	translation	11783	12623	0.10	0.925	rpsG
1600105-1600357	translation	10753	6378	-0.75	0.439	rpsT
1659192-1659501	translation	10585	10572	0.00	0.999	rplU
1734576-1735179	translation	10422	10869	0.06	0.955	rpsD
2182601-2182856	translation	10254	9372	-0.13	0.895	rpmE2
2627401-2627620	conserved hypothetical	9651	9374	-0.04	0.965	2853
1161540-1161891	translation	9390	8857	-0.08	0.933	rplS
2821009-2821147	translation	9131	6259	-0.54	0.546	rpmH
360629-360926	translation	8764	9270	0.08	0.935	rpsF
2299326-2299725	translation	8524	8664	0.02	0.982	rpsI
360946-361450	translation	8500	8593	0.02	0.988	349
1658570-1658855	translation	8332	8814	0.08	0.935	rpmA
1658866-1659187	conserved hypothetical	8329	9101	0.13	0.897	1756
529607-529862	translation	8262	8630	0.06	0.949	526
1882163-1882508	conserved hypothetical	7817	3588	-1.12	0.250	1977
2299738-2300176	translation	7664	8713	0.19	0.856	rplM
2315349-2316183	translation	7159	8692	0.28	0.794	rplB
1180819-1181588	translation	6897	9291	0.43	0.689	rpsB
1682851-	translation	6236	9140	0.55	0.560	rpmI

1683052						
2361997-2362801	conserved hypothetical	6209	3634	-0.77	0.451	2571
1159734-1160010	translation	6076	5102	-0.25	0.790	rpsP
2316215-2317114	translation	6054	6813	0.17	0.859	rplW

Table A.3a: Single Nucleotide Polymorphisms previously identified as unique to RN4220, but also present in RNA-seq data for NCTC8325-4

Genome position	Putative genes and functions	8325	8325-4 /RN4220	Amino acid change	Locus tag
22181	---	C	A	---	intergenic
47652	---	T	.	---	intergenic
110019	phosphonate ABC transporter, substrate-binding protein, putative	C	A	16 V-F	00105
392716	---	G	A	---	intergenic
412762	---	.	T	---	intergenic
412765	---	C	G	---	intergenic
590402	Conserved hypothetical protein	G	.	Frameshift	00591
653801	---	C	T	---	intergenic
751285	SecA	A	T	449 E-V	00769
827849	TauE, sulfite exporter	A	T	164 T-S	00860
926213	hypothetical protein	G	C	235 P-A	00952
939304	ComK	C	T	53 E-K	00961
947899	hypothetical protein	C	.	232 P-X	00973
1016979	putative ABC transporter	G	A	220 E-K	01048
1020577	Conserved Hypothetical protein	G	T	286 S-*	01053
1042000	ribosomal protein L32	T	.	34 G-X	01078
1123048	PyrE pyramadine biosynthesis	G	A	42 G-S	01
1160513	rimM	G	A	106 A-T	rimM
1160531	rimM	A	G	112 K-E	rimM
1180886	rpsB 30S ribosomal protein S2	G	.	10 V-X	01232
1283784	hypothetical protein	C	.	943 D-X	01342
1358230	Kdg alpha-ketogluterate decarboxylase	C	T	590 D-N	01418
1562913	rhomboid family protein	A	T	337 *-K	01649
1636255	hypothetical protein	T	.	100 Q-X	01732
1683491	infC translation initiation factor IF-3	T	C	40 K-E	01786
1733515	EzrA	G	T	73 T-N	01827
2087725	GroEL	A	T	218 F-I	Transcript
2106539	ABC transporter, putative	A	T	602 L-F	Transcript
2166163	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	G	C	124 T-R	02337
2221850	hypothetical protein	C	A	244 S-Y	02401
2244467	---	A	G	---	intergenic
2244495	---	G	A	---	intergenic
2244932	hypothetical protein	C	.	296 *-X	02417
2244933	hypothetical protein	T	.	296 *-X	02417

2318272	---	G	A	---	intergenic
2318274	---	G	T	---	intergenic
2318290	---	C	A	---	intergenic 02512 <--> 02515
2349916	---	G	T	---	intergenic
2420618	hypothetical protein;-hypothetical protein	.	T	---	02632;02633
2447620	---	G	A	---	intergenic
2592012	hypothetical protein;-hypothetical protein	A	.	---	02813;02814
2689048	hypothetical protein	G	T	353 V-L	02923
2762204	hypothetical protein	T	C	1514 N-S	02990
2782820	imidazole glycerol phosphate synthase subunit hisF, putative	C	.	247 G-X	03008
649126	hypothetical protein	G	T	202 G-G	00661
841103	hypothetical protein	G	T	39 G-G	00877
841139	hypothetical protein	G	T	51 G-G	00877
1653482	tgt queuine tRNA-ribosyltransferase	G	A	91 I-I	01748
2166183	UDP-N-acetylglucosamine 1-carboxyvinyltransferase	C	A	117 G-G	02337
2383630	hypothetical protein	G	T	68 G-G	02591
2383660	hypothetical protein	G	T	78 G-G	02591
2446246	PTS system sucrose-specific IIBC component	C	T	42 E-E	02662
2446641	hypothetical protein	T	A	95 I-I	02663
2678563	hypothetical protein	T	C	107 P-P	02911

Table A.3b: Single Nucleotide Polymorphisms newly identified in NCTC8325-4

Genome Position	**8325**	**8325-4**	**Function**	**amino acid change**	**codon change**	**locus tag**
1009713	G	C	Pyruvate dehydrogenase	G64A	gGt/gCt	01041
1137863	T	C	Phosphatase, putative	I134T	aTt/aCt	01186
1653482	G	A	Queuine tRNA-ribosyltransferase	Syn	atC/atT	01747
1733515	G	T	EzrA	T73N	aCc/aAc	01827
1967009	C	T	Conserved hypothetical (YfkB-like)	syn	aaC/aaT	02090
2733480	G	T	aureolysin (putative)	Q231K	Caa/Aaa	02971

Table A.4a: Genes significantly upregulated in RN4220 relative to NCTC8325-4 by RNA-seq (listed by fold change)

Genome position	Gene	NCTC 8325-4	RN 4220	2ln(fold change)	p value	locus SAUOHSC
551104-555154	sdrD	19.6431	593.981	4.9183	0.000	sdrD
141657-142398	putative nitrate transporter	1.51517	44.2299	4.8674	0.000	00136
424709-425735	ABC transporter	3.37815	75.2257	4.4769	0.000	metN1
425738-426398	putative ABC transporter	2.33495	53.3995	4.5153	0.001	00424
896023-898037	putative ABC transporter	3.87502	69.9488	4.1740	0.001	00926
894010-896007	oppB putative	3.58305	62.2538	4.1189	0.001	00924
140803-141316	conserved hypothetical	23.6997	321.892	3.7636	0.001	00135
898055-899711	ABC transporter OppA, putative	8.17295	103.689	3.6652	0.002	00927
1263418-1264699	dhoM (thrC operon)	11.2979	117.789	3.3820	0.005	01320
624571-627080	putative ABC transporter	510.859	6489.18	3.6670	0.008	00636
2722967-2725601	clfB	130.692	1781.62	3.7689	0.009	clfB
1338096-1339086	asd (dap operon)	6.33142	68.3555	3.4324	0.009	01395
547607-550738	hypothetical	22.2434	233.87	3.3942	0.010	sdrC
142411-144144	srpL ABC transporter, putative	3.42187	36.4878	3.4145	0.015	00137
1339087-1340694	amino acid metabolism	9.4465	84.6008	3.1628	0.018	dapA
426434-427277	ABC transporter, putative	12.8884	105.589	3.0343	0.021	00426
1073723-1074074	conserved hypothetical	44.3692	332.065	2.9038	0.023	01115
2484751-2485438	hypothetical	238.492	1921.97	3.0105	0.024	gpmA
73428-74979	cell wall	643.985	7448.74	3.5319	0.027	spa
2794365-2795214	conserved hypothetical, putative SAM adeno transferase	2.12122	20.5846	3.2786	0.027	03021
813119-814833	MetN: Methionine transporter	16.4666	123.011	2.9011	0.030	00843
569401-569569	conserved hypothetical	458.178	3001.92	2.7119	0.031	00561
814850-815672	conserved hypothetical	31.9421	208.052	2.7034	0.032	00844
1736777-1739529	conserved hypothetical	13.7354	96.2486	2.8088	0.032	01832
144156-145188	conserved hypothetical	6.04009	48.4601	3.0041	0.032	00139

276307-280009	conserved hypothetical	15.999	98.0291	2.6152	0.038	00258
2115681-2116686	ilvC, valine / isoleucine metabolism	2.37121	19.9625	3.0736	0.040	ilvC
2581382-2584524	hypothetical	11.1302	63.99	2.5233	0.048	fnbA
2788278-2789097	tRNA synthetase	1.15579	9.79585	3.0832	0.049	hisZ
938929-939460	comK, competence transcription factor	3.63073	29.5025	3.0225	0.049	00961
1264704-1265766	thrC, threonine synthetase	26.608	152.859	2.5222	0.049	01321

Table A.4b: Genes significantly downregulated in RN4220 relative to NCTC8325-4 by RNA-seq (listed by fold change)

Genome position	**Gene**	**NCTC 8325-4**	**RN 4220**	**2ln(fold change)**	**p value**	**locus** (SAUOHSC)
2093157-2093673	RNAIII	1304.97	41.2993	-4.981	0.0000	RNAIII
2244538-2244724	conserved hypothetical	431.246	41.6507	-3.372	0.0002	02416
2290322-2291027	putative alpha-acetolactate decarboxylase	334.645	39.6627	-3.076	0.0006	02467
2291063-2292728	AlsS: Stationary phase metabolism	323.345	41.4389	-2.964	0.0010	02468

References:

(1925). The Bacteriophage. Science *61*, x.

(2010). Malaria Fact Sheet (World Health Organization).

Ades, S.E. (2007). Control of the alternative sigma factor σ^E in Escherichia coli. Current Opinion in Microbiology *7*, 156-162.

Aiyar, S.E., McLeod, S.M., Ross, W., Hirvonen, C.A., Thomas, M.S., Johnson, R.C., and Gourse, R.L. (2002). Architecture of Fis-activated transcription complexes at the *Escherichia coli rrnB* P1 and *rrnE* P1 promoters. Journal of Molecular Biology *316*, 501-516.

Allison, D.P., Ganesan, A.T., Olson, A.C., Snyder, C.M., and Mitra, S. (1977). Electron microscopic studies of bacteriophage M13 DNA replication. Journal of virology *24*, 673-684.

Artsimovitch, I., Svetlov, V., Murakami, K.S., and Landick, R. (2003). Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. Journal of Biological Chemistry *278*, 12344-12355.

Baba, T., Bae, T., Schneewind, O., Takeuchi, F., and Hiramatsu, K. (2008). Genome sequence of Staphylococcus aureus strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. Journal of Bacteriology *190*, 300-310.

Barker, M.M., Gaal, T., and Gourse, R.L. (2001a). Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. Journal of Molecular Biology *305*, 689-702.

Barker, M.M., Gaal, T., Josaitis, C.A., and Gourse, R.L. (2001b). Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation in vivo and in vitro. Journal of Molecular Biology *305*, 673-688.

Barne, K.A., Bown, J.A., Busby, S.J.W., and Minchin, S.D. (1997). Region 2.5 of the Escherichia coli RNA polymerase s70 subunit is responsible for the recognition of the 'extended -10' motif at promoters. EMBO Journal *16*, 4034-4040.

Bartlett, M.S., Gaal, T., Ross, W., and Gourse, R.L. (1998). RNA polymerase mutants that destabilize RNA polymerase-promoter complexes alter NTP-sensing by *rrn* P1 promoters. Journal of Molecular Biology *279*, 331-345.

Benner, E.J., and Morthlan.V (1967). Methicillin-Resistant Staphylococcus Aureus - Antimicrobial Susceptibility. New Engl J Med *277*, 678-&.

Berdygulova, Z., Esyunina, D., Miropolskaya, N., Mukhamedyarov, D., Kuznedelov, K., Nickels, B.E., Severinov, K., Kulbachinskiy, A., and Minakhin, L. (2012). A novel phage-encoded transcription antiterminator acts by suppressing bacterial RNA polymerase pausing. Nucleic Acids Research.

Bergh, O., Borsheim, K., Bratbak, G., and M, H. (1989). High abundance of viruses found in aquatic environments. Nature *340*, 467-468.

Berscheid, A., Sass, P., Weber-Lassalle, K., Cheung, A.L., and Bierbaum, G. (2012). Revisiting the genomes of the Staphylococcus aureus strains NCTC 8325 and RN4220. Int J Med Microbiol.

Black, L.W., and Peng, G. (2006). Mechanistic coupling of bacteriophage T4 DNA packaging to components of the replication-dependent late transcription machinery. Journal of Biological Chemistry *281*, 25635-25643.

Blatter, E., Ross, W., Tang, H., Gourse, R., and Ebright, R. (1994). Domain organization of RNA polymerase slpha subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. Cell *78*, 889-896.

Bohannan, B.J.M., and Lenski, R.E. (2000). Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage. Ecol Lett *3*, 362-377.

Borukhov, S., Polyakov, A., Nikiforov, V., and Goldfarb, A. (1992). GreA protein: A transcription elongation factor from *Escherichia coli*. Proceedings of the National Academy of Sciences USA *89*, 8899-8902.

Bown, J.A., Barne, K.A., Minchin, S.D., and Busby, S.J.W. (1997). Extended -10 promoters. In Nucleic Acids and Molecular Biology Mechanisms of Transcription, F. Eckstein, and D.M.J. Lilley, eds. (Berlin Heidelberg New York, Springer), pp. 41-52.

Bronfenbrenner, J.J., and Korb, C. (1925). Studies on the Bacteriophage of D'herelle : I. Is the Lytic Principle Volatile? J Exp Med *41*, 73-79.

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

Burian, M., Wolz, C., and Goerke, C. (2010). Regulatory adaptation of Staphylococcus aureus during nasal colonization of humans. PLoS One *5*, e10040.

Callaci, S., Heyduk, E., and Heyduk, T. (1999). Core RNA polymerase from E. coli induces a major change in the domain arrangement of the sigma70 subunit. Molecular Cell *3*, 229-238.

Camarero, J.A., Shekhtman, A., Campbell, E.A., Chlenov, M., Gruber, T.M., Bryant, D.A., Darst, S.A., Cowburn, D., and Muir, T.W. (2001). Autogregulation of a bacterial sigma factor explored using segmental isotopic labeling and NMR. Proceedings of the National Academy of Sciences USA *99*, 8536-8541.

Campbell, E.A., and Darst, S.A. (2000). The anti-σ factor SpoIIAB forms a 2:1 complex with σ^F, contacting multiple conserved regions of the σ factor. Journal of Molecular Biology *300*, 17-28.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell *104*, 901-912.

Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma factor. Molecular Cell *9*, 527-539.

Campbell, E.A., Westblade, L.F., and Darst, S.A. (2008). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. Current Opinion in Microbiology *11*, 121-127.

Chamberlin, M., McGrath, J., and Waskell, L. (1970). New RNA polymerase from Escherichia coli infected with bacteriophage T7. Nature *228*, 227-231.

Chatterji, D., and Ojha, A.K. (2001). Revisiting the stringent response, ppGpp and starvation signaling. Current Opinion in Microbiology *4*, 160-165.

Chaturvedi, D., Goswami, A., Saikia, P.P., Barua, N.C., and Rao, P.G. (2010). Artemisinin and its derivatives: a novel class of anti-malarial and anti-cancer agents. Chem Soc Rev *39*, 435-454.

Chen, J., Darst, S.A., and Thirumalai, D. (2010). Promoter melting triggered by bacterial RNA polymerase occurs in three steps. Proceedings of the National Academy of Sciences USA, Epub ahead of print.

Churchman, L.S., and Weissman, J.S. (2011). Nascent transcript sequencing visualizes transcription at nucleotide resolution. Nature *469*, 368-+.

Corrigan, R.M., and Foster, T.J. (2009). An improved tetracycline-inducible expression vector for Staphylococcus aureus. Plasmid *61*, 126-129.

Cox, F.E.G. (2010). History of the discovery of the malaria parasites and their vectors. Parasite Vector *3*.

Craig, M.L., Tsodikov, O.V., McQuade, K.L., Schlax, P.E.J., Capp, M.W., Saecker, R.M., and Record, M.T.J. (1998). DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase-promoter open complex: Evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA. Journal of Molecular Biology *283*, 741-756.

Dahl, E.L., and Rosenthal, P.J. (2007). Multiple antibiotics exert delayed effects against the Plasmodium falciparum anicoplast. Antimicrobial Agents and Chemotherapy *51*, 3485-3490.

Darst, S.A. (2001). Bacterial RNA polymerase. Current Opinion in Structural Biology *11*, 155-162.

Darst, S.A., Kubalek, E.W., and Kornberg, R.D. (1989). Three-dimensional structure of Escherichia coli RNA polymerase holoenzyme determined by electron crystallography. Nature *340*, 730-732.

Datta, I., Sau, S., Sil, A.K., and Mandal, N.C. (2005). The bacteriophage lambda DNA replication protein P inhibits the oriC DNA- and ATP-binding functions of the DNA replication initiator protein DnaA of Escherichia coli. J Biochem Mol Biol *38*, 97-103.

deHaseth, P.L., Zupancic, M.L., and Record, M.T.J. (1998). RNA polymerase-promoter interactions: The comings and going of RNA polymerase. Journal of Bacteriology *180*, 3019-3025.

Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Kwan, T., Liu, J., McCarty, J., Williams, D., DuBow, M.*, et al.* (2008). Inhibition of transcription in Staphylococcus aureus by a primary sigma factor-binding polypeptide from phage G1. Journal of Bacteriology *191*, 3763-3771.

Deighan, P., Diez, C.M., Leibman, M., Hochschild, A., and Nickels, B.E. (2008). The bacteriophage lambda Q antiterminator protein contacts the eta-flap domain of RNA polymerase. Proceedings of the National Academy of Sciences USA *105*, 15305-15310.

Deitsch, K.W., Driskill, C.L., and Wellems, T.E. (2001). Transformation of malaria parasites by the spontaneous uptake and expression of DNA from human erythrocytes. Nucleic Acids Research *29*, 850-853.

Deora, R., and Misra, T.K. (1996). Characterization of the primary sigma factor of Staphylococcus aureus. Journal of Biological Chemistry *271*, 21828-21834.

Deutscher, M.P. (2003). Degradation of stable RNA in bacteria. Journal of Biological Chemistry *278*, 45041-45044.

Dondorp, A.M., Nosten, F., Yi, P., Das, D., Phyo, A.P., Tarning, J., Lwin, K.M., Ariey, F., Hanpithakpong, W., Lee, S.J.*, et al.* (2009). Artemisinin Resistance in Plasmodium falciparum Malaria. New Engl J Med *361*, 455-467.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. Nature *391*, 251-258.

Echols, H. (1972). Developmental pathways for the temperate phage: lysis vs lysogeny. Annual Reviews of Genetics *6*, 157-190.

Epp, C., Raskolnikov, D., and Deitsch, K.W. (2008). A regulatable transgene expression system for cultured Plasmodium falciparum parasites. Malaria J *7*.

Estrem, S.T., Gaal, T., Ross, W., and Gourse, R.L. (1998). Identification of an UP element consensus sequence for bacterial promoters. Proceedings of the National Academy of Sciences USA *95*, 9761-9766.

Estrem, S.T., Ross, W., Gaal, T., Chen, Z.W.S., Niu, W., Ebright, R.H., and Gourse, R.L. (1999). Bacterial promoter architecture: Subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes and Development *13*, 2134-2147.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Brechemmier-Baey, D., and Putzer, H. (2005). Ribonucleases J1 and J2: two novel endoribonucleases in B.subtilis with functional homology to E.coli RNase E. Nucleic Acids Research *33*, 2141-2152.

Ezekiel, D.H., and Hutchins, J.E. (1968). Mutations affecting RNAP associated with rifampicin resistance in *Escherichia coli*. Nature London *220*, 276-277.

Feklistov, A., Barinova, N., Sevostyanova, A., Heyduk, E., Bass, I., Vvedenskaya, I., Kuznedelov, K., Merkiene, E., Stavrovskaya, E., Klimasauskas, S.*, et al.* (2006). A basal promoter element recognized by free RNA polymerase sigma subunit determines promoter recognition by RNA polymerase holoenzyme. Mol Cell *23*, 97-107.

Feklistov, A., and Darst, S.A. (2009). Promoter recognition by bacterial alternative sigma factors: the price of high selectivity? Genes and Development *23*, 2371-2375.

Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. Cell *147*, 1257-1269.

Feklistov, A., Mekler, V., Jiang, Q., Westblade, L.F., Irschik, H., Jansen, R., Mustaev, A., Darst, S.A., and Ebright, R.H. (2008). Rifamycins do not function by allosteric modulation of binding of Mg2+ to the RNA polymerase active center. Proc Natl Acad Sci U S A *105*, 14820-14825.

Felden, B., Vandenesch, F., Bouloc, P., and Romby, P. (2011). The Staphylococcus aureus RNome and its commitment to virulence. PLoS Pathogens *7*, e1002006.

Finn, R.D., Orlova, E.V., Gowen, B., Buck, M., and van Heel, M. (2000). Escherichia coli RNA polymerase core and holoenzyme structures. EMBO Journal *19*, 6833-6844.

Fischetti, V.A. (2008). Bacteriophage lysins as effective antibacterials. Current Opinion in Microbiology *11*, 393-400.

Fischetti, V.A. (2010). Bacteriophage endolysins: a novel anti-infective to control Gram-positive pathogens. International Journal of Medical Microbiology *300*, 357-362.

Fox, K.R. (1997). DNase I footprinting. Methods Mol Biol *90*, 1-22.

Friesen, J.D., Parker, J., Watson, R.J., Fill, N.P., Pedersen, S., and Pedersen, F.S. (1976). Isolation of a lambda transducing bacteriophage carrying the relA gene of Escherichia coli. Journal of Bacteriology *127*, 917-922.

Gaal, T., Barkei, J., Dickson, R.R., deBoer, H.A., deHaseth, P.L., Alavi, H., and Gourse, R.L. (1989). Saturation mutagenesis of an Escherichia coli rRNA promoter and initial characterization of promoter variants. Journal of Bacteriology *171*, 4852-4861.

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

Gaal, T., Ross, W., Estrem, S.T., Nguyen, L.H., Burgess, R.R., and Gourse, R.L. (2001). Promoter recognition and discrimination by EsigmaS RNAP. Molecular Microbiology *42*, 939-954.

Gardella, T., Moyle, T., and Susskind, M.M. (1989). A mutant *Escherichia coli* sigma 70 subunit of RNA polymerase with altered promoter specificity. Journal of Molecular Biology *206*, 579-590.

Gardner, M.J., Williamson, D.H., and Wilson, R.J.M. (1991). A Circular DNA in Malaria Parasites Encodes an Rna-Polymerase Like That of Prokaryotes and Chloroplasts. Mol Biochem Parasit *44*, 115-123.

Geiduschek, E.P. (1991). Regulation of expression of the late genes of bacteriophage T4. Annual Review of Genetics *25*, 437-460.

George, E.A., and Muir, T.W. (2007). Molecular mechanisms of agr quorum sensing in virulent staphylococci. Chembiochem *8*, 847-855.

Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: An RNA polymerase II elongation complex at 3.3 Å resolution. Science *292*, 1876-1882.

Gourse, R.L., Gaal, T., Aiyar, S.E., Barker, M.M., Estrem, S.T., Hirvonen, C.A., and Ross, W. (1998). Strength and regulation without transcription factors: lessons from bacterial rRNA promoters. Cold Spring Harb Symp Quant Biol *63*, 131-139.

Gourse, R.L., Gaal, T., Bartlett, M.S., Appleman, J.A., and Ross, W. (1996). rRNA transcription and growth rate-dependent regulation of ribosome synthesis in Escherichia coli. Annual Reviews of Microbiology *50*, 645-677.

Grandgirard, D., Loeffler, J.M., Fischetti, V.A., and Leib, S.L. (2008). Phage lytic enzyme Cpl-1 for antibacterial therapy in experimental pneumococcal meningitis. Journal of Infectious Diseases *197*, 1519-1522.

Gribskov, M., and Burgess, R.R. (1986). Sigma factors from *E. coli*, *B. subtilis*, phase SPO1, and phage T4 are homologous proteins. Nucleic Acids Research *14*, 6745-6763.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Research *35*, W52-W57.

Gruber, T.M., and Gross, C.A. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. Annual Review of Microbiology *57*, 441-466.

Harley, C.B., and Reynolds, R.P. (1987). Analysis of *E. coli* promoter sequences. Nucleic Acids Research *15*, 2343-2361.

Hartmann, G., Honikel, K.O., Knusel, F., and Nuesch, J. (1967). The specific inhibition of the DNA-directed RNA synthesis by rifamycin. Biochimica Biophysica Acta *145*, 843-844.

Haugen, S.P., Ross, W., Michele, M., and Gourse, R.L. (2008). Fine structure of the promoter-sigma region 1.2 interaction. Proceedings of the National Academy of Sciences USA *105*, 3292-3297.

Heep, M., Rieger, U., Beck, D., and Lehn, N. (2000). Mutations in the beginning of the rpoB gene can induce resistance to rifamycins in both Helicobacter pylori and Mycobacterium tuberculosis. Antimicrobial Agents and Chemotherapeutics *44*, 1075-1077.

Helmann, J.D. (1999). Anti-sigma factors. Current Opinion in Microbiology *2*, 135-141.

Hinkle, D.C., and Chamberlin, M.J. (1972a). Studies of the binding of Escherichia coli RNA polymerase to DNA. I. The role of sigma subunit in site selection. J Mol Biol *70*, 157-185.

Hinkle, D.C., and Chamberlin, M.J. (1972b). Studies of the binding of Escherichia coli RNA polymerase to DNA. II. The kinetics of the binding reaction. J Mol Biol *70*, 187-195.

Hinkle, D.C., Mangel, W.F., and Chamberlin, M.J. (1972a). Studies of the binding of Escherichia coli RNA polymerase to DNA. IV. The effect of rifampicin on binding and on RNA chain initiation. J Mol Biol *70*, 209-220.

Hinkle, D.C., Mangel, W.F., and Chamberlin, M.J. (1972b). Studies of the binding of Escherichia coli RNA polymerase to DNA. IV. The effect of rifampicin on binding and on RNA chain initiation. Journal of Molecular Biology *70*, 209-220.

Hinton, D.M., Pande, S., Wais, N., Johnson, X.B., Vuthoori, M., Makela, A., and Hook-Barnard, I. (2005). Transcriptional takeover by sigma appropriation: remodeling of the sigma70 subunit of Escherichia coli RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. Microbiology *151*, 1729-1740.

Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J., Moore, C.E., Hurst, L., Atkin, R.*, et al.* (2004). Complete genomes of two clinical Staphylococcus aureus strains: evidence for the rapid evolution of virulence and drug resistance. Proc Natl Acad Sci U S A *101*, 9786-9791.

Howden, B.P., Davies, J.K., Johnson, P.D., Stinear, T.P., and Grayson, M.L. (2010). Reduced vancomycin susceptibility in Staphylococcus aureus, including vancomycin-intermediate and heterogeneous vancomycin-intermediate strains: resistance mechanisms, laboratory detection, and clinical implications. Clinical Microbiology Reviews *23*, 99-139.

Hubscher, J., Jansen, A., Kotte, O., Schafer, J., Majcherczyk, P.A., Harris, L.G., Bierbaum, G., Heinemann, M., and Berger-Bachi, B. (2007). Living with an imperfect cell wall: compensation of femAB inactivation in Staphylococcus aureus. BMC Genomics *8*, 307.

Iandolo, J.J., Worrell, V., Groicher, K.H., Qian, Y., Tian, R., Kenton, S., Dorman, A., Ji, H., Lin, S., Loh, P.*, et al.* (2002). Comparative analysis of the genomes of the temperate bacteriophages phi 11, phi 12 and phi 13 of Staphylococcus aureus 8325. Gene *289*, 109-118.

Jain, D., Kim, Y., Maxwell, K.L., Beasley, S., Zhang, R., Gussin, G.N., Edwards, A.M., and Darst, S.A. (2005). Crystal structure of bacteriophage λ cII and its DNA complex. Molecular Cell *19*, 259-269.

Jishage, M., and Ishihama, A. (1999). Transcriptional organization and in vivo role of the Escherichia coli rsd gene, encoding the regulator of RNA polymerase sigma D. Journal of Bacteriology *181*, 3768-3776.

Jishage, M., Kvint, K., Shingler, V., and Nystrom, T. (2002). Regulation of sigma factor competition by the alarmone ppGpp. Gene Dev *16*, 1260-1270.

Kadesch, T.R., Rosenberg, S., and Chamberlin, M.J. (1982). Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of binding at bacteriophage T7 promoter A1 using a template competition assay. Journal of Molecular Biology *155*, 1-29.

Kang, J.-G., Paget, M.S., Seok, Y.-J., Hahn, M.-Y., Bae, J.-B., Hahn, J.-S., Kleanthous, C., Buttner, M.J., and Roe, J.-H. (1999). RsrA, and anti-sigma factor regulated by redox change. EMBO Journal *18*, 4292-4298.

Kar, S., and Kar, S. (2010). Control of malaria. Nat Rev Drug Discov *9*, 511-512.

Kassavetis, G.A., and Geiduschek, E.P. (1984). Defining a bacteriophage T4 late promoter: bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. Proceedings of the National Academy of Sciences USA *81*, 5101-5105.

Kelch, B.A., and Agard, D.A. (2007). Mesophile versus thermophile: Insights into the structural mechanisms of kinetic stability. Journal of Molecular Biology *370*, 784-795.

Kelly, D., McAuliffe, O., Ross, R.P., O'Mahony, J., and Coffey, A. (2011). Development of a broad-host-range phage cocktail for biocontrol. Bioengineered Bugs *2*, 31-37.

Kettenberger, H., Armache, K.-J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. Molecular Cell *16*, 955-965.

Klein, E., Smith, D.L., and Laxminarayan, R. (2007a). Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. Emerg Infect Dis *13*, 1840-1846.

Klein, E., Smith, D.L., and Laxminarayan, R. (2007b). Hospitalizations and deaths caused by methicillin-resistant Staphylococcus aureus, United States, 1999-2005. Emerging Infectious Diseases *13*, 1840-1846.

Koller, T., Nelson, D., Nakata, M., Kreutzer, M., Fischetti, V.A., Glocker, M.O., Podbielski, A., and Kreikemeyer, B. (2008). PlyC, a novel bacteriophage lysin for compartment-dependent proteomics of group A streptococci. Proteomics *8*, 140-148.

Krasny, L., and Gourse, R.L. (2004). An alternative strategy for bacterial ribosome synthesis: Bacillus subtilis rRNA transcription regulation. EMBO Journal *23*, 4473-4483.

Kumar, A., Malloch, R.A., Fujita, N., Smillie, D.A., Ishihama, A., and Hayward, R.S. (1993). The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. Journal of Molecular Biology *232*, 406-418.

Kwan, T., Liu, J., DuBow, M., Gros, P., and Pelletier, J. (2005). The complete genomes and proteomes of 27 Staphylococcus aureus bacteriophages. Proceedings of the National Academy of Sciences USA *102*, 5174-5179.

Lambert, L.J., Wei, Y., Schirf, V., Demeler, B., and Werner, M.H. (2004). T4 AsiA blocks DNA recognition by remodeling sigma70 region 4. EMBO Journal *23*, 2952-2292.

Laptenko, O., Lee, J., Lomakin, I., and Borukhov, S. (2003). Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. EMBO Journal *23*, 6322-6334.

Laurie, A.D., Bernardo, L.M., Sze, C.C., Skarfstad, E., Szalewska-Palasz, A., Nystrom, T., and Shingler, V. (2003). The role of the alarmone (p)ppGpp in sigma N competition for core RNA polymerase. Journal of Biological Chemistry *278*, 1494-1503.

Lemke, J.J., Sanchez-Vazquez, P., Burgos, H.L., Hedberg, G., Ross, W., and Gourse, R.L. (2011). Direct regulation of Escherichia coli ribosomal protein promoters by the transcription factors ppGpp and DksA. Proc Natl Acad Sci U S A *108*, 5712-5717.

Letalaer, J.Y., and Jeanteur, P. (1971). Purification and Base Composition Analysis of Phage Lambda Early Promoters. P Natl Acad Sci USA *68*, 3211-&.

Liu, J., Dehbi, M., Moeck, G., Arhin, F., Bauda, P., Bergeron, D., Callejo, M., Ferretti, V., Ha, N., Kwan, T.*, et al.* (2004). Antimicrobial drug discovery through bacteriophage genomics. Nature Biotechnology *22*, 185-191.

Lowy, F.D. (1998). Staphylococcus aureus infections. New England Journal of Medecine *339*, 520-532.

Maeda, H., Fujita, N., and Ishihama, A. (2000). Competition among seven Escherichia coli sigma subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Research *28*, 3497-3503.

Marraffini, L.A., and Sontheimer, E.J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet *11*, 181-190.

McConkey, G.A., Rogers, M.J., and McCutchan, T.F. (1997). Inhibition of Plasmodium falciparum protein synthesis - Targeting the plastid-like organelle with thiostrepton. Journal of Biological Chemistry *272*, 2046-2049.

McLeod, S.M., Aiyar, S.E., Gourse, R.L., and Johnson, R.C. (2002). The C-terminal domains of the RNA polymerase α subunits: Contact site with Fis and localization during co-activation with CRP at the *Escherichia coli proP* P2 promoter. Journal of Molecular Biology *316*, 517-529.

Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B.*, et al.* (2009). Genome-Wide Identification of Transcription Start Sites, Promoters and Transcription Factor Binding Sites in E. coli. PLoS One *4*.

Miller, L.H., and Su, X.Z. (2011). Artemisinin: Discovery from the Chinese Herbal Garden. Cell *146*, 855-858.

Mills, A.E. (1956). Staphylococcus bacteriophage lysate aerosol therapy of sinusitis. Laryngoscope *66*, 846-858.

Minakhin, L., and Severinov, K. (2005). Transcription regulation by bacteriophage T4 AsiA. Protein Expression and Purification *41*, 1-8.

Munson, L.M., and Reznikoff, W.S. (1981). Abortive initiation and long ribonucleic acid synthesis. Biochemistry *20*, 2081-2085.

Murakami, K., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Current Opinion in Structural Biology *13*, 31-39.

Murakami, K., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: An RNA polymerase holoenzyme/DNA complex. Science *296*, 1285-1290.

Murakami, K., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. Science *296*, 1280-1284.

Murray, C.J.L., Rosenfeld, L.C., Lim, S.S., Andrews, K.G., Foreman, K.J., Haring, D., Fullman, N., Naghavi, M., Lozano, R., and Lopez, A.D. (2012). Global malaria mortality between 1980 and 2010: a systematic analysis. Lancet *379*, 413-431.

Mustaev, A., and Korzheva, N. (2001). Transcription elongation complex: Structure and function. Current Opinion in Microbiology *4*, 119-125.

Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., Francois, P., and Cheung, A.L. (2011). Whole-genome sequencing of Staphylococcus aureus strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. Journal of Bacteriology *193*, 2332-2335.

Nair, S.C., and Striepen, B. (2011). What Do Human Parasites Do with a Chloroplast Anyway? Plos Biol *9*.

Nechaev, S., and Severinov, K. (1999). Inhibition of Escherichia coli RNA polymerase by bacteriophage T7 gene 2 protein. J Mol Biol *289*, 815-826.

Newlands, J.T., Ross, W., Gosink, K.K., and Gourse, R.L. (1991). Factor-independent activation of *Escherichia coli* rRNA transcription. II. Characterization of complexes of rrnB P1 promoters containing or lacking the upstream activator region with *Escherichia coli* RNA polymerase. Journal of Molecular Biology *220*, 569-583.

Newman, J.A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C.R., and Lewis, R.J. (2011). Unusual, Dual Endo- and Exonuclease Activity in the Degradosome Explained by Crystal Structure Analysis of RNase J1. Structure *19*, 1241-1251.

Nordmann, P., Naas, T., Fortineau, N., and Poirel, L. (2007). Superbugs in the coming new decade; multidrug resistance and prospects for treatment of Staphylococcus aureus, Enterococcus spp. and Pseudomonas aeruginosa in 2010. Current Opinion in Microbiology *10*, 436-440.

Nudler, E. (1999). Transcription elongation: structural basis and mechanisms. Journal of Molecular Biology *288*, 1-12.

Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: Protein-DNA interactions holding together the elongation complex. Science *273*, 211-217.

O'Neill, A.J. (2010). Staphylococcus aureus SH1000 and 8325-4: comparative genome sequences of key laboratory strains in staphylococcal research. Lett Appl Microbiol *51*, 358-361.

O'Neill, A.J., Lindsay, J.A., Gould, K., Hinds, J., and Chopra, I. (2009). Transcriptional signature following inhibition of early-stage cell wall biosynthesis in Staphylococcus aureus. Antimicrob Agents Chemother *53*, 1701-1704.

Ohta, T., Hirakawa, H., Morikawa, K., Maruyama, A., Inose, Y., Yamashita, A., Oshima, K., Kuroda, M., Hattori, M., Hiramatsu, K.*, et al.* (2004). Nucleotide substitutions in Staphylococcus aureus strains, Mu50, Mu3, and N315. DNA Research *11*, 51-56.

Opalka, N., Brown, J., Lane, W.J., Twist, K.-A.F., Landick, R., Asturias, F.J., and Darst, S.A. (2010). Complete structural model of Escherichia coli RNA polymerase from a hybrid approach. Plos Biol, in press.

Oppenheim, A.B., Kobiler, O., Stavans, J., Court, D.L., and Adhya, S. (2005). Switches in bacteriophage lambda development. Annual Reviews of Genetics *39*, 409-429.

Orsini, G., Igonet, S., Pene, C., Sclavi, B., Buckle, M., Uzan, M., and Kolb, A. (2004). Phage T4 early promoters are resistant to inhibition by the anti-sigma factor AsiA. Molecular Microbiology *52*, 1013-1028.

Orsini, G., Ouhammouch, M., Le Caer, J.-P., and Brody, E.N. (1993). The *asiA* gene of bacteriophage T4 codes for the anti-σ^{70} protein. Journal of Bacteriology *175*, 85-93.

Ouhammouch, M., Adelman, K., Harvey, S.R., Orsini, G., and Brody, E.N. (1995). Bacteriophage T4 MotA and AsiA proteins suffice to direct *Escherichia coli* RNA polymerase to initiate transcription at T4 middle promoters. Porceedings of the National Academy of Sciences USA *92*, 1451-1455.

Paget, M.S., and Helmann, J.D. (2003). The sigma 70 family of sigma factors. Genome Biology *4*, 203.

Pastagia, M., Euler, C., Chahales, P., Fuentes-Duculan, J., Krueger, J.G., and Fischetti, V.A. (2011). A novel chimeric lysin shows superiority to mupirocin for skin decolonization of methicillin-resistant and -sensitive Staphylococcus aureus strains. Antimicrobial Agents and Chemotherapy *55*, 738-744.

Patikoglou, G.A., Westblade, L.F., Campbell, E.A., Lamour, V., Lane, W.J., and Darst, S.A. (2007). Crystal structure of the Escherichia coli regulator of sigma70, Rsd, in complex with sigma70 domain 4. Journal of Molecular Biology *372*, 649-659.

Paul, B.J., Barker, M., Ross, W., Schneider, D., Webb, C., Foster, J., and Gourse, R. (2004). DksA: A critical component of the transcription intiiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. Cell *118*, 311-322.

Paul, V.D., Rajagopalan, S.S., Sundarrajan, S., George, S.E., Asrani, J.Y., Pillai, R., Chikkamadaiah, R., Durgaiah, M., Sriram, B., and Padmanabhan, S. (2011). A novel bacteriophage Tail-Associated Muralytic Enzyme (TAME) from Phage K and its development into a potent antistaphylococcal protein. BMC Microbiol *11*, 226.

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

Pene, C., and Uzan, M. (2000). The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. Molecular Microbiology *35*, 1180-1191.

Perederina, A., Svetlov, V., Vassylyeva, M., Tahirov, T., Yokoyama, S., Artsimovitch, I., and Vassylyev, D. (2004). Regulation through the secondary channel - structural framework for ppGpp-DksA synergism during transcription. Cell *118*, 297-309.

Polyakov, A., Severinova, E., and Darst, S.A. (1995). Three-dimensional structure of *Escherichia coli* core RNA polymerase: Promoter binding and elongation conformations of the enzyme. Cell *83*, 365-373.

Pomerantz, R.T., and O'donnell, M. (2008). The replisome uses mRNA as a primer after colliding with RNA polymerase. Nature *456*, 762-U751.

Pomerantz, R.T., and O'donnell, M. (2010). Direct Restart of a Replication Fork Stalled by a Head-On RNA Polymerase. Science *327*, 590-592.

Ptashne, M. (1992). A Genetic Switch, Phage Lambda and Higher Organisms (Cambridge, Cell Press).

Pukrittayakamee, S., Viravan, C., Charoenlarp, P., Yeamput, C., Wilson, R.J.M., and White, N.J. (1994). Antimalarial Effects of Rifampin in Plasmodium-Vivax Malaria. Antimicrobial Agents and Chemotherapy *38*, 511-514.

Quail, M.A., Otto, T.D., Gu, Y., Harris, S.R., Skelly, T.F., McQuillan, J.A., Swerdlow, H.P., and Oyola, S.O. (2012). Optimal enzymes for amplifying sequencing libraries. Nat Methods *9*, 10-11.

Raivio, T.L., and Silhavy, T.J. (2001). Periplasmic stress and ECF sigma factors. Annual Review of Microbiology *55*, 591-624.

Ramaswamy, S., and Musser, J.M. (1998). Molecular genetic basis of antimicrobial agent resistnace in *Mycobacterium tuberculosis*: 1998 update. Tubercle and Lung Disease *79*, 3-29.

Rao, L., Karls, R.K., and Betley, M.J. (1995). In vitro transcription of pathogenesis-related genes by purified RNA polymerase from Staphylococcus aureus. Journal of Bacteriology *177*, 2609-2614.

Rao, L., Ross, W., Appleman, J.A., Gaal, T., Leirmo, S., Schlax, P.J., and Gourse, R.L. (1994). Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. Journal of Molecular Biology *235*, 1421-1435.

Ray, D.S., Dueber, J., and Suggs, S. (1975). Replication of bacteriophage M13 IX. Requirement of the Escherichia coli dnaG function for M13 duplex DNA replication. Journal of virology *16*, 348-355.

Reiss, S., Pane-Farre, J., Fuchs, S., Francois, P., Liebeke, M., Schrenzel, J., Lindequist, U., Lalk, M., Wolz, C., Hecker, M.*, et al.* (2012). Global Analysis of the Staphylococcus aureus Response to Mupirocin. Antimicrobial Agents and Chemotherapy *56*, 787-804.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. Science *314*, 1139-1143.

Reynolds, J., and Wigneshweraraj, S. (2011). Molecular Insights into the Control of Transcription Initiation at the Staphylococcus aureus agr operon. Journal of Molecular Biology *412*, 862-881.

Roberts, J.W. (1969). Promoter Mutation in Vitro. Nature *223*, 480-&.

Roos, D.S., Crawford, M.J., Donald, R.G.K., Kissinger, J.C., Klimczak, L.J., and Striepen, B. (1999). Origin, targeting, and function of the apicomplexan plastid. Current Opinion in Microbiology *2*, 426-432.

Rosenberg, S., Kadesch, T.R., and Chamberlin, M.J. (1982). Binding of *Escherichia coli* RNA polymerase holoenzyme to bacteriophage T7 DNA. Measurements of the rate of open complex formation at T7 promoter A. Journal of Molecular Biology *155*, 31-51.

Ross, W., Gosink, K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. Science *262*, 1407-1413.

Rossolini, G.M., Mantengoli, E., Montagnani, F., and Pollini, S. (2010). Epidemiology and clinical relevance of microbial resistance determinants versus anti-Gram-positive agents. Current Opinion in Microbiology *13*, 582-588.

Ruiz de Gopegui, E., Juan, C., Zamorano, L., Perez, J.L., and Oliver, A. (2012). Transferable multidrug resistance plasmid carrying cfr associated with tet(L), ant(4')-Ia and dfrK genes from a clinical methicillin-resistant Staphylococcus aureus ST125 strain. Antimicrobial Agents and Chemotherapy.

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.

Saecker, R.M., Tsodikov, O.V., McQuade, K.L., Schlax, P.E.J., Capp, M.W., and Record, M.T.J. (2002). Kinetic studies and structural models of the association of E. coli sigma(70) RNA polymerase with the lambda P(R) promoter: large scale conformational changes in forming the kinetically significant intermediates. Journal of Molecular Biology *319*, 649-671.

Sanderson, A., Mitchell, J.E., Minchin, S.D., and Busby, S.J.W. (2003). Substitutions in the Escherichia coli RNA polymerase sigma70 factor that affect recognition of extended -10 elements at promoters. FEBS Letters *544*, 199-205.

Saxena, V., Garg, S., Kochar, D.K., and Das, A. (2010). The Apicoplast Genome of Plasmodium Vivax. Am J Trop Med Hyg *83*, 220-220.

Schenk, S., and Laddaga, R.A. (1992). Improved method for electroporation of Staphylococcus aureus. FEMS Microbiol Lett *73*, 133-138.

Sen, R., Nagai, H., Hernandez, V.J., and Shimamoto, N. (1998). Reduction in abortive transcription from the lambdaPR promoter by mutations in region 3 of the sigma70 subunit of Escherichia coli RNA polymerase. Journal of Biological Chemistry *273*, 9872-9877.

Severinov, K., and Darst, S.A. (1997). A mutant RNA polymerase that forms unusual open promoter complexes. Proceedings of the National Academy of Sciences USA *94*, 13481-13486.

Severinov, K., and Goldfarb, A. (1994). Topology of the product binding site in RNA polymerase revealed by transcript slippage at the phage λ P_L promoter. Journal of Biological Chemistry *269*, 31701-31705.

Severinova, E., Severinov, K., and Darst, S.A. (1998). Inhibition of Escherichia coli RNA polymerase by bacteriophage T4 AsiA. Journal of Molecular Biology *279*, 9-18.

Siebenlist, U. (1979). RNA polymerase unwinds an 11-base pair segment of a phage T7 promoter. Nature *279*, 651-652.

Siebenlist, U., and Gilbert, W. (1980). Contacts between Escherichia coli RNA polymerase and an early promoter of phage T7. Proceedings of the National Academy of Sciences USA *77*, 122-126.

Siegele, D.A., Hu, J.C., and Gross, C.A. (1988). Mutations in *rpoD*, the gene encoding the sigma 70 subunit of *Escherichia coli* RNA polymerase, that increase expression of the lac operon in the absence of CAP-cAMP. Journal of Molecular Biology *203*, 29-37.

Smith, J. (1924). The Bacteriophage in the Treatment of Typhoid Fever. British Medical Journal *2*, 47-49.

Snyder, L., Gold, L., and Kutter, E. (1976). A gene of bacteriophage T4 whose product prevents true late transcription on cytosine-containing T4 DNA. Proceedings of the National Academy of Sciences USA *73*, 3098-3102.

Stallings, C.L., and Glickman, M.S. (2011). CarD: a new RNA polymerase modulator in mycobacteria. Transcription *2*, 15-18.

Stallings, C.L., Stphanou, N.C., Chu, L., Hochschild, A., Nickels, B.E., and Glickman, M.S. (2009). CarD is an essential regulator of rRNA transcription required for Mycobacterium tuberculosis persistence. Cell *138*, 149-159.

Stebbins, C.E., Borukhov, S., Orlova, M., Polyakov, A., Goldfarb, A., and Darst, S.A. (1995). Crystal structure of the GreA transcript cleavage factor from *Escherichia coli*. Nature *373*, 636-640.

Stevens, A. (1977). Inhibition of DNA-enzyme binding by an RNA polymerase inhibitor from T4 phage-infected *Escherichia coli*. Biochimica Biophysica Acta *475*, 193-196.

Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from Mycobacterium tuberculosis. Proc Natl Acad Sci U S A *103*, 8060-8065.

Sun, L., Dove, S.L., Panaghie, G., deHaseth, P.L., and Hochschild, A. (2004). An RNA polymerase mutant deficient in DNA melting facilitates study of activation mechanism: Application to an artificial activator of transcription. Journal of Molecular Biology *343*, 1171-1182.

Takebe, S., Witola, W.H., Schimanski, B., Gunzl, A., and Ben Mamoun, C. (2007). Purification of components of the translation elongation factor complex of Plasmodium falciparum by tandem affinity purification. Eukaryot Cell *6*, 584-591.

Tangpukdee, N., Duangdee, C., Wilairatana, P., and Krudsood, S. (2009). Malaria Diagnosis: A Brief Review. Korean J Parasitol *47*, 93-102.

Tehranchi, A.K., Blankschien, M.D., Zhang, Y., Halliday, J.A., Srivatsan, A., Peng, J., Herman, C., and Wang, J.D. (2010). The Transcription Factor DksA Prevents Conflicts between DNA Replication and Transcription Machinery. Cell *141*, 595-605.

Testerman, T.L., Vazquez-Torres, A., Xu, Y., Jones-Carson, J., Libby, S.J., and Fang, F.C. (2002). The alternative sigma factor sigmaE controls antioxidant defences required for Salmonella virulence and stationary-phase survival. Molecular Microbiology *43*, 771-782.

Tiemann, B., Depping, R., Gineikiene, E., Kaliniene, L., Nivinskas, R., and Ruger, W. (2004). ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. J Bacteriol *186*, 7262-7272.

Tong, L., and Rossmann, M.G. (1997). Rotation function calculations with GLRF program. Methods in Enzymology *276*, 594-611.

Tong, S.Y., Chen, L.F., and Fowler, V.G., Jr. (2012). Colonization, pathogenicity, host susceptibility, and therapeutics for Staphylococcus aureus: what is the clinical relevance? Semin Immunopathol *34*, 185-200.

Traber, K., and Novick, R. (2006). A slipped-mispairing mutation in AgrA of laboratory strains and clinical isolates results in delayed activation of agr and failure to translate delta- and alpha-haemolysins. Molecular Microbiology *59*, 1519-1530.

Trampuz, A., Jereb, M., Muzlovic, I., and Prabhu, R.M. (2003). Clinical review: Severe malaria. Crit Care *7*, 315-323.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

Travers, A.A. (1984). Conserved features of coordinately regulated E. coli promoters. Nucleic Acids Research *12*, 2605-2618.

Traxler, M.F., Summers, S.M., Nguyen, H.T., Zacharia, V.M., Hightower, G.A., Smith, J.T., and Conway, T. (2008). The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. Molecular Microbiology *68*, 1128-1148.

Truong-Bolduc, Q.C., Dunman, P.M., Eidem, T., and Hooper, D.C. (2011). Transcriptional profiling analysis of the global regulator NorG, a GntR-like protein of Staphylococcus aureus. J Bacteriol *193*, 6207-6214.

Tuchscherr, L., Medina, E., Hussain, M., Volker, W., Heitmann, V., Niemann, S., Holzinger, D., Roth, J., Proctor, R.A., Becker, K.*, et al.* (2011). Staphylococcus aureus phenotype switching: an effective bacterial strategy to escape host immune response and establish a chronic infection. EMBO Molecular Medicine *3*, 129-141.

Twist, K.A., Husnain, S.I., Franke, J.D., Jain, D., Campbell, E.A., Nickels, B.E., Thomas, M.S., Darst, S.A., and Westblade, L.F. (2011a). A novel method for the production of in vivo-assembled, recombinant Escherichia coli RNA polymerase lacking the alpha C-terminal domain. Protein Sci *20*, 986-995.

Twist, K.A.F., Campbell, E.A., Deighan, P., Nechaev, S., Jain, V., Geiduschek, E.P., Hochschild, A., and Darst, S.A. (2011b). Crystal structure of the bacteriophage T4 late-transcription coactivator gp33 with the beta-subunit flap domain of Escherichia coli RNA polymerase. P Natl Acad Sci USA *108*, 19961-19966.

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 Å resolution. Nature *417*, 712-719.

Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. Nature *448*, 157-162.

Vrentas, C.E., Gaal, T., Berkmen, M.B., Rutherford, S.T., Haugen, S.P., Vassylyev, D.G., Ross, W., and Gourse, R.L. (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. Journal of Molecular Biology *377*, 551-564.

Wacker, S.A., Houghtaling, B.R., Elemento, O., and Kapoor, T.M. (2012). Using transcriptome sequencing to identify mechanisms of drug action and resistance. Nat Chem Biol *8*, 235-237.

Wada, A., Ohta, H., Kulthanan, K., and Hiramatsu, K. (1993). Molecular cloning and mapping of 16S-23S rRNA gene complexes of Staphylococcus aureus. J Bacteriol *175*, 7483-7487.

Wade, H.E., Lovett, S., and Robinson, H.K. (1964). The autodegradation of 32-P-labelled ribosomes from Escherichia coli. Biochem J *93*, 121-128.

Wagner, P.L., and Waldor, M.K. (2002). Bacteriophage control of bacterial virulence. Infection and Immunity *70*, 3985-3993.

Waller, R.F., Reed, M.B., Cowman, A.F., and McFadden, G.I. (2000). Protein trafficking to the plastid of Plasmodium falciparum is via the secretory pathway. EMBO Journal *19*, 1794-1802.

Walter, R.F., and McFadden, G.I. (2005). The apicoplast: A review of the derived plastid of apicomplexan parasites. Curr Issues Mol Biol *7*, 57-79.

Wellems, T.E. (2002). Plasmodium chloroquine resistance and the search for a replacement antimalarial drug. Science *298*, 124-126.

Wernsdorfer, W.H. (1994). Epidemiology of Drug-Resistance in Malaria. Acta Trop *56*, 143-156.

Wernsdorfer, W.H., and Noedl, H. (2003). Molecular markers for drug resistance in malaria: use in treatment, diagnosis and epidemiology. Curr Opin Infect Dis *16*, 553-558.

Westblade, L.F., Ilag, L.L., Powell, A.K., Kolb, A., Robinson, C.V., and Busby, S.J.W. (2004). Studies of the *Escherichia coli* Rsd-σ^{70} complex. Journal of Molecular Biology *335*, 685-692.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: the unseen majority. Proc Natl Acad Sci U S A *95*, 6578-6583.

WHO (2009). Global Tuberculosis Control.

Wilson, R.J.M., Denny, P.W., Preiser, P.R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D.J., Moore, P.W.*, et al.* (1996). Complete gene map of the plastid-like DNA of the malaria parasite Plasmodium falciparum. Journal of Molecular Biology *261*, 155-172.

Witte, W., and Cuny, C. (2011). Emergence and spread of cfr-mediated multiresistance in staphylococci: an interdisciplinary challenge. Future Microbiology *6*, 925-931.

Wongsrichanalai, C., Barcus, M.J., Muth, S., Sutamihardja, A., and Wernsdorfer, W.H. (2007). A review of malaria diagnostic tools: Microscopy and rapid diagnostic test (RDT). Am J Trop Med Hyg *77*, 119-127.

Xu, H.H., Trawick, J.D., Haselbeck, R.J., Forsyth, R.A., Yamamoto, R.T., Archer, R., Patterson, J., Allen, M., Froelich, J.M., Taylor, I.*, et al.* (2010). Staphylococcus aureus TargetArray: comprehensive differential essential gene expression as a mechanistic tool to profile antibacterials. Antimicrob Agents Chemother *54*, 3659-3670.

Yano, S.T., and Rothman-Denes, L.B. (2011). A phage-encoded inhibitor of Escherichia coli DNA replication targets the DNA polymerase clamp loader. Molecular Microbiology *79*, 1325-1338.

Yeeles, J., and Marians, K. (2011). The Escherichia coli Replisome Is Inherently DNA Damage Tolerant. Science *334*, 235-238.

Yeh, E., and DeRisi, J.L. (2011). Chemical Rescue of Malaria Parasites Lacking an Apicoplast Defines Organelle Function in Blood-Stage Plasmodium falciparum. Plos Biol *9*.

Young, B.A., Gruber, T.M., and Gross, C.A. (2002). Views of transcription initiation. Cell *109*, 417-420.

Yuan, A.H., Nickels, B.E., and Hochschild, A. (2009). The bacteriophage T4 AsiA protein contacts the beta-flap domain of RNA polymerase. P Natl Acad Sci USA *106*, 6597-6602.

Zenkin, N., Kulbachnskiy, A., Bass, I., and Nikiforov, V. (2005). Different rifampin sensitivities of Escherichia coli and Mycobacterium tuberculosis RNA polymerases are not explained by the difference in the beta-subunit rifampin regions I and II. Antimicrobial Agents and Chemotherapy *49*, 1587-1590.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. Cell *98*, 811-824.

Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., and Schneider, G. (2001). Deciphering apicoplast targeting signals - feature extraction from nuclear-encoded precursors of Plasmodium falciparum apicoplast proteins. Gene *280*, 19-26.