CHARACTERIZATION OF NOVEL DNA ELEMENTS NECESSARY FOR σ-DEPENDENT PROMOTER PROXIMAL TRANSCRIPTIONAL PAUSING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jeremy Gilbert Bird

May 2013

CHARACTERIZATION OF NOVEL DNA ELEMENTS NECESSARY FOR σ-DEPENDENT PROMOTER PROXIMAL TRANSCRIPTIONAL PAUSING

Jeremy Gilbert Bird, Ph. D.

Cornell University 2013

Transcription is not necessarily uniform in rate, but instead consists of multiple periods of continuous elongation by RNA polymerase (RNAP) interrupted by occasional pausing events. These pauses have several determined origins, including the interaction of protein co-factors, the influence of transcribed RNA secondary structure, and a direct effect of the sequence composition of the DNA being transcribed. Promoter proximal pauses, in which RNAP pauses tens of nucleotides downstream of the transcription start site, have been observed in systems ranging from bacteria to more complex eukaryotes such as *D. melanogaster* and humans, and play important regulatory roles in organisms of all levels. In *E. coli*, σ70-dependent promoter proximal pauses are of particular interest because lambdoid phage late gene promoters require these pauses for the loading of the anti-terminator Q protein, which is necessary for transcription of the phage late genes during lytic growth.

The primary interaction that induces σ70-dependent pauses is between the non-template strand of the -10-like sequence and σ70 region 2. However, recent evidence identifies other important sequences including several nucleotides around the pause site itself that are required for pausing (Perdue and Roberts 2010). Mutational studies of lambdoid phage 82 promoter *pR'* show that a 3-bp GC-rich sequence

primarily promotes pausing through the template DNA strand, but also acts through a non-template DNA strand interaction with σ70. I show that this template strand sequence determines where the pause occurs. It is likely that in σ-dependent promoter-proximal pausing, the σ70/DNA interaction anchors the elongation complex upstream of the pause site, requiring it to "scrunch" (by analogy to scrunching during initial transcription (Kapanidis et al. 2006)) and to create an enlarged DNA bubble as the active center reaches the pause site. "Scrunched" complexes are energetically strained, but stable at the pause site; my data suggest that the G/C-rich template sequence is a critical element in stabilizing the paused, scrunched structure.

Mutational studies of the A/T-rich region demonstrate that the terminal nucleotide of the pause site plays an important role in pause formation, and that the base composition of the A/T-rich sequence determines the likelihood that a transcription complex escapes the pause. My data also suggest that a σ70-dependent paused transcription complex enters the paused state through a pretranslocated state of the enzyme. Through these studies, I have expanded the current understanding of the protein/nucleotide and nucleotide/nucleotide interactions that constitute the σ70-dependent promoter proximal paused complex.

BIOGRAPHICAL SKETCH

Jeremy Gilbert Bird was born to Robert Bird Ph.D. and Barbara Walker Ph.D. Esq. on September 6, 1983 in Nashville, Tennessee. He has two sisters Gabrielle Bagley and Susan Walker. During his childhood he developed a love of the natural world, spending much of his time playing in the creek behind his parents home in Rockville, Maryland, catching minnows, crayfish and bugs. Despite his parents' best efforts to dissuade him from a career in science with long dinnertime conversations about their own scientific work and difficulties, he became enamored with molecular biology, after attending an HHMI Holiday lecture series during his junior year at Rockville High School in 1999. He went on to major in Biochemistry at Loyola University in Baltimore (formerly College) while spending his summers studying protein chaperones in the laboratory of Sue Wickner at the National Institutes of Health in Bethesda, MD. During this time he furthered his love of the outdoors by working as a trip leader and equipment manager for the Outdoor Education Experience at Loyola. In 2006, after completing a year post-baccalaureate fellowship in Dr. Wickner's laboratory, he moved to Ithaca, New York to pursue his Ph.D. in Biochemistry, Molecular and Cell Biology in the laboratory of Jeffery Roberts at Cornell University. While studying at Cornell he spent his free time fishing, hiking and generally enjoying the beauty of the Finger Lakes.

ACKNOWLEDGMENT

First, I would like to thank my advisor Jeffery Roberts who has taught me a tremendous amount about what it takes to be a successful scientist. Why he gave me the chance I may never know, especially after I fell asleep in his office when interviewing for the BMCB program. He is the most knowledgeable person I have ever met about how to effectively communicate one's project to a wide audience. In addition, I greatly appreciate the independence he afforded me to pursue my own ideas in his lab and his not very subtle way of telling me when I was are not making any sense or pursuing something irrelevant.

Next I want to thank all of the science teachers I have had over the years that have helped foster my interest in molecular biology. Specifically, I thank Mrs. Hamilton, my high school Biology teacher who chose me to go to the HHMI Holiday Lectures on Science, which sparked my curiosity in doing lab work. I also thank Dr. Danielle Brabazon, who taught me general and physical chemistry and told me she did not think I was cut out for biochemistry my freshman year of college. Her frankness forced me to get serious about my studies. She was pleased to eat her words when I stuck with it and got As in her physical chemistry class. I also want to thank my college advisor Dr. Brian Barr who encouraged me to come to his alma mater Cornell.

I have to also thank all of the members of the Roberts lab in my time there: Asma, Smita, Chris, Josh, Sarah, Hee-Yun, Xiao-Qiu, Man-Hee, Rada, and Eric. Chris was one of the most wonderful people I have ever met. She encouraged me to keep working hard and was a true Lab Mom. She will be missed. Josh, Man-Hee and Sarah were always there to get coffee, and quick with criticism, which, while frustrating was many times deserved and was a lesson to be learned. Josh has been a wonderful fishing buddy (a bad influence), knew just about every protocol one might need and a

great colleague to vent my frustrations to. Man-Hee taught me more than I ever thought possible about protein purification and he, Josh and Eric were always up for going out to lunch (every day practically).

I must also thank my group of friends here at Cornell and in the Ithaca community without whom I would not have had such an enjoyable experience in graduate school. I must thank all of my fishing buddies I have acquired and convinced to give up their prime fishing spots. I must also single out my friend Mike Guertin for convincing me to get Cornell hockey season tickets my first year. I have now held the tickets for 7 years. I have been part of the Lynah faithful every Friday and Saturday night for so long it will be hard to give up and it has inspired me to take up hockey as both exercise and stress relief. I also must thank my good friend Zach Via for forcing me to get off my butt and begin running, which has also been a great pleasure.

Lastly I must thank my family and the light of my life Amber Krauchunas. My parents have always been supportive of me, which I appreciate so much more than they know. I must thank my niece Emily who has taught me how important it is to always have a positive point of view no matter how difficult the problem I am tackling. If had not come to Cornell, I never would have met Amber Krauchunas. I thank her for sticking with me through thick and thin. She has helped me through some of the toughest of times and her support and encouragement means the world to me. Without her I would not have succeeded at this endeavor.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Transcription in *Escherichia coli (E. coli)* by the multi-subunit RNA polymerase complex (RNAP) is processive, but is not uniform in rate, consisting of multiple periods of continuous elongation interrupted by occasional pausing events. These pausing events can be attributed to a number of factors including protein co-factors, transcribed RNA secondary structure and the sequence composition of the DNA being transcribed. Such pausing events have been observed in systems ranging from bacteria to more complex eukaryotes such as *D. melanogaster* and humans (Core et al., 2008; Landick, 2006; Saunders et al., 2006). While some pauses are natural consequences of the mechanism by which transcription proceeds and currently have no known function, many of them serve important regulatory roles. Examples of regulatory roles for pausing in bacteria include: an RNA sequence-dependent pause in the *his* and *trp* operons of *E. coli* and *Salmonella* that synchronizes transcription and translation; Rho-dependent terminators, where pausing is believed to allow sufficient time for Rho factor to interact with the RNA and RNAP to cause termination; and promoter-proximal pauses, which allow the loading of anti-termination proteins such as lambdoid phage Q protein and the *E.coli* protein RfaH (Artsimovitch and Landick, 2000, 2002; Belogurov et al., 2007; Grayhack et al., 1985; Richardson, 2002; Toulokhonov et al., 2001). Also, pausing may more generally function to synchronize the rate of transcription with utilization of the RNA.

σ70-dependent promoter proximal pauses are of particular interest because lambdoid phage late gene promoters require these pauses for the loading of the anti-terminator Q protein. Q, through its anti-termination activity, is necessary for transcription and expression of the phage late genes during lytic growth (Deighan and Hochschild, 2007; Yarnell and Roberts, 1999). Previously, it was thought that σ70 interaction with a repeat of the -10 promoter element is sufficient for pause formation, but, only recently, the sequence present in the RNA/DNA hybrid at the site of the pause was also found to be necessary for pause formation (Perdue and Roberts, 2010; Ring et al., 1996). This also led to the realization that the sequence at the site of nearly all σ70 dependent pauses has a conserved motif: a 3 to 4 base pair GC-rich region followed by a longer AT-rich region (Perdue and Roberts, 2010).

The purpose of my research, as discussed herein, is to determine which sequences in addition to the -10-like element repeat are necessary for σ70- dependent pausing and to determine the exact role these sequences play in pause formation. This research has allowed us to better understand the mechanism of σ70 dependent pausing, but also has broader implications in understanding a general pausing mechanism: a newly recognized "pausing sequence" that I describe, a 3 to 4 base pair GC-rich region followed by a longer AT-rich region, appears to be present in a number of different identified transcriptional pauses (Herbert et al., 2006). As background to my experiments and the σ70-dependent promoter-proximal pause, I present here a discussion of the process of transcription and previous understanding of transcription pausing.

In addition to my research on pausing, I also studied the mechanism of transcription termination by looking at ATP hydrolysis by the transcription repair coupling factor Mfd, which utilizes ATP to release transcription complexes that have become irreversibly stalled at sites of DNA damage. I have included some background on transcription termination here, but I give a much more in depth overview of Mfd structure and function in the Appendix.

Transcription

The structure of RNA polymerase

Transcription of DNA into RNA in organisms ranging from bacteria to the most complicated eukaryote is performed by a multi-subunit complex called the RNA polymerase (RNAP). The structure of RNAP is highly conserved between bacteria, archaea and eukaryotes and, not surprisingly, much of the enzymatic mechanism is also conserved (Cramer, 2002). The two primary subunits of RNAP in *E. coli* are β and β' (RPB1 and RPB2 in eukaryotes), which form a claw-like structure around a central cleft that contains the active site of the enzyme, the DNA transcription bubble during initiation elongation, and the RNA/DNA hybrid (Cramer et al., 2001; Gnatt et al., 2001; Korzheva et al., 2000; Zhang et al., 1999). The α dimer (RPB3-RPB11 in eukaryotes) subunits each contains two domains: the NTD portion that interacts with and holds together the β/β'structure, and the CTD domain involved in recognition of UP elements during transcription initiation (Chen et al., 2003; Cramer, 2002; Ross et al., 1993). Another subunit present in *E. coli* is ω (RBP6 in eukaryotes), which largely plays a structural role in the complex (Minakhin et al., 2001). Lastly, the initiation

factor σ largely interacts with β and β' in the cleft and the DNA at the site of initiation before leaving the complex sometime during the elongation phase. There are a number of different σ factors responsible for transcribing genes under a variety of conditions, but I discuss the housekeeping σ factor σ70, which is responsible for a class of promoter-proximal pauses (Hatoum and Roberts, 2008; Ring et al., 1996). In eukaryotes there are a number of other subunits that are also part of the complex, but as my research is directed to mechanisms of the bacterial enzyme I will not mention them here.

Initiation

Initiation is instigated when RNAP holoeznyme, (σ70 in complex with core RNAP) recognizes two sequences, the -35 and -10 elements at the promoter of a gene. σ70 region 4 and region 2 respectively interact with these sequences and recruit RNAP to the promoter. An α-CTD interaction with UP elements, a recognition DNA sequence upstream of the -35 and -10 promoter elements, as well as other transcription factors interacting with subunits of the holoenzyme, also play a role in recruitment (Benoff et al., 2002; Busby and Ebright, 1999; Chen et al., 2003). This and subsequent steps in initiation are modeled in Figure 1.1. The binding of σ70 to the promoter and subsequent loading of the promoter DNA into the β/β' cleft forms a complex referred to as the closed promoter complex, in a process that is believed to involve β' acting as a clamp holding the DNA in place in the enzyme (Mukhopadhyay et al., 2008). This in

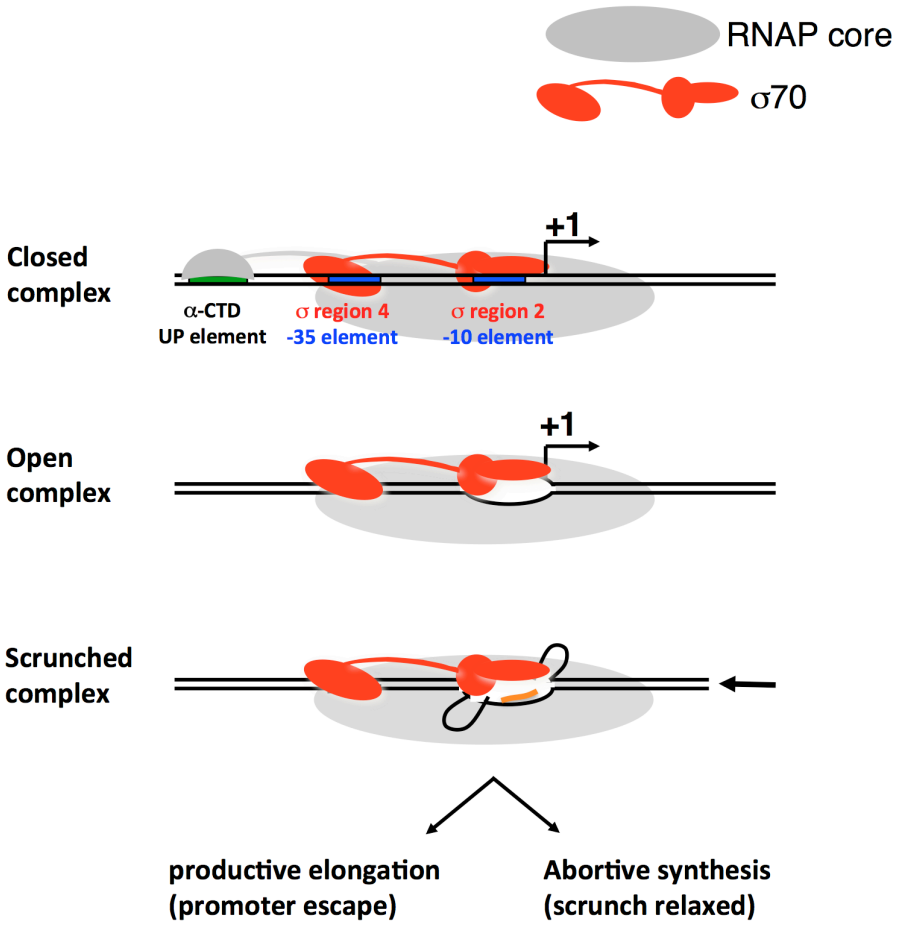


Figure 1.1. Initiation of transcription in *E. coli.* 1) σ70 interaction with the -35 and -10 promoter elements, in conjunction with α-CTD/UP element interactions and other transcription initiation factors/DNA interactions, recruits RNAP to the promoter and forms closed complex. 2) RNAP unwinds the DNA to form open complex, inserting the single stranded template DNA in the active site of the enzyme. 3) Synthesis begins, but RNAP cannot escape the promoter due to σ70 interactions with the -35 and -10 elements causing the complex to elongate the transcription bubble or "scrunch". 4) Either the scrunch collapses causing abortive initiation or the energy formed in the scrunched complex allows RNAP to escape the promoter and begin elongation.

turn allows the enzyme to begin melting the DNA to form the transcription bubble, an approximately 14 base pair unwinding of the DNA, in which RNA synthesis initiates (Chen et al., 2010).

Initial transcription occurs by a process called "scrunching." Scrunching is associated with a secondary process called abortive initiation that occurs on some promoters. Synthesis of a new RNA begins, but the complex is initially unable to disrupt the interaction of σ70 with the promoter elements. Synthesis past the initial transcription bubble length only occurs when the complex "scrunches" or pulls downstream DNA into itself, elongating the transcription bubble (Kapanidis et al., 2006; Marr et al., 2001; Revyakin et al., 2006). According to the model, this scrunched state stores energy, presumably partly represented by melted DNA in the elongated transcription bubble that is used to break the σ70-promoter element bonds and begin elongation (Kapanidis et al., 2006; Revyakin et al., 2006; Straney and Crothers, 1987). If scrunching is unable to break the σ70/promoter element interactions, synthesis of the RNA aborts and the short RNA product is released from the complex, allowing synthesis and scrunching to begin again. Interestingly, new evidence suggests that DNA scrunching can be eliminated without abolishing abortive initiation, at least for a specific promoter. This suggests the elongated transcription bubble is not solely responsible for the escape from the promoter, but instead that changes to RNAP in this state also are responsible (Martin, Personal Communication). There is structural data to support this as σ70 region 3 sits in the β/β' cleft of an initiation complex in such a way that it would prevent elongation and must be moved for RNA synthesis to proceed (Murakami and Darst, 2003). Most likely, both the

DNA scrunch and the structure of the complex are involved and the partitioning of energy between these elements may vary with the promoter.

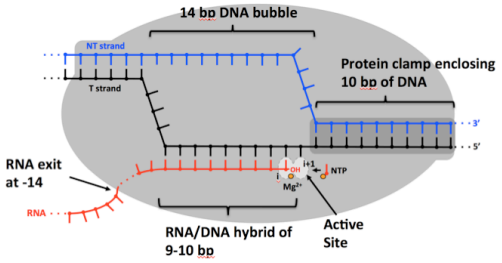
Elongation

Once RNAP escapes the promoter, it is able to actively synthesize RNA in the elongation complex (EC)(Figure 1.2A). This complex is composed of core RNAP (β, β', α^2, ω), and the 14 base pair melted DNA transcription bubble encompassing a 9 to 10 nucleotide long RNA/DNA hybrid enclosed in the β/β' cleft (Korzheva et al., 2000; Nudler et al., 1996; Vassylyev et al., 2007a). Additional contacts are made between the protein and downstream duplex DNA that is yet to be melted as well as 5 nucleotides of RNA upstream of the hybrid. The growing RNA leaves the exit channel of the complex 14 nucleotides from the hybrid, according to both structural analysis and nuclease protection studies.

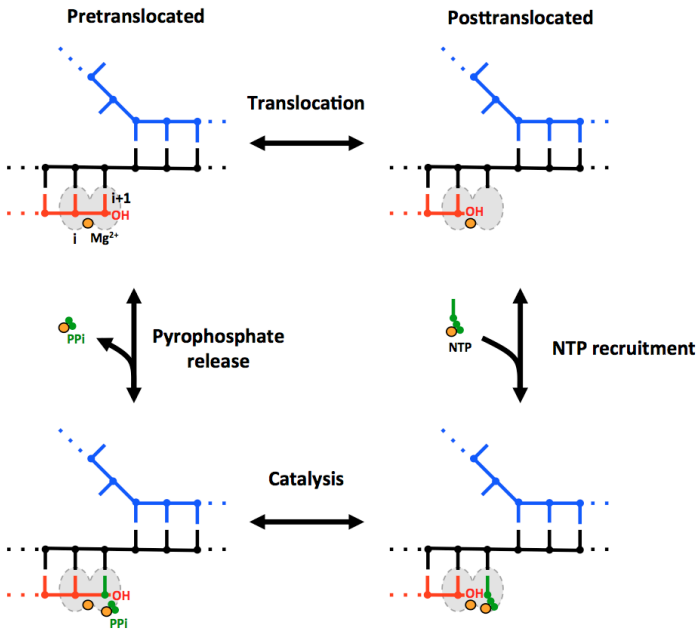
The active site of RNAP, which contains two subsites, i and i+1, contains a bound Mg^{2+} ion that is partly responsible for catalysis of the addition of new nucleotides to the growing RNA chain (Figure 1.2B)(Vassylyev et al., 2007b). A second catalytic Mg^{2+} ion is brought in by the NTP substrate of the EC. Elongation of the RNA chain occurs in a four step process controlled by a ratcheting mechanism of a helical bundle/loop present at the active site of the enzyme referred to as the trigger loop/helices (Bai et al., 2007; Hein and Landick, 2010; Kireeva et al., 2010; Zhang et al., 2010). A second adjacent structure, the bridge helix, coordinates the ratcheting of the trigger loop (Hein and Landick, 2010). First, the enzyme must translocate forward (to a position referred to as the post-translocated state), moving the previously added

Figure 1.2. The mechanism of transcription elongation. (A) Cartoon model of a transcription elongation complex showing all relevant structures and interactions. Modified from Roberts et al. 2008. (B) The nucleotide addition cycle: 1) The transcription complex translocates forward one base pair from a pretranslocated to a posttranslocated state, freeing the i+1 of the active site for the incoming nucleotide by moving the 3' end of the RNA to the i site (this is a reversible process); 2) During translocation, the correct nucleotide triphosphate enters the i+1 site, bringing with it a second Mg^{2+} ion; 3) Catalysis instigated by a conformational change of the active site; 4) Release of pyrophosphate. The cycle is then able to start again for the addition of the next nucleotide. Modified from Hein et al. 2011.

A) Transcription Elongation Complex



B) Nucleotide Addition Cycle



nucleotide from the i+1 site to the i site and opening the i+1 for recruitment of the new NTP, which is step two. There is a large body of evidence suggesting that these two steps occur at the same time (Abbondanzieri et al., 2005; Gong et al., 2005; Temiakov et al., 2005; Tuske et al., 2005). Next, the 3'-OH of the RNA chain attacks the NTP α-PO_4, in a reaction catalyzed by the two Mg^{2+} ions in the active site and instigated by the conformational change of the trigger loop to its helical form (Zhang et al., 2010). Finally PPi is released as a product of this reaction and the complex is in a pretranslocated state ready for the cycle to begin for the next nucleotide.

Transcriptional pausing

Transcription elongation, although highly processive, is not uniform in rate. Elongation is punctuated by a number of pausing events at approximately every 100 to 200 base pairs for anywhere in the range of one to six seconds (Adelman et al., 2002; Neuman et al., 2003). These events, called ubiquitous pauses, are caused by specific DNA sequences encountered by the RNAP and induce what are thought to be non-backtracked pauses. Another sequence specific form of pausing is backtracking, in which in a sequence dependent manner an elongation complex translocates backwards leaving a 3' overhanging RNA that prevents further elongation, until either the complex alone or an external factor is able to relieve the pause (Komissarova and Kashlev, 1997a; Nudler et al., 1997). In addition to these two types of pauses, there are also another class of pausing events that occur directly after transcription initiation, called promoter-proximal pauses, that are induced by either RNA secondary structure that interacts with RNAP or the re-engagement of σ70 with repeats of promoter

elements downstream of the transcription start site. An additional class of pauses is caused by the interaction of other transcription factors such as RfaH with the EC (Belogurov et al., 2007; Perdue and Roberts, 2011; Ring et al., 1996; Toulokhonov et al., 2001). In eukaryotic systems promoter-proximal pauses 30 to 50 base pairs downstream of the transcription start site are prevalent and appear to be caused by interaction of transcription factors with the EC (Core et al., 2008; Nechaev et al., 2010).

The current understanding of pausing is that it occurs as the result of a structural rearrangement of the EC, called an elemental pause, as a direct result of the complex encountering a pause inducing element, such as a specific DNA sequence, protein co-factor or RNA secondary structure (Landick, 2006). The nucleotide addition cycle described above is important to understanding how a rearrangement of the EC can induce pausing. The paused state inhibits the addition of new nucleotides to the nascent RNA chain until the polymerase either by itself or with the assistance of a cofactor is able to reverse the active site structural rearrangement and escape back into productive elongation. There is strong evidence that this elemental pause originates from the pretranslocated state of the nucleotide addition cycle, as a posttranslocated complex would be more likely to continue synthesis (Hein et al., 2011; Landick, 2006, 2009).

ECs are highly stable, but freely allow the lateral translocation of RNA and DNA through the complex, which is necessary for the enzyme's function. Occasionally the EC will encounter situations that cause a weak or disrupted RNA/DNA hybrid such as AT-rich sequence, base misincorporation, or chemical

DNA damage. The instability of the hybrid can cause the complex to reverse translocate in a process called backtracking (Nudler et al., 1997). As it does so it leaves the 3' end of the RNA chain overhanging from the RNA/DNA hybrid protruding into a structure called the secondary channel, which is believed to be the entry point of NTP substrate (Komissarova and Kashlev, 1997a; Nudler et al., 1997). The paused state of backtracked pausing is the least difficult to understand: This overhang prevents elongation of the RNA chain and must be overcome for transcription to resume, leaving the EC stalled. The current thought is that each different pausing mechanism, backtracking, RNA hairpin, accessory factor and σ70-dependent pausing, all induce a general elemental pause state that prevents continued synthesis (Landick, 2006).

The changes to the elongation complex that constitute the elemental pause state for all of non-backtracked pauses are not completely understood yet, but recent findings are beginning to shed more light on a general mechanism of pausing. One particularly interesting finding is that while all of these different forms of pausing are instigated by different mechanisms, they all have one thing in common. They all have a similar sequence that encompasses the RNA/DNA hybrid at site of the pause composed of a short, approximately three to four base pair, GC-rich region followed by a largely AT-rich region (Herbert et al., 2006; Perdue and Roberts, 2010). A similar sequence has been implicated in backtracking, but also appears to play an important role in non-backtracked pauses as a GC-rich element followed by AT-richness appears at sites of ubiquitous pauses, the *ops* pause, and the *his* pause, as well as being a

necessary part of σ70 dependent promoter-proximal pauses (Artsimovitch and Landick, 2000; Herbert et al., 2006; Perdue and Roberts, 2010).

σ70-dependent promoter proximal pausing

Biological role

σ70 dependent promoter-proximal pauses, found associated with lambdoid phage late gene promoters, as well as a number of *E. coli* genes, are particularly interesting from both a regulatory and a mechanistic standpoint. These pauses are caused by a repeat of the either the promoter -10 element or -35 element downstream of the promoter start site that interacts with σ70 and induces a pause at a synthesized RNA length of 15-25 nucleotides (Brodolin et al., 2004; Hatoum and Roberts, 2008; Nickels et al., 2004; Ring et al., 1996). In the case of phage λ, phage 82 and other lambdoid phages a σ70 region 2-dependent promoter-proximal pause on the late gene promoter *pR'* allows the anti-terminator protein Q to load into the transcription complex through a series of protein-protein and protein-DNA interactions (Deighan et al., 2008; Nickels and Hochschild, Unpublished data; Santangelo et al., 2003). This Q modified complex is able to prevent termination and allow transcription of genes downstream of a terminator (Roberts et al., 1998). The λ pR' σ70 dependent pause exists for approximately 30 seconds *in vivo* before escaping into active elongation, although the lifetime of the pause is significantly shortened in the presence of Q protein (Grayhack et al., 1985; Kainz and Roberts, 1995). σ70 region 2 dependent promoter-proximal pauses also exist in many *E. coli* chromosomal genes including

lacZ, although a regulatory role has not been ascribed to these particular pauses (Brodolin et al., 2004; Hatoum and Roberts, 2008; Liu and Roberts, Unpublished data; Nickels et al., 2004).

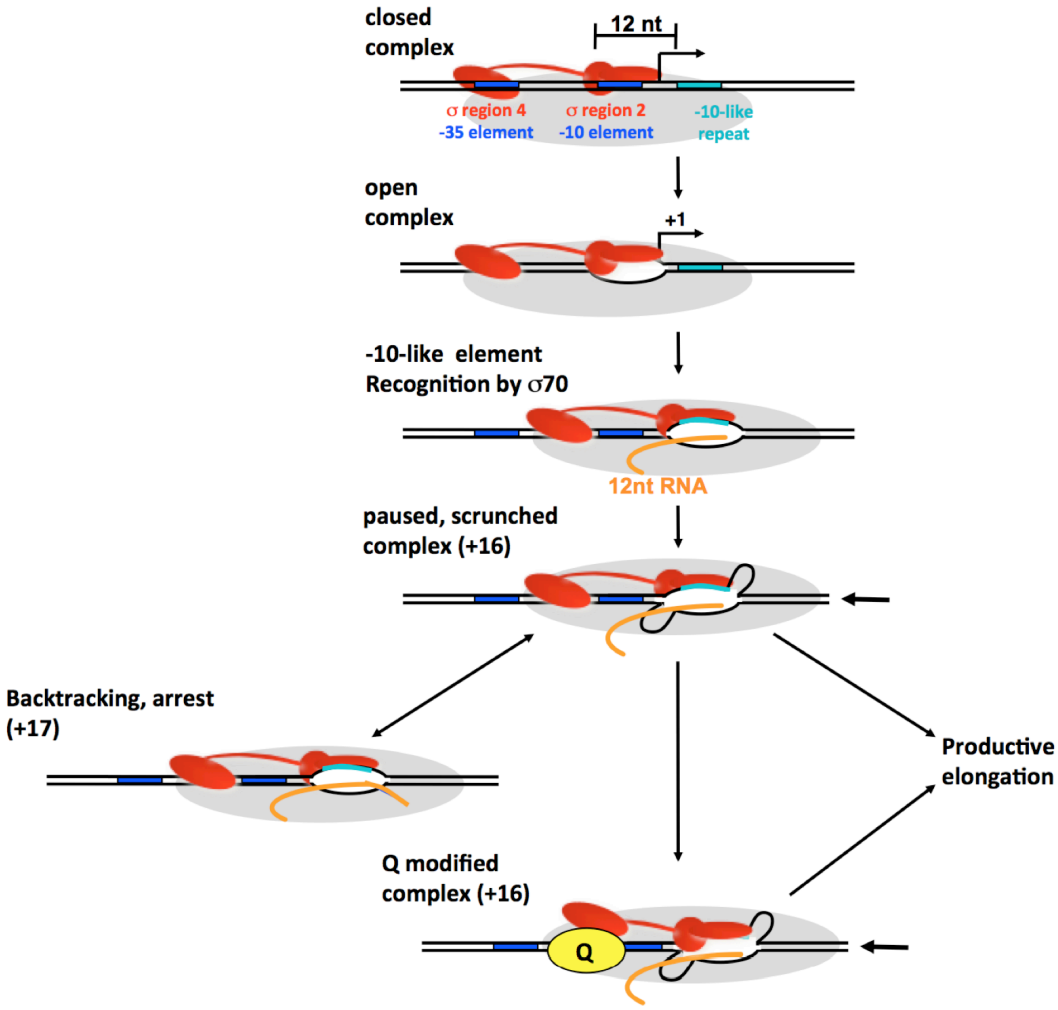
Mechanism

The mechanism of σ70-dependent pausing is understood in some detail. After a transcription complex escapes the promoter and enters elongation, σ70 remains a part of the complex for an unspecified distance (Kapanidis et al., 2005; Mooney et al., 2005). On the λ phage late gene promoter *pR'*, the pause is induced when region 2 of σ70, which is still associated with the EC, interacts with a -10-like element in the non-template strand of the DNA two basepairs downstream of the transcription start site (Figure 1.3)(Ring et al., 1996). Unable to break this interaction, the complex stops forward translocation and pauses at +16 and +17 (Grayhack et al., 1985; Kainz and Roberts, 1992). In the case of phage 82 *pR'*, the -10-like repeat is 9 basepairs downstream of the transcription start site at positions +9 to +14 and is responsible for a pause at +25 (Roberts et al., 1998).

The majority of known σ70-dependent pauses are caused by the interaction of σ70-region 2 with a -10-like sequence. Not surprisingly, a σ70 region 4 interaction with a -35 promoter element-like sequence can also cause pausing (Perdue and Roberts, 2010). Phage 82 pR', in addition to the σ70 region 2 dependent +25 pause, produces a second pause at positions +14/15 that is dependent on a σ70 interaction with a -35-like element located between the promoter -35 and -10 elements. Further

Figure 1.3. σ70 region 2-dependent promoter proximal pausing on λ *pR'*.

1) Transcription initiation occurs as described in Figure 1, but after clearing the promoter the transcription complex maintains its contacts with σ70. 2) σ70 region 2 recognizes and interacts with a -10 promoter element-like sequence after synthesis of a 12 nt RNA. 3) Anchored by the σ70 region 2/-10-like element interaction, the complex synthesizes out to an RNA length of 16 nt by scrunching in a mechanism analogous to initiation 4) One of three outcomes occurs: synthesis of 1 nt to +17 and the scrunch relaxes causing the complex to backtrack in a reversible process; or breaks the σ70/-10-like interaction by synthesizing past +17, entering active elongation; or the complex is modified by the antiterminator Q protein that disrupts the σ70/DNA interaction and allows synthesis to continue and the complex to enter elongation.


closed
complex
12 nt
σ region 4
-35 element
σ region 2
-10 element
-10-like
repeat
open
complex
+1
-10-like element
Recognition by σ70
12nt RNA
paused, scrunched
complex (+16)
Backtracking, arrest
(+17)
Productive
elongation
Q modified
complex (+16)
Q

confirming this, additional σ70 region 4 pauses have been identified in *E. coli* using RNA sequencing of short RNAs (Liu and Roberts, Unpublished data).

If translocation stopped exactly where the σ70 interaction starts during pausing on λ *pR',* then the pause would occur when an RNA of 12 nucleotides has been synthesized. Instead, pausing occurs 4 and 5 basepairs downstream at +16/17. Similarly, the -10-like element on phage 82 pR' should induce a pause at position +21, but pausing is seen at position +25. Most likely, a process analogous to the scrunching mechanism of initial transcribing complexes at the promoter is needed for the continued synthesis to +16/17 on λ *pR'* and +25 on phage 82 *pR'* to occur. Due to the inability of the complex to break the interaction of σ70 region 2 with the -10-like element, it instead draws the following 4 or 5 basepair segment of the DNA into the complex, forming a larger transcription bubble and synthesizing RNA out to the site of the pause (Kainz and Roberts, 1992; Kapanidis et al., 2006; Marr and Roberts, 2000; Revyakin et al., 2006).

The similarity to the initial scrunching mechanism has strong implications for how the pause functions. Interestingly, *in vitro* transcription experiments on λ *pR'* reveal that the pause at +17 is sensitive to cleavage by the transcription factor GreB whereas the pause at +16 remains unaffected (Marr and Roberts, 2000). One of the functions of GreB is to stimulate RNAP to cleave the 3' overhanging RNA of a backtracked EC, which indicates that the +17 pause is backtracked whereas +16 is not (Borukhov et al., 1993; Marr and Roberts, 2000; Orlova et al., 1995). The most likely explanation of this result is that the +16 pause is in a stable scrunched state and that synthesizing one more nucleotide to +17 relaxes the scrunch, forcing the complex to

backtrack due to the continued interaction of σ70 region 2 with the -10-like element. A paused complex must synthesize past +17 in order to disrupt the σ70 region 2/-10-like element interaction and escape the pause. For the lambdoid phage pauses, this occurs by two mechanisms: either the complex eventually escapes by scrunching past the pause, which disrupts the σ70 interaction and releases the complex or Q protein modifies the complex in such a way that the σ70 interaction is similarly disrupted.

Function of lambdoid phage Q protein

The σ70 dependent promoter-proximal pause found in lambdoid late gene promoters serves to allow modification of the complex by Q protein. Phage λ Q is a 22.5 kDa protein responsible for allowing ECs to read through a terminator approximately 200 base pairs downstream of the promoter (Roberts et al., 1998). This activity is necessary for proper expression of late genes during the lytic cycle (Deighan and Hochschild, 2007; Yarnell and Roberts, 1999). Phage 82, and a number of other lambdoid phages such as ϕ80, phage 21 and P22 all contain similar systems that require Q protein (Roberts et al., 1998). Recently, the prophage operon *essDP* in *E.coli* MG1655, necessary for biofilm formation, has been suggested to require a similar antitermination activity for proper expression (Ruggerberg and Hay, 2013).

In order for lambdoid phage Q protein to be activated as an antiterminator, it must first be properly loaded onto the elongation complex during the late gene σ70 region 2-dependent promoter proximal pause. This process involves a complex series of protein/protein and protein/DNA interactions. Proper Q modification of an EC for

antitermination first requires the Q Binding Element (QBE), which is a DNA sequence between the -35 and -10 elements of the promoter, and partly overlapping the -10 element. Q protein is able to bind QBE either on free DNA or as part of the σ70 dependent paused complex (Bartlett, 1998; Yarnell and Roberts, 1992). Surprisingly, Exo III DNA protection experiments show that the QBE is found in the same position on both λ and phage 82 *pR'* despite the pause occurring 7 base pairs downstream on the phage 82 *pR'* (Yarnell and Roberts, 1992). No consensus exists for the QBE, as phage Q proteins all appear to bind different sequences, which makes studying this interaction difficult. In addition to the DNA binding site Q also must contact RNAP through the β-flap, α-CTD and potentially though more contacts in the core of the enzyme (Deighan et al., 2008; Nickels and Hochschild, Unpublished data; Santangelo et al., 2003).

The presence of σ70 in the paused complex is also required, as Q will not modify complexes artificially stalled at the pause site in the absence of σ70 (Yarnell and Roberts, 1992). From this result, it can be inferred that most likely Q can only modify a complex that is in this specialized scrunched pause state. In addition to its role in pause formation, σ70 also interacts with Q during modification and it is thought that Q may disrupt the region 2/-10-like interaction, possibly releasing σ70 from the complex and allowing the complex to escape the pause (Deighan et al., 2008; Roberts et al., 2008).

A properly Q-modified complex can antiterminate by one of two mechanisms. In the first 82Q-modified transcription complexes read through intrinsic terminators because Q and the transcription factor NusA in complex with the EC protect an

additional 10 bases of RNA emerging from the exit channel and prevent the formation of the terminating hairpin (explained below)(Shankar et al., 2007). In the absence of NusA, Q is thought to antiterminate by its second activity of antipausing. Essentially Q does not allow the EC to dwell at a termination site long enough for the termination event to occur, which is presumably a kinetically sensitive step in the termination process (Shankar et al., 2007). Either mechanism is thought to be sufficient to prevent intrinsic termination, and presumably Rho-dependent termination as well.

Transcription Termination

There are currently three known mechanisms of transcription termination in *E. coli*; intrinsic or hairpin-mediated termination, Rho-mediated termination, and Mfd-mediated termination (also referred to as transcription coupled repair)(Nudler and Gottesman, 2002). For these different mechanisms to cause termination they must overcome the inherent high stability of an actively elongating transcription complex. Intrinsic terminators do this by forming a specific RNA structure that destabilizes the RNA/DNA hybrid, whereas both Rho and Mfd are proteins that use the energy of ATP hydrolysis to exert a force on the RNAP complex to dissociate it.

Intrinsic terminators are encoded in the DNA sequence and are responsible for termination of most bacterial transcription units. These sequences have two important regions: an RNA containing a G/C rich hairpin, and a 7 to 9 nucleotide U-rich region that follows the hairpin. The formation of the G/C rich hairpin is thought to facilitate the separation of the U-rich region from the template DNA, causing dissociation of the complex. Interestingly, artificially creating a hairpin stem-like structure, by annealing

a DNA oligonucleotide upstream of a U-rich region is sufficient to induce termination (Yarnell and Roberts, 1999). The dissociation of the complex is thought to happen either through a shortening of the RNA/DNA hybrid due to forward translocation or through slippage of the hybrid due to its homopolymeric nature (Komissarova et al., 2002; Larson et al., 2008; Santangelo and Roberts, 2004; Yarnell and Roberts, 1999). It is possible that both of these mechanisms are contributing to the termination event, as they are not mutually exclusive.

Rho-dependent termination occurs when Rho, a hexameric AAA+ ATPase encounters the unoccupied RNA being synthesized by an active elongation complex (Lau et al., 1983; Lowery-Goldhammer and Richardson, 1974; Morgan et al., 1985). Rho is believed to act by translocating to the EC along the RNA and exerting a force on the complex perhaps similar to that induced by formation of the GC-rich hairpin of the intrinsic terminator (Ciampi, 2006; Skordalakes and Berger, 2006). Rho functions as a terminator for a number of bacterial transcription units, as shown by treating *E. coli* with the antibiotic bicyclomycin, which inhibits Rho function. Without Rho function there is a large increase in the transcription of intergenic regions (Cardinale et al., 2008). Rho is also responsible for inducing operon polarity and preventing the accumulation of unutilized RNA in a cell (Adhya et al., 1974; Franklin, 1974; Richardson et al., 1975).

The third mechanism of termination is that of the protein Mfd, also called Transcription Repair Coupling Factor (TCRF). Mfd serves a different role than the other two modes of termination in that it acts on transcription complexes stalled by DNA damage and also promotes DNA repair by recruiting enzymes of DNA excision

repair to the stalled enzyme. By recruiting these proteins, Mfd acts as a link between transcription and DNA repair. UV-radiation often causes photochemical lesions in DNA bases that cannot be transcribed by RNAP polymerase such as thymine or cyclobutane dimers. In an ATP hydrolysis-dependent manner, Mfd is capable of recognizing and dissociating ECs that have been stalled by these bulky DNA lesions and other DNA damage, by nucleotide starvation, or by other factors inhibiting forward translocation (Park et al., 2002; Selby and Sancar, 1995a, b). Interestingly, as other forms of DNA damage do not irreversibly stall ECs, Mfd does not appear capable of causing termination at these lesions (Smith and Savery, 2008).

It is important to note that all three mechanisms of termination are thought to act through a similar mechanism. Dissociation of the complex at an intrinsic terminator occurs through either a slippage or forward translocation mechanism. Mfd, and most likely Rho as well, act through the forward translocation mechanism. This forward translocation has been demonstrated by Mfd's ability to alleviate a backtracked pause, allowing continued synthesis in the presence of NTPS, but causing dissociation of the complex in the absence of NTPs (Park et al., 2002). Additionally a EcoRI Gln111 mutant protien bound directly downstream of an EC requires three times as long to be dissociated than complexes stalled at DNA damage lesions, which is consistent with it preventing forward translocation of the complex and needing to be displaced before release can occur (Park et al., 2002).

CHAPTER 2

A NOVEL GC-RICH ELEMENT DETERMINES THE SITE OF PHAGE 82 pR' σ70-DEPENDENT PAUSING

Introduction

Transcription by multi-protein subunit RNA polymerases (RNAPs) is not uniform in rate, but instead consists of multiple stops and starts, induced by a number of elements including protein co-factors of RNAP, transcribed RNA secondary structure, and particular features of the sequence composition of the DNA being transcribed. Such stoppage events, referred to as pauses, have been identified in different species ranging from bacteria to more complicated organisms such as *D. melanogaster* and humans (Core et al., 2008; Landick, 2006; Nechaev et al., 2010). In prokaryotes, these pauses can serve important regulatory roles during transcription including: an RNA sequence dependent pause in the *his* and *trp* operons of *E. coli* and *Salmonella* that synchronizes transcription and translation; Rho dependent terminators, where pausing is believed to allow sufficient time for Rho factor to interact with the RNA and RNAP to cause termination; and promoter-proximal pauses, which allow the loading of anti-termination proteins such as lambdoid phage Q protein and the *E.coli* protein RfaH. Pausing may also generally function to synchronize the rate of transcription with utilization of the RNA (Artsimovitch and Landick, 2000, 2002; Belogurov et al., 2007; Grayhack et al., 1985; Henkin and Yanofsky, 2002; Richardson, 2002; Toulokhonov et al., 2001). In eukaryotic systems, there are a number of genes, including the exemplar heat shock genes, that contain promoter-

proximal pauses 30-50 basepairs downstream of the transcription start site that are implicated in regulation of the gene (Core et al., 2008; Nechaev et al., 2010).

One particularly interesting set of pauses from a regulatory standpoint is the σ70 region 2-dependent promoter proximal pauses of lambdoid phage late genes, as well as a number of *E. coli* genes. These pauses are caused by a repeat of the promoter -10 element downstream of the transcription start site that induces a pause at a synthesized RNA length of 15-25 nucleotides (Ring et al., 1996). In the case of phage λ, phage 82 and other lambdoid phages a σ70 region 2-dependent promoter-proximal pause on transcription complexes from the late gene promoter *pR'* allows the anti-terminator protein Q to load into the transcription complex through a series of protein-protein and protein-DNA interactions (Deighan et al., 2008; Nickels et al., 2002; Santangelo et al., 2003; Yarnell and Roberts, 1992). This Q modified complex is able to prevent termination and allow transcription of genes downstream of a terminator (Roberts et al., 1998). σ70 region 2-dependent promoter proximal pauses also exist in many *E. coli* chromosomal genes including *lacZ*, although a regulatory role has not been ascribed to these particular pauses (Brodolin et al., 2004; Hatoum and Roberts, 2008; Liu and Roberts, Unpublished data; Nickels et al., 2004).

The mechanism of sigma-dependent pausing is understood in some detail. After a transcription complex escapes the promoter and begins transcription, σ70 remains a part of the complex for an unspecified distance (Kapanidis et al., 2005; Mooney and Landick, 2003). On the λ phage late gene *pR'*, the pause is induced when σ70 region 2 interacts with a repeat of a -10 element in the non-template strand of the DNA 2 basepairs downstream of the transcription start site (Ring et al., 1996). Unable

to break σ70 region 2/-10-like interaction, the complex cannot continue forward translocation and pauses at +16 and +17 (Grayhack et al., 1985; Kainz and Roberts, 1992). In the case of phage 82 *pR'* the -10-like repeat is 10 basepairs downstream of the transcription start site and a pause is induced at +25 (Marr and Roberts, 2000; Yarnell and Roberts, 1999).

If stability of the σ70/-10 element complex interaction is the primary cause of pausing on the λ phage late gene *pR'* promoter, pausing would be expected at position +12. Instead, pausing occurs 4 and 5 basepairs downstream, at +16 and +17. This most likely happens in a process analogous to "scrunching" that occurs during transcription initiation. During initiation, the transcription complex "scrunches" or pulls downstream DNA into itself, elongating the transcription bubble in order to synthesize past the initial transcription bubble length (Kapanidis et al., 2006; Marr et al., 2001; Revyakin et al., 2006). This scrunched state is necessarily high energy and is believed to provide energy for the complex to break the σ70 promoter element bonds and begin elongation (Kapanidis et al., 2006; Revyakin et al., 2006; Straney and Crothers, 1987). A similar process happens during σ70-dependent pausing, where the complex cannot break the interaction with σ70 region 2 and instead draws the following 4 or 5 basepair segment of the DNA into the complex, forming a larger transcription bubble and synthesizing RNA to +16/+17 (Kainz and Roberts, 1992; Marr et al., 2001; Perdue and Roberts, 2010). Phage 82 *pR'* promoter-proximal pausing acts by the same mechanism and should pause at position +21, but in fact pauses at +25 (Perdue and Roberts, 2010). The fact that the scrunch stops at a defined position suggests that some

other DNA sequence or property of the RNAP transcription complex must play a role in σ70 dependent pausing.

Two other sequences near the -10-like element, including several nucleotides around the pause site itself, were identified by an alignment of known σ70 dependent pauses; the first is a natural A/T richness of the terminal sequence region, and the second, a 3 to 4 basepair GC-rich element located between the AT-rich sequence and the -10-like element (Perdue and Roberts, 2010). These two sequences in the absence of the -10-like element induce a backtracked pause on phage 82 *pR'*, although their role in the σ70 dependent pause is mechanistically distinct because (as we describe below) the major element of the σ70 dependent pause is not backtracked. (Perdue and Roberts, 2010). Increasing the GC content of the AT-rich region, corresponding to the 3' end of the RNA/DNA hybrid at the pause, greatly reduced the amount of σ70 dependent pausing seen during *in vitro* transcription experiments (Perdue and Roberts, 2010). This mutation, in fact, was nearly as strong as eliminating the -10-like element that induces the pause implicating it in pause formation, but not necessarily pause site determination.

The 3 to 4 basepair GC-rich element bears a striking similarity to another known transcription promoter element called the discriminator, which is present on a number of promoters including particularly those of rRNA and tRNA genes, and acts through a non-template strand interaction with σ70 region 1.2 (Haugen et al., 2006; Haugen et al., 2008). There is a strong preference for G bases in the non-template strand element as discriminators containing G residues, especially at position 2 of the 3 bases, appear to most strongly interact with σ70 region 1.2 (Feklistov et al., 2006).

Interestingly, the non-template strand of this GC-rich discriminator-resembling sequence on λ phage *pR'* is necessary for pause formation. The sequence of bases +7-9 of λ *pR'* is GGG. If any of them are mutated to A:T base pairs the amount of pausing seen is greatly reduced, which implies that the discriminator interaction is playing a role in pause formation (Ring and Roberts, 1994). This discriminator sequence, like the extended -10 sequence/σ70 region 3 interaction present at the 82 *pR'* promoter-proximal pause (but not the λ pause), most likely acts as an extension of the σ70 region 2/-10 element interaction, further strengthening the σ70 DNA interaction and stabilizing the open complex (Barne et al., 1997; Ring et al., 1996). The 82 *pR'* sequence is GCC at positions +15-17, which is theoretically less strong as a discriminator element. The experiments on this sequence in λ *pR'* showed that the template strand of positions +7-9 appeared to play almost no role in pausing (Ring and Roberts, 1994). However, there is also growing evidence that this sequence and the AT-rich region that follows it play an important role in pause formation not only for σ70 dependent promoter-proximal pauses, but also a general pausing mechanism through what appears to be a template strand effect (Herbert et al., 2006; Perdue and Roberts, 2010, 2011). Similar sequences are present at a number of other pauses including the *ops* pause and ubiquitous pauses, which do not require σ70 for pause formation (Herbert et al., 2006).

We have studied the effect on pause formation of mutations to the GC-rich sequence downstream of the -10-like element of the phage 82 *pR'* promoter in order to probe the function of this element. Using σ70 mutants deficient for binding either the -

10-like element (region 2 mutation) or a discriminator interaction (region 1.2 mutation), in addition to heteroduplex DNA templates, we further explored how the GC-rich sequence is contributing to pause formation. Our major findings are that the phage 82 *pR'* GC-rich discriminator-like sequence is necessary for pause formation not only through the nontemplate strand but also through a template strand specific effect. Furthermore, we find that moving this sequence moves the site of the pause. From this we propose that this sequence is what determines the length of scrunching that occurs during σ70 region 2-dependent promoter proximal pausing. Surprisingly, this is an entirely new role for sequences at the discriminator site, even though there is some precedent for the importance of this sequence in pausing (Herbert et al., 2006).

Materials and Methods

Plasmids and DNA templates: Linear DNA templates for transcription were amplified from plasmids by PCR and gel purified using a gel extraction kit from Qiagen. All of the DNA templates were created from derivatives of the p82a+2G -35-like mutant plasmid by Agilent quickchange mutagenesis (Guo and Roberts, 1990; Perdue and Roberts, 2010). Heteroduplex templates were constructed as described in Ring et al. 1996.

Proteins: RNAP protein was the gift of M. Suh (Cornell University, Ithaca, NY). The following proteins were purified as described: 6xHis-σ70, M102A, and 402F mutants (Marr and Roberts, 1997), GreB (Borukhov and Goldfarb, 1996).

In vitro transcription: Reaction mixtures containing 2 nM DNA template, 10 nM HoloRNAP (10 nM CoreRNAP reconstituted with 50 nM WT or mutant σ70), plus 200 mM each of ATP, GTP and CTP, 50 mM UTP (supplemented with 0.5 mCi/ml [α-^{32}P]-UTP) in Transcription buffer (20 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 10% glycerol) were incubated at 37^{o}C for 10 min to form open complexes. For Figure 2.3E, 100 nM GreB was added to the reaction mix. A single round of transcription was initiated by the addition of Start mix (50 mM $MgCl_2$ and 10 mg/ml Rifampacin) and reactions were allowed to proceed at 37^{o}C.Alliquots were taken at indicated times after addition of start mix and mixed with 5 volumes Stop solution (600 mM Tris-HCl (pH 8.0), 12 mM EDTA, and 0.16 mg/ml tRNA) on ice to stop transcription. RNA was extracted by mixing with a equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). 2.5 volumes of 100% Ethanol was added to the aqueous layer to precipitate samples overnight at -20^{o}C. Samples were run on 12% polyacrylimide gel containing 6M Urea. RNA species were detected using a Molecular Dynamics Typhoon 9400 and images were analyzed using GE Imagequant software. Pause efficiency was determined by first normalizing transcript bands based on the number of radiolabeled residues, then the adjusted intensity of the pause bands (+25 and +26) are divided by the total of all of the intensities present in a lane (paused bands + terminated band).

Results

Promoter-proximal Pausing on λ phage and phage 82 pR' occurs in a similar manner

Promoter-proximal pausing on the phage λ and phage 82 late gene promoter *pR'* is necessary for expression of the phage late genes during the lytic cycle, through

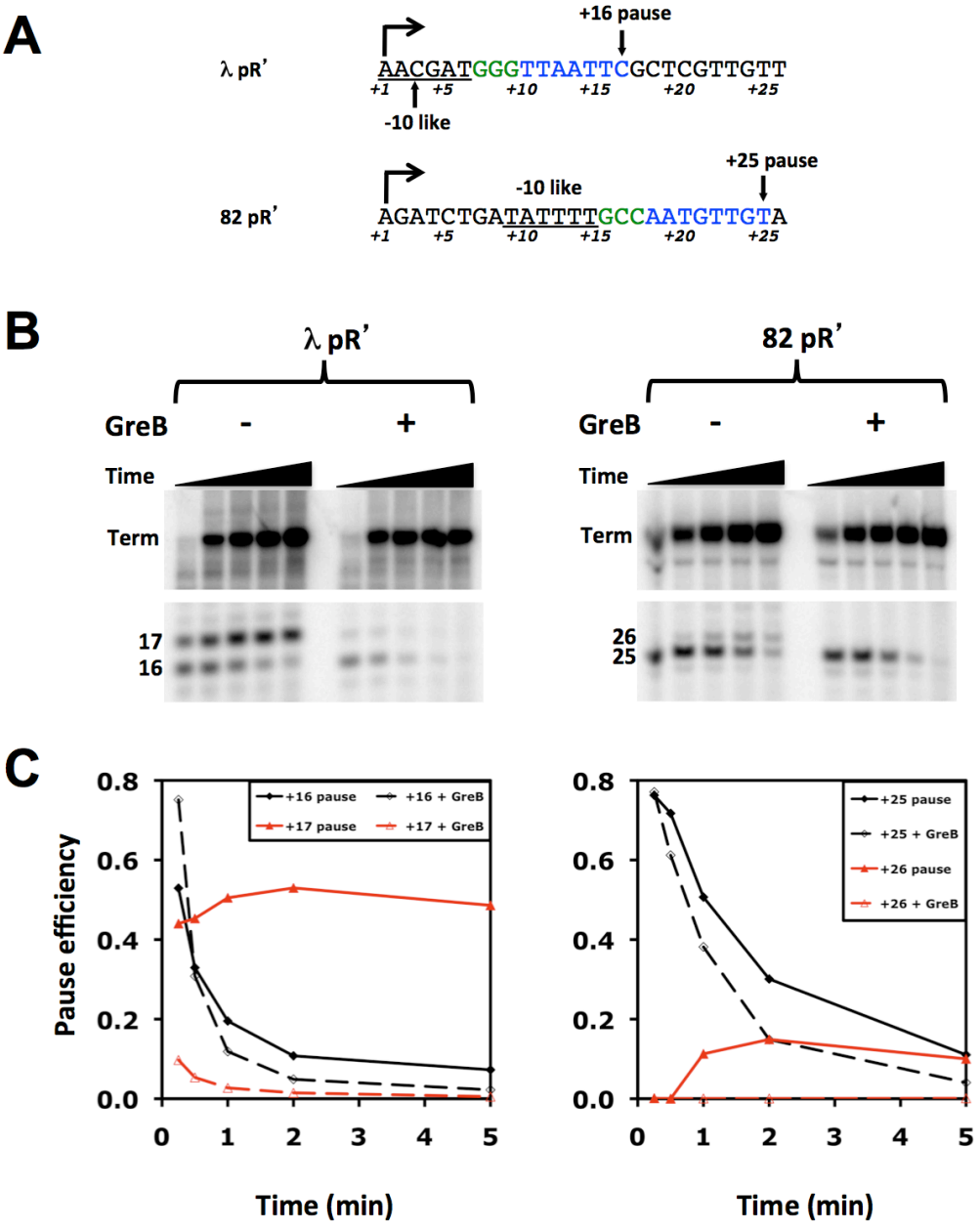
its requirement for Q loading (Deighan and Hochschild, 2007; Yarnell and Roberts, 1999). We have studied this process in great detail and we know the specific location of pauses on both λ and 82 *pR'* and that the pause, in both phage DNAs, requires primarily an interaction between σ70 region 2 and a repeat of the -10 promoter element downstream of the promoter start site (Ring et al., 1996; Roberts et al., 1998). The pause inducing sequences for both are slightly different, as are the locations of the pauses (Figure 2.1A). Transcription of the λ *pR'* template pauses at positions +16/17, whereas the pause occurs at +25/26 on the phage 82 *pR'* template; the pauses thus show similar displacement from the -10-like element (Figure 2.1B).

Interestingly, when *in vitro* transcription on λ *pR'* is performed in the presence of excess transcription cleavage factor GreB, the pause at +17 disappears whereas the pause at +16 remains unaffected (Figure 2.1B) (Marr and Roberts, 2000). One of the functions of GreB is to cleave the 3' overhanging RNA of a backtracked EC, which suggests that the +17 pause is backtracked and +16 is not (Borukhov et al., 1993; Orlova et al., 1995). The most likely explanation of this result is that the +16 pause is in a stable scrunched state and that synthesizing one more nucleotide to +17 relaxes the scrunch, forcing the complex to backtrack due to the continued interaction of σ70 region 2 with the -10-like element. A paused complex most likely must synthesize past +17 in order to disrupt the σ70 region 2/-10-like element interaction and escape the pause.

This pausing behavior has not been previously reported for the +25 pause on phage 82 *pR'*, but in fact does occur (Figure 2.1B, C). Both +25 and +26 RNA species

Figure 2.1. Pausing behavior of phage λ versus phage 82 *pR'* promoters.

(A) Phage λ and phage 82 *pR'* sequences from the transcription start site to position +26. The pause-inducing -10-like element is indicated by underline. The subsequent GC-rich and AT-rich sequences are labeled in green and Blue respectively. The site of the promoter proximal pause is marked for each sequence. (B) Single-round *in vitro* transcription time courses were performed on M650 and p82a -35-like mutant template either in the presence or absence of GreB. Relevant paused and termination product RNAs are marked. (C) Band intensities of +16/17 pause on the M650 template and +25/26 pause on p82a -35-like mutant template as well as terminated product were quantified and amount of pause seen was plotted as % of total RNA over time.

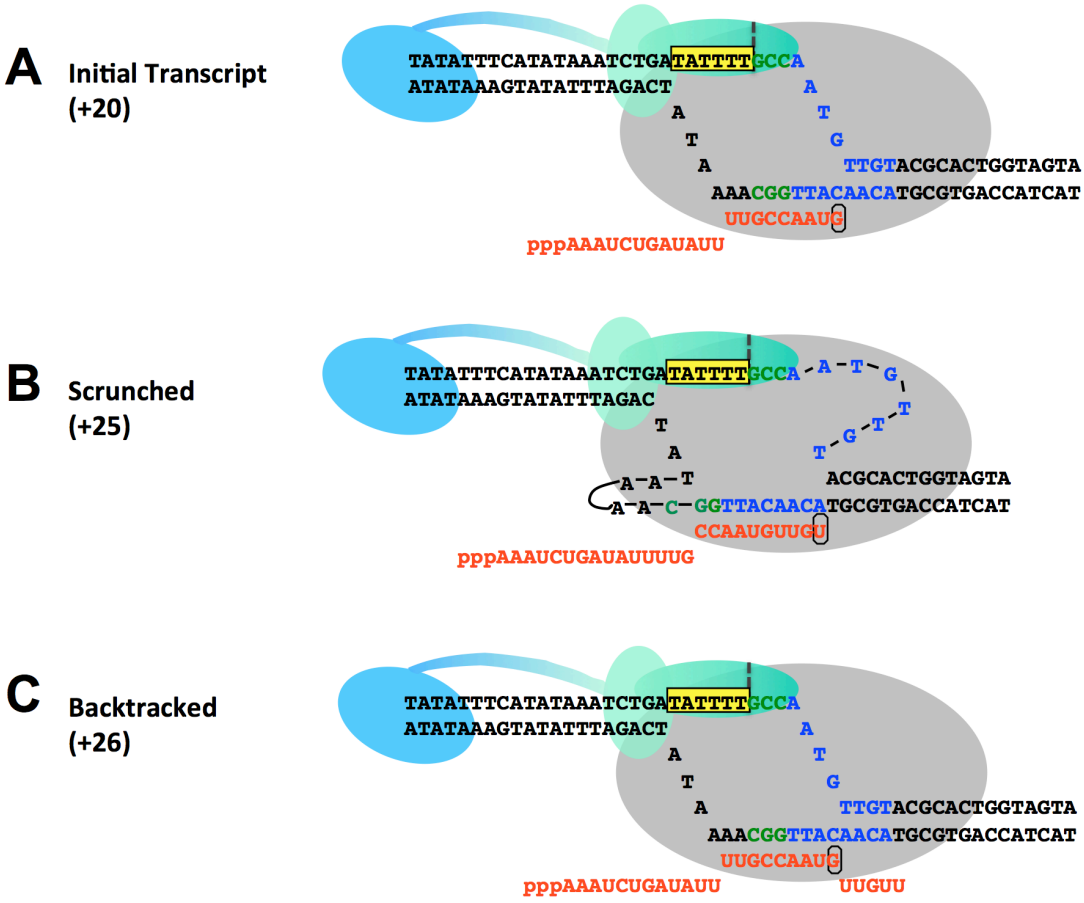

A
λ pR'
+16 pause
AACGATGGGTTAATTCGCTCGTTGTT
+1 +5 +10 +15 +20 +25
-10 like
82 pR'
+25 pause
-10 like
AGATCTGATATTTTGCCAATGTTGTA
+1 +5 +10 +15 +20 +25
B
λ pR'
82 pR'
GreB - + GreB - +
Time
Term
17
16
26
25
C
Pause efficiency
0.8
0.6
0.4
0.2
0.0
+16 pause
+16 + GreB
+17 pause
+17 + GreB
+25 pause
+25 + GreB
+26 pause
+26 + GreB
0 1 2 3 4 5
Time (min)
Time (min)

are present in the absence of GreB. When reactions are treated with excess GreB the pause at +26 disappears in the same manner as the +17 pause of λ *pR'*. The behavior of +26 pause on phage 82 *pR'* is not exactly the same as that of the +17 pause on λ *pR'* during *in vitro* transcription. λ *pR'* +17 pause is present even as early as 15 seconds from the start of the reaction. Phage 82 *pR'* +26 pause only begins to form later in the time course, at about 1 to 2 minutes. This could be due to complexes more quickly escaping the +25 pause and immediately backtracking after the addition of the 26th nucleotide to the RNA chain. The fact that they behave the same supports the model that the first site, 16 or 25, is special in representing a stable, forward scrunched complex, and that the addition of one more nucleotide destabilizes it, causing backtracking to the original position of σ70 binding (12 or 20)(Figure 2.2).

The discriminator-like sequence is necessary for promoter proximal pause formation on 82 pR', but surprisingly acts primarily through the template strand

The three bases directly downstream of the -10-like element of both λ *pR'* and phage 82 *pR'* (+7-9GGG in λ, +15-17GCC in 82; numbering is bases from the start of transcription) is remarkably similar to the promoter element called the discriminator. We tested the role of this sequence in pause formation by performing single round *in vitro* transcription on phage 82 *pR'* templates containing one of the following mutations: +15 G:C to A:T, +16 C:G to A:T, +17 C:G to A:T, or 15-17 GCC:CGG to AAA:TTT (Figure 2.3A). A:T base pairs were chosen because of the lower preference for interaction with A rather than G seen in the promoter discriminator sequence (Feklistov et al., 2006). Mutating each base individually slightly reduced the amount

Figure 2.2. Models of the different possible pause states of the phage 82 promoter-proximal pause. (A) The transcription complex first encounters the -10-like sequence and σ70 engages the DNA. (B) The GreB insensitive active +25 pause. Transcription continues to an RNA length of 5 bases from (A) despite the σ70 region 2/-10-like interaction, which anchors the complex and forces the complex to scrunch. Due to hybrid sequence and Protein/Nucleotide interactions this complex pauses and does not continue to scrunch. (C) Once the transcription complex is able to escape the +25 pause, either by the incorporation of Q protein into the complex or a natural escape it can possibly fall into a backtracked confirmation at +26 presumably because of the the AT-richness of the RNA/DNA hybrid.


A
Initial Transcript
(+20)
TATATTTCATATAAATCTGATATTTTGCCA
ATATAAAGTATATTTAGACT
TTGTACGCACTGGTAGTA
AAACGGTTACAACATGCGTGACCATCAT
UUGCCAAUG
pppAAAUCUGAUAUU
B
Scrunched
(+25)
TATATTTCATATAAATCTGATATTTTGCCA
ATATAAAGTATATTTAGAC
ACGCACTGGTAGTA
GGTTACAACATGCGTGACCATCAT
CCAAUGUUGU
pppAAAUCUGAUAUUUUG
C
Backtracked
(+26)
TATATTTCATATAAATCTGATATTTTGCCA
ATATAAAGTATATTTAGACT
TTGTACGCACTGGTAGTA
AAACGGTTACAACATGCGTGACCATCAT
UUGCCAAUG
pppAAAUCUGAUAUU
UUGUU

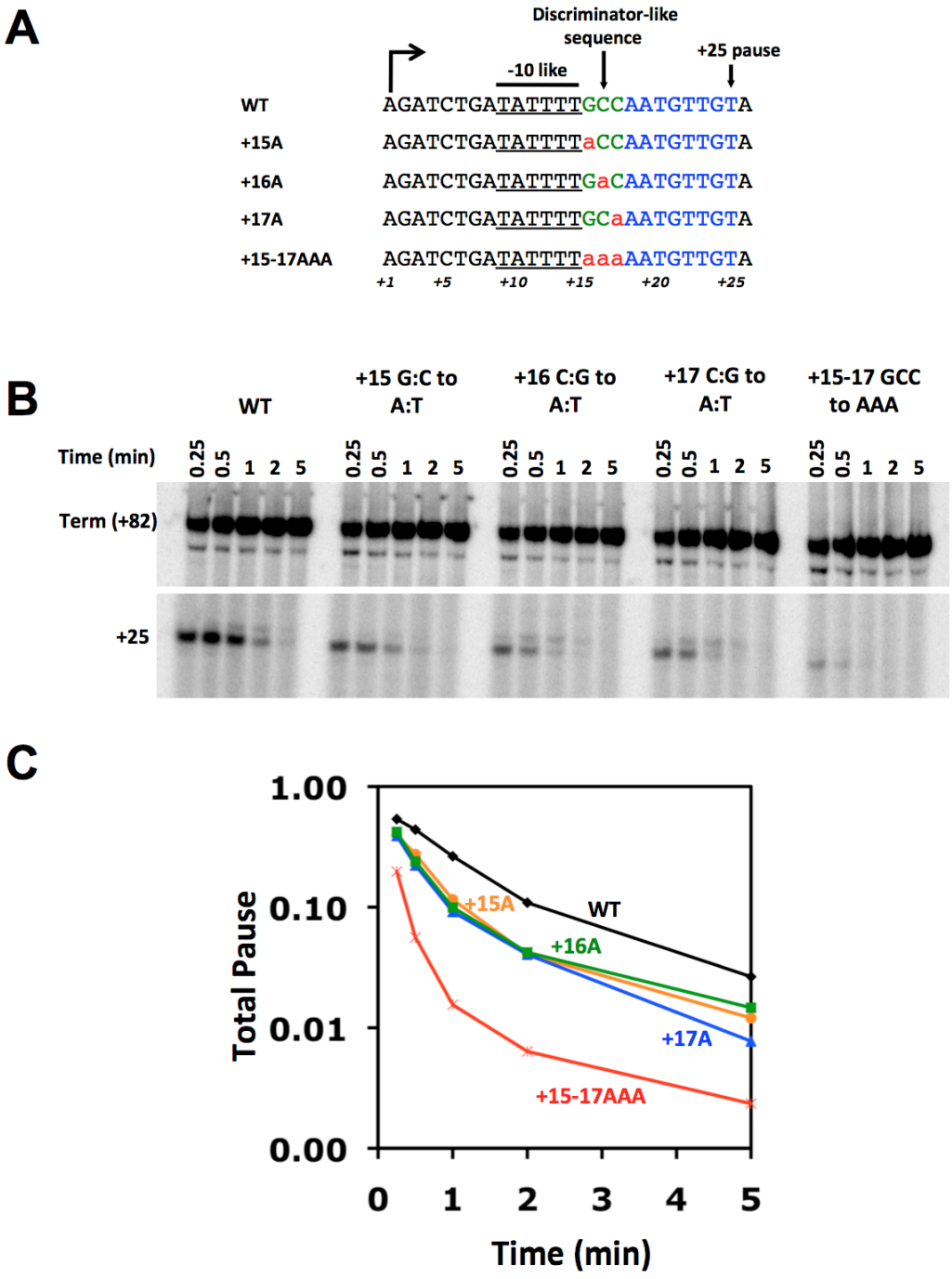
of pause formed and reduced the half-life of the pause. Mutating all three bases drastically reduced the amount of pause seen, nearly eliminating it. This drastic effect indicates that this sequence is necessary for pause formation (Figure 2.3B, C).

Next, we wanted to determine if the discriminator-like sequence is acting in the same manner as a promoter discriminator sequence, i.e. through a non-template strand interaction with σ70 as has been also shown for λ (Haugen et al., 2006; Haugen et al., 2008; Ring and Roberts, 1994). To do this we performed heteroduplex analysis of the mutant templates: templates containing either a mutant non-template strand and wild type template strand or the opposite were constructed and transcription was performed comparing each heteroduplex DNA to the homoduplex wild type and mutant templates (Figure 2.4).

Interestingly and surprisingly, in the case of the triple mutant and the +15 and +16 single mutants, the non-template strand mutant template showed identical levels of pausing to the wild type template (Figure 2.4A, B, C, D). In all three cases, the amount of pausing seen on the template strand mutant template most closely resembled the homoduplex mutant. This result is the opposite of what one would expect if this sequence functions through an interaction with σ70 region 1.2, and different from the result with the λ pause, for which discriminator site mutants act primarily through the non-template strand (Ring and Roberts, 1994). The 82 pR' +17 C:G to A:T heteroduplex mutants act differently from +15 and +16 mutants, as both strands contribute to the reduction in pause (Figure 2.4E).

Assuming that the paused complex is in a pretranslocated state similar to other known pauses, which we discuss in Chapter 3, the length of the RNA/DNA hybrid in

Figure 2.3. The discriminator-like sequence is necessary for pause formation. (A) Sequences of WT and discriminator-like sequence mutants. The discriminator-like sequence is marked in green; The AT-rich sequence in blue; The -10-like element is underlined; Mutations are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A). The +25 and +26 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pausesand terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; +15G:C to A:T is yellow circles; +16C:G to A:T is green squares; +17C:G to A:T is blue triangles; +15-17GCC:CGG to AAA:TTT is red Xs.

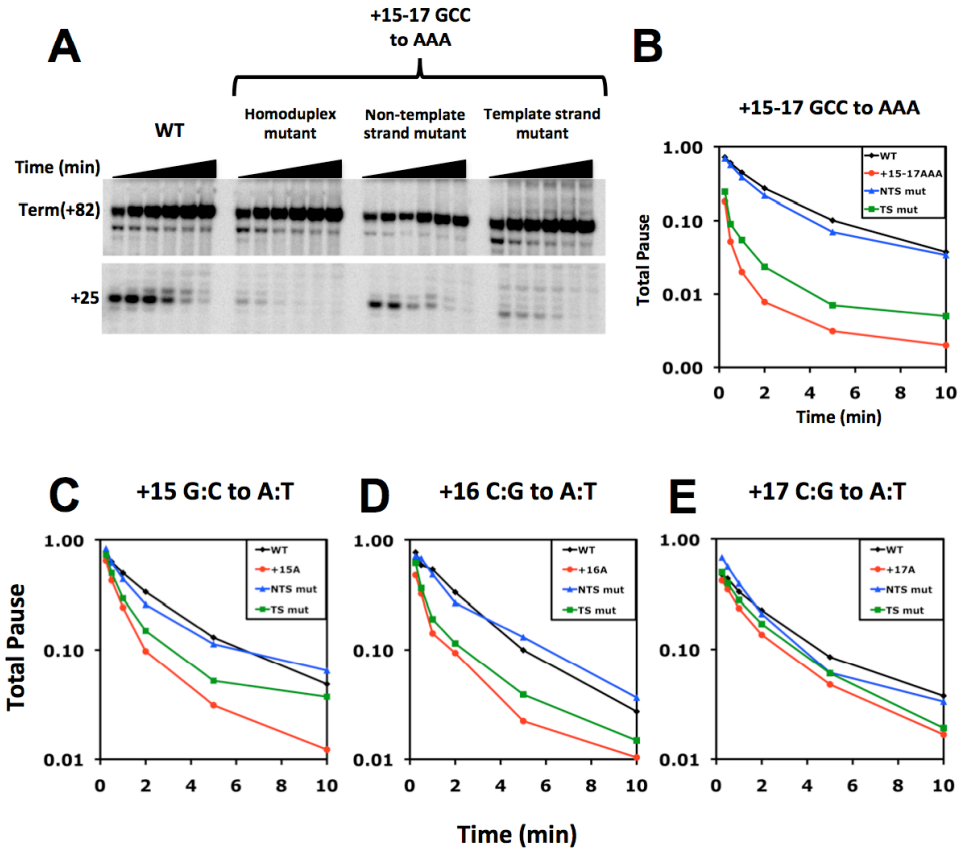

A
Discriminator-like sequence
-10 like
+25 pause
WT AGATCTGATATTTTGCCAATGTTGTA
+15A AGATCTGATATTTTaCCAATGTTGTA
+16A AGATCTGATATTTTGaCAATGTTGTA
+17A AGATCTGATATTTTGCaAATGTTGTA
+15-17AAA AGATCTGATATTTTaaaAATGTTGTA
+1 +5 +10 +15 +20 +25
B
WT
+15 G:C to A:T
+16 C:G to A:T
+17 C:G to A:T
+15-17 GCC to AAA
Time (min)
0.25 0.5 1 2 5
Term (+82)
+25
C
Total Pause
1.00
0.10
0.01
0.00
WT
+15A
+16A
+17A
+15-17AAA
0 1 2 3 4 5
Time (min)

the paused complex is expected to be ten bases. This means the G residues at positions +16 and +17 of the template strand would be paired with RNA at the 3' of the hybrid, whereas the C at +15 would not. It has been considered that the hybrid might be longer, and in fact specifically elongated in pauses, but this has not been demonstrated (Hein et al., 2011).

The surprising result that this sequence acts through a template strand mechanism stands in contrast to previous work on the discriminator-like sequence of λ *pR'*, for which heteroduplex analysis of position +7-9 base mutations showed that the reduction in pause seen is largely caused by mutations in the non-template strand (Ring and Roberts, 1994). Furthermore σ70 region 1.2 is required for this non-template strand function, as expected (Filter and Roberts, Unpublished data). However, there is a minor amount of template strand effect seen for mutations to positions +7 and +8 of λ *pR'* (Ring and Roberts, 1994).

The sequence of the λ *pR'* discriminator-like sequence is GGG, which is also the consensus sequence of the promoter discriminator element (Feklistov et al., 2006). The 82 *pR'* sequence is a less favored GCC. There is some evidence that the second of the three bases is most important for the interaction (Filter and Roberts, Unpublished data). It is therefore possible that the Cs in position +16 and +17 of the non-template strand of 82 *pR'* in this position means there is only a weak interaction with σ70 region 1.2, and therefore any template stand effect would be more prominent. We propose that this explains why the template strand effect is more prominent in case of 82 *pR'*.

Figure 2.4. Heteroduplex analysis reveals that the discriminator-like mutants act through a template-strand effect. (A) Single round *in vitro* transcription time-courses on WT, +15-17GCC:CGG to AAA:TTT mutant homoduplex and constructed heteroduplex templates. The +25 and +26 pauses and terminated product RNAs are labeled. (B) Band intensities of +25/26 pauses and terminated product from (A) were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; Homoduplex mutant is red circles; Non-template strand mutant heteroduplex is blue triangles; Template strand mutant heteroduplex is green squares. (C, D, E) Single round *in vitro* transcription time-courses were also run for the three single base mutant homoduplex and heteroduplex templates: +15G:C to A:T, +16C:G to A:T and +17C:G to A:T. Gels are not shown, but % pause was determined as in (B) and graphed. The different templates are labeled as in (B).


A
+15-17 GCC
to AAA
WT
Homoduplex
mutant
Non-template
strand mutant
Template strand
mutant
Time (min)
Term(+82)
+25
B
+15-17 GCC to AAA
Total Pause
Time (min)
WT
+15-17AAA
NTS mut
TS mut
C
+15 G:C to A:T
+15A
D
+16 C:G to A:T
+16A
E
+17 C:G to A:T
+17A

The s70 region 1.2/non-template strand interaction is still necessary for pause formation

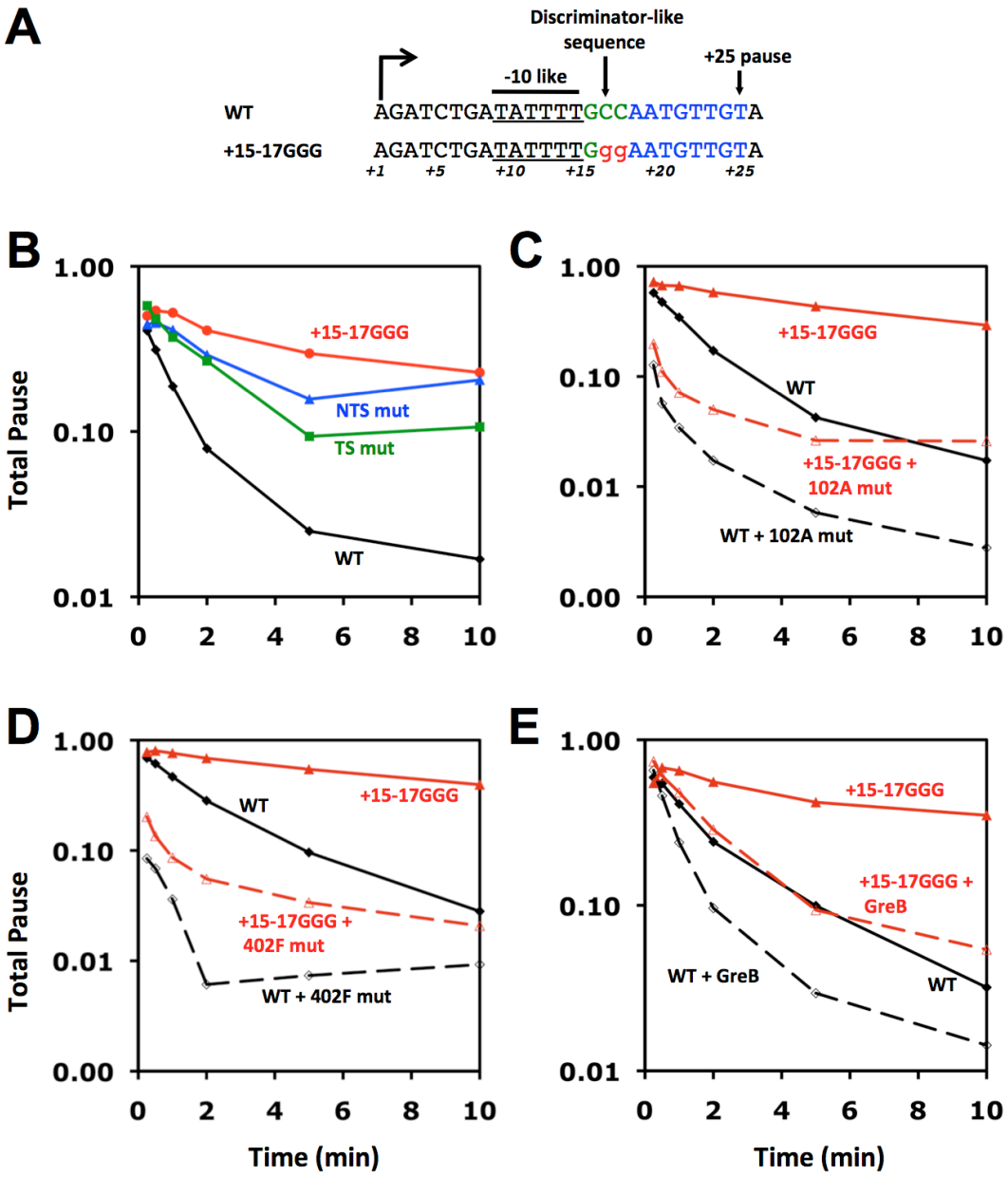
The effect of mutating λ +7-9GGG was primarily due to a non-template strand effect, which most likely was due to the interruption of the σ70 region 1.2/discriminator interaction, whereas mutating phage 82 +15-17GCC has mostly a template strand effect and cannot be due to this interaction. We next mutated phage 82 +15-17 to the λ sequence GGG, a consensus expected to bind σ70 region 1.2 strongly, to determine if we could replicate the non-template strand effect seen on λ pR'. Mutating the discriminator sequence to consensus does not greatly affect pause capture, but increases the pause duration significantly (Figure 2.5A, B). Transcription of the heteroduplex templates showed that when the discriminator-like sequence is changed to the consensus, the non-template strand contributes more significantly to the pause effect, similar to what was previously seen in λ pR' This effect is indicative of an increased affinity for the interaction of σ70 region 1.2 with the sequence. However, both strands are contributing to the increase in pause duration. Mutating just the template strand also increases the half-life of the pause suggesting that the CCC sequence at the 5' end of the DNA/RNA hybrid is still playing a role in pause formation.

Changing residue 102 of σ70 from methionine to alanine in region 1.2 disrupts the discriminator interaction of the protein (Haugen et al., 2008). We performed *in vitro* transcription of both our wild type 82 *pR'* and +15-17 GGG mutant templates using either wild type σ70 or this mutant to determine if the σ70 region

1.2/discriminator-like element interaction is necessary for pausing on either the wild type or mutant template (Figure 2.5C). The σ70 region 1.2 mutation reduces the pause capture of both templates about ten fold while having little to no effect on pause half-life. The susceptibility of both templates to the M102A mutation shows that the σ70 region 1.2/discriminator-like element interaction still makes an important contribution to pause formation on phage 82 *pR'*, even though the sequence is different from consensus. Transcription with σ70 M102A RNAP of the three single base mutants +15A, +16A and +17A used in the experiment of Figure 1 reduces pause capture in a similar manner (Figure 2.6; data not shown).

Disrupting the σ70 region 2/-10-like element has an almost identical effect on pausing on both wild type and consensus discriminator-like sequence templates as disrupting the discriminator interaction (Figure 2.5D). Mutating residue 402 from leucine to phenylalanine in region 2 of σ70 weakens the protein's interaction with -10 promoter elements and also with our pause inducing -10-like element (Hatoum and Roberts, 2008; Ko et al., 1998). Little to no pause remains if the L402F mutant is used instead of wild type σ70. This is consistent with the importance of the -10-like element in inducing the pause. The fact that the σ70 region 1.2 mutation M102A has the same effect implies that the discriminator interaction also plays an important role in the initial σ70/DNA interaction necessary to induce the pause on 82 pR'. This is true despite the fact that the 82 pR' has a C-rich non-consensus discriminator that is less than optimal. Changes to the DNA sequence of the pause discriminator region affect both pause duration as evidenced by the consensus mutant data, and also pause

Figure 2.5. A consensus discriminator-like sequence reveals a role for σ70 region 1.2 interaction in pause formation. (A) Sequences of WT and consensus discriminator-like sequence mutant. The discriminator-like sequence is marked in green; the AT-rich sequence in blue; the -10-like element is underlined; Mutations are marked in red. (B), (C), (D) and (E) Band intensities of +25/26 pauses and terminated product from single round *in vitro* transcription time-courses were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; homoduplex mutant is red circles; non-template strand mutant heteroduplex is blue triangles; template strand mutant heteroduplex is green squares. Dashed lines are pausing during transcription using the indicated σ70 mutant (C and D) or in the presence of GreB (E).


A
Discriminator-like sequence
-10 like
+25 pause
WT
AGATCTGATATTTTGCCAATGTTGTA
+15-17GGG
AGATCTGATATTTTGggAATGTTGTA
+1 +5 +10 +15 +20 +25
B
C
D
E
Total Pause
Time (min)
1.00
0.10
0.01
0.00
0 2 4 6 8 10
+15-17GGG
NTS mut
TS mut
WT
+15-17GGG + 102A mut
WT + 102A mut
+15-17GGG + 402F mut
WT + 402F mut
+15-17GGG + GreB
WT + GreB

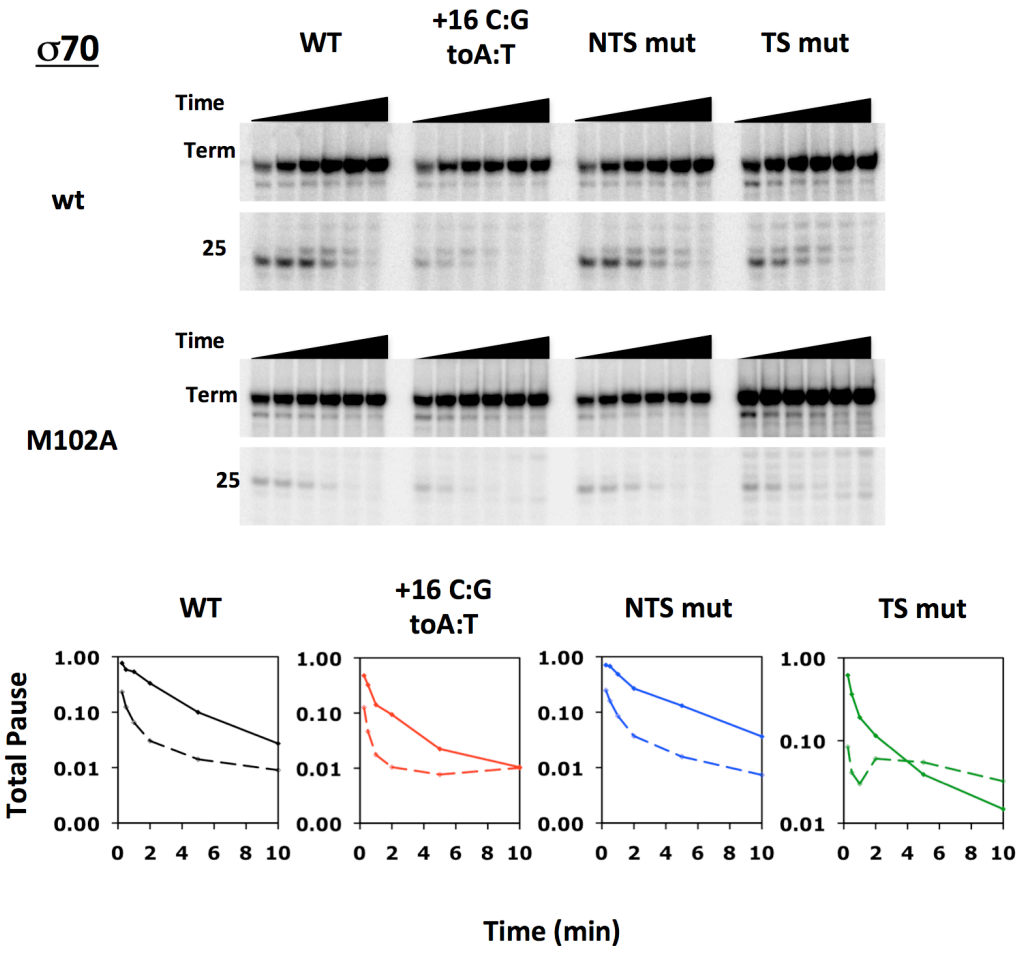


Figure 2.6. σ70 region 1.2 interaction is still important for pausing on deficient discriminator-like mutant templates. (A) Single round *in vitro* transcription time-courses on WT, +16C:G to A:T mutant homoduplex and constructed heteroduplex templates either with wt σ70 or σ70 M102A mutant. The +25 pauses and terminated product RNAs are labeled. (B) Band intensities of +25/26 pauses and terminated product were quantified and amount of pause seen was plotted as % of total RNA over time. Solid lines indicate transcription performed with wt σ70; dashed lines with σ70 M102A mutant.

capture as the +15-17AAA template strand mutant reduces the amount of pause seen (Figure 2.3, 2.4). However, changing the protein by disrupting the σ70 interaction only appears to affect pause capture. These observations suggest that pausing occurs in a two step mechanism, consistent with the scrunching model of σ70 dependent pausing. σ70 engagement with the DNA is the first step and then a second template strand sequence dependent process is required for pausing.

When transcription is performed on both the wild type and discriminator consensus templates in the presence of GreB, the amount of pause capture is not affected, but the rate that transcription complexes can escape the pause is increased. This suggests that a portion of the pause seen, +26 pause in the wild type case, but some of both +25 and +26 pause in the consensus mutant case, can more quickly escape the paused state (Figure 2.5E). It is particularly interesting that such a large portion of pause seen on the consensus template at both +25 and +26 appears to be GreB sensitive as it suggests some of the +25 pause in addition to the +26 pause is backtracked. Some the +25 pause is collapsing back into a backtracked state similar to the +26 state as a consequence of the strengthened σ70 interaction on the consensus template. This increased backtracking also implies that the odds of escaping sigma binding, which must occur for escape into downstream elongation, is reduced on the consensus template.

The discriminator-like element of the template strand determines the site of the pause

We have established that the GC-rich sequence of the 82 *pR'* promoter-associated pause directly downstream of the -10-like site of the pause is necessary for

σ70 interaction in the non-template strand, but the sequence of the template strand plays a larger role. To probe the function that this sequence plays in pause formation, we made three sets of sequence insertion mutant templates to determine the effect of displacing the GC-rich sequence. The first set of templates consisted of the addition of an A:T, AA:TT, or AAT:TTA sequence at position +9 of the *pR'* template. These insertions move all of the sequence necessary for the pause downstream. The second set of mutant templates had the same sequences inserted at position +15 of the template between the -10-like element and the discriminator-like sequence, and the third set inserted those sequences at +18, downstream of the discriminator-like sequence (Figure 2.7A). If this sequence were only acting as a nontemplate discriminator, it would not make sense to separate it from the -10-like element because they are part of the same recognition region, but because this sequence primarily acts through the template strand, the operation is reasonable.

As one would expect, inserting sequence at position +9 and thereby moving the entire pause inducing sequence downstream, moves the pause formed on the template 1, 2 or 3 bases downstream to positions +26, +27 or +28 depending on how many bases are inserted in the sequence (Figure 2.7B). Inserting sequence at +18 downstream of the discriminator-like sequence does not move the pause in a similar manner. In this case the primary pause continues to exist at +25, no matter how many bases are inserted, although the fine structure of the paused transcripts are modulated

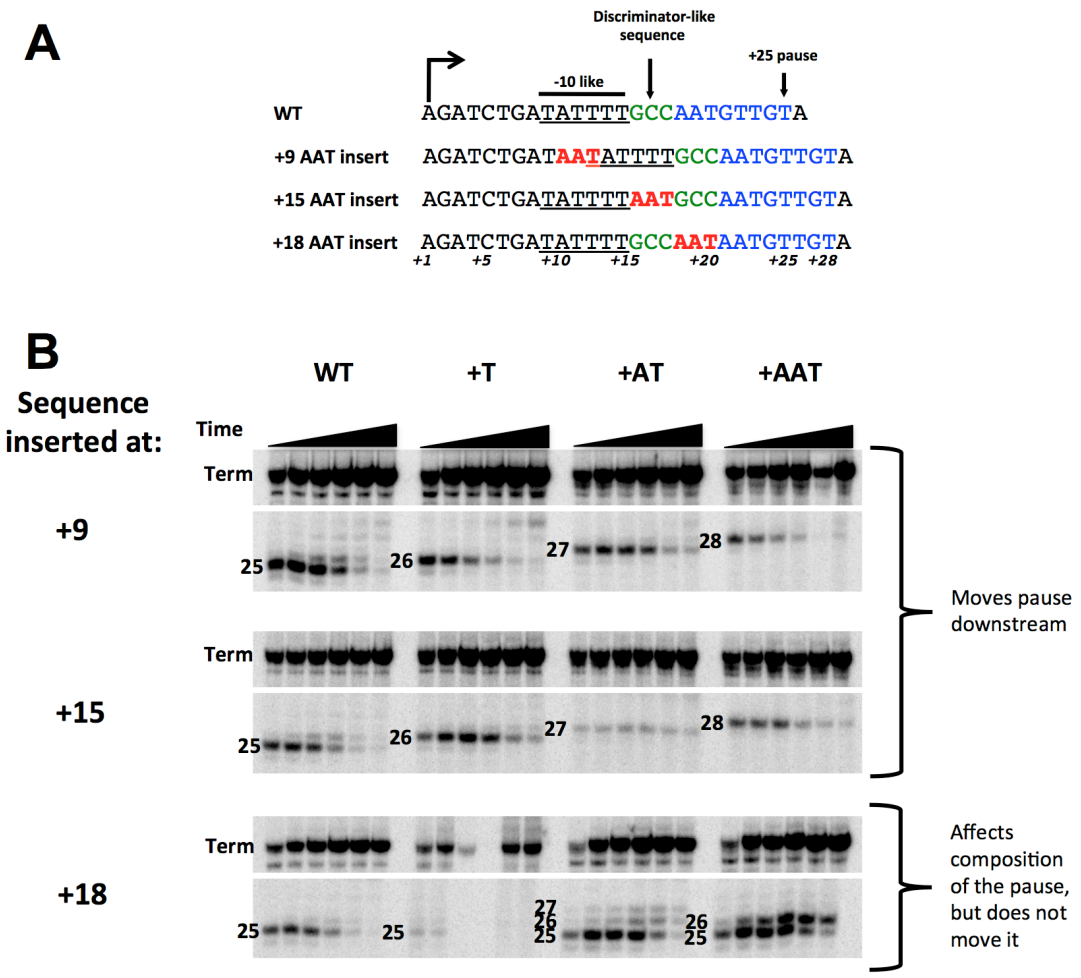


Figure 2.7. Moving the discriminator-like sequence moves the pause downstream. (A) Sequences of WT and insertion mutant templates. The discriminator-like sequence is marked in green; the AT-rich sequence in blue; the -10-like element is underlined; base insertions are marked in red. (B) Single round *in vitro* transcription time-courses run on A, AA and AAT sequence insertions at positions +9, +15 and +18 as indicated in (A). Relevant pauses and terminated bands are as labeled.

by this sequence change at the actual site of the pause. This result shows that the AT-rich sequence, which we previously knew affected the amount of pause formed, does not determine the site of the pause by itself (Perdue 2010).

The striking result is that inserting sequence at +15 between the -10-like element and the discriminator-like sequence has the same effect as the +9 insertions, moving the pause downstream 1,2 or 3 bases. This implies that the discriminator-like sequence either alone or in concert with the AT-rich sequence downstream (because this AT-rich sequence is also moved in the insertion mutants), determines the position of the pause. Also the positioning of the -10-like element, while necessary for inducing the pause event, is not important for the determination of the specific location of the pause. It is primarily the positioning of the discriminator-like sequence, with possibly some contribution by the AT-rich sequence following it that determines the specific position where a pause occurs. Based on the current model of promoter-proximal pause formation being homologous to scrunching during transcription initiation, it must follow that by inserting sequence between the -10-like element and the discriminator sequence we are increasing the amount of DNA that must scrunch before the complex pauses, as modeled in Figure 2.10.

Interestingly, inserting more sequence between the -10-like site and the discriminator-like sequence can move out the pause farther than 3 basepairs. Pausing still occurs when 5 or even 10 bases are inserted between the two sequences (insert sequences of AACAC or AACACAACAC respectively) (Figure 2.8). However, these new pauses are much less robust than the wild type pause. Transcription of the 5 base pair insert template shows pauses that form at positions +26, +27, +28, +29 and +30

(Figure 2.8). Most of these species are partially sensitive to the RNA cleavage factor GreB that relieves backtracked transcription complexes, although the +30 pause is relatively unaffected, suggesting that there is stable population of paused complexes at +30. A paused species is also seen at position +35 on the 10 base insert template when treated with GreB, implying that scrunching of the pause can be increased at least 10 basepairs. The small amount of pausing seen at positions leading up to the final paused position may be due to the complexes being able to stably pause at these positions due to the similarity of insert sequence to a necessary pause sequence and the difficulty of a complex scrunching for such a large distance.

Pauses that have been extended downstream by moving the discriminator-like element are still dependent on the σ70/-10-like site interaction

Due to disruption of the distance between important elements for pausing, it could be suggested that σ70 region 2 is not interacting with the canonical -10-like element on the +15 insert templates. These changes could move all of the necessary portions of the pause inducing sequence downstream or eliminate the need for the -10-like site, explaining the similarity between the pauses seen on +9 insert and +15 insert templates. To address this we looked at the σ70 region 2/-10-like element interaction during transcription on our +15 AAT insert template (Figure 2.9). Mutating residue 402 from leucine to phenylalanine in region 2 of σ70 weakens the protein's interaction with -10 promoter elements and also our pause inducing -10-like element (Ko et al., 1998). By performing transcription using core RNAP and mutant σ70 L402F we nearly eliminated pausing as compared to transcription using core RNAP plus wild

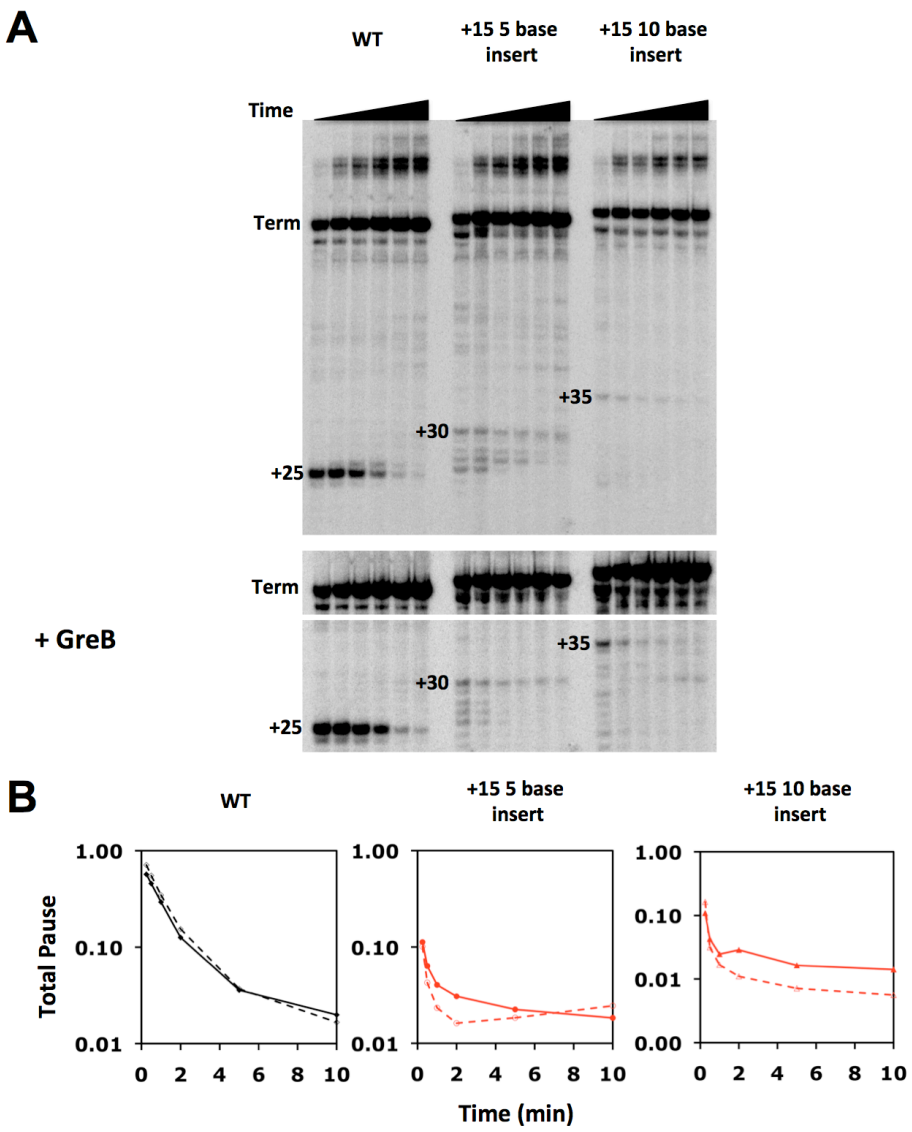


Figure 2.8. The pause can be moved at least 10 bases downstream by moving the discriminator-like sequence. (A) Single round *in vitro* transcription time-courses run on position +15 AATGT and AATGTAATGT insert templates in the absence or presence of GreB. Relevant pauses and terminated bands are as labeled. (B) Band intensities of +25/26 pauses and terminated product were quantified and amount of pause seen was plotted as % of total RNA over time. Solid lines indicate pausing without GreB; dashed lines indicate pausing in the presence of GreB.

type σ70 on both a wild type 82 *pR'* template and the +15AAT insertion template showing that an σ70 region 2/DNA interaction is necessary for pause formation on both templates (Figure 2.9B).

A possible consequence of the insertion sequence used on the 3 base insert mutant in these experiments is that a new -10-like sequence has been constructed starting three base pairs downstream of the first. This is due to the sequence being inserted, AAT, resembling the second half of a -10 like element including the necessary T at the final position. To directly test this we disrupted the -10-like sequence of both the wild type and +15AAT insert template DNAs, making the following sequence changes: +8 T:A to G:C, +10 A:T to G:C, +12 T:A to G:C and +14 T:A to G:C to address whether the known -10-like site was still being used on the +15 insert template (Figure 2.9A). On both the wild type and +15 insert templates the mutation of the -10-like element nearly eliminates the pause, similar to the effect of σ70 L402F on pausing (Figure 2.9C). These two results taken together show that σ70 region 2 must interact with the known -10-like element in order for pausing to occur on templates where the pause site has been displaced downstream. This supports our model that inserting sequence between the -10-like element and discriminator-like element increases the size of the scrunched DNA bubble in a transcription complex in the paused state.

The role of the template strand scrunched DNA in promoter-proximal pausing

Other sequences in addition to those necessary for σ70 interaction and those that take part in the RNA/DNA hybrid may play important roles in stabilizing paused

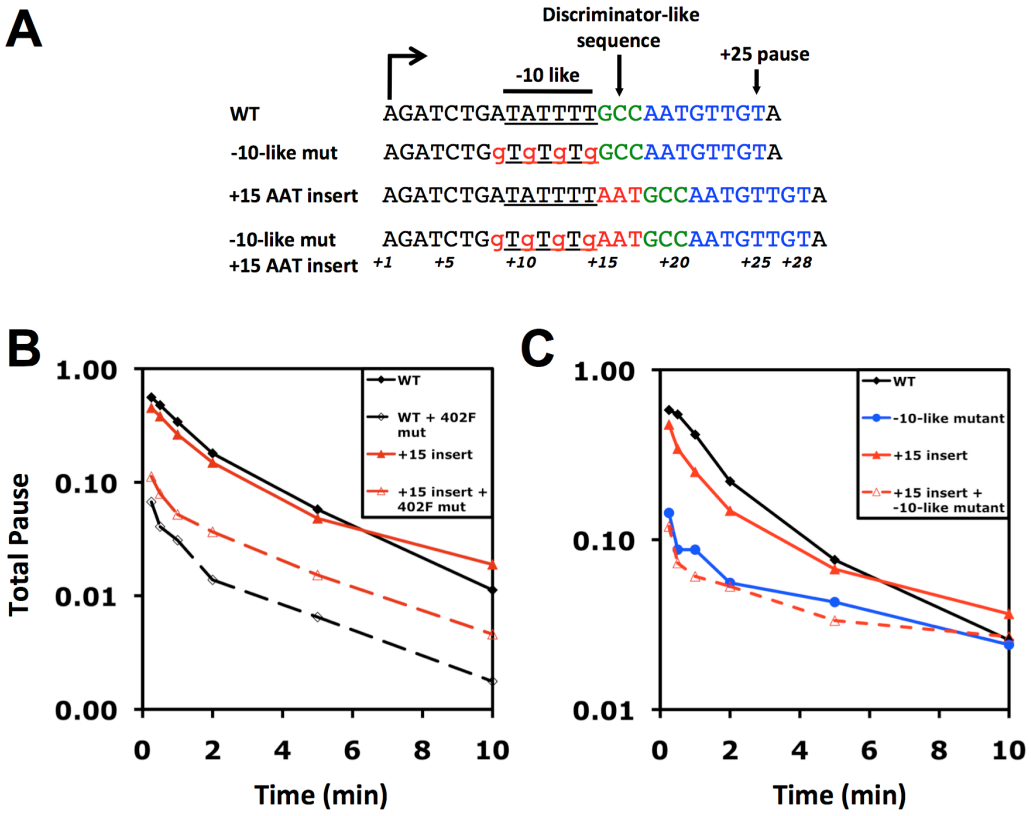


Figure 2.9. Pauses that have been extended downstream still require a σ70/-10-like element interaction. (A) Sequences of WT and mutant templates. The discriminator-like sequence is marked in green; The AT-rich sequence in blue; The -10-like element is underlined; Mutations are marked in red. (B) and (C) Band intensities of +25/26 pauses and terminated product from single round *in vitro* transcription time-courses were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; +15 AAT insert mutant is red triangles; -10-like Mutant is blue circles; dashed lines are pausing during transcription using σ70 402F mutant (B) or +15 AAT insert template with the -10-like mutant (C).

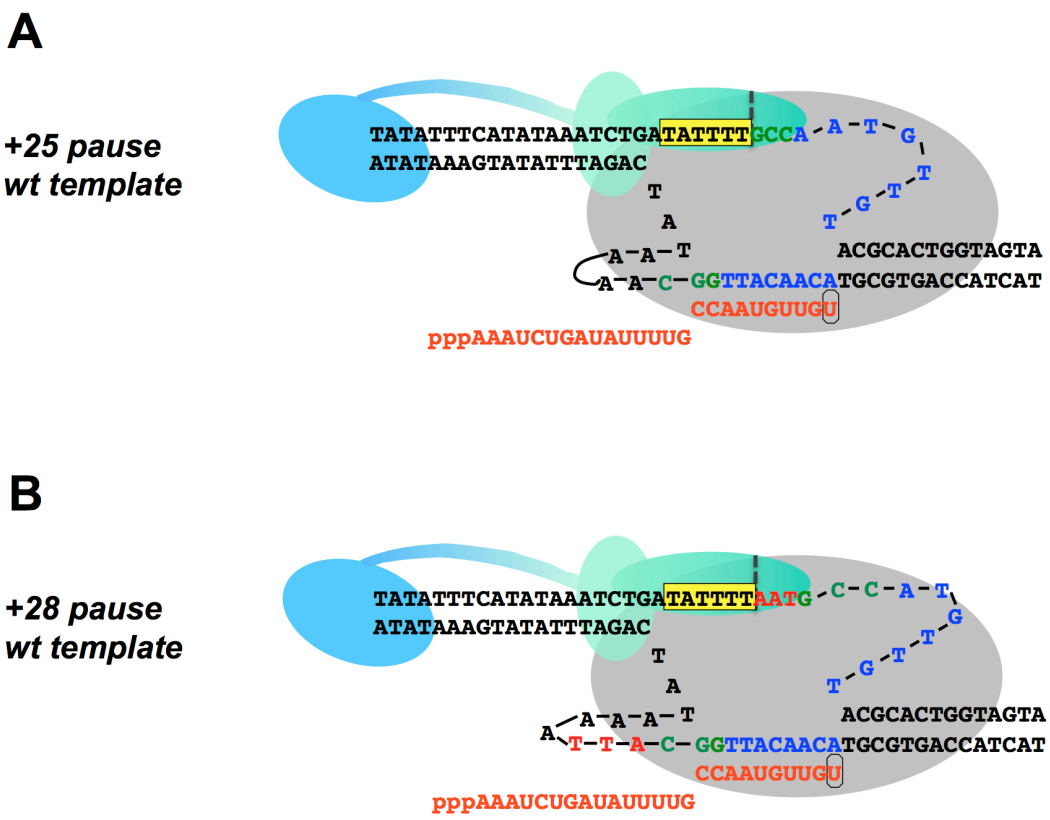


Figure 2.10. Models of the paused states of WT and insert template promoter proximal pauses. (A) Model of a transcription complex at the +25 promoter-proximal pause on phage 82 *pR'*. (B) The insert sequence moves the TempGC sequence downstream, thus moves the site of the pause an equivalent distance downstream. σ70 (light blue/green) is still interacting with the -10-like site requiring the complex to create a larger scrunched bubble to reach the pause site.

transcription complexes. One particular question we wanted to answer is what role the DNA in the transcription bubble that has been scrunched plays in the pause, specifically DNA scrunched in the template strand. To answer this we made templates with single base mutations in positions +12 to +14, changing T:A base pairs to G:C and a triple mutant +12-14 TTT:AAA to GGG:CCC (Figure 2.11A).

Not surprisingly, mutating position +14 from T:A to G:C reduces the amount of pause seen significantly from the wild type (Figure 2.11B, C); +14 corresponds to the last and highly conserved base of the -10 promoter element in which mutation strongly reduce promoter function (Feklistov and Darst, 2011; Liu et al., 2011; Ring and Roberts, 1994). Mutating the equivalent base of λ pR' (+6) eliminates pausing on that promoter. Thus, this mutation disrupts the σ70 region 2 interaction. Interestingly, mutating positions +12 and +13 changes the character of the pause in a manner different than the +14 mutant. In both cases, pause capture is not affected, but the rate of pause escape is greatly reduced (Figure 2.11B, C). Surprisingly, if all three positions are mutated at once, pause capture is only slightly reduced but the rate of escape is like the +12 or +13 individual mutants. This means that mutating positions +12 and +13 has a significantly different effect than the +14 mutation.

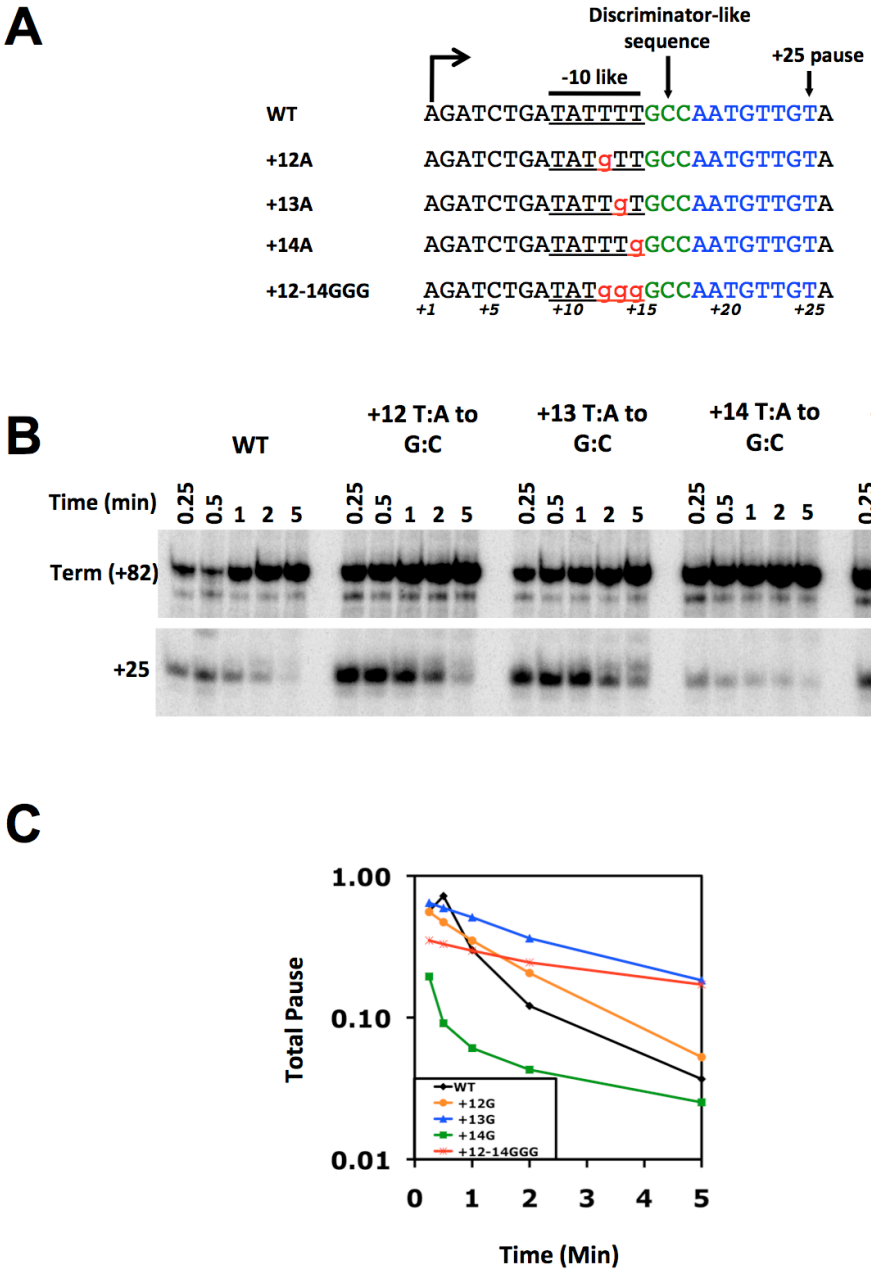
We next performed transcription on heteroduplex templates of the +12-14 triple mutant (Figure10D). The mutant effect seen in the triple mutant is due to the mutation to the template strand. The non-template strand mutant showed greatly reduced pausing that corresponds with the +14 mutation in figure10B and C. This makes sense because σ70 region 2 interacts with the non-template strand of the -10-like element (Feklistov and Darst, 2011; Liu et al., 2011; Marr and Roberts, 1997). So

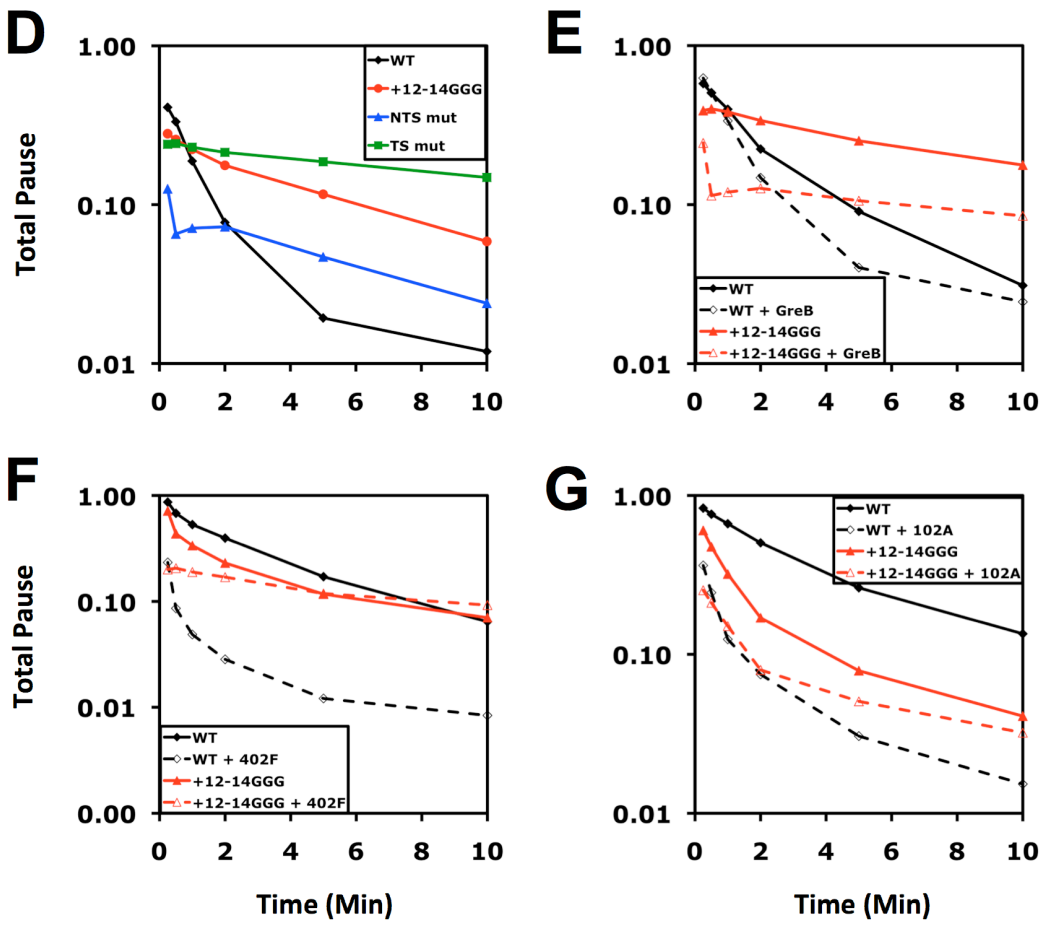
the triple mutant in the non-template strand negatively effects pause formation. The ability of the template strand mutation to overcome that effect is even more evident when transcription is performed using the σ70 L402F mutant. The region 2 mutant σ70 has a modest effect on transcription of the +12-14GGG template as compared to the effect on wild type template, which has a strong effect on capture but nothing on halflife (Figure 2.11F). Presumably this effect is due to some unanticipated but interesting effect of the template G's on halflife. This means at least one of two things: one that mutating +12-14GGG in the non-template strand is not sufficient for completely disrupting the σ70 region 2/-10-like element interaction, which is unlikely from all previous mutant results, or two, that the triple mutant in the template strand must be overcoming the weakening of the σ70 region 2/-10-like site caused by mutating +14 in the template strand by inducing a different type of pause. The pause seen during transcription on the +12-14GGG template is also more GreB sensitive than the wild type pause most likely due to more of the pause being backtracked at position +26(Figure 2.11E).

Transcription of the mutant +12-14 template using the discriminator interaction disrupting σ70 M102A mutant reduces the pause to the same level that is seen when using the mutant on a wild type template (Figure 2.11G). This contrasts with the result seen with σ70 L402F. Even though the +12-14GGG mutation increases the size of the GC-rich sequence from the original 3 bp of the discriminator-like element to 6 bases long it does not affect DNA binding by σ70 region 1.2.

Figure 2.11. Extending the GC-rich sequence effects pause duration.

(A) Sequences of WT and mutant tempates. The discriminator-like sequence is marked in green; The AT-rich sequence in blue; The -10-like element is underlined; Mutations are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A). The +25 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; +12T:A to G:C is yellow circles; +13T:A to G:C is green squares; +14T:A to G:C is blue triangles; +12-14TTT:AAA to GGG:CCC is red Xs. (D), (E), (F) and (G) Band intensities of +25/26 pauses and terminated product from single round *in vitro* transcription time-courses were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; homoduplex mutant is red circles; non-template strand mutant heteroduplex is blue triangles; template strand mutant heteroduplex is green squares. Dashed lines are pausing during transcription in the presence of GreB (E) or using the indicated σ70 mutant (F and G).


A
Discriminator-like sequence
-10 like
+25 pause
WT AGATCTGATATTTTGCCAATGTTGTA
+12A AGATCTGATATgTTGCCAATGTTGTA
+13A AGATCTGATATTgTGCCAATGTTGTA
+14A AGATCTGATATTTgGCCAATGTTGTA
+12-14GGG AGATCTGATATgggGCCAATGTTGTA
+1 +5 +10 +15 +20 +25
B
WT
+12 T:A to G:C
+13 T:A to G:C
+14 T:A to G:C
+12-14 TTT to GGG
Time (min)
0.25 0.5 1 2 5
Term (+82)
+25
C
Total Pause
1.00
0.10
0.01
WT
+12G
+13G
+14G
+12-14GGG
0 1 2 3 4 5
Time (Min)


D
E
F
G
Total Pause
Time (Min)
1.00
0.10
0.01
0.00
0
2
4
6
8
10
WT
+12-14GGG
NTS mut
TS mut
WT + GreB
+12-14GGG + GreB
WT + 402F
+12-14GGG + 402F
WT + 102A
+12-14GGG + 102A

Discussion

A novel GC-rich sequence is necessary for determining the site of pausing

We have presented evidence of a novel sequence necessary for pause formation. It has long been understood that a σ70 region 2/DNA interaction with the -10-like site is necessary for promoter-proximal pausing on lambdoid phage promoter *pR'*, but what has not been known is the role the surrounding sequence plays in pause formation (Perdue and Roberts, 2010, 2011; Ring et al., 1996). Many different pauses including these lambdoid phage promoter-proximal pauses contain a 3-4 basepair GC-rich sequence followed by an AT-rich sequence that comprise the sequence in the DNA bubble and the RNA/DNA hybrid in the transcription complex (Perdue and Roberts, 2011). Here we have shown that this GC-rich sequence, which greatly resembles the promoter element called the discriminator, is necessary for promoter proximal pausing at the phage 82 *pR'* promoter. This sequence is responsible for determining the specific site of the pause and primarily acts through a template strand effect. We have decided to name this sequence TempGC for template strand GC-rich sequence.

Inserting sequence between the -10-like element and the TempGC in phage 82 pR' transcription templates moves the pause seen downstream an equivalent number of bases (Figures 2.7 and 2.8). This allows us to conclude that the TempGC either alone or in conjunction with the AT-rich sequence that makes up the RNA/DNA hybrid at the site of the pause is responsible for determining the specific base at which the pause occurs. This even holds true when a large sequence of 10 base pairs is inserted between the two elements; the pause is weak, but its resistance to the cleavage

factor GreB indicates that it is an authentic, forward, scrunched paused complex (Figure 2.5).

Pausing on the +15 insertion template still requires an interaction of σ70 with the -10-like element even though the TempGC and the pause are displaced downstream from this site (Figure 2.9). This implies that a transcription complex at the pause is still anchored by the -10-like element at positions +9 to +14 but is extended further downstream. This result is consistent with our pausing model whereby a paused transcription complex is anchored by the σ70/DNA interaction and must scrunch to reach the site of the pause in an analogous mechanism to that seen during abortive initiation (Perdue and Roberts, 2010, 2011).

Abortive products from initiation have been reported as long as 15 to 19 bases (Chander et al., 2007; Hsu et al., 1995). This means that a transcription complex trying to escape the promoter would have to scrunch at least 11 bases to in order to synthesis an abortive product of 19 nt in length. A complex at the promoter-proximal pause on phage 82 *pR'* contains a scrunch of at least 4 base pairs. By moving the TempGC 10 basepairs downstream I have moved the GreB insensitive pause the same distance downstream, resulting in a scrunch of 14 bases, which is consistent with the size of scrunches at the promoter necessary to form longer abortive products (Figure 2.10).

A model for the function of the TempGC in pausing

The nature of the template strand effect of the GC rich region on phage 82 *pR'* promoter-proximal pausing is unknown, although a strong possibility is that the GC-richness of the 5' end of the RNA/DNA hybrid has a stabilizing effect on the

transcription complex. Hybrid lengths of 9 base pairs to 10 base pairs have been reported for different RNAP complexes (Kyzer et al., 2007; Vassylyev et al., 2007a). Most likely the hybrid length is 10 base pairs after addition of a 3' terminal nucleotide and 9 base pairs in a post-translocated complex that has moved forward to clear the active site for the addition of the next nucleotide to the end of the growing chain, as reported for the *his* operon promoter-proximal pause (Kyzer et al., 2007). This would mean that positions +16 and +17 of 82 *pR'* are the 5' terminal nucleotides of the hybrid in a post-translocated complex. +15 may also be part of the hybrid if it is in fact longer, but experiments to determine the exact length of the hybrid have not done.

It has been suggested that a GC-rich segment in this region could make the unwinding of the 5' end of the hybrid more difficult and thus contribute to pause formation (Gilbert et al., 1974). This hypothesis for function of these residues is further supported by the findings of Herbert et al. (2006) who found a strong G:C base pair bias for positions -10 and -11 from the site of pauses for a number of pauses including *ops* and *his* pauses. These pauses all are caused by different mechanisms from the σ70-dependent lambdoid phage promoter-proximal pauses, but still have TempGC-like sequences either prior to or as the 5' end of the RNA/DNA hybrid.

This model could explain the effect of mutating this sequence to A:T base pairs. A distinct although unlikely model is that these nucleotides are making contact with the protein at the back edge of the hybrid-binding cleft and thus acting through a protein binding mechanism. Our data supports Gilbert's hypothesis in that individual mutations of positions +15 to +17 from a G:C or C:G base pair to A:T have similar slight effects on the duration of pausing. Thus, disrupting one of the three G:C

basepairs is marginally weakening the unwinding effect, whereas mutating all three to A:T base pairs allows the hybrid to unwind much more freely and drastically lowers the amount of pause seen.

A σ70 region 1.2/DNA interaction is still necessary for pause formation

The strong effect of mutating the TempGC sequence is more evident on templates containing phage 82 *pR'* than templates containing λ *pR'*. On λ *pR'*, this sequence is GGG as opposed to GCC for the phage 82 sequence. GGG is the consensus sequence for σ70- region 1.2/discriminator interaction and transcription experiments on templates containing λ *pR'* show that mutating this sequence has a large non-template strand effect in heteroduplex template analysis. The discriminator interaction is the primary role of this GC-rich sequence on λ *pR'* as shown previously in heteroduplex template transcription experiments performed by Ring and Roberts (1994). They found that the majority of the reduction in pause seen by mutating the discriminator-like region was due to the change in the non-template strand. There is, however, a slight effect of mutating the template strand, which presumably corresponds to the much stronger effect on phage 82 *pR'* templates. It is also worth noting that Ring performed these experiments before the role of σ70 binding was understood to cause the promoter-proximal pause.

Changing the discriminator-like sequence of phage 82 *pR'* to GGG increases the duration of promoter-proximal pausing on the template and also causes an increased non-template strand effect, as seen by heteroduplex analysis (figure 2.3B). This implies that a consensus discriminator-like sequence is able to better interact with

σ70 region 1.2 than the wild type sequence, partially explaining the increased duration of pausing. However, the template strand consensus mutant also showed increased pause duration as compared to the wild type template, so that there is a contribution to pausing from both strands.

The σ70 region 1.2/DNA interaction on wild type phage 82 *pR'* template is clearly important, however. Pausing on wild type, pause deficient mutants and consensus discriminator templates are all decreased when a σ70 region 1.2 mutant is used during transcription (Figure 2.3C, Figure 2.6). This implies that σ70 region 1.2 is still necessary for pause formation even when a less than ideal binding site is present in the DNA. Most likely σ70 region 1.2 can interact with any single-stranded DNA presented to it, but prefers the consensus GGG sequence to other sequences.

σ70 dependent promoter-proximal pausing

Our findings that the TempGC plays a major role in pause site determination, and that the σ70 region 1.2 interaction with the non-template strand of the transcription bubble also is important, help us refine the model for promoter-proximal pausing on the lambdoid phage *pR'* late gene promoter. First, transcription is initiated at *pR'* by the recruitment of core RNAP by σ70 to the promoter start site and the formation of open complex. The complex escapes the promoter, and in the case of phage 82 *pR'* is able to synthesize either a 20 or 21 base RNA before σ70, which is still part of the transcription complex, forms a region 2/-10-like DNA element interaction (bases +10-14 of the promoter). Further contacts are made by σ70 region

1.2 with the discriminator-like region (+15-17) and σ70 region 3 with the extended -10 (if present), essentially extending and strengthening the interaction formed by σ70 region 2 and the -10-like element. Although the order that these contacts occur in is not known. These interactions occur with the non-template strand of the transcription bubble and prevent the forward translocation of the transcription complex, although not RNA synthesis. Interestingly, the σ70 region 1.2 DNA interaction occurs irrespective of the presence of a strong discriminator-like sequence.

The σ70/DNA interaction initiates a process analogous to that of scrunching during initial transcript formation at the promoter. Because the back end of the complex is anchored by the σ70, in order to continue transcription the transcription bubble is extended in size by movement of downstream DNA into the enzyme cleft, scrunching the newly unwound DNA. This process continues until the complex encounters the TempGC sequence followed by an AT-rich sequence that acts through the RNA/DNA hybrid to stabilize the complex and cause the pause at position +25. The pause persists until one of two events occur: either the complex overcomes the pause by itself or Q protein is loaded into the transcription complex after interacting with its DNA binding site, also releasing the complex from the pause. The pause will then either continue transcribing or become backtracked at position +26 due to the backtrack-stabilizing effect of the TempGC and AT-rich sequence in the hybrid.

CHAPTER 3

THE ROLE OF THE TERMINAL AT-RICH SEQUENCE IN PHAGE 82 σ70-DEPENDENT PAUSING

Introduction

Transcriptional pausing occurs through a variety of different and variably characterized mechanisms, including: 1) protein/DNA interactions such as the σ70 dependent promoter-proximal pauses discussed here; 2) emerging RNA secondary structure, in "hairpin" –dependent pausing; 3) elongation complex backtracking at certain specific sequence elements; and 4) ubiquitous pausing, a less defined category which also is DNA sequence dependent (Adelman et al., 2002; Artsimovitch and Landick, 2000; Herbert et al., 2006; Komissarova and Kashlev, 1997b; Neuman et al., 2003; Ring et al., 1996). Interestingly, while these different forms of pausing are instigated by different mechanisms, they tend to have in common a short GC-rich region at the back of the RNA/DNA hybrid at the pause, followed by a largely AT-rich region (Herbert et al., 2006; Perdue and Roberts, 2011). Sequences with this general characteristic of a relatively GC-rich (or hybrid-stabilizing) region followed by an AT-rich (or hybrid-destabilizing) region are generally considered to provide the essential structure for backtracking, but also, as we further describe here, appear to play an important role in non-backtracked pauses (Artsimovitch and Landick, 2000; Herbert et al., 2006; Perdue and Roberts, 2010).

The existence of a universal transcriptional pausing sequence is a very intriguing idea considering the number of different mechanisms by which pausing can be initiated. It implies that although all of these different pauses are induced or stabilized by different mechanisms such as σ70/DNA engagement or RNA secondary structure, they fundamentally make the same or similar changes to the transcription elongation complex to induce a pause. This structural rearrangement of the RNA polymerase has been termed the elemental pause, and it is postulated to inhibit the addition of new nucleotides to the nascent RNA chain until the polymerase either by itself or through the mediation of a cofactor is able to reverse the rearrangement and escape the pause (Landick, 2006, 2009).

Transcriptional pause mechanisms have been separated into two classes: 1) Backtrack inducing pauses and 2) non-backtracking pauses where the pause state is the result of some other conformation change of the complex in which pausing is caused by two distinctly different mechanisms (Artsimovitch and Landick, 2000). The structural change that constitutes the paused state of backtracking pauses is easy to understand. The RNA/DNA hybrid is too unstable at the site of the pause and in order to deal with the instability the complex reverse translocates to a more stable position (Nudler et al., 1997). As it does so it leaves the 3' end of the RNA chain overhanging from the RNA/DNA hybrid. This overhang causes the elongation complex to stall or pause and requires the removal of the overhang or forward translocation of the complex to reverse the backtrack in order for transcription to resume (Borukhov et al., 1992; Komissarova and Kashlev, 1997a; Nudler et al., 1997; Orlova et al., 1995; Park et al., 2002). The changes to the elongation complex that constitute the paused state of

non-backtracking pauses are not understood yet. However, it is interesting that the same sequence that can cause a backtracked pause when present by itself, can act with a secondary pause-inducing sequence to create a different non-backtracked transcriptional pause as evidenced by the work of Perdue and Roberts (2010), which I discuss in more detail in chapter 2.

The σ70 dependent promoter-proximal pause seen on lambdoid phage late gene promoters and also some endogenous *E. coli* gene promoters is particularly interesting as it contains both non-backtracked and backtracked states (Marr and Roberts, 2000; Perdue and Roberts, 2010). As previously described, this pause is formed when σ70, which does not immediately leave the RNAP complex after initiation, recognizes a repeat of the -10 promoter element downstream of the start site. σ70 interacts with this -10-like element and induces a pause at a sequence composed of 3-4 basepairs of GC-rich sequence followed by an AT-rich stretch (Perdue and Roberts, 2011; Ring et al., 1996). These pauses exist in two states *in vitro*: the first is believed to be in a "scrunched" conformation similar to the process of abortive initiation; and the second is a backtracked state that occurs if and when the complex synthesizes one nucleotide further and collapses the scrunched DNA bubble (Marr and Roberts, 2000; Perdue and Roberts, 2011). *In vivo*, the Gre proteins are present, and the predominant state should be the first as evidenced by small RNAs sequencing experiments to identify promoter proximal pauses (Liu and Roberts, Unpublished data).

The role of this general pausing sequence, containing a short GC-rich region followed by an AT-rich region, in σ70 dependent promoter-proximal pausing was

only recently identified from experiments on the phage 82 *pR'* late gene promoter. Perdue and Roberts (2010) noted that mutating the AT-rich region of the pause sequence, corresponding to the 3' end of the RNA/DNA hybrid, greatly reduced the amount of pausing seen during *in vitro* transcription experiments. This mutation, in fact, was nearly as strong as eliminating the -10-like element that is believed to be the major element required for the pause. They further identified this pausing sequence in a number of other σ70 dependent promoter-proximal pauses including phage λ *pR'* and the promoter of *lacZ* (Perdue and Roberts, 2010).

In the previous chapter, I discussed the function of the GC-rich element and the role it plays both as an extension of the σ70 region 2/-10-like interaction of the non-template strand of the DNA bubble by interacting with σ70 region 1.2 and a novel template strand function that determines the site of the pause. In this chapter, I describe experiments designed to answer the question of how this pausing sequence acts in σ70 dependent promoter-proximal pausing by exploring the role of AT-rich region of the pausing sequence. By perturbing the AT content of the region and looking at the effect on pausing during *in vitro* transcription reactions, I will show that this sequence has a strong effect on the lifetime of the pause and also on the efficiency of capture of elongation complexes in the secondary backtracking pause that occurs subsequent to escape from the σ70 dependent pause.

One question that still remains about elemental pauses is the exact mechanism by which the paused state forms. RNA chain elongation is thought to occur through a four step process that involves a ratcheting mechanism with the active site of RNAP existing in one of two states: a pretranslocated state in which the previous nucleotide

has been added to the chain, but the complex has yet to translocate forward to allow addition of the next nucleotide; and a post-translocated state, in which translocation has occurred and the active site of the enzyme is poised to receive the next nucleotide and catalyze the addition of it to the growing RNA (Abbondanzieri et al., 2005; Gong et al., 2005; Landick, 2006; Temiakov et al., 2005; Tuske et al., 2005). There is substantial evidence that pausing occurs when the complex is in the pre-translocated register (Hein et al., 2011; Landick, 2006, 2009). Here I intend to show that σ70 dependent pauses also occur when the complex is in a pre-translocated state as is true for other pauses using a series of mutant RNAPs deficient for interaction with a specific part of the non-template strand of the transcription bubble.

Materials and Methods

Plasmids and DNA templates: Linear DNA templates for transcription were amplified from plasmids by PCR and gel purified using a gel extraction kit from Qiagen. All of the DNA templates were created from derivatives of the p82a+2G -35-like mutant plasmid by Agilent quickchange mutagenesis (Guo and Roberts, 1990; Perdue and Roberts, 2010). Heteroduplex templates were constructed as described in Ring et al. 1996.

Proteins: Core RNAP protein was the gift of M. Suh (Cornell University, Ithaca, NY). β-pocket mutant Core RNAPs (R151A, D446A, R451A and W183A) were a gift of Richard Ebright (Rutgers University, Piscataway, NJ). The following proteins were

purified as described: 6xHis-σ70, M102A, and 402F mutants (Marr and Roberts, 1997), GreB (Borukhov and Goldfarb, 1996).

In vitro transcription: Reaction mixtures containing 2 nM DNA template, 10 nM HoloRNAP (10 nM CoreRNAP reconstituted with 50 nM WT or mutant σ70), plus 200 mM each of ATP, GTP and CTP, 50 mM UTP (supplemented with 0.5 mCi/ml [α-^{32}P]-UTP) in Transcription buffer (20 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 10% glycerol) were incubated at 37^{o}C for 10 min to form open complexes. For Figure 3.3E, 100 nM GreB was added to the reaction mix. A single round of transcription was initiated by the addition of Start mix (50 mM $MgCl_2$ and 10 mg/ml Rifampacin) and reactions were allowed to proceed at 37^{o}C.Alliquots were taken at indicated times after addition of start mix and mixed with 5 volumes Stop solution (600 mM Tris-HCl (pH 8.0), 12 mM EDTA, and 0.16 mg/ml tRNA) on ice to stop transcription. RNA was extracted by mixing with a equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). 2.5 volumes of 100% Ethanol was added to the aqueous layer to precipitate samples overnight at -20^{o}C. Samples were run on 12% polyacrylimide gel containing 6M Urea. RNA species were detected using a Molecular Dynamics Typhoon 9400 and images were analyzed using GE Imagequant software. Pause efficiency was determined by first normalizing transcript bands based on the number of radiolabeled residues, then the adjusted intensity of the pause bands (+25 and +26) are divided by the total of all of the intensities present in a lane (paused bands + terminated band).

Results

Increasing the GC content of the RNA/DNA hybrid reveals a role for this sequence in pause formation

Phage 82 *pR'* promoter-proximal pausing requires the engagement of σ70 with a -10-like element to induce pausing, but beyond this interaction little was known about what other factors may be involved in the formation of the pause until recently (Ring et al., 1996). Perdue and Roberts (2010) showed that this -10-like element is not sufficient for pause formation by itself and that a sequence at the site of the pause, encompassing the RNA/DNA hybrid was also required. They demonstrated that this sequence was able to induce a backtracked pause at position +26 in the absence of a -10-like element, which lead them to call it the "backtrack inducing sequence." By making the 3' end of the RNA/DNA hybrid more GC-rich by mutating residues +22, +23, and +25, as I show again in Figure 3.1A, they were able to eliminate this backtracked pause in the -10-like element mutant and also greatly reduce the amount of non-backtracked +25 pause when a wild type -10-like element was present (Perdue and Roberts, 2010). From this they inferred that the AT-rich sequence, which can induce a backtrack by itself, was also necessary for formation of the primary pause at +25 through a distinct, but unknown process (Perdue and Roberts, 2010).

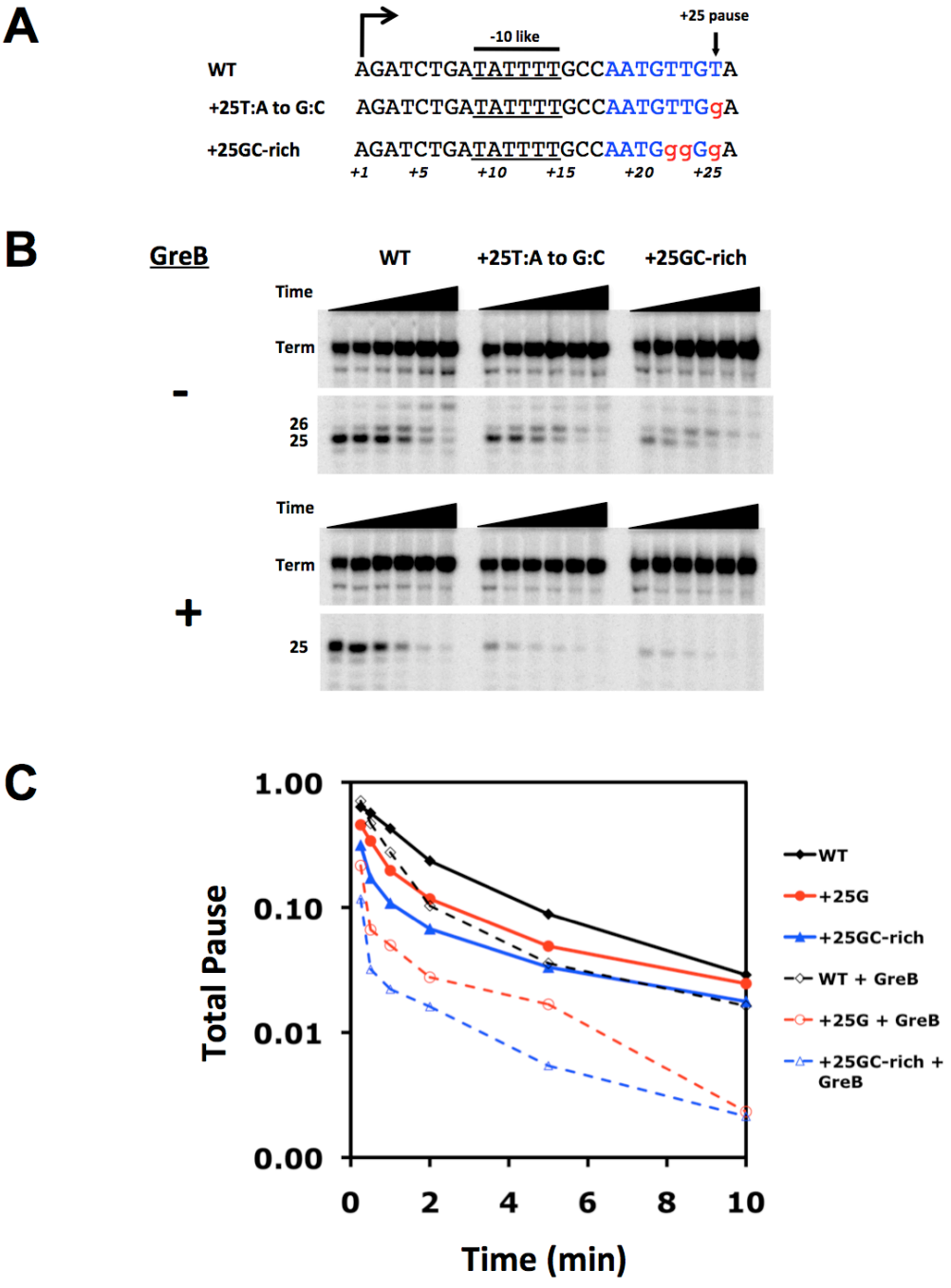
To further explore the role of the AT-rich sequence that makes up the majority of the RNA/DNA hybrid at the site of the pause, I performed *in vitro* transcription experiments on phage 82 *pR'* DNA templates with changes in individual bases of the AT-rich sequence, as well as the +25 GC-rich multiple mutant. The +25 GC-rich mutant drastically reduces the amount of pausing seen as compared to wild type,

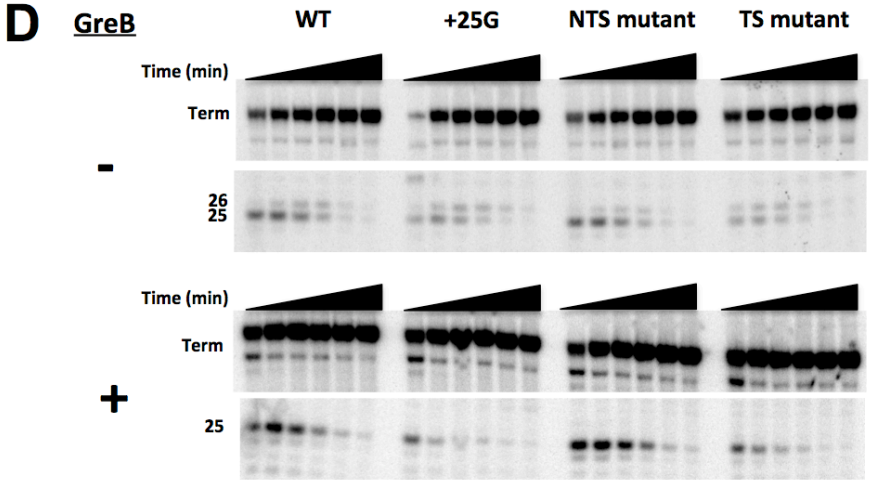
especially in the presence of the cleavage factor GreB, which eliminates the +26 backtracked pause, leaving only the putatively scrunched +25 pause (Figure 3.1B, C). Interestingly, mutating only the terminal nucleotide of the pause at position +25 from a T:A basepair to a G:C recreates much of the pausing defect seen in the +25GC-rich mutant (Figure 3.1A, B, C).

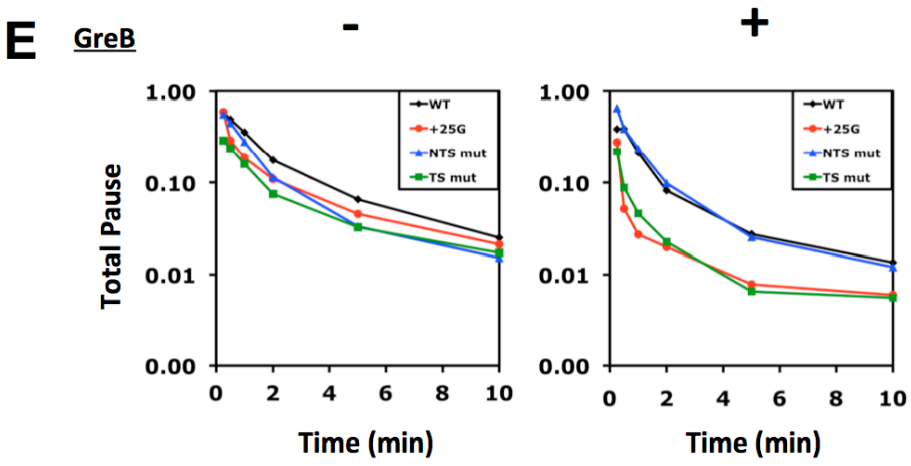
It is entirely possible that these mutations are reducing pausing through an unexpected mechanism, different from affecting the stability of the 3' end of the RNA/DNA hybrid as was assumed in Perdue and Roberts (2010); for example, they could reduce interaction of the non-template strand of the transcription bubble with the σ70, β or β' subunits of RNAP. To test this possibility, in vitro transcription was performed on homoduplex +25T:A to G:C template in addition to heteroduplex templates that were wild type in one strand and mutant in the other. In transcription experiments without GreB it is not clear which strand of the DNA is contributing more to the defect seen in pausing, although the data do suggest that the template strand effect is dominant. However, when GreB is added to the reactions, the template strand mutant heteroduplex clearly shows the same pause defect as the homoduplex mutant and the non-template strand heteroduplex looks like wild type template (Figure 3.1D, E). Transcription on heteroduplex templates for the +25GC-rich mutation also shows the same result (data not shown). This is consistent with the proposed mechanism of affecting the RNA/DNA hybrid.

One perplexing result of these experiments is that there still is a significant amount of +26 backtracked pause with both the +25GC-rich and +25T:A to G:C mutations. Perdue and Roberts (2010) noted that the +25GC-rich mutation was able to

Figure 3.1. A single G mutation to the terminal nucleotide of the pause reduces the amount of pause almost as much as the GC-rich mutant template. (A) Sequences of WT, mutant templates from the transcription start site to the site of the promoter-proximal pause. The -10-like element is underlined; transcription start site and the site of the pause are as marked; the AT-rich region is marked in blue; mutated bases are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A) either in the absence or presence of the cleavage factor GreB. The +25 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. WT pausing is black diamonds; +25T:A to G:C mutant template pause is red squares and +25GC-rich mutant template pause is blue triangles. Solid lines and markers indicate transcription in the absence of GreB; dashed lines and open markers, transcription with GreB. (D) Single round *in vitro* transcription time-courses on the WT, +25T:A to G:C homoduplex mutant and heteroduplex mutant templates in the absence or presence of the cleavage factor GreB. The +25 pauses and terminated product RNAs are labeled. (E) Band intensities of +25/26 pauses and terminated product from (D) were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; homoduplex mutant is red circles; non-template strand mutant heteroduplex is blue triangles; template strand mutant heteroduplex is green squares.


A
-10 like
+25 pause
WT
AGATCTGATATTTTGCCAATGTTGTA
+25T:A to G:C
AGATCTGATATTTTGCCAATGTTGgA
+25GC-rich
AGATCTGATATTTTGCCAATGggGgA
+1 +5 +10 +15 +20 +25
B
GreB
WT
+25T:A to G:C
+25GC-rich
Time
Term
-
26
25
Time
Term
+
25
C
Total Pause
1.00
0.10
0.01
0.00
0 2 4 6 8 10
Time (min)
WT
+25G
+25GC-rich
WT + GreB
+25G + GreB
+25GC-rich + GreB

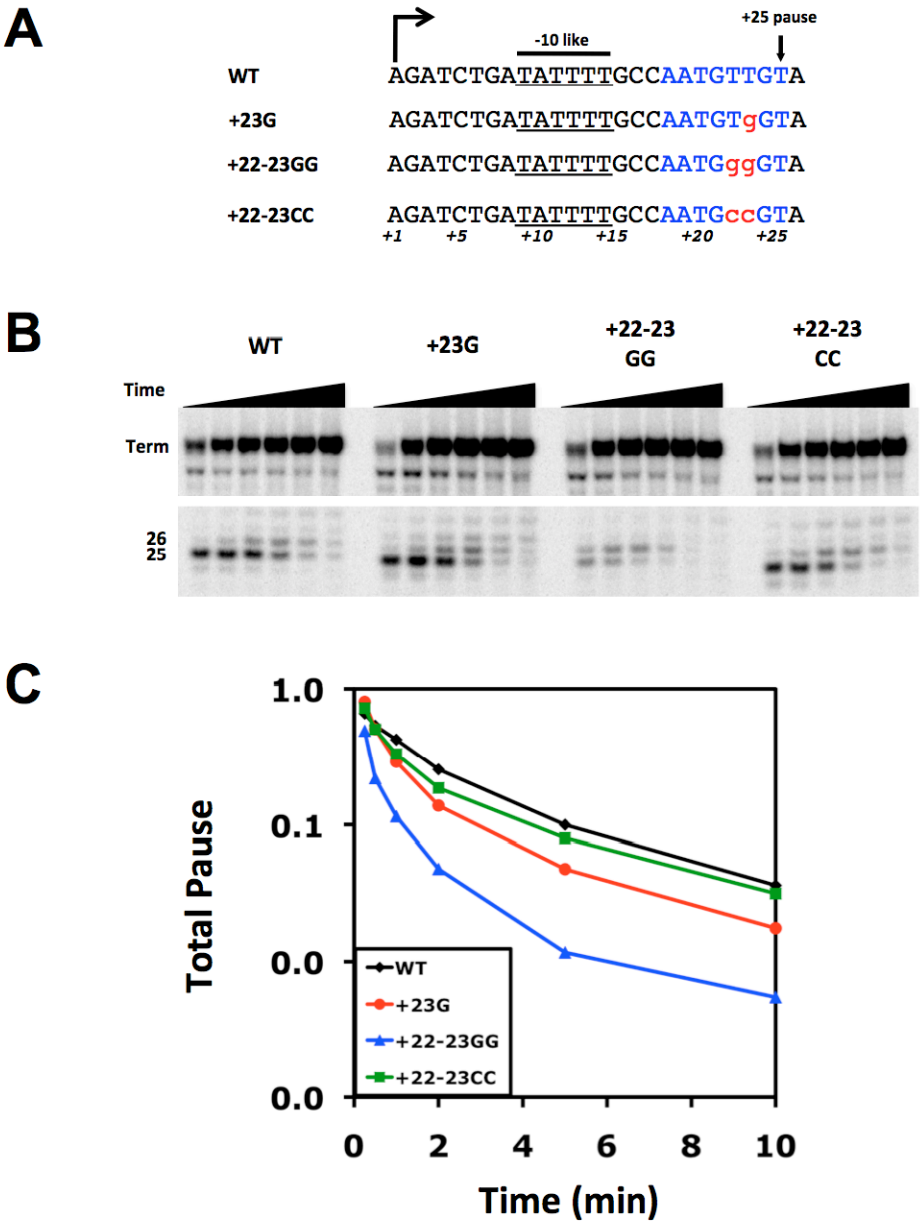

D
GreB
WT
+25G
NTS mutant
TS mutant
Time (min)
Term
-
26
25
Time (min)
Term
+
25


E
GreB
-
+
Total Pause
1.00
0.10
0.01
0.00
0
2
4
6
8
10
Time (min)
WT
+25G
NTS mut
TS mut

eliminate the residual +26 pause seen on a -10-like element mutant template and, yet, when the -10-like element is present with this mutation, the +26 backtracked pause still exists. The apparent paradox is that, by themselves, these mutations should *reduce* backtracking. The difference is presumably due to the difference between a simple backtracked complex, which presumably is what occurs in the absence of a -10-like site and the σ70 stabilized backtracked +26 paused complex that occurs when the scrunched pause complex collapses. In σ70 dependent pausing, the "backtrack-inducing" sequence is stabilizing the +25 forward scrunched pause in some way quite distinct from inducing backtracking. The GC content mutants are reducing the amount of +25 non-backtracked pause, without affecting the tendency of the σ70-stabilized pause to undergo backtracking from the +26 position.

Mutating the terminal nucleotide alone does not explain the entire effect of the +25GC-rich mutation, however. In fact, mutating positions +22 and +23 from T:A to G:C also decreases the amount of pause seen (Figure 3.2A, B, C). Mutating both +22 and +23 to G:Cs may slightly decrease the amount of pause capture as quantified in Figure 3.2C, but acts primarily to increase the rate that complexes can escape the pause as compared to wild type. Interestingly, this does not appear to be the case if both +22 and +23 are changed to C:G rather than G:C . This mutation had little to no effect on the promoter-proximal pause, and a similar effect is seen when mutating position +25 (Figure 3.5). It is not clear why this difference exists, but a potential cause is a difference in the strength of a dGTP:rCTP and dCTP:rGTP interaction strength in the hybrid (Sugimoto et al., 1995).

Figure 3.2. Increasing the GC-richness of the RNA/DNA hybrid without changing the terminal nucleotide still decreases pausing. (A) Sequences of WT, +23T:A to G:C, +22-23TT:AA to GG:CC and +22-23TT:AA to CC:GG mutant templates from the transcription start site to the site of the promoter-proximal pause. The -10-like element is underlined; transcription start site and the site of the pause are as marked; the AT-rich region is marked in blue; mutated bases are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A). The +25 and +26 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. WT is black diamonds; +23T:A to G:C is red circles; +22-23TT:AA to GG:CC is blue triangles; +22-23TT:AA to CC:GG is green squares.


A
+25 pause
-10 like
WT
AGATCTGATATTTTGCCAATGTTGTA
+23G
AGATCTGATATTTTGCCAATGTgGTA
+22-23GG
AGATCTGATATTTTGCCAATGggGTA
+22-23CC
AGATCTGATATTTTGCCAATGccGTA
+1 +5 +10 +15 +20 +25
B
WT
+23G
+22-23 GG
+22-23 CC
Time
Term
26
25
C
Total Pause
1.0
0.1
0.0
0.0
WT
+23G
+22-23GG
+22-23CC
0 2 4 6 8 10
Time (min)

Changing just position +23 to a G:C basepair had a modest effect on the amount of +25/+26 pause seen. What is interesting though is that more of a +27 species can be seen when transcribing this template. It appears that this mutation made it more likely for complexes to backtrack at +27, possibly through retaining the AT-richness at positions +25/26 while stabilizing the sequence upstream of those positions. This result could have implications for understanding the process of pause release. One important question is when σ70 actually is released from the EC during pause escape and the presence of a paused species at +27 could indicate that it can be released some distance downstream of the pause.

Making the RNA/DNA hybrid more AT-rich affects backtracking at +26 but not pausing at +25

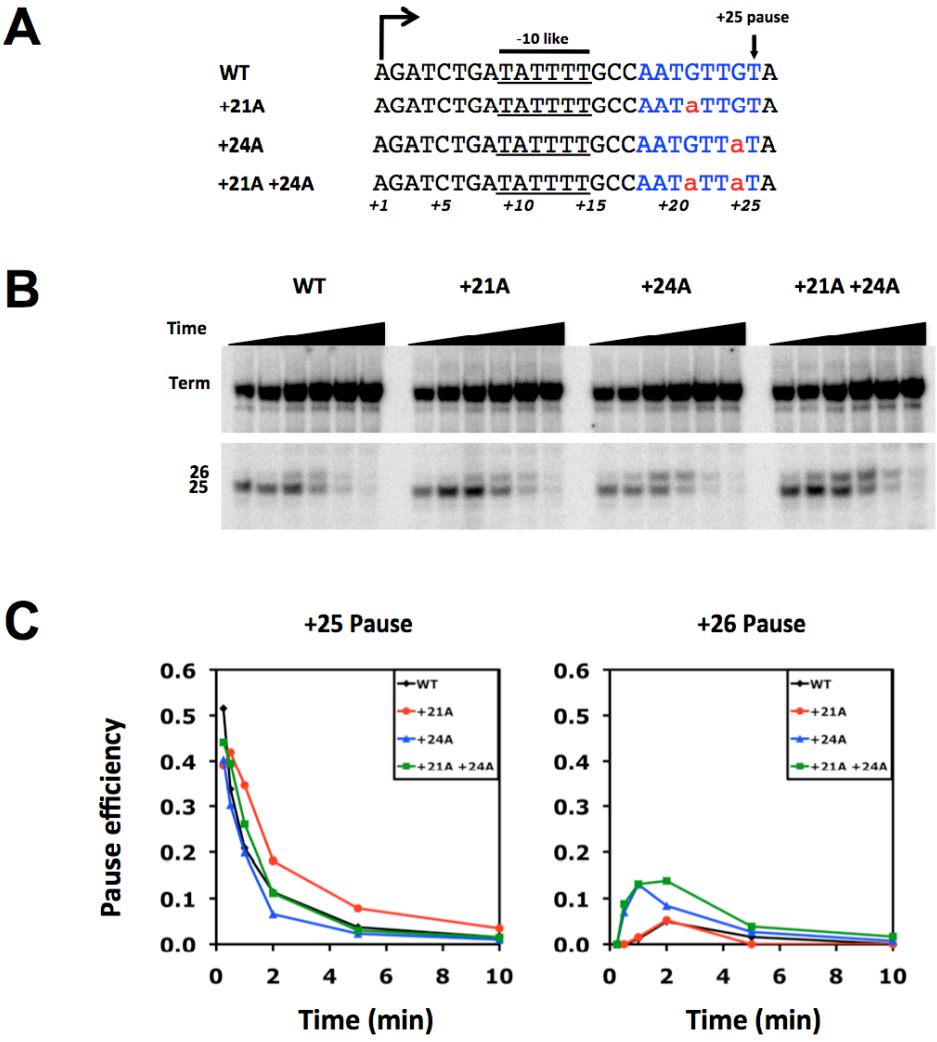
The AT-rich sequence that composes most of the RNA/DNA hybrid at the site of the phage 82 pR' promoter-proximal pause contains only two G:C basepairs at positions +21 and +24, not including the TempGC sequence at the 5' end of the RNA/DNA hybrid discussed in Chapter 2 (Figure 3.3A). As described above, changing the terminal sequence to be more GC-rich reduces the amount of pausing seen (Figure 3.1B, C)(Perdue and Roberts, 2010). So in principle making this sequence more AT-rich might help strengthen the contribution the AT-rich sequence is making towards pause formation.

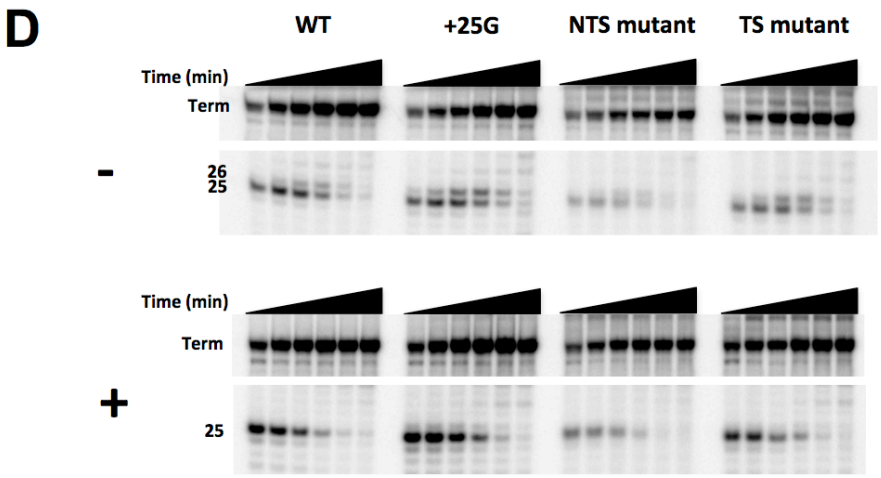
We decided to look at what would happen if we changed these positions individually to A:T basepairs or if we made a double mutant changing both positions together to A:T basepairs on the 82 *pR'* pause sequence (Figure 3.3B, C). These

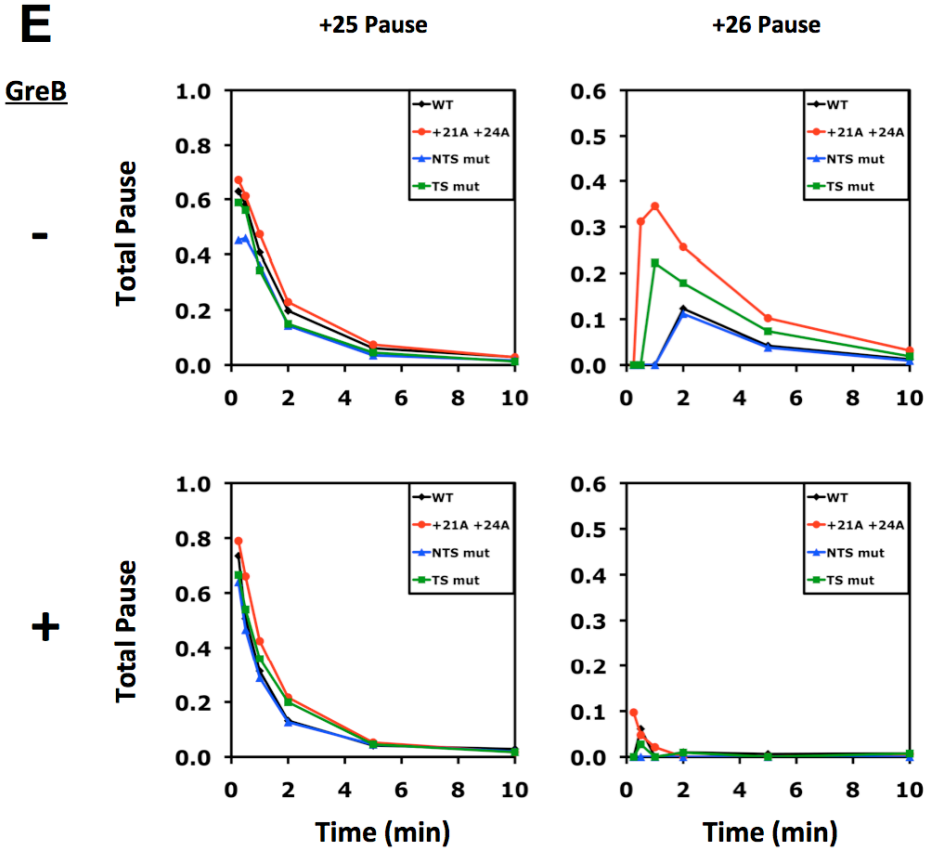
changes had little to no effect on the +25 pause (Figure 3.3C), suggesting that these changes did not substantially affect the formation of the σ70-dependent pause. However, both individual mutations at +21 and +24 increased the amount of pause seen at position +26. The double mutant increased the amount of +26 seen even more than either of the single mutations (Figure 3.3D,E).

To rule out the possibility that these mutations could be working through a mechanism other than affecting the stability or structure of the RNA/DNA hybrid, transcription was performed on heteroduplex templates containing the +21G:C to A:T +24G:C to A:T mutation. As expected the template strand mutation increased amount of backtracked pause seen like the homoduplex mutant whereas the non-template strand mutation showed a similar amount of +26 to the wild type template (Figure 3.3D, E). This suggests these mutations affect the stability of the RNA/DNA hybrid after escape from the +25 pause. Also, the increased amount of +26 pause seen on the double mutant template and heteroduplex templates is eliminated in the presence of GreB, indicating that the increased pausing seen at +26 is still backtracked just as in the wild type case (Figure 3.3D, E). The GreB sensitivity of the mutants fit well with what is known about the pause sequence. As described, the terminal sequences of the sigma-dependent pauses are similar to the "Class II" backtrack inducing pause sequence (Artsimovitch and Landick, 2000; Perdue and Roberts, 2010). Making the sequence more AT-rich affects the relative probability of escape versus collapse at +26 by reducing the stability of RNA/DNA hybrid in the forward position and increasing the likelihood that the scrunched DNA bubble would collapse.

Figure 3.3. Making the RNA/DNA hybrid more AT-rich increases the amount of backtracked +26 pause without affecting +25 pausing. (A) Sequences of WT, +21G:C to A:T, +24G:C to A:T and double mutant templates from the transcription start site to the site of the promoter-proximal pause. The -10-like element is underlined; the AT-rich region is marked in blue; transcription start site and the site of the pause are as marked; mutated bases are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A). The +25 and +26 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen at each position individually was plotted as % of total RNA over time. WT is black diamonds; +21G:C to A:T is red circles; +24G:C to A:T is blue triangles; the double mutant is green squares. (D) Single round *in vitro* transcription time-courses on the WT, +21A +24A homoduplex mutant and heteroduplex mutant templates in the absence or presence of the cleavage factor GreB. The +25 pauses and terminated product RNAs are labeled. (E) Band intensities of +25/26 pauses and terminated product from (D) were quantified and amount of pause seen seen at each position individually was plotted as % of total RNA over time. WT is black diamonds; Homoduplex mutant is red circles; non-template strand mutant heteroduplex is blue triangles; template strand mutant heteroduplex is green squares.


A
-10 like
+25 pause
WT AGATCTGATATTTTGCCAATGTTGTA
+21A AGATCTGATATTTTGCCAATaTTGTA
+24A AGATCTGATATTTTGCCAATGTTaTA
+21A +24A AGATCTGATATTTTGCCAATaTTaTA
+1 +5 +10 +15 +20 +25
B
WT
+21A
+24A
+21A +24A
Time
Term
26
25
C
+25 Pause
+26 Pause
Pause efficiency
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10
Time (min)
WT
+21A
+24A
+21A +24A


D
WT
+25G
NTS mutant
TS mutant
Time (min)
Term
-
26
25
Time (min)
Term
+
25


E
+25 Pause
+26 Pause
GreB
-
+
Total Pause
Total Pause
WT
+21A +24A
NTS mut
TS mut
Time (min)
Time (min)

There is a clear base preference at positions +25 and +26 for pause formation

Given the strong effect on the +25 pause of mutating only the terminal G at the 3' end of the RNA/DNA hybrid to U, I wanted to determine which nucleotide is preferred at this terminal position (Figure 3.1B, C). To do this I made four 82 pR' templates containing the four different nucleotides at the terminal position (+25), and also a second set of four with the different nucleotides at +26, in order to determine the effect of single nucleotide changes at this position on the backtracked pause.

Mutating position +26 from the wild type A to either T or G did not have a large effect on pausing (Figure 3.4A, B). At most both changes only slightly increased the amount of backtracked pause at +26 as evidenced by the slightly increased amount of total +25/+26 pause seen by the two mutants in the absence of GreB that disappears when GreB is present (Figure 3.4B). Interestingly, mutating +26 to C had the largest effect on pausing, reducing the amount of pause seen both with and without GreB. Since the +26C mutation affects +25, it can't be the hybrid, and must affect either unwinding of the downstream DNA or some interaction with core RNAP. If it were an unwinding mechanism, changing a A:T basepair in the RNA/DNA hybrid to a C:G should have the same effect as changing it to a G:C. Possibly this effect has to do with C fitting in the β-pocket, discussed below, in some way that we don not yet understand.

Having an A, C or G instead of the wild type T at the terminal position (+25) of the pause all reduced the +25/26 pause, with both A and G having a greater effect than C (Figure 3.4C, D). Transcription in the presence of GreB reveals the effect of these mutations on the scrunched +25 pause and not the backtracked +26 pause,

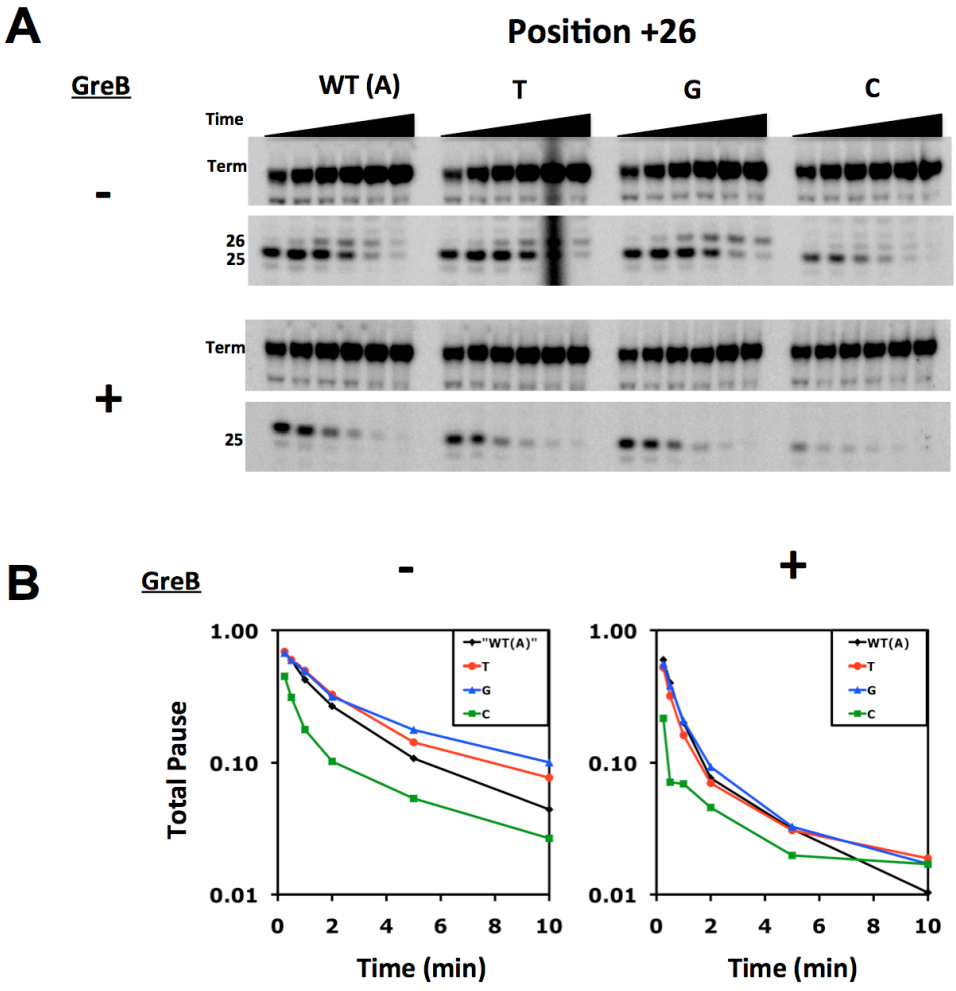
(Figure 3.4C, D), and shows a clear hierarchy of base preference. Both the A and C mutations do not greatly affect pause capture, but both cause the pause to not persist as long as T, with the pause on A disappearing sooner than C. The G mutation acts as previously described having a small to non-existent effect on pause capture as compared to the wild type T, but the pause persists for a significantly shorter period of time than on any of the other templates. This gives us a base preference of T>C>A>G for the terminal nucleotide of the pause.

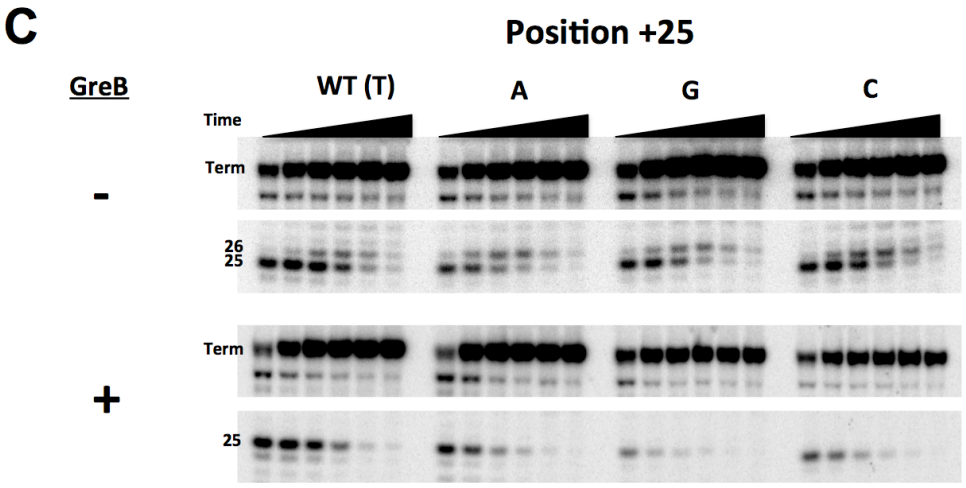
The base preference I see for the terminal nucleotide of the promoter-proximal pause of phage 82 *pR'* is particularly interesting in light of the recent results of Hein et al. (2011); they saw the same base preference of U>C>A>G for the terminal nucleotide of the RNA/DNA hybrid during pyrophosphorolysis experiments that determined the translocation state of transcription elongation complexes during the nucleotide addition cycle. In addition, they found that both U and C at the terminal nucleotide were biased towards existing in a pretranslocated state and that proclivity can increase the likelihood of pause formation (Hein et al., 2011). The finding of the same order of base preference strongly suggests that the phage 82 σ70-dependent +25 pause is pretranslocated as modeled in Figure 3.5A, although, more evidence will be needed to confirm the possibility.

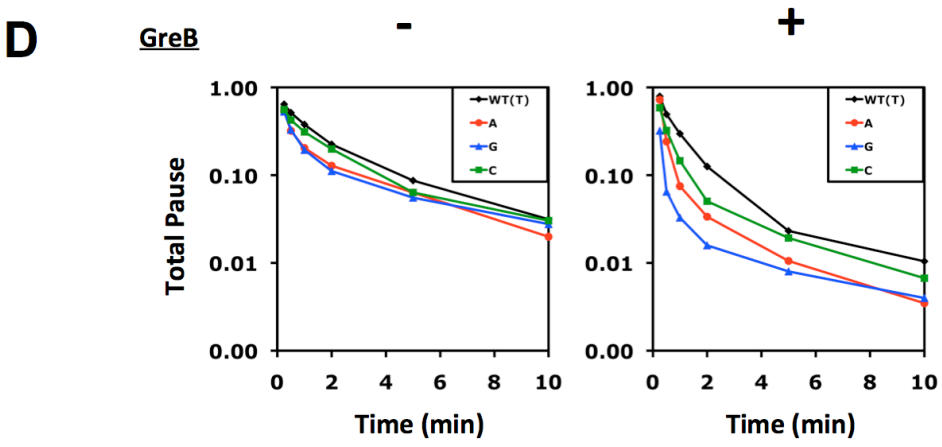
RNAP mutants reveal that the paused transcription complex is pretranslocated

A recent structure of *T. thermophilus* initiation complex showed that the non-template strand nucleotide corresponding to the nucleotide in the active site of the enzyme is flipped out of normal stacking position and interacts with a pocket in the

Figure 3.4. Base preference of the terminal nucleotide of the RNA/DNA hybrid and the subsequent base. (A) Single round *in vitro* transcription time-courses on WT(+26A:T), +26T:A, +26G:C and +26C:G mutant templates either in the absence or presence of the cleavage factor GreB to determine the preferred basepair downstream of the terminal position of the promoter-proximal pause. The +25/+26 pauses and terminated product RNAs are labeled. (B) Band intensities of +25/26 pauses and terminated product from (A) were quantified and amount of pause seen was plotted as % of total RNA over time. WT pausing is black diamonds; +25A:T mutant template pause is red squares; +25G:C mutant template pause is blue triangles and +25 C:G mutant template pause is green squares. (C) Single round *in vitro* transcription time-courses on WT(+25T:A), +25A:T, +25G:C and +25C:G mutant templates either in the absence or presence of the cleavage factor GreB to determine the preferred basepair at the terminal position of the promoter-proximal pause. The +25/+26 pauses and terminated product RNAs are labeled. (D) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. WT pausing is black diamonds; +26T:A mutant template pause is red squares; +26G:C mutant template pause is blue triangles and +26 C:G mutant template pause is green squares.


A
Position +26
GreB
WT (A)
T
G
C
Time
Term
-
26
25
Term
+
25
B
GreB
-
+
Total Pause
1.00
0.10
0.01
"WT(A)"
T
G
C
0
2
4
6
8
10
Time (min)
WT(A)
T
G
C
Time (min)



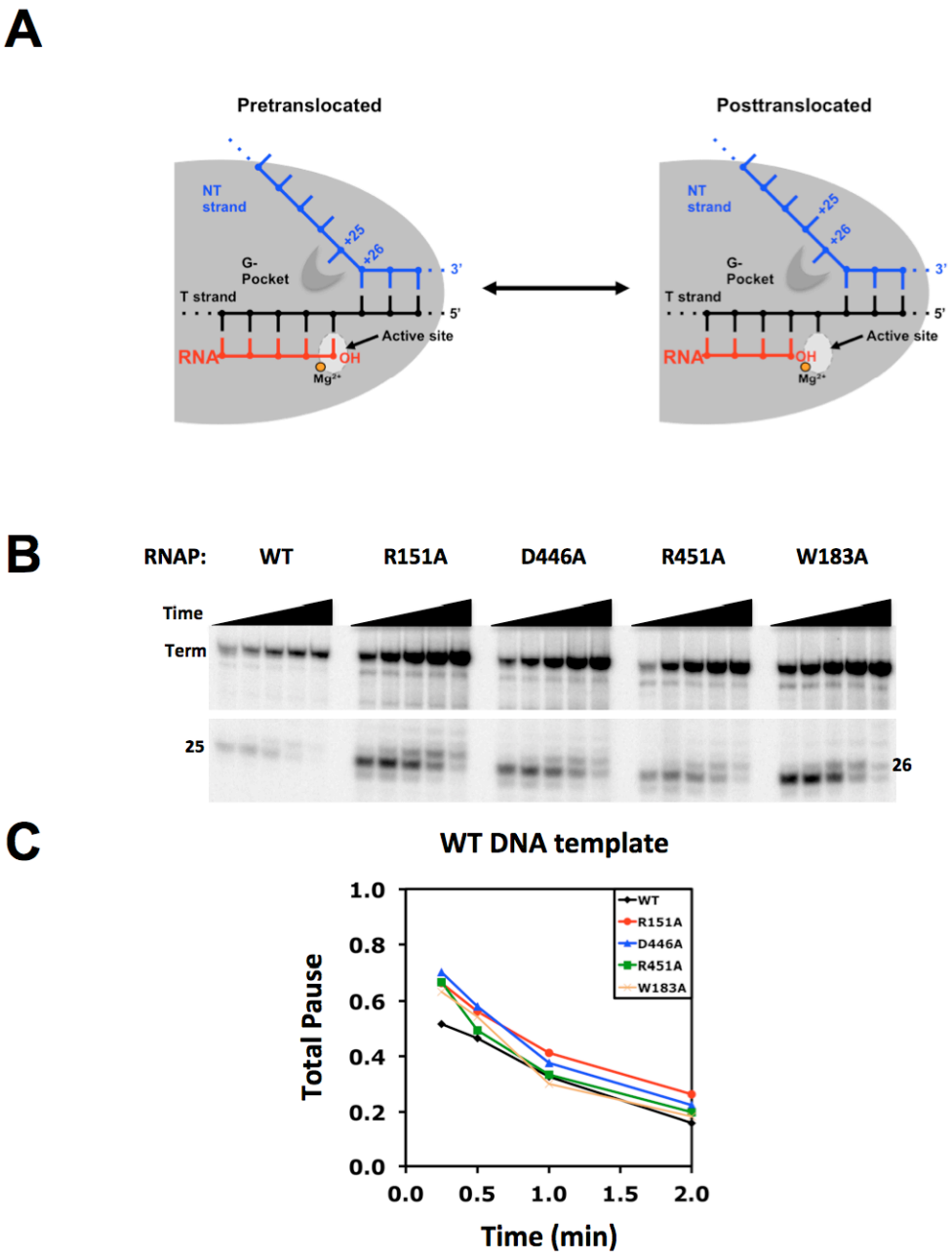


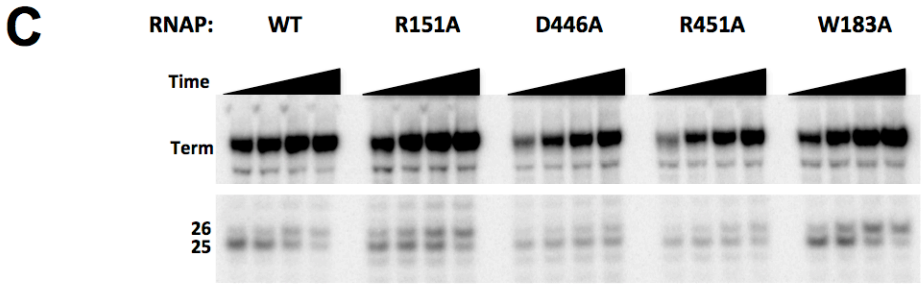
Base preference: T > C > A > G

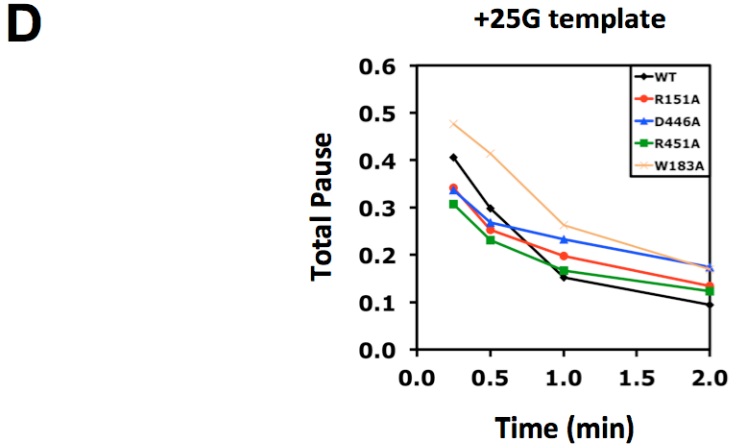
β-subunit called the "β-pocket" (Zhang et al., 2012). The β-pocket preferentially interacts with a G base (Zhang et al., 2012). I set out to use this knowledge to test the translocation state of the transcription complex at the promoter proximal pause.

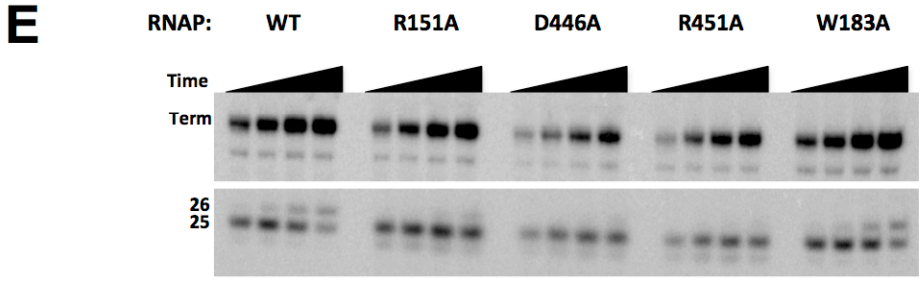
Given the preference of the β-pocket for binding to a G base, I performed *in vitro* single round transcription on the wild type 82 *pR'* template and mutant templates containing a G base at either +25 or 26 with a series of RNAPs containing mutations to the β-pocket R151A, D446A, and R451A, as well as one mutant deficient for interaction with the base directly upstream of the pocket, W183A, as a control (Figure 3.5)(Zhang et al., 2012). These mutations interfere with the binding of the β-pocket with nucleotides in the non-template strand, especially a G base, and have varied effects on transcriptional efficiency as a result; β R151A and D446A reduce efficiency, while R451A has little effect (Zhang et al., 2012). If the base at +25 interacts with the β-pocket as I would expect if the complex is in fact pretranslocated I would expect a stronger effect of the mutant RNAPs on the +25G template than the +26G template. Not surprisingly, none of the mutants had a significant effect on transcription of the wild type template that lacks a G base in the non-template strand of the DNA bubble at either position +25 or +26 (Figure 3.5B, C). The three β-pocket mutants showed reduced pause capture as compared to the wild type polymerase on the +25G mutant template, while β W183A showed slightly more pausing (Figure 3.5C, D). It is important to note again that the +25G template shows less pausing than the wild type +25T; the β-pocket mutant polymerases reduce that amount of pause

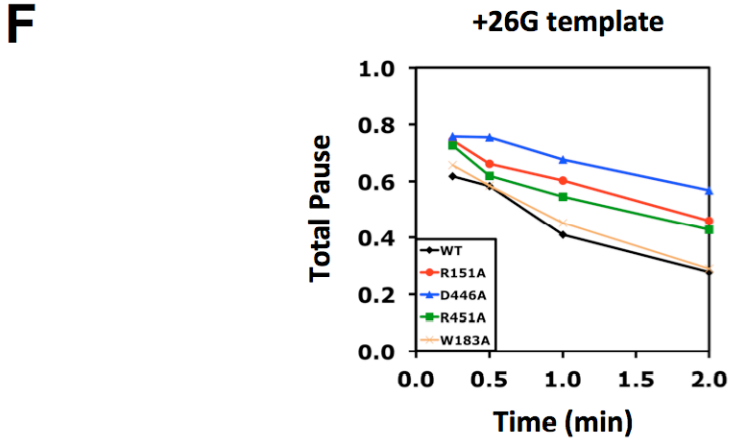
Figure 3.5. β-pocket mutations reveal that σ70 dependent promoter-proximal pauses occur when a transcription complex is in a pre-translocated state. (A) Model of possible translocation states at the phage 82 pR' +25 pause. RNAP is in Grey; template DNA in black; non-template DNA in blue; RNA in red and Mg^{2+} in orange. State 1: After the addition of UTP to the 3' end of the RNA strand at +25, the pause is in a pretranslocated state and the nucleotide at +25 of the non-template strand makes contacts with the residues of the β-pocket. State 2: After the addition of nucleotide at +25 the complex forward translocates opening the active site for the addition of the next nucleotide. In this posttranslocated state. The base at +25 of the non-template strand leaves the β -pocket and base +26 enters (adapted from Hein et al. 2011 and Zhang et al. 2012). (B, D, F) Single round *in vitro* transcription time-courses on WT, +25G:C and +26G:C mutant templates respectively, using either WT holoenzyme RNAP or one of the following mutant polymerases: RNAP β R151A, D446A, R451A and W183A. The +25/+26 pauses and terminated product RNAs are labeled. (C, E, G) Band intensities of +25/26 pauses and terminated product from (B, D, F respectively) were quantified and amount of pause seen was plotted as % of total RNA over time. WT RNAP pausing is black diamonds; β R151A mutant pausing is red squares; β D446A mutant pausing is blue triangles; β R451A mutant pausing is green squares and β W183A mutant pausing is yellow Xs.


A
Pretranslocated
Posttranslocated
NT strand
+25
+26
G-Pocket
3'
T strand
5'
Active site
RNA
OH
Mg2+
B
RNAP:
WT
R151A
D446A
R451A
W183A
Time
Term
25
26
C
WT DNA template
Total Pause
Time (min)
WT
R151A
D446A
R451A
W183A
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.5
1.0
1.5
2.0


C
RNAP:
WT
R151A
D446A
R451A
W183A
Time
Term
26
25


D
+25G template
Total Pause
Time (min)
WT
R151A
D446A
R451A
W183A
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.5
1.0
1.5
2.0


E
RNAP:
WT
R151A
D446A
R451A
W183A
Time
Term
26
25


F
+26G template
Total Pause
Time (min)
WT
R151A
D446A
R451A
W183A
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0

even further suggesting that there is in fact an interaction between the +25G in the non-template strand and the β-pocket during pausing on the +25G template.

The β W183A mutant RNAP had no effect on pausing activity of the +26G template, but pausing was actually increased with the three β-pocket mutant RNAPs, possibly due to the decreased transcriptional activity of the three mutants (Figure 3.5E, F). Interestingly, the three β-pocket mutants also did not form any +26 backtracked pause. This implies that by interfering with the interaction of the β-pocket with position +26G on the non-template strand, these mutants inhibit the formation of the backtracked pause. The increase in +25 pausing could be due to the inhibition of translocation to 26 that results from the G base not fitting in the β-pocket. If the complex escapes +25 it is able to escape the pause completely instead of backtracking as in the wild type case. Taken together with the effect of the mutants on the +25G template this indicates that nt position +25 is in the β-pocket at the +25 pause and nt position +26 is in the β-pocket at the +26 pause suggesting that the transcription complex is in a pretranslocated state when pausing is induced.

Sequences downstream of the pause play a modest role in pause formation

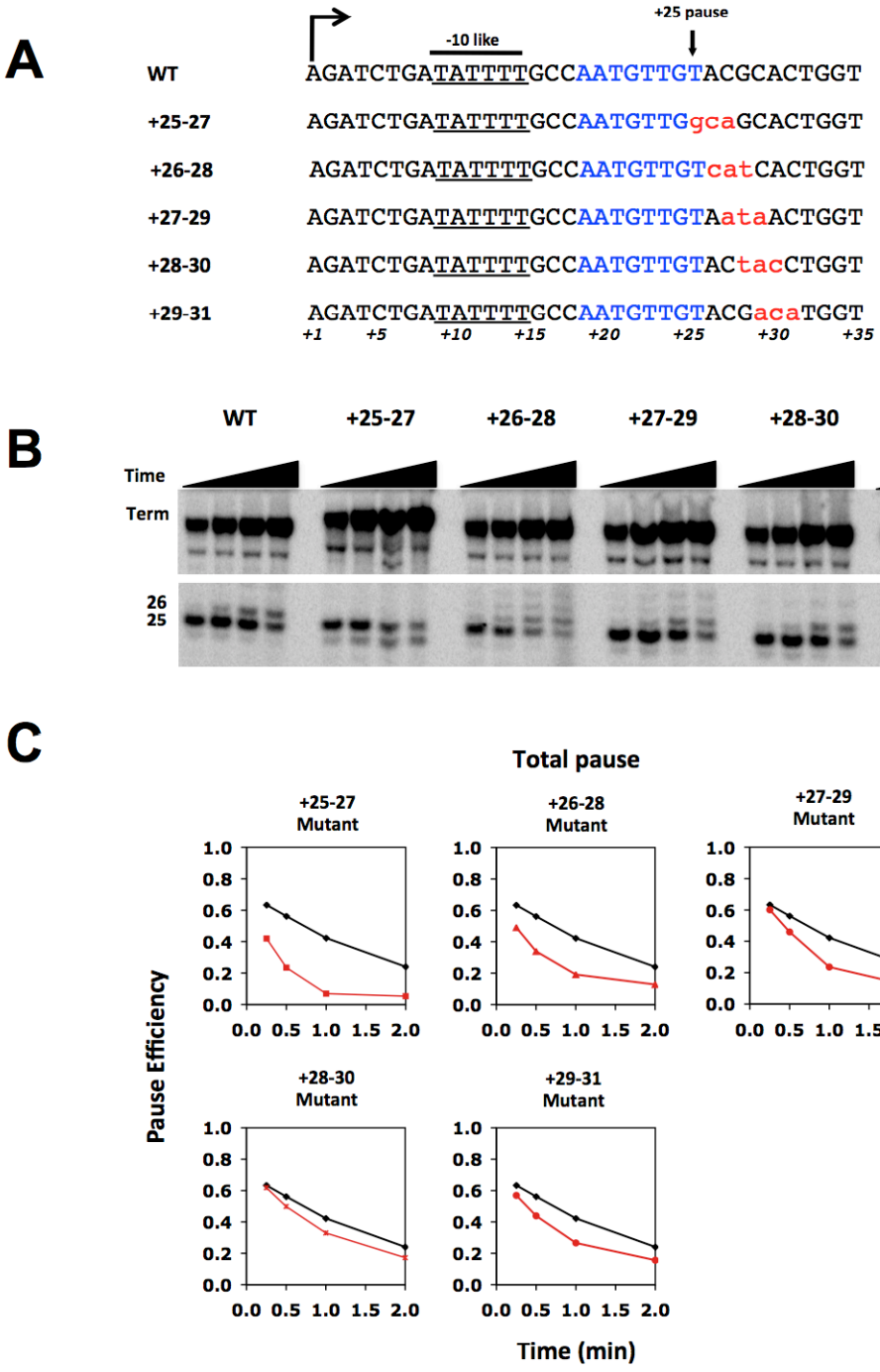
Mutating position +26 from an A:T basepair to a C:G basepair significantly reduces the 82 pR' promoter-proximal pause at +25 even though the base is not transcribed during pausing. This lead me to believe that the sequence downstream of the pause is possibly affecting pause formation, possibly through then energy requirement necessary to unwind the DNA downstream of the pause as the complex

forward translocates or through an interaction of the downstream DNA with core RNAP. To test this I created a series of mutant templates in which I mutated three consecutive bases to the opposite purine/pyrimidine; A to C, T to G, G to T and C to A. The mutations were made scanning through a region from the site of the pause at +25 to residue +31 6 bases downstream (Figure 3.6A). One particular feature of the DNA that made this region interesting was a GC-rich sequence from +27 to +29, which might affect pausing due to the G:C base pairs being harder to unwind than A:T base pairs as the transcription complex escapes past the +26 backtracked pause. The sequence of +27 to +29 is CGC which is a motif also present in the sequence of λ pR'. However, in the λ case the CGC is in positions +16 to +18, the first two bases of which constitute the promoter-proximal pause, which may suggest they are not performing the same function (Chapter 2 Figure 2.1).

Not surprisingly, the first mutant template, with mutated bases +25-27 greatly reduces pausing (Figure 3.6B, C). This is most likely due to both position +25 being mutated to a G and position +26 being mutated to a C. Both of these nucleotide changes are the least favored at either position as previously mentioned (Figure 3.4). Similarly the +26-28 mutant template also shows reduced pausing on an order similar to just mutating position +26 to C alone (Figure 3.4C, D and 3.6B, C). It is still possible that mutations to +27 and +28 are also contributing to the effect, but are not as strong as the +26C mutation, as the subsequent three scanning mutant templates all show a slight reduction in pausing. All three of these mutants reduce the GC content of the DNA directly downstream of the pause; +29-31 reduces it by only one base pair

as compared to either the +27-29 or +28-30 mutants that reduce it by 3 and 2 base pairs respectively (Figure 3.6A). It is likely that DNA unwinding does not play a

Figure 3.6. Scanning mutations at the site of the pause and immediately downstream of the pause affect pause formation. (A) Sequences of WT and 3 basepair scanning mutant templates from the transcription start site to the site of the promoter-proximal pause. The -10-like element is underlined; the AT-rich region is marked in blue; transcription start site and the site of the pause are as marked; mutated bases are marked in red. (B) Single round *in vitro* transcription time-courses on the WT and mutant templates from (A). The +25 and +26 pauses and terminated product RNAs are labeled. (C) Band intensities of +25/26 pauses and terminated product from (B) were quantified and amount of pause seen was plotted as % of total RNA over time. Pausing from each mutant was plotted individually against the WT pause. WT pausing is black diamonds; mutant template pause is red squares.


A
+25 pause
-10 like
WT AGATCTGATATTTTGCCAATGTTGTACGCACTGGT
+25-27 AGATCTGATATTTTGCCAATGTTGgcaGCACTGGT
+26-28 AGATCTGATATTTTGCCAATGTTGTcatCACTGGT
+27-29 AGATCTGATATTTTGCCAATGTTGTAataACTGGT
+28-30 AGATCTGATATTTTGCCAATGTTGTACtacCTGGT
+29-31 AGATCTGATATTTTGCCAATGTTGTACGacaTGGT
+1 +5 +10 +15 +20 +25 +30 +35
B
WT
+25-27
+26-28
+27-29
+28-30
+29-31
Time
Term
26
25
C
Total pause
+25-27 Mutant
+26-28 Mutant
+27-29 Mutant
+28-30 Mutant
+29-31 Mutant
Pause Efficiency
Time (min)
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.5 1.0 1.5 2.0

significant role in pause formation as the +27-29 mutation has a particularly strong effect. The last two mutants, 28-30 and 29-31, shouldn't have to unwind to proceed from 25 to 26, so it would seem that the effect they have must be as dsDNA, perhaps moving through or binding the downstream binding region.

Discussion

Mutations to the AT-rich sequence reveal a mechanism for σ70 dependent promoter-proximal pausing

Our investigation of phage 82 *pR'* promoter-proximal pause-inducing sequences has revealed many mechanical details of how the pausing mechanism works including the role of the AT-rich region of the pausing sequence and the translocation state of a transcription complex at the pause. Both the GC-rich element, discussed in Chapter 2, and the AT-rich region at the terminus of the pausing sequence, are necessary for pause formation and play important roles in mechanism of pausing.

I have shown in this thesis that the σ70 region 2-dependent promoter proximal pausing relies on two distinct elements. The first is a repeat of the -10 element, along with its associated extended -10 and non-template strand discriminator elements, which allows for the re-engagement of σ70 region 2 with the DNA at a site downstream of the original promoter contact (Ring et al., 1996). The second is a template strand GC-rich element, encoded by the same DNA as the non-template strand discriminator element, followed by a template strand active AT-rich sequence at the terminal region of the paused RNA. The latter was originally identified as a backtrack inducing sequence (Perdue and Roberts, 2010). However, identifying this

sequence only for it ability to induce backtracking is not correct in this context: its role here is to stabilize the forward, scrunched pause and not to stabilize the backtracked configuration. Sequences similar to these two template strand-active elements have been shown to be involved in a number of different transcriptional pauses including ubiquitous transcriptional pausing, the *ops* pause and RNA dependent promoter-proximal pausing on the *his* operon, implicating the sequence in a mechanism that is general to all pauses (Herbert et al., 2006).

In the absence of a secondary pause inducing element such as σ70-binding element the 82 *pR'* promoter-proximal pausing sequence does in fact cause a backtracked pause as shown by Perdue and Roberts (2010), and increasing the GC-content of this sequence eliminates that backtracking. Curiously, though, in the presence of a -10-like site the same GC-rich mutation does not eliminate the naturally occurring backtracking pause seen at +26, but rather reduces the non-backtracked +25 pause instead (Figure 3.1B, C). Thus transcription complexes that escape the scrunched pause at +25 and become backtracked after extending the transcript by one base to +26 are distinct in their origin from complexes that fall into a backtracked position solely based on encountering the pausing sequence.

A significant portion of the effect of the +25GC-rich mutation appears to be coming from the single mutation of the terminal nucleotide of the pause as mutating this base alone significantly reduces pausing (Figure 3.1). This suggests a prominent role for the terminal nucleotide in the formation of the pause, which is supported by the clear hierarchy of base preference at that terminal position (Figure 3.4). However, general increases in the GC-content of the AT-rich region also are contributing to the

reduction in pausing (Figure 3.2), as shown by increasing the GC content of positions +22 and +23.

In Chapter 2, I showed that the 3 basepair GC-rich portion of the pause-inducing sequence has two distinct functions in pause formation. The first is through a non-template strand interaction with σ70 region 1.2 that acts as an extension of the -10-like element/σ70 region 2 interaction, and the second is through function in positioning the 5' end of the RNA/DNA hybrid. The GC-rich sequence (TempGC) alone or in concert with the AT-rich sequence in the template strand is what is responsible for pause site determination. My research has shown that there are two separate functions to the GC-rich and AT-rich regions of the pausing sequence as the first appears to determine the site of pause and the second the amount of pausing and the amount of backtracked versus scrunched pause; however, it is likely that the two regions may work together to determine the site of the pause (Chapter 2).

In the current scrunching model, after engagement of σ70 with the -10-like element and discriminator-like sequence, the complex continues synthesizing RNA by extending the size of the DNA bubble by four base pairs (Kapanidis et al., 2006; Perdue and Roberts, 2010; Revyakin et al., 2006). At the beginning of this process the RNA/DNA hybrid contains the 3 GC base pairs of the TempGC sequence, which has a fairly strong hybrid (Perdue and Roberts, 2011). As it synthesizes past this sequence into the AT-rich region the strength of the hybrid is reduced until it either reaches +25 and changes into a paused state or synthesizes one base further to +26 where a specific hybrid strength threshold is crossed making the complex unstable, collapsing the DNA

scrunch and causing the RNA to backtrack, similarly to the release of abortive products during initiation.

Interestingly, changing the only two GC basepairs found in the AT-rich region of the pausing sequence only affects the backtracked pause at +26, but not the scrunched +25 pause (Figure 3). This reaffirms the original assumption that this sequence is responsible for backtracking. This would support my model of pause formation as well, since one to two additional AT basepairs in the AT-rich region theoretically should weaken the hybrid further as a transcription complex reaches an RNA length of +26 nt, making it more likely for the scrunch it undergoes to collapse and backtracking to occur.

The base preference of the terminal nucleotide and β-pocket mutant RNAP data suggest that the pause occurs in a pretranslocated state

RNA synthesis during transcription elongation by RNA polymerase proceeds through a four step mechanism: translocation, NTP binding, catalysis, and pyrophosphate release (Bai et al., 2007; Gong et al., 2005; Kireeva et al., 2010). A ratcheting mechanism by the trigger structure at the active site of RNAP, alternating between a loop confirmation and a three helix bundle, is believed to drive the process (Bai et al., 2007; Hein and Landick, 2010; Kireeva et al., 2010; Zhang et al., 2010). This structure switches the RNAP between a pretranslocated state with a previously synthesized nucleotide occupying the active site of the enzyme and a posttranslocated state, in which a new NTP can be added to the growing chain (Figure 3.5A)(Bai et al., 2007; Kireeva et al., 2010). Not surprisingly, the 3' terminal nucleotide of the

RNA/DNA hybrid strongly affects the conformation bias between the pre and translocated state, with U and C bases showing a strong bias towards the pretranslocated state and A and G a bias towards the posttranslocated state (Hein et al., 2011).

Our data shows that there is a base preference of T>C>A>G for the terminal nucleotide in determining the strength of the phage 82 *pR'* promoter-proximal +25 pause. Hein et al. (2011) found the same base preference order of the terminal nucleotide for the pyrophosphorolysis rate of the terminal nucleotide bond, which measures the fraction of complex in the pretranslocated state. This state is suggested to be a preliminary stage for certain categories of transcriptional pausing that involve a structural rearrangement of RNAP into a distorted state that prevents further nucleotide addition (Hein et al., 2011; Landick, 2006, 2009). Our findings taken together with the pyrophosphorlysis data thus suggest that a transcription elongation complex at the 82 *pR'* +25 pause is most likely in a pretranslocated state when pausing occurs. This could also imply that the paused state of σ70 dependent pauses is like these other pauses in this respect.

At position +26 the only base change that has a strong effect on pausing is a change to a C:G basepair from the wild type A:T (Figure 3.4C, D). If this effect were due to it being more difficult to unwind a GC basepair than an A:T basepair downstream of the transcription bubble during forward translocation, the same effect would be expected from the +26G:C mutant template, but that is not the case. The only likely explanation left is that either the C in the non-template strand or the G in the template strand at the active site is somehow changing the interaction of the DNA

with the polymerase itself, thereby affecting pausing. Consistent with this, mutating the sequence immediately downstream of the pause site only has a modest effect on pause formation, suggesting that downstream DNA binding is not playing a prominent role in the pause formation (Figure 3.6). This is in direct contrast to what was previously found for the promoter-proximal hairpin-induced pause on *trpL*, suggesting a significant difference in the pause state of the two processes. Mutating the sequence downstream of the pause on *trpL* significantly reduced the halflife of the pause, although a clear sequence specific effect was not evident (Lee et al., 1990).

The effects of the RNAP β-pocket mutations on +25G and +26G mutant transcription further support the idea that the +25 pause forms from a transcription complex in a pretranslocated state. The β-pocket mutant RNAPs are deficient for pausing on the +25G template despite having no effect on a wild type template, and increasing +25 pausing on the +26G template. This suggests that they are having a negative impact of the interaction of the polymerase with the non-template strand DNA, presumably that G nucleotide (Figure 3.5B, C, D, E). This result also implies that an artificial interaction of the non-template strand +25G with the β-pocket is occurring in the mutant template, when it does not seem to be playing any role during pausing on the wild type template and also does not contribute in heteroduplex analysis of the +25G template, which showed that 100% of the mutant effect came from the template strand (Figure 3.1D, E). Still, while the exact mechanism of interaction between the b-pocket mutant polymerases and the +25G template is confusing, it is clear that an interaction is occurring at the pause, which would only happen if the complex is in a pretranslocated register.

Again this G nucleotide/β-pocket interaction has a significant effect on the +26G mutant template. None of the three β-pocket mutant RNAPs were able to form +26 backtracked pause on this template despite the wild type and W183A polymerases still being able to (Figure 3.5D). This same effect is not seen on the +25G template despite the +25 pause being reduced by the β-pocket mutants (Figure 3.5B). The interaction of the G base with the mutant β-pocket appears to have significant effects on the formation of the pause and may be an artificial disruption to proper RNAP function as the natural base at position +26 is an A and should not interact favorably with the β-pocket. The lack of pausing at +26 still implies that the base at +26 position of the non-template strand is making contact with the β-pocket before it becomes backtracked consistent with the theory that paused complexes formed from elongation complexes in a pretranslocated register. The interaction at +26 affects the +25 pause even though the +26 position is only unwound in the postranslocated state. Maybe pausing at +25 is enhanced by the mutants because the postranslocated state is destabilized and therefore residence at +25 is enhanced. This result is consistent with the lateral mobility of RNAP complexes and the fact that the complex does not dwell at +26 on its way to elongation.

It is most likely that all σ70-dependent promoter-proximal pauses enter the paused state through a pretranslocated conformation. The terminal nucleotide of the promoter-proximal pause on phage 82 *pR'* is a T residue, which is the most likely to be in the pretranslocated state (Hein et al., 2011). The terminal nucleotide of the λ phage pR' promoter-proximal pause is a C, which also according to Hein et al. (2011) prefers to

be in the pretranslocated state, suggesting that it too is also acts by the same mechanism. Many other known and predicted σ70-dependent promoter-proximal pauses including *lacZ, tnaA*, phage 80 *pR'*, *rpsUp2* and cspD all have either T or C as a terminal nucleotide as well (Perdue and Roberts, 2011).

CHAPTER 4

DISCUSSION

σ70-dependent promoter proximal pauses have been identified in a number of transcriptional units from the *E. coli lacZ* gene to the late gene promoters of lambdoid phages (Brodolin et al., 2004; Hatoum and Roberts, 2008; Liu and Roberts, Unpublished data; Nickels et al., 2004). These pauses were originally discovered and characterized on the phage λ *pR'* promoter in studies looking at the activity of the antiterminator Q protein (Grayhack et al., 1985; Marr et al., 2001; Ring et al., 1996; Yarnell and Roberts, 1992). Antitermination activity by Q protein requires the loading of the protein into a transcription complex at the σ70-dependent pause directly downstream of the *pR'* promoter. Transcription complexes that are modified by Q are able to read through terminators and transcribe the phage late genes necessary for the final stages of lytic growth (Deighan and Hochschild, 2007; Yarnell and Roberts, 1999). To date, the Q antitermination system is the only known regulatory role for σ70-dependent pausing, although it is likely that these pauses are part of other regulatory mechanisms, considering the breadth of transcriptional units in which they are found.

Promoter-proximal pausing on lambdoid phage promoter *pR'* requires a σ70 region 2/non-template strand DNA interaction with the -10-like element, but what has not been known is the role the surrounding sequence plays in pause formation (Perdue and Roberts, 2010, 2011; Ring et al., 1996). These pauses occur approximately 5 base pairs downstream of where it would be expected; a discrepancy that has been

attributed to involve a process similar to that of "scrunching" during transcription initiation (Marr et al., 2001; Perdue and Roberts, 2010). Directly downstream of the -10-like element there is a 3-4 base pair GC-rich sequence followed by an AT-rich sequence in many know σ70-dependent pauses (Perdue and Roberts, 2011). As originally shown by Perdue and Roberts (2010) changing the AT-rich sequence to have more GC-content significantly reduces σ70-dependent pausing, implicating this sequence in pause formation. In this thesis I have explored the role that these two elements, the GC-rich and AT-rich sequences, play in σ70-dependent pause formation on phage 82 *pR'*.

In Chapter 2 I specifically looked at the role of the GC-rich element. I found that this sequence plays two roles. First this sequence is important for pause formation through an interaction with σ70, mediated by a region 1.2/non-template strand DNA contact. This contact is equivalent to the interaction that occurs at the discriminator element found on some prokaryotic promoters and it serves to extend the s70/-10-like interaction that induces the pause. This interaction still occurs during pausing on phage 82 *pR',* despite the GC-rich sequence being non-ideal for discriminator interaction. Changing the sequence to the consensus GGG actually increases pausing on this promoter. The GC-rich sequence on phage λ *pR'* is the consensus GGG and the discriminator interaction plays a large role in pause formation (Filter and Roberts, Unpublished data; Ring and Roberts, 1994).

It is the fact that the 82 pause discriminator-like sequence is less than ideal for σ70 interaction that it allows us to see the second and possibly more important role that this sequence plays. Transcription experiments on λ *pR'* pause discriminator-like

sequence heteroduplex mutant templates showed that majority of the effect was through the non-template strand. However, the same experiments on 82 *pR'* sequence reveals that the GC-rich element also contributes strongly to pausing through a template strand mechanism. Futhermore, the template strand GC-rich sequence, called the TempGC, either alone, or in conjunction with the downstream AT-rich sequence, is what determines the site of the pause. If both of these sequences are moved downstream in the DNA, transcriptional pausing is similarly moved downstream. Additionally, pauses moved in this manner are still in dependent on the σ70/-10-like element interaction and are resistant to cleavage by GreB, suggesting that they are in the same state as the non-backtracked wild type pause.

It was previously noted that the TempGC and the AT-rich sequence that follows it resembles a backtrack-inducing sequence, and if these elements are present without a preceding -10-like element they do in fact induce backtracking (Perdue and Roberts, 2010). It is interesting that the same properties that make this sequence ideal for backtracking alone appear to allow it to determine the site of the σ70-dependent pause. However, mutating the only two guanine residues in the AT-rich sequence to adenine only affects the +26 backtracked pause and not the +25 pause. These mutations do not affect the stable scrunched pause, but make it more likely to escape into a collapsed backtracked state after the addition of one more nucleotide. This state is distinctly different than a simple backtrack, as it is still mediated by the σ70/-10-like interaction and may indicate that the AT-rich sequence's ability to induce backtracking is separate from the role it is playing in pause formation.

Similar sequences are present in a number of other transcription pauses including ubiquitous transcriptional pausing, the *ops* pause and RNA dependent promoter-proximal pausing on the *his* operon. The presence of these similar sequences in all of these pauses implicates them in a mechanism general to pausing (Herbert et al., 2006). The template strand GC-rich element, which corresponds to the back edge of the RNA/DNA hybrid in a paused complex, and the AT-rich sequence to a lesser extent, are present in all of these pauses and may be necessary to induce a conformational change in RNAP, called the elemental pause (Landick, 2006). The elemental pause is a set of changes in the transcription complex that induces an offline conformation that inhibits continued synthesis and cause pausing (Landick, 2006). Further experiments are required to see if these sequences play the same role in pause formation at these other sites.

In Chapter 3, I also performed experiments that suggest that transcription complexes enter σ70-dependent pauses from a pretranslocated state in which the previously added nucleotide to the RNA strand still occupies the active site of the enzyme. This state would ensure that synthesis could not continue until the complex could forward translocate, freeing up the active site for the addition of the next nucleotide; the elemental pause is thought to be induced from this state (Landick, 2006). *In vitro* transcription on mutants of the terminal nucleotide of the pause (+25) showed that there is a base preference of T>C>A>G for the terminal nucleotide of the phage 82 *pR'* promoter-proximal +25 pause. This corresponds with the same preference order for the rate of pyrophosphorylsis, which measures the fraction of a complex in the pretranslocated state (Hein et al., 2011). Experiments using RNAP

mutants in a region called the β-pocket in conjunction with DNA templates containing G residues at +25 and +26 further support the suggestion that the complex enters the +25 pause from a pretranslocated state. The mutant RNAPs are deficient for pausing on the +25G template and show increased pausing on the +26G. However, these mutant RNAPs did not show any +26 backtracked pause on the +26G template. This indicates that the disruption of the β-pocket by the mutant RNAPs in the presence of the +26G nucleotide prevented the formation of the collapsed backtracked pause at +26 when the complex was in a pretranslocated state.

Taken together these results have shed additional light on the mechanism by which σ70-dependent pausing occurs on lambdoid phage *pR'* late gene promoters. What is particularly interesting is that my data are consistent with our current model of pausing occurring by a scrunching mechanism, although, none of my experiments directly addressed the issue of scrunching. Pauses that were moved downstream by inserting sequence between the -10-like element and the TempGC still react in the same manner as the wild type pause, even if they are moved 5 or 10 bases downstream. This leads to questions that are still unanswered regarding how this pause works. First and foremost is how to directly show that the pause occurs by a scrunching mechanism. This will most likely require studying the pause through single molecule methods such as FRET and lazer-tweezer force experiments, which were used to show scrunching during initiation of transcription (Kapanidis et al., 2006; Revyakin et al., 2006).

A second question that remains is why the paused complex exists in two states. The backtracked pauses at +26 (phage 82) and +17 (λ) are easily understandable in the

context of a scrunching mechanism because both result from the scrunch relaxing back to the initial σ70 bound state. In terms of scrunching during initiation, this would be equivalent to abortive initiation, except that the RNA is sufficiently long that it backtracks rather than being released from the complex. The +25 (phage 82) and +16 (λ) non-backtracked pause is not as easily understandable. These pauses are not backtracked, but are stable in a scrunched state, which is thought to be energetically strained (Kapanidis et al., 2006; Revyakin et al., 2006). The question is what protein/nucleotide interactions stabilize this paused state. This fundamental question is still not well understood, except that these pauses must result from a conformational change of the complex that prevents further nucleotide addition. This change is likely to be the elemental pause that has been suggested as a general mechanism of pausing (Landick, 2006). Perhaps the elemental pause is the result of some yet unknown conformation of the trigger loop/helices that catalyze the addition of nucleotides to the growing RNA chain. This state may be better understood by using known mutant RNAPs deficient for pausing in addition to future screens for additional mutations with the same effect.

APPENDIX

ATP USE BY THE TRANSCRIPTION COUPLED REPAIR ENZYME MFD DURING TRANSCRIPTION COMPLEX RELEASE

Introduction

Termination is an important step in the mechanism of transcription, with distinct implications for the regulation of gene expression. There are currently three known mechanisms of transcription termination in *E. coli*: intrinsic or hairpin termination, Rho-mediated termination and Mfd-mediated termination (a component of transcription coupled repair)(Nudler and Gottesman, 2002). Intrinsic terminators use a specific RNA structure encoded by the DNA to terminate, whereas both Rho and Mfd are proteins that use the energy of ATP hydrolysis to dissociate RNAP transcription complexes. Much effort has been applied to study how these mechanisms dissociate transcription complexes and cause termination.

Intrinsic terminators are encoded in the DNA sequence and are responsible for termination of most bacterial transcription units. Intrinsic termination is caused by the formation of a G/C rich RNA hairpin that is thought to facilitate the separation of a U-rich region from the template DNA and dissociation of the complex. Interestingly, artificially creating a hairpin, by annealing a DNA oligonucleotide upstream of a U-rich region is also able to induce termination (Yarnell and Roberts, 1999). The dissociation of the complex is thought to happen either through a shortening of the RNA/DNA hybrid due to forward translocation or through slippage of the hybrid due

to its homopolymeric nature (Komissarova et al., 2002; Larson et al., 2008; Santangelo and Roberts, 2004; Yarnell and Roberts, 1999). It is possible that both of the mechanisms are contributing to the termination event, as they are not mutually exclusive.

Rho-dependent termination occurs when Rho, a hexameric AAA+ ATPase encounters the unoccupied RNA being synthesized by an active elongation complex (Lau et al., 1983; Lowery-Goldhammer and Richardson, 1974; Morgan et al., 1985). Rho is believed to act by translocating to the EC along the RNA and exerting a force on the complex similar to that of the GC-rich hairpin of the intrinsic terminator (Ciampi, 2006; Skordalakes and Berger, 2006). Rho functions as a terminator for a number of bacterial transcription units and is also responsible for inducing operon polarity and preventing the accumulation of unutilized RNA in a cell (Adhya et al., 1974; Cardinale et al., 2008; Franklin, 1974; Richardson et al., 1975).

The third mechanism of termination is that of the protein Mfd, also called Transcription Repair Coupling Factor (TCRF). Mfd serves a different role from those of the other two modes of termination in that it acts on transcription units stalled by DNA damage, and also connects DNA repair to transcription by recruiting enzymes of excision repair. UV-radiation causes photochemical lesions in DNA bases that cannot be transcribed by RNAP polymerase. In an ATP hydrolysis-dependent manner, Mfd is capable of recognizing and dissociating ECs that have been stalled by these bulky DNA lesions and other DNA damage, by nucleotide starvation, or by other factors inhibiting forward translocation (Park et al., 2002; Selby and Sancar, 1995a, b). In the case of DNA damage such as thymine or cyclobutane dimers, Mfd also recruits

proteins of the Nucleotide Excision Repair pathway (NER) to begin repairing the damaged DNA (Selby and Sancar, 1993b). However, as other forms of DNA damage do not irreversibly stall ECs, Mfd does not appear capable of causing termination at these lesions (Smith and Savery, 2008).

Mfd is a 130-kDa protein that is a member of the Helicase superfamily 2, composed of 8 domains (Figure 2)(Deaconescu et al., 2006). The crystal structure of Mfd has been solved. That structure together with previous studies of the function of Mfd provides insight into the distinct functions of each domain (Deaconescu et al., 2006). Domains 1a, 2 and 1b have a high degree of sequence and structural similarity to NER protein UvrB. This domain is responsible for recruiting the proteins of the NER pathway by interacting with UvrA, the protein responsible for recruiting UvrB to the site of DNA damage (Manelyte et al., 2010; Selby and Sancar, 1993a, 1995a). Domain 4, called RID (RNAP Interaction Domain) is responsible for the interaction of Mfd with RNAP (Selby and Sancar, 1995b). Domains 5 and 6 are composed of seven conserved helicase/translocase motifs of superfamily 2 helicases. These two domains are responsible for both the ATPase activity and DNA interaction of Mfd (Chambers et al., 2003; Deaconescu et al., 2006). Domain 7 has been shown to be involved in regulating the translocase and ATPase activities of Mfd. A deletion of this domain constitutively activates Mfd, removing the requirement of binding to a RNAP/DNA complexes for the stimulation of Mfd's ATPase and translocase activities (Smith et al., 2007). The function of Domain 3 is not understood at this time.

Mfd, as a DNA translocase, uses the same motor domains as helicases to travel along DNA, but does not contain the "wedge" domain used by helicases for separating

DNA strands. Mfd has high sequence similarity to another prokaryotic helicase RecG, another motor protein of helicase superfamily 2 involved in DNA repair and recombination. Both proteins contain an ATP dependent DNA translocation motif (TRG for Translocation in RecG) that is only found in Mfd and RecG, (Mahdi et al., 2003). RecG does contain a "wedge" domain, while Mfd contains domains that allow interaction with the β-subunit of RNAP (Park et al., 2002; Smith and Savery, 2005).

Park and Roberts (2002) showed that Mfd requires contact with both RNA polymerase and the DNA comprising 26 basepairs upstream of the complex in order to act on the complex, and they presented evidence supporting the forward translocation model of termination by Mfd. Actively transcribing ECs sometimes encounter sequences that cause them to "backtrack" or translocate backwards past the catalytic configuration, extruding the 3' primer end of the RNA from the active site. Normally complexes in this state hydrolyze or phosphorolyze the phosphodiester bonds of the RNA at the active site, a reaction catalyzed by stimulatory factors such as *E. coli* GreA or GreB, thereby releasing the 3' over hanging RNA and allowing transcription to continue (Borukhov and Goldfarb, 1996; Borukhov et al., 1992). Mfd is able to reverse backtracking by forward translocating the backtracked complex and reincorporating the 3' overhang into the RNA/DNA hybrid. Under NTP starvation conditions the EC dissociates. This means Mfd must be forcing the EC forward in order either to induce a return to transcription or dissociation (Park et al., 2002).

To study the mechanism of complex release in more detail, I looked for a correlation of ATP usage by Mfd and its activity of releasing elongation complexes (ECs) that have been artificially stalled by nucleotide starvation in a manner similar to

previous studies looking at ATP use during protein unfolding and degradation by the *E. coli* chaperone/protease complex ClpXP (Kenniston et al., 2003). By looking at ATP use during protein unfolding/translocation, Kenniston et al. (2003) were able to find a correlation between ATP use and protein function, in addition to information about which step was rate limiting. Hopefully applying similar methods to Mfd activity will reveal more about the mechanism of complex release and possibly elucidate the exact conformational changes to the transcription complex is necessary for termination.

Materials and Methods

Plasmids and DNA templates: Linear DNA transcription templates were amplified from pBY409 plasmid by PCR and gel purified using a gel extraction kit from Qiagen. pMfdΔ1,2 and pMfdΔ7 were constructed by introducing NdeI and EcoRI sites respectively internal to the gene on pAD6 Mfd expression vector by quickchange mutagenesis (Agilent). Plasmids were then digested by each enzyme respectively, and the resulting product ligated and transformed into DH5α cells. Resulting plasmids from colonies were sequenced ensure the presence correct deletions.

Proteins: Core RNAP protein was the gift of M. Suh (Cornell University, Ithaca, NY). The following proteins were purified as described: 6xHis-σ70 (Marr and Roberts, 1997), GreB (Borukhov and Goldfarb, 1996), 6xHis-Mfd (Deaconescu and Darst, 2005).

Transcript release assay: Assays were performed as in Park and Roberts (2002). Open complexes were made by incubating 250 nM HoloRNAP with 50 nM BY409 template DNA bound to streptavidin-coated magnetic beads (Promega) for 10 min at 37^{o}C in transcription buffer (20 mM Tris-HCl (pH 8.0), 50 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 10% glycerol) and 5 mM $MgCl_2$. Stalled elongation complexes were then formed by the addition of 75 μM ApU, 100 μM ATP, GTP, 25 μM UTP (supplemented with 0.5 μCi/μl [α-^{32}P]-UTP) and incubating for 10 min at 37^{o}C. Complexes were washed 3x with (20 mM Tris-HCl (pH 8.0), 500 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.1 mg/ml BSA and 10% glycerol) to remove excess nucleotide and free protein and resuspended in transcription buffer with 50 μM ATP. Reactions were started with the addition of 250 nM Mfd or mutant and incubated at 37^{o}C. Alliquots were taken at indicated times after addition of Mfd and mixed with 5 volumes Stop solution (600 mM Tris-HCl (pH 8.0), 12 mM EDTA, and 0.16 mg/ml tRNA) on ice to stop transcription. The supernatant was separated from the beads by magnets and the beads were then washed 3x with the wash buffer above, before being resuspended in 150 μL stop solution. RNA was extracted by mixing with a equal volume of phenol:chloroform:isoamyl alcohol (25:24:1). 2.5 volumes of 100% Ethanol was added to the aqueous layer to precipitate samples overnight at -20^{o}C. Samples were run on 15% polyacrylimide gel containing 6M Urea. RNA species were detected using a Molecular Dynamics Typhoon 9400 and images were analyzed using GE Imagequant software.

ATP hydrolysis assay: Reactions were carried out in TB supplemented with 5 mM $MgCl_2$ and 50 nM ATP or analog (supplemented with 0.5 μCi/μl [α-^{32}P]-ATP or 1250 mCi/mmol [γ-^{35}S]-ATPγS as indicated) at 37^{o}C. 100 nM DNA template, 100 nM core RNAP, 300 nM GreB and 500 nM GreB were added as indicated. Substrate stalled Elongation Complexes were formed as above, with the exclusion of [α-^{32}P]-UTP. Reactions were started by the addition of 250 nM Mfd and 2 μl aliquots were spotted on to PEI-cellulose TLC plates. Nucleotides were resolved using TLC buffer (1.5 M formic acid, 0.5 M $LiCl_2$). Species were detected using a Molecular Dynamics Typhoon 9400 and images were analyzed using GE Imagequant software.

Native-PAGE electrophoresis: RNAP complexes were formed as above (in the absence of radiolabeled nucleotide) and incubated for 10 min at 37^{o}C in TB supplemented with 5 mM $MgCl_2$ and 2.5 mg/mL Rifampicin. Holo and core RNAP was added at 200 nM; DNA template at 500 nM. Holo and core RNAP was added at 400 nM; DNA template at 500 nM and Mfd at 300 nM for Figure A.1D. Samples were run on 5% Tris pH 8.0 native gels supplemented with 5 mM $MgCl_2$ at 150V and stained with either SYBR-Green (Invitrogen) or by Bradford protein stain (200 mg Coomassie brilliant blue G250 (biorad), 5% Ethanol, 8.5% o-Phosphoric acid).

Results

Determining ATP usage by Mfd during transcription complex release

In order to study the mechanism of Mfd activity, in a manner similar to studies on ClpXP, it was necessary first to construct a stable substrate for the Mfd release

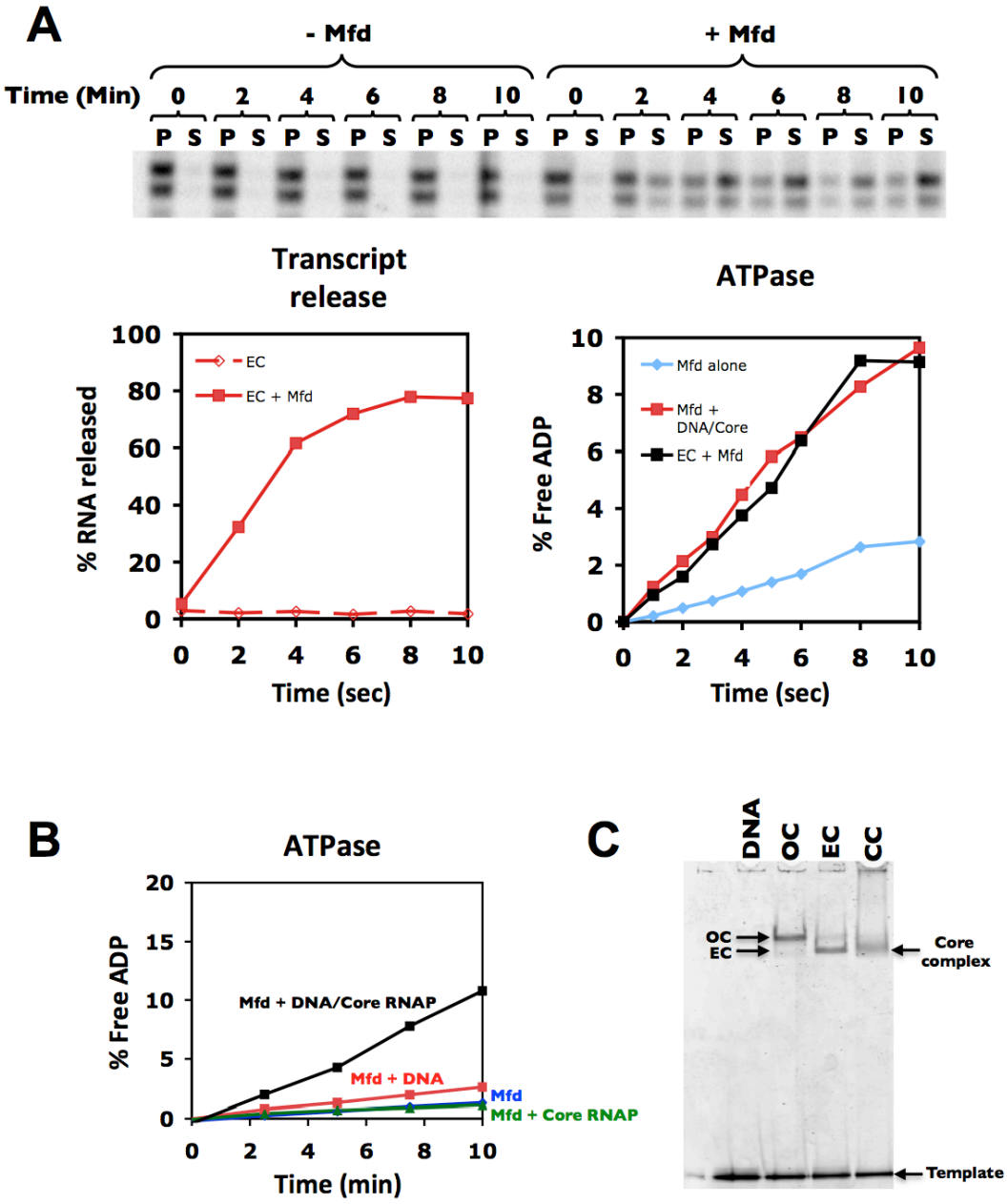
activity. Previously, Park and Roberts (2002) demonstrated that Mfd is capable of dissociating ECs stalled by nucleotide starvation on BY409 DNA templates. This template was derived from phage 82 *pR'* DNA, which has a naturally occurring σ70 dependent pause at +25 (discussed in chapters 2 and 3), but has been modified in two important ways: first the initial transcribed sequence has been mutated so that ECs stall at position +25 when CTP is omitted from the transcription reaction; and it contains a mutation to residue +14, which allows formation of stalled complexes at +25 that do not contain σ70, which inhibits Mfd activity (Park et al., 2002).

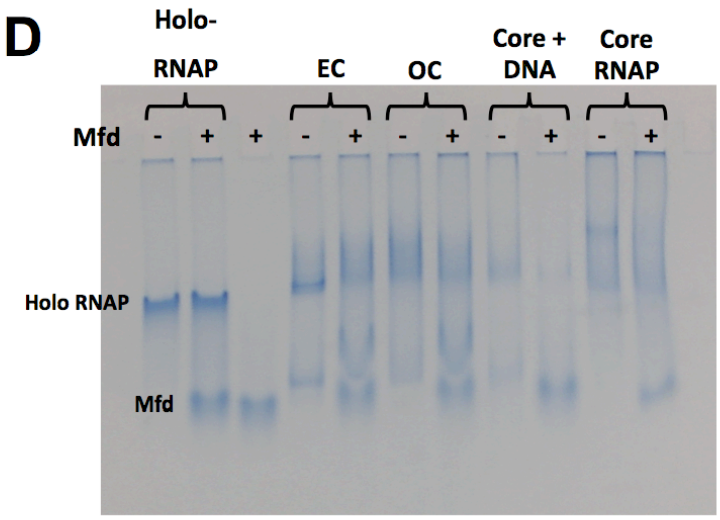
To look specifically at ATP hydrolysis during complex release, stalled complexes were formed on DNA template affixed to magnetic beads as described in the Methods. This allowed for washing away excess nucleotide and separation of the stalled EC substrates for the two assays for complex release and ATP hydrolysis activities (Figure A.1A). EC release was measured by comparing the amount of radiolabeled-RNA still present in stalled complexes in the pellet fraction versus the amount of free labeled-RNA in the supernatant of the reaction and is presented as percentage of free RNA for each time point. When 100 nM stalled ECs were treated with 250 nM Mfd, the majority of transcript is released by about 4 minutes and release plateaus at 80% at about 6 minutes. This activity is associated with a 3-fold stimulation of ATP hydrolysis by Mfd, increasing from a basal rate of 153.3 ATP hydrolyzed/min to 494.6 ATP hydrolyzed per minute. Based on this, Mfd is hydrolyzing 48.5 ATP to release a stalled elongation complex, which seems likely to be too high a number to reflect a direct correlation of ATP hydrolysis with energy usage in the reaction.

One particularly perplexing feature is that this stimulation of Mfd ATP hydrolysis is not closely correlated to complex release. The majority of complexes are released by about 5 minutes in the transcript release assay, but there is not a corresponding drop in the rate of ATP hydrolysis. Instead, Mfd continues hydrolyzing ATP and the same stimulated rate. This continued ATPase activity appears to largely be due to Mfd reacting with free core RNAP and DNA in solution as demonstrated in Figure A.1B. Mfd ATPase activity is slightly stimulated by the presence of DNA template (1.5 fold), but is significantly stimulated by the presence of both core RNAP and DNA template at similar concentrations (100 nM each) as those used for forming stalled ECs (3 fold). Interestingly, ATP hydrolysis is not stimulated by the presence of Core RNAP without DNA, even though interaction with core RNAP is necessary for Mfd translocase and stimulated ATPase activities (Smith et al., 2007).

Both open complexes (OCs) and stalled elongation complexes are easily visualized by native-PAGE with either sybr-GREEN staining for DNA or Bradford staining for protein components of the complex (Figure A.1C, D). Free core RNAP forms a complex with DNA, called core complex here (CC) that has a similar mobility to the stalled ECs; likely, this complex is causing the background stimulation of ATP hydrolysis in complex. In fact this complex disappears in the presence of Mfd, just as do stalled ECs (Figure A.1D). Mfd has no noticeable effect on OC, which is consistent with previous data that showed Mfd cannot act on complexes that contain σ70, most likely because Mfd, which requires 26 base pairs of DNA upstream of the complex, must interact with DNA that is protected by σ70 in the complex (Figure A.3A) (Deaconescu et al., 2006; Park et al., 2002). Additionally, neither core nor holo

Figure A.1. The stimulation of Mfd ATPase activity does not correlate with transcription complex release activity. (A) Release of stalled transcripts by Mfd and associated ATP hydrolysis. Band intensities of pellet and supernatant fractions of +25 stalled complexes were quantified and amount of pause seen was plotted as % of complexes released over time. Associated ATP hydrolysis is plotted as % of ADP released. ECs alone are open red squares/dashed lines; ECs with Mfd solid red squares; Mfd alone is blue diamonds; Mfd + Core RNAP and DNA template is black squares. (B) ATPase activity of Mfd in the presence of free DNA template (100 nM) and core RNAP (100nM). Mfd alone is blue; Mfd or mutant + DNA template is red; Mfd or mutant + Core RNAP is green; Mfd or mutant + Core RNAP and DNA template is black (C) Native-PAGE electrophoresis of transcription complexes. Visualized by Sybr-GREEN I. (D) Native-PAGE of open, stalled elongation and core transcription complexes formed either in the presence or absence of Mfd; Holo RNAP and core RNAP were also incubated and tested for interaction with Mfd. Visualized by Bradford protein stain.

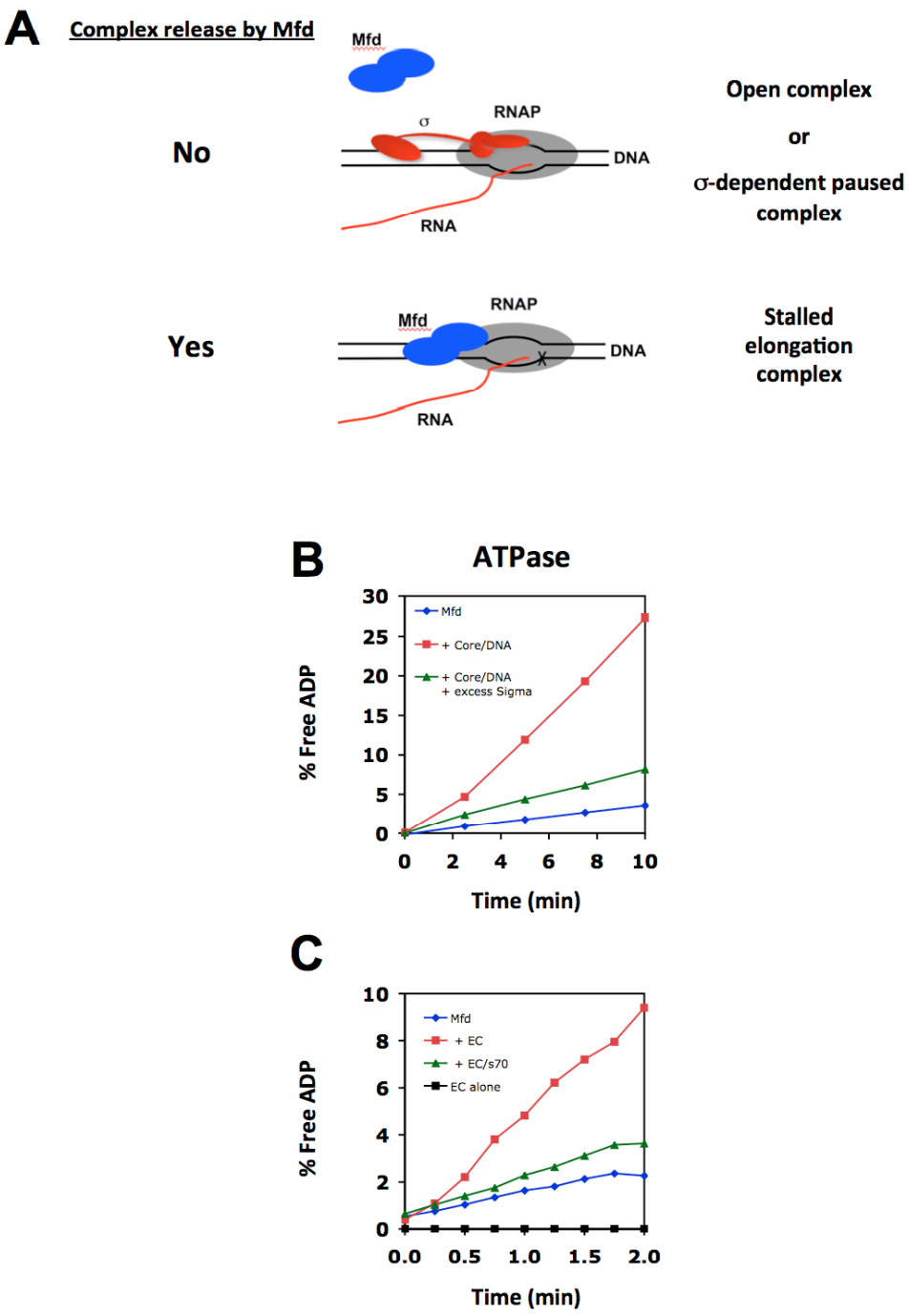

A
- Mfd
+ Mfd
Time (Min)
0 2 4 6 8 10 0 2 4 6 8 10
P S P S P S P S P S P S P S P S P S P S P S P S
Transcript release
% RNA released
EC
EC + Mfd
Time (sec)
ATPase
% Free ADP
Mfd alone
Mfd + DNA/Core
EC + Mfd
Time (sec)
B
ATPase
% Free ADP
Mfd + DNA/Core RNAP
Mfd + DNA
Mfd
Mfd + Core RNAP
Time (min)
C
DNA
OC
EC
CC
OC
EC
Core complex
Template


D
Holo-
RNAP
EC
OC
Core +
DNA
Core
RNAP
Mfd
- + + - + - + - + - +
Holo RNAP
Mfd

RNAP in the absence of DNA is able to form a stable complex with Mfd; this is consistent with the inability of either to stimulate Mfd's ATPase activity.

The background stimulation of ATPase activity needs to be removed in order to get accurate measurements of the amount of ATP hydrolyzed per complex released. One possible origin of the core complex formation is binding of core RNAP to ends of the linear DNA template. To deal with this issue I tried a number of different methods including using EcoRI Gln111 mutant protein, and binding streptavidin to biotin-substituted nucleotides, to block the ends of the DNA, without finding any effect on the ATPase background (data not shown). The one method for inhibiting background ATPase activity that was most effective was to take advantage of the ability of σ70 to inhibit Mfd activity (Figure A.2A). Adding σ70 to the reaction forms holoenzyme, which should then bind to the promoter region of the template as OC, which Mfd cannot act on. By adding an excess of σ70 (500 nM) to the ATPase reactions, I was able to reduce stimulation of background activity by Mfd in the presence of core RNAP/DNA template by 5 fold (Figure A.2B). The same effect is seen when adding 500 nM σ70 to reactions containing stalled ECs, which causes a 3 fold reduction in rate of ATP hydrolysis (Figure A.2C). This potentially could solve the background problem, but more experiments, including σ70 titration into transcript release assays/ATPase assays, would be required to find the optimal concentration of σ70 to add for inhibition.

Figure A.2. Excess σ70 inhibits the stimulation of Mfd ATPase active by core RNAP and free DNA. (A) Schematic of transcription complexes can (lower) and cannot (upper) terminate transcription on. (B) ATPase activity assays on Mfd alone (blue), Mfd with 100 nM Core RNAP/DNA template (red) and Mfd with 100 nM Core RNAP/DNA template in the presence of excess s70 (500nM)(green). (C) ATPase activity assays on Mfd alone (blue), Mfd with 100 nM stalled ECs on BY409 template (red) and Mfd with 100 nM stalled ECs on BY409 template in the presence of excess s70 (500nM)(green)


A
Complex release by Mfd
Mfd
RNAP
σ
DNA
RNA
No
Open complex
or
σ-dependent paused
complex
Mfd
RNAP
DNA
RNA
Yes
Stalled
elongation
complex
B
ATPase
% Free ADP
Time (min)
Mfd
+ Core/DNA
+ Core/DNA
+ excess Sigma
C
% Free ADP
Time (min)
Mfd
+ EC
+ EC/s70
EC alone

Mfd can release stalled transcription complexes by hydrolyzing ATPγS, but cannot utilize ADP

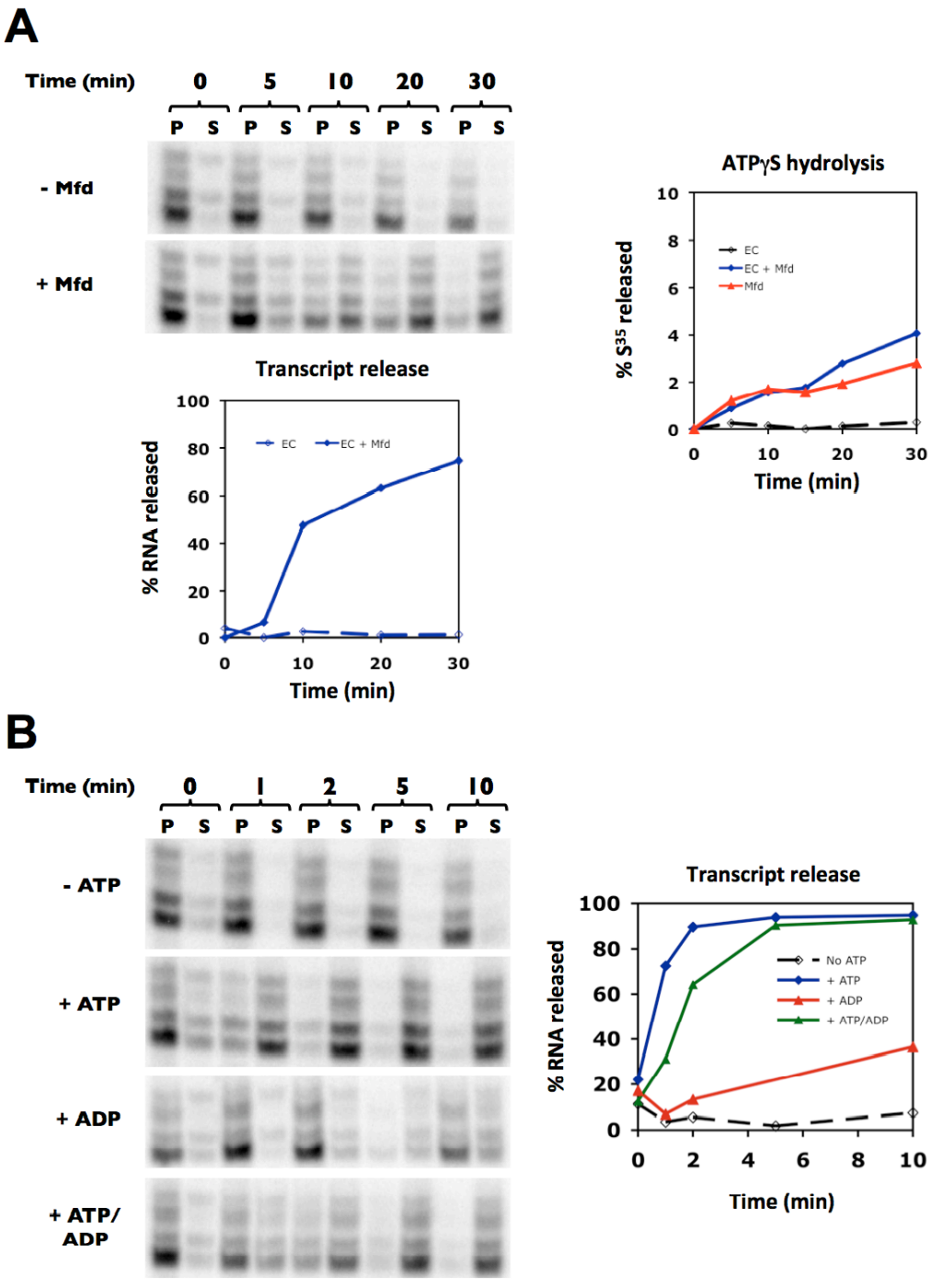
My initial experiment does not show a correlation between the two activities, but the presence of core RNAP and DNA in the reaction does stimulate Mfd ATPase activity. This is despite the fact that, previously, it was shown that Mfd must hydrolyze ATP to release stalled transcription complexes and that ATP hydrolysis by Mfd is greatly stimulated by the presence of stalled complexes as a substrate (Chambers et al., 2003; Selby and Sancar, 1995a; Smith et al., 2007). It is entirely possible that Mfd may require multiple rounds of ATP hydrolysis to dissociate a complex or it may only require one hydrolysis event to trigger dissociation. However unlikely, the ATP hydrolysis and transcription termination activities also may not be correlated. To further address how ATP use affects complex release, transcript release assays were performed with the adenosine nucleotide analogs ATPγS, AMP-PNP and ADP.

Mfd is able to cause transcript release when provided ATPγS instead of ATP (Figure A.3A). Previously, it was shown that Mfd can stably interact with DNA, resulting in a band shift by native-PAGE, in the presence of the slowly hydrolyzed ATP analog ATPγS, but not in the presence of ATP; the authors believed this was due to how quickly Mfd hydrolyses ATP versus ATPγS (Selby and Sancar, 1995a). It is somewhat surprising that Mfd is able to cause transcript release with ATPγS, even though it occurs much more slowly than release with ATP, taking 20 minutes or nearly 4 times as long to release the majority of complexes. Complex release also only occurs after a lag time of at least 5 minutes, which may be correlated with to the significantly

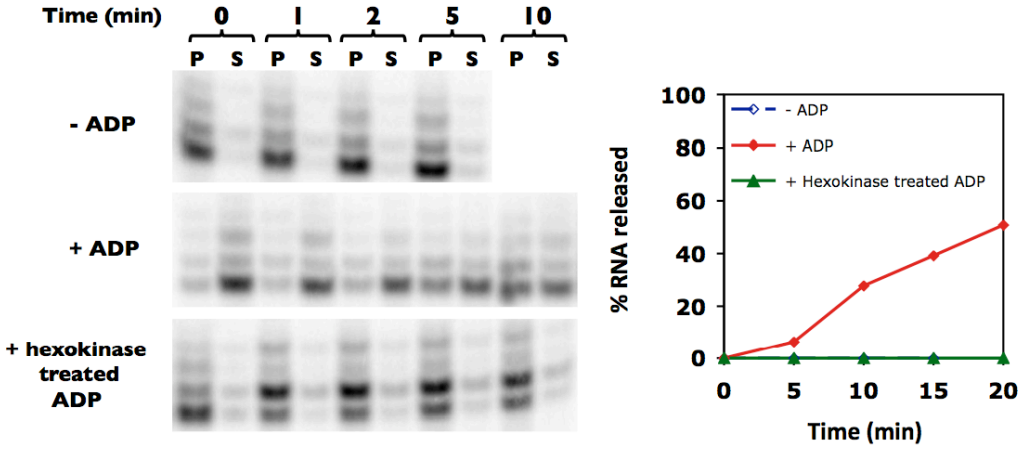
reduced rate of hydrolysis. The addition of ECs to Mfd in ATPase assays only slightly stimulates the rate of hydrolysis from 51.3 ATPγS/min to 65.9 ATPγS/min. Based on this, Mfd is hydrolyzing 6.1 ATPγS to release a stalled elongation complex, which indicates that very little nucleotide hydrolysis is actually required for Mfd to dissociate a stalled EC.

Unlike ATPγS, ADP does not support Mfd activity to dissociate transcription complexes (Figure A.3B). In the assay shown, transcript release in the presence of only ADP proceeds at a rate significantly slower that with ATP alone or a mix of ATP and ADP. The small amount of complex release seen is due to ATP contamination of the ADP used in the assay. When the transcript release assay was performed with ADP that was treated with hexokinase to hydrolyze any contaminating ATP, Mfd was unable to release any transcript (Figure A.3C). Mixing ATP and ADP slightly reduces the rate of transcript release presumably because of competition of ATP and ADP for binding the ATP binding pocket located in the translocase motif of the protein. Not surprisingly, the ADP analog AMP-PNP also is unable to stimulate complex dissociation by Mfd (data not shown). The effects of ATP analogs on Mfd transcription complex dissociation are consistent the current model of function, whereby Mfd uses the energy of ATP hydrolysis to act as a translocase and exert a force on the transcription complex and cause dissociation. The NTP bound form of Mfd is not the active form, as for some ATPases, rather actual hydrolysis and recycling is required.

Figure A.3. Mfd can utilize ATPγS but not ADP to release stalled elongation complexes. (A) Release of stalled transcripts by Mfd in the presence of ATPγS and associated ATPγS hydrolysis. Band intensities of pellet and supernatant fractions of +25 stalled complexes were quantified and amount of pause seen was plotted as % of complexes released over time. Associated ATP hydrolysis is plotted as % of ADP released. ECs alone are open blue diamonds/dashed lines; ECs with Mfd solid blue diamonds; Mfd alone is red. (B) Release of stalled transcripts by Mfd in the presence of ADP or ATP. Band intensities of pellet and supernatant fractions of +25 stalled complexes were quantified and amount of pause seen was plotted as % of complexes released over time. Release in the absence of nucleotide is open black diamonds; ATP is blue diamonds; ADP is red triangles and ADP/ATP mix is green triangles. (C) Release of stalled transcripts by Mfd in the presence of hexokinase treated ADP. Band intensities of pellet and supernatant fractions of +25 stalled complexes were quantified and amount of pause seen was plotted as % of complexes released over time. Release in the absence of nucleotide is open black diamonds; ADP is red diamonds; hexokinase treated ADP is green triangles.


A
Time (min)
0
5
10
20
30
P S P S P S P S P S
- Mfd
+ Mfd
Transcript release
% RNA released
EC
EC + Mfd
Time (min)
ATPγS hydrolysis
% S35 released
EC
EC + Mfd
Mfd
Time (min)
B
Time (min)
0
1
2
5
10
P S P S P S P S P S
- ATP
+ ATP
+ ADP
+ ATP/ ADP
Transcript release
% RNA released
No ATP
+ ATP
+ ADP
+ ATP/ADP
Time (min)

C


Time (min)
0
1
2
5
10
P S P S P S P S P S
- ADP
+ ADP
+ hexokinase treated ADP
% RNA released
100
80
60
40
20
0
0
5
10
15
20
Time (min)
- ADP
+ ADP
+ Hexokinase treated ADP

The UvrB homology region in addition to the autoinhibitory domain 7 plays roles in regulating the activity of Mfd

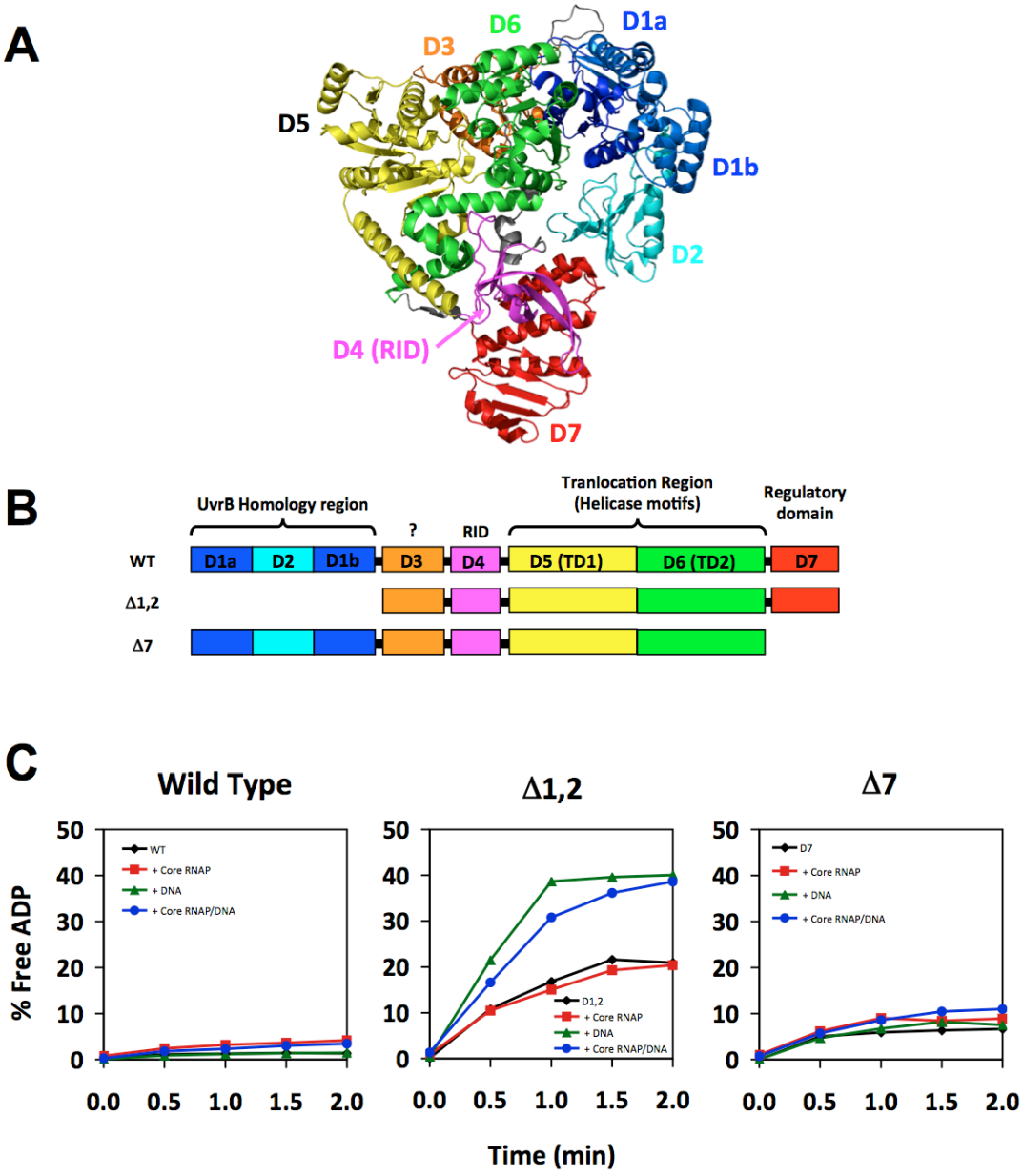
The structure and function of Mfd have been well characterized based on the known structure of the protein and similarities of individual domains to other proteins (Figure A.4A)(Chambers et al., 2003; Deaconescu et al., 2006; Selby and Sancar, 1995a, b). One particularly interesting function of the protein is its ability to self regulate. Domain 7 appears to be responsible for this regulation, as it inhibits the translocase and ATPase activities of Mfd (Smith et al., 2007). A deletion of this domain constitutively activates Mfd, removing the requirement of binding to a RNAP/DNA complex for the stimulation of Mfd's ATPase and translocase activities (Smith et al., 2007). A second interesting motif is Domains 1 and 2, which have high sequence and structural similarity to the nucleotide excision repair protein UvrB (Deaconescu et al., 2006). This motif is necessary for recruitment of the NER proteins to the site of DNA damage by Mfd (Deaconescu et al., 2006; Selby and Sancar, 1993a, 1995a, b).

In an ongoing collaboration with Michelle Wang (Cornell University), Mfd function is being probed by single molecule experiments that measure the force exerted on an elongation complex by Mfd. To aid in this, I have cloned and purified both a Domain 7 deletion mutant (MfdΔ7) and a UvrB homology motif deletion mutant (MfdΔ1,2)(Figure A.4B). Basal and stimulated ATPase activities for these two mutants were tested; consistent with the previous report for MfdΔ7, its basal ATPase activity was 5 fold higher than the wild type protein (Figure A.4C)(Smith et al., 2007). The stimulation of this activity by DNA is not as pronounced as previously reported,

but interestingly, the removal of Domain 7 allows a 2 fold stimulation of ATP hydrolysis by RNAP alone This is similar to what is seen when both DNA and RNAP are added to the reaction suggesting that the previous model of Domain 7 function is not correct. Based on the stimulated activity of the MfdΔ7 mutant, it was proposed that binding of RNAP to the RID displaces Domain 7 and allows DNA to interact with the translocase domains (Smith et al., 2007). The mechanism of action may actually require both interactions for wild type Mfd activity to be correctly stimulated.

The basal level of ATP hydrolysis by the MfdΔ1,2 mutant is significantly higher than the wild type protein by about 14 fold, and even higher than MfdΔ7 activity by approximately 2.5 fold (Figure A.4C). This suggests that Domains 1 and 2 in addition to Domain 7 are also playing a role in regulation of the protein's activity. This makes sense as Domains 1 and 2 make contacts to Domain 6 (part of the translocase motif) and Domain 7 (Figure A.4A)(Deaconescu et al., 2006). This is consistent with previous work showing that a Mfd Domain 1, 2 and 3 deletion mutant had elevated ATPase activity and was capable of acting as a DNA translocase in the absence of RNAP similar to the MfdΔ7 mutant (Manelyte et al., 2010; Murphy et al., 2009). This result implicates the UvrB homology region and its interaction with UvrA as also playing a part in the regulation of Mfd activity. MfdΔ1,2 ATPase activity is greatly stimulated by the presence of template DNA alone and by a core RNAP/DNA mixture, suggesting that displacement of the motif eases the access of the translocase motif by DNA. Follow up studies of these mutants is required to better understand how these motifs are involved in regulating the function of the protein. Hopefully their use in single molecule studies will help elucidate the role that they are playing.

Figure A.4. Mfd Domains 1 and 2 play a similar role to Domain 7 in regulating the mechanism of Mfd action. (A) Crystal structure of Mfd protein. Structure shown as ribbons. Domains are as follows: domains 1a and 1b, blue; domain 2, light blue; Domain 3, orange; domain 4 (RNAP Interaction domain), pink; translocase module: domain 5, yellow and domain 6, green; domain 7, red. Structure made using Pymol (Deaconescu et al. 2006) PDB file: 2EYQ**.** (B) Domain diagram of full-length Mfd and Δ1,2 and Δ7 domain deletion mutants. Domains are labeled as in (A). (C) ATPase activity assays of Wild type, Δ1,2 and Δ7 Mfd proteins. Mfd or mutant alone is black; Mfd or mutant + 100 nM Core RNAP is red; Mfd or mutant + 100 nM DNA template is green; Mfd or mutant + 100 nM Core RNAP and 100 nM DNA template is blue.


A
D6
D1a
D3
D5
D1b
D2
D4 (RID)
D7
B
UvrB Homology region
?
RID
Tranlocation Region
(Helicase motifs)
Regulatory
domain
WT
D1a
D2
D1b
D3
D4
D5 (TD1)
D6 (TD2)
D7
Δ1,2
Δ7
C
Wild Type
Δ1,2
Δ7
% Free ADP
WT
+ Core RNAP
+ DNA
+ Core RNAP/DNA
D1,2
D7
0
10
20
30
40
50
0.0
0.5
1.0
1.5
2.0
Time (min)

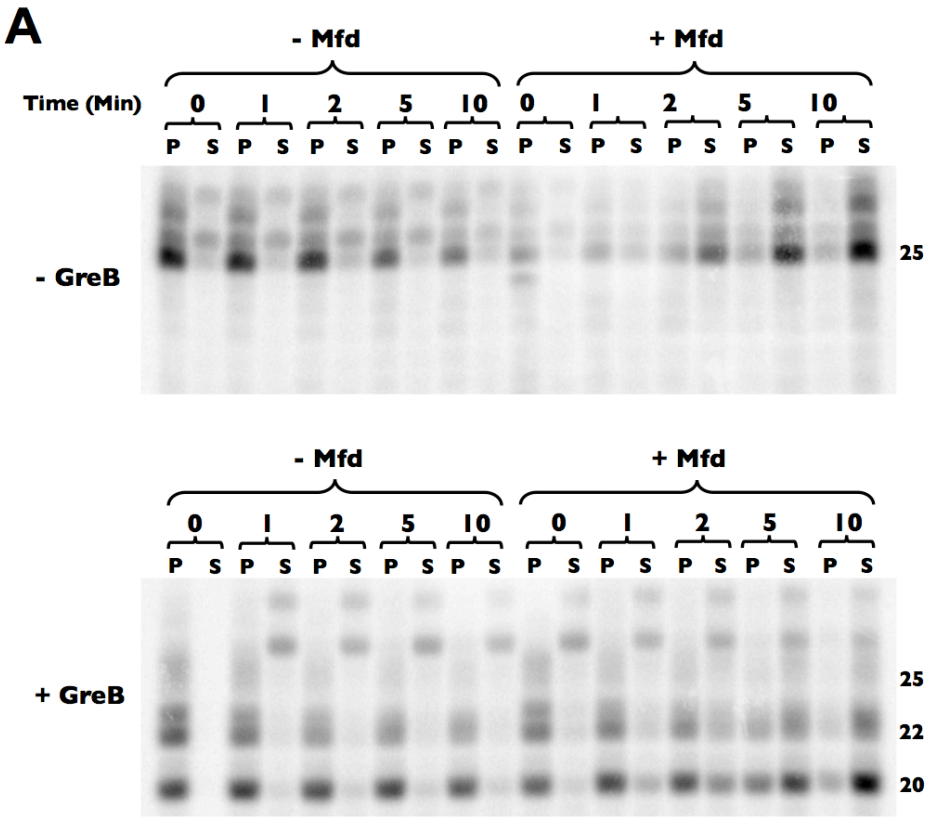
Treatment of BY409 stalled ECs with GreB reveals that they are not an ideal substrate for these studies

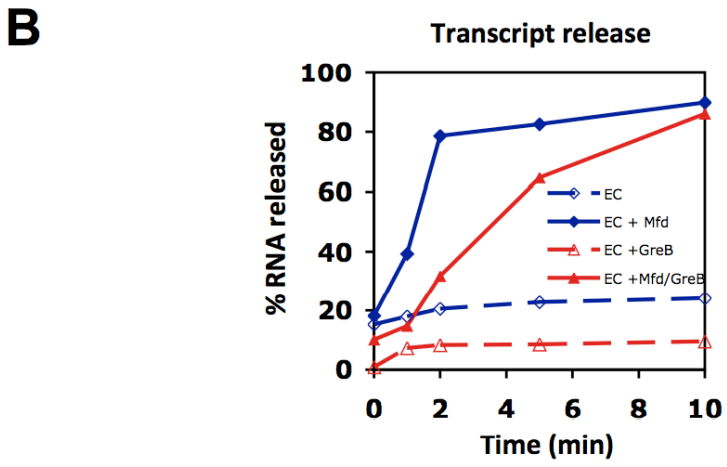
Park and Roberts (2002) previously looked at the effect of GreB on Mfd release of complexes on BY412 template (based on phage 82 *pR'*), which presumably retains σ70 at the site of the stall. Mfd was able to release complexes at +24 that were not cleaved by GreB, but not cleaved products at +21 and +22, that had resulted from backtracking. Based on this result, I would postulate that the complexes at +24 did not contain σ70, but were only stalled due to nucleotide starvation. Both the +21 and +22 complexes were originally backtracked due to σ70 forming a paused complex, similar to the +26 pause I have discussed in chapters two and three.

On the BY409 template, that contains a mutation to the -10-like element to disrupt σ70 interaction, Mfd is able to release complexes both in the absence and presence of GreB (Figure A.5A, B). Particularly, Mfd can release complexes at +20 and +22 that result from GreB cleaving the 3' overhanging RNA of backtracked complexes. This backtracking is due to the 3 basepair GC-rich sequence followed by an AT-rich region at the site of the pause and may affect the activity of Mfd on the substrate ECs (Perdue and Roberts, 2010). Interestingly, nearly all of the +25 stalled complexes formed on BY409 template appears to be backtracked as little to none of the +25 species remains when GreB is added to the reaction (Figure A.5A).

Due to the complicated nature of the σ70-dependent pause sequence at the site of these nucleotide starved complexes, the BY409 and BY412 templates are not a good substrate for determining ATP use by Mfd during complex release. GreB cleavage of complexes stalled at +25 by nucleotide starvation shows this, as the

Figure A.5. The effect of GreB on transcription complex release. (A) Release of stalled transcripts by Mfd in the absence and presence of GreB as labeled. (B) Band intensities of pellet and supernatant fractions of +25 stalled complexes (- GreB) and +20 complexes (+ GreB) from (A) were quantified and amount of pause seen was plotted as % of complexes released over time. Complex release in the absence of GreB is blue diamonds; Complex release in the presence of GreB is red triangles. Solid lines and markers indicate time courses with Mfd; Dashed lines and open markers, time courses in the absence of Mfd.


A
- Mfd
+ Mfd
Time (Min)
0 1 2 5 10 0 1 2 5 10
P S P S P S P S P S P S P S P S P S P S
- GreB
25
- Mfd
+ Mfd
0 1 2 5 10 0 1 2 5 10
P S P S P S P S P S P S P S P S P S P S
+ GreB
25
22
20


B
Transcript release
% RNA released
100
80
60
40
20
0
0 2 4 6 8 10
Time (min)
EC
EC + Mfd
EC +GreB
EC +Mfd/GreB

complexes are cleaved at different lengths, suggesting that there is a mixed population of complexes. In addition, Mfd is able to forward translocate backtracked complexes, which potentially requires energy from ATP hydrolysis as does complex release, due to its resulting from the same "pushing" or forward translocation activity. Also, the mutation that was used to disrupt the σ70/-10-like element was a single base pair mutation to +10 from A:T to G:C. This single base pair mutation significantly reduces the σ70-dependent pause, but does not completely eliminate it (data not shown).

The ideal substrate for Mfd joint complex release/ATPase assays for this investigation would be non-backtracked complexes stalled at a sequence that is not a previously determined pause site. Most likely, the easiest way to achieve this is to construct complexes using what is referred to as a DNA/RNA scaffold. Complexes can be constructed by incubating purified individual strands of DNA and RNA with core RNAP, which are capable of transcription and are assumed to be in a state similar to an EC. This method has been used both to study elongation and construct a number of *T. thermophilus* initiation complexes for X-ray crystallography (Zhang et al., 2012). Constructing such a template also is possible using Quickchange mutagenesis to substitute a different non-pausing sequence downstream of the promoter on pBY409. This would allow for more accurate measurements of ATP use during complex release.

Discussion

There does not appear to be a correlation between ATP hydrolysis and complex dissociation by Mfd, since the stimulation of Mfd ATPase activity continues well after

Mfd has released all of the substrate stalled ECs in a transcript release assay (Figure A.1A). This appears to be due to core RNAP forming complexes on free DNA in the absence of a σ factor that are in such a conformation that Mfd can act on them (Figure A.1C). ATP hydrolysis by Mfd is stimulated 5 fold by the presence of core RNAP and DNA in similar concentrations to that present in the complex release assay. This background stimulation is a difficulty that must be overcome in order to accurately measure ATP usage during transcription complex release. One possibility to do this is by adding excess σ70 to capture free RNAP at the promoter in an initiation complex that Mfd does not act on (Figure A.2)(Park et al., 2002). Further experiments including a s70 titration experiment to find the ideal amount to add to properly inhibit background activity will be required. Ideally, a newly designed DNA template for stalled complex formation should also help make joint complex release/ATPase assays more feasible. Mechanistically, Mfd is tough to work on, but with careful experimentation, finding a correlation between ATP hydrolysis and transcription complex release should be possible.

A second interesting finding about Mfd activity is that it can hydrolyze ATPgS and use the energy of that to release stalled transcription complexes. What is particularly surprising is that transcript release with ATPγS occurs only 3 fold more slowly than with ATP. This is in contrast to other known ATPase such as ClpX, which shows a 20x slower rate of ATPγS hydrolysis than ATP hydrolysis (Baker and Sauer, 2012).

Another interesting aspect of complex release by Mfd in the presence of ATPγS is that there is a lag time of about 5 minutes in the assay from the addition of

Mfd to the release of the first complexes. Based on the results of Figure A.1, Mfd hydrolyzes approximately 48.5 ATP per elongation complex released. The rate of ATPγS hydrolysis by Mfd acting on ECs is not dramatically different from the rate of Mfd alone (65.9 versus 51.3 ATPγS/min), which is significantly lower than what was seen with ATP. Also the results seen when performing the EC release with ATPγS indicate that relatively few hydrolysis events are required for the release of a complex (6.1 ATPγS/complex release). It is possible that Mfd needs less than the 6.1 ATPγS , and possibly one hydrolysis event to cause complex release. Although, considering that the mechanism of release appears to be due to the translocase activity, multiple hydrolysis events are most likely required, especially considering different substrate complexes may be in different conformational states that require more or less energy to dissociate.

In addition to looking at the enzymatic activities of Mfd, there is more to be learned about function of the protein from structural studies. Previously, Mfd Domain 7 was shown to be necessary for regulating the translocase and ATP hydrolysis activities of the protein (Smith et al., 2007). It is clear that Domains 1 and 2 are also are playing a role in this self regulation because the MfdΔ1,2 deletion mutant shows significantly more basal ATPase activity than either the wild type or MfdD7 mutant (Figure A.4). This is consistent with what has been observed with similar deletion mutants that have been tested (Murphy et al., 2009). Hopefully these mutants can help reveal important mechanistic details of Mfd translocase activity as it acts on RNAP complexes in single molecule experiments.

There are also a whole set of questions the single molecule studies performed in the laboratory of Michelle Wang can answer about Mfd action. One particular question is what happens after Mfd releases a transcription complex. It must remain near the site of DNA damage in order to recruit NER proteins to repair it, but if it is properly loaded onto DNA it is entirely possible that it will continue translocating downstream in the absence of the transcription complex. The translocase motif may be able to recognize the DNA damage and stop the protein at the site until UvrA can be recruited. Considering the role the UvrB homology role plays in regulating the protein it is possible that Mfd will not release a DNA damage stalled EC until the NER proteins are recruited through contacts with UvrA and as Mfd clears away the RNAP complex, the NER proteins go to work (Manelyte et al., 2010).

REFERENCES

Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. Nature *438*, 460-465.

Adelman, K., La Porta, A., Santangelo, T.J., Lis, J.T., Roberts, J.W., and Wang, M.D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. Proc Natl Acad Sci U S A *99*, 13538-13543.

Adhya, S., Gottesman, M., and De Crombrugghe, B. (1974). Release of polarity in Escherichia coli by gene N of phage lambda: termination and antitermination of transcription. Proc Natl Acad Sci U S A *71*, 2534-2538.

Artsimovitch, I., and Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci U S A *97*, 7090-7095.

Artsimovitch, I., and Landick, R. (2002). The transcriptional regulator RfaH stimulates RNA chain synthesis after recruitment to elongation complexes by the exposed nontemplate DNA strand. Cell *109*, 193-203.

Bai, L., Fulbright, R.M., and Wang, M.D. (2007). Mechanochemical kinetics of transcription elongation. Phys Rev Lett *98*, 068103.

Baker, T.A., and Sauer, R.T. (2012). ClpXP, an ATP-powered unfolding and protein-degradation machine. Biochim Biophys Acta *1823*, 15-28.

Barne, K.A., Bown, J.A., Busby, S.J., and Minchin, S.D. (1997). Region 2.5 of the Escherichia coli RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. EMBO J *16*, 4034-4040.

Bartlett, E. (1998). Characterization of the lambda Q binding site. In Molecular Biology and Genetics (Ithaca, NY: Cornell).

Belogurov, G.A., Vassylyeva, M.N., Svetlov, V., Klyuyev, S., Grishin, N.V., Vassylyev, D.G., and Artsimovitch, I. (2007). Structural basis for converting a general transcription factor into an operon-specific virulence regulator. Mol Cell *26*, 117-129.

Benoff, B., Yang, H., Lawson, C.L., Parkinson, G., Liu, J., Blatter, E., Ebright, Y.W., Berman, H.M., and Ebright, R.H. (2002). Structural basis of transcription activation: the CAP-alpha CTD-DNA complex. Science *297*, 1562-1566.

Borukhov, S., and Goldfarb, A. (1996). Purification and assay of Escherichia coli transcript cleavage factors GreA and GreB. Methods Enzymol *274*, 315-326.

Borukhov, S., Polyakov, A., Nikiforov, V., and Goldfarb, A. (1992). GreA protein: a transcription elongation factor from Escherichia coli. Proc Natl Acad Sci U S A *89*, 8899-8902.

Borukhov, S., Sagitov, V., and Goldfarb, A. (1993). Transcript cleavage factors from E. coli. Cell *72*, 459-466.

Brodolin, K., Zenkin, N., Mustaev, A., Mamaeva, D., and Heumann, H. (2004). The sigma 70 subunit of RNA polymerase induces lacUV5 promoter-proximal pausing of transcription. Nat Struct Mol Biol *11*, 551-557.

Busby, S., and Ebright, R.H. (1999). Transcription activation by catabolite activator protein (CAP). J Mol Biol *293*, 199-213.

Cardinale, C.J., Washburn, R.S., Tadigotla, V.R., Brown, L.M., Gottesman, M.E., and Nudler, E. (2008). Termination factor Rho and its cofactors NusA and NusG silence foreign DNA in E. coli. Science *320*, 935-938.

Chambers, A.L., Smith, A.J., and Savery, N.J. (2003). A DNA translocation motif in the bacterial transcription--repair coupling factor, Mfd. Nucleic Acids Res *31*, 6409-6418.

Chander, M., Austin, K.M., Aye-Han, N.N., Sircar, P., and Hsu, L.M. (2007). An alternate mechanism of abortive release marked by the formation of very long abortive transcripts. Biochemistry *46*, 12687-12699.

Chen, H., Tang, H., and Ebright, R.H. (2003). Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription. Mol Cell *11*, 1621-1633.

Chen, J., Darst, S.A., and Thirumalai, D. (2010). Promoter melting triggered by bacterial RNA polymerase occurs in three steps. Proc Natl Acad Sci U S A *107*, 12523-12528.

Ciampi, M.S. (2006). Rho-dependent terminators and transcription termination. Microbiology *152*, 2515-2528.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science *322*, 1845-1848.

Cramer, P. (2002). Multisubunit RNA polymerases. Curr Opin Struct Biol *12*, 89-97.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. Science *292*, 1863-1876.

Deaconescu, A.M., Chambers, A.L., Smith, A.J., Nickels, B.E., Hochschild, A., Savery, N.J., and Darst, S.A. (2006). Structural basis for bacterial transcription-coupled DNA repair. Cell *124*, 507-520.

Deaconescu, A.M., and Darst, S.A. (2005). Crystallization and preliminary structure determination of Escherichia coli Mfd, the transcription-repair coupling factor. Acta Crystallogr Sect F Struct Biol Cryst Commun *61*, 1062-1064.

Deighan, P., Diez, C.M., Leibman, M., Hochschild, A., and Nickels, B.E. (2008). The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. Proc Natl Acad Sci U S A *105*, 15305-15310.

Deighan, P., and Hochschild, A. (2007). The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription elongation complex. Mol Microbiol *63*, 911-920.

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

Filter, J.J., and Roberts, J.W. Unpublished data.

Franklin, N.C. (1974). Altered reading of genetic signals fused to the N operon of bacteriophage lambda: genetic evidence for modification of polymerase by the protein product of the N gene. J Mol Biol *89*, 33-48.
Gilbert, W., Maizels, N., and Maxam, A. (1974). Sequences of controlling regions of the lactose operon. Cold Spring Harb Symp Quant Biol *38*, 845-855.
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 A resolution. Science *292*, 1876-1882.
Gong, X.Q., Zhang, C., Feig, M., and Burton, Z.F. (2005). Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. Mol Cell *18*, 461-470.
Grayhack, E.J., Yang, X.J., Lau, L.F., and Roberts, J.W. (1985). Phage lambda gene Q antiterminator recognizes RNA polymerase near the promoter and accelerates it through a pause site. Cell *42*, 259-269.
Guo, H.C., and Roberts, J.W. (1990). Heterogeneous initiation due to slippage at the bacteriophage 82 late gene promoter in vitro. Biochemistry *29*, 10702-10709.
Hatoum, A., and Roberts, J. (2008). Prevalence of RNA polymerase stalling at Escherichia coli promoters after open complex formation. Mol Microbiol *68*, 17-28.
Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. Cell *125*, 1069-1082.
Haugen, S.P., Ross, W., Manrique, M., and Gourse, R.L. (2008). Fine structure of the promoter-sigma region 1.2 interaction. Proc Natl Acad Sci U S A *105*, 3292-3297.
Hein, P.P., and Landick, R. (2010). The bridge helix coordinates movements of modules in RNA polymerase. BMC Biol *8*, 141.
Hein, P.P., Palangat, M., and Landick, R. (2011). RNA transcript 3'-proximal sequence affects translocation bias of RNA polymerase. Biochemistry *50*, 7002-7014.
Henkin, T.M., and Yanofsky, C. (2002). Regulation by transcription attenuation in bacteria: how RNA provides instructions for transcription termination/antitermination decisions. Bioessays *24*, 700-707.
Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block, S.M. (2006). Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell *125*, 1083-1094.
Hsu, L.M., Vo, N.V., and Chamberlin, M.J. (1995). Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. Proc Natl Acad Sci U S A *92*, 11588-11592.
Kainz, M., and Roberts, J. (1992). Structure of transcription elongation complexes in vivo. Science *255*, 838-841.
Kainz, M., and Roberts, J.W. (1995). Kinetics of RNA polymerase initiation and pausing at the lambda late gene promoter in vivo. J Mol Biol *254*, 808-814.

Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science *314*, 1144-1147.

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

Kenniston, J.A., Baker, T.A., Fernandez, J.M., and Sauer, R.T. (2003). Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. Cell *114*, 511-520.

Kireeva, M., Kashlev, M., and Burton, Z.F. (2010). Translocation by multi-subunit RNA polymerases. Biochim Biophys Acta *1799*, 389-401.

Ko, D.C., Marr, M.T., Guo, J., and Roberts, J.W. (1998). A surface of Escherichia coli sigma 70 required for promoter function and antitermination by phage lambda Q protein. Genes Dev *12*, 3276-3285.

Komissarova, N., Becker, J., Solter, S., Kireeva, M., and Kashlev, M. (2002). Shortening of RNA:DNA hybrid in the elongation complex of RNA polymerase is a prerequisite for transcription termination. Mol Cell *10*, 1151-1162.

Komissarova, N., and Kashlev, M. (1997a). RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J Biol Chem *272*, 15329-15338.

Komissarova, N., and Kashlev, M. (1997b). Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc Natl Acad Sci U S A *94*, 1755-1760.

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. Science *289*, 619-625.

Kyzer, S., Ha, K.S., Landick, R., and Palangat, M. (2007). Direct versus limited-step reconstitution reveals key features of an RNA hairpin-stabilized paused transcription complex. J Biol Chem *282*, 19020-19028.

Landick, R. (2006). The regulatory roles and mechanism of transcriptional pausing. Biochem Soc Trans *34*, 1062-1066.

Landick, R. (2009). Transcriptional pausing without backtracking. Proc Natl Acad Sci U S A *106*, 8797-8798.

Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. Cell *132*, 971-982.

Lau, L.F., Roberts, J.W., and Wu, R. (1983). RNA polymerase pausing and transcript release at the lambda tR1 terminator in vitro. J Biol Chem *258*, 9391-9397.

Lee, D.N., Phung, L., Stewart, J., and Landick, R. (1990). Transcription pausing by Escherichia coli RNA polymerase is modulated by downstream DNA sequences. J Biol Chem *265*, 15145-15153.

Liu, X., Bushnell, D.A., and Kornberg, R.D. (2011). Lock and key to transcription: sigma-DNA interaction. Cell *147*, 1218-1219.

Liu, X.Q., and Roberts, J.W. Unpublished data.
Lowery-Goldhammer, C., and Richardson, J.P. (1974). An RNA-dependent nucleoside triphosphate phosphohydrolase (ATPase) associated with rho termination factor. Proc Natl Acad Sci U S A *71*, 2003-2007.
Mahdi, A.A., Briggs, G.S., Sharples, G.J., Wen, Q., and Lloyd, R.G. (2003). A model for dsDNA translocation revealed by a structural motif common to RecG and Mfd proteins. EMBO J *22*, 724-734.
Manelyte, L., Kim, Y.I., Smith, A.J., Smith, R.M., and Savery, N.J. (2010). Regulation and rate enhancement during transcription-coupled DNA repair. Mol Cell *40*, 714-724.
Marr, M.T., Datwyler, S.A., Meares, C.F., and Roberts, J.W. (2001). Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc Natl Acad Sci U S A *98*, 8972-8978.
Marr, M.T., and Roberts, J.W. (1997). Promoter recognition as measured by binding of polymerase to nontemplate strand oligonucleotide. Science *276*, 1258-1260.
Marr, M.T., and Roberts, J.W. (2000). Function of transcription cleavage factors GreA and GreB at a regulatory pause site. Mol Cell *6*, 1275-1285.
Martin, C.T. Personal Communication.
Minakhin, L., Bhagat, S., Brunning, A., Campbell, E.A., Darst, S.A., Ebright, R.H., and Severinov, K. (2001). Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. Proc Natl Acad Sci U S A *98*, 892-897.
Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? Mol Cell *20*, 335-345.
Mooney, R.A., and Landick, R. (2003). Tethering sigma70 to RNA polymerase reveals high in vivo activity of sigma factors and sigma70-dependent pausing at promoter-distal locations. Genes Dev *17*, 2839-2851.
Morgan, W.D., Bear, D.G., Litchman, B.L., and von Hippel, P.H. (1985). RNA sequence and secondary structure requirements for rho-dependent transcription termination. Nucleic Acids Res *13*, 3739-3754.
Mukhopadhyay, J., Das, K., Ismail, S., Koppstein, D., Jang, M., Hudson, B., Sarafianos, S., Tuske, S., Patel, J., Jansen, R.*, et al.* (2008). The RNA polymerase "switch region" is a target for inhibitors. Cell *135*, 295-307.
Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the wholo story. Curr Opin Struct Biol *13*, 31-39.
Murphy, M.N., Gong, P., Ralto, K., Manelyte, L., Savery, N.J., and Theis, K. (2009). An N-terminal clamp restrains the motor domains of the bacterial transcription-repair coupling factor Mfd. Nucleic Acids Res *37*, 6042-6053.
Nechaev, S., Fargo, D.C., dos Santos, G., Liu, L., Gao, Y., and Adelman, K. (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science *327*, 335-338.
Neuman, K.C., Abbondanzieri, E.A., Landick, R., Gelles, J., and Block, S.M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. Cell *115*, 437-447.

Nickels, B.E., and Hochschild, A. Unpublished data.
Nickels, B.E., Mukhopadhyay, J., Garrity, S.J., Ebright, R.H., and Hochschild, A. (2004). The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. Nat Struct Mol Biol *11*, 544-550.
Nickels, B.E., Roberts, C.W., Sun, H., Roberts, J.W., and Hochschild, A. (2002). The sigma(70) subunit of RNA polymerase is contacted by the (lambda)Q antiterminator during early elongation. Mol Cell *10*, 611-622.
Nudler, E., Avetissova, E., Markovtsov, V., and Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. Science *273*, 211-217.
Nudler, E., and Gottesman, M.E. (2002). Transcription termination and anti-termination in E. coli. Genes Cells *7*, 755-768.
Nudler, E., Mustaev, A., Lukhtanov, E., and Goldfarb, A. (1997). The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. Cell *89*, 33-41.
Orlova, M., Newlands, J., Das, A., Goldfarb, A., and Borukhov, S. (1995). Intrinsic transcript cleavage activity of RNA polymerase. Proc Natl Acad Sci U S A *92*, 4596-4600.
Park, J.S., Marr, M.T., and Roberts, J.W. (2002). E. coli Transcription repair coupling factor (Mfd protein) rescues arrested complexes by promoting forward translocation. Cell *109*, 757-767.
Perdue, S.A., and Roberts, J.W. (2010). A backtrack-inducing sequence is an essential component of Escherichia coli sigma(70)-dependent promoter-proximal pausing. Mol Microbiol *78*, 636-650.
Perdue, S.A., and Roberts, J.W. (2011). Sigma(70)-dependent transcription pausing in Escherichia coli. J Mol Biol *412*, 782-792.
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.
Richardson, J.P. (2002). Rho-dependent termination and ATPases in transcript termination. Biochim Biophys Acta *1577*, 251-260.
Richardson, J.P., Grimley, C., and Lowery, C. (1975). Transcription termination factor rho activity is altered in Escherichia coli with suA gene mutations. Proc Natl Acad Sci U S A *72*, 1725-1728.
Ring, B.Z., and Roberts, J.W. (1994). Function of a nontranscribed DNA strand site in transcription elongation. Cell *78*, 317-324.
Ring, B.Z., Yarnell, W.S., and Roberts, J.W. (1996). Function of E. coli RNA polymerase sigma factor sigma 70 in promoter-proximal pausing. Cell *86*, 485-493.
Roberts, J.W., Shankar, S., and Filter, J.J. (2008). RNA polymerase elongation factors. Annu Rev Microbiol *62*, 211-233.
Roberts, J.W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D.C., Sun, H., and Roberts, C.W. (1998). Antitermination by bacteriophage lambda Q protein. Cold Spring Harb Symp Quant Biol *63*, 319-325.

Ross, W., Gosink, K.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.

Ruggerberg, K.G., and Hay, A.G. (2013). personal communication.

Santangelo, T.J., Mooney, R.A., Landick, R., and Roberts, J.W. (2003). RNA polymerase mutations that impair conversion to a termination-resistant complex by Q antiterminator proteins. Genes Dev *17*, 1281-1292.

Santangelo, T.J., and Roberts, J.W. (2004). Forward translocation is the natural pathway of RNA release at an intrinsic terminator. Mol Cell *14*, 117-126.

Saunders, A., Core, L.J., and Lis, J.T. (2006). Breaking barriers to transcription elongation. Nat Rev Mol Cell Biol *7*, 557-567.

Selby, C.P., and Sancar, A. (1993a). Molecular mechanism of transcription-repair coupling. Science *260*, 53-58.

Selby, C.P., and Sancar, A. (1993b). Transcription-repair coupling and mutation frequency decline. J Bacteriol *175*, 7509-7514.

Selby, C.P., and Sancar, A. (1995a). Structure and function of transcription-repair coupling factor. I. Structural domains and binding properties. J Biol Chem *270*, 4882-4889.

Selby, C.P., and Sancar, A. (1995b). Structure and function of transcription-repair coupling factor. II. Catalytic properties. J Biol Chem *270*, 4890-4895.

Shankar, S., Hatoum, A., and Roberts, J.W. (2007). A transcription antiterminator constructs a NusA-dependent shield to the emerging transcript. Mol Cell *27*, 914-927.

Skordalakes, E., and Berger, J.M. (2006). Structural insights into RNA-dependent ring closure and ATPase activation by the Rho termination factor. Cell *127*, 553-564.

Smith, A.J., and Savery, N.J. (2005). RNA polymerase mutants defective in the initiation of transcription-coupled DNA repair. Nucleic Acids Res *33*, 755-764.

Smith, A.J., and Savery, N.J. (2008). Effects of the bacterial transcription-repair coupling factor during transcription of DNA containing non-bulky lesions. DNA Repair (Amst) *7*, 1670-1679.

Smith, A.J., Szczelkun, M.D., and Savery, N.J. (2007). Controlling the motor activity of a transcription-repair coupling factor: autoinhibition and the role of RNA polymerase. Nucleic Acids Res *35*, 1802-1811.

Straney, D.C., and Crothers, D.M. (1987). A stressed intermediate in the formation of stably initiated RNA chains at the Escherichia coli lac UV5 promoter. J Mol Biol *193*, 267-278.

Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T., Yoneyama, M., and Sasaki, M. (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. Biochemistry *34*, 11211-11216.

Temiakov, D., Zenkin, N., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N.*, et al.* (2005). Structural basis of transcription inhibition by antibiotic streptolydigin. Mol Cell *19*, 655-666.

Toulokhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. Science *292*, 730-733.

Tuske, S., Sarafianos, S.G., Wang, X., Hudson, B., Sineva, E., Mukhopadhyay, J., Birktoft, J.J., Leroy, O., Ismail, S., Clark, A.D., Jr.*, et al.* (2005). Inhibition of bacterial RNA polymerase by streptolydigin: stabilization of a straight-bridge-helix active-center conformation. Cell *122*, 541-552.

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

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. Nature *448*, 163-168.

Yarnell, W.S., and Roberts, J.W. (1992). The phage lambda gene Q transcription antiterminator binds DNA in the late gene promoter as it modifies RNA polymerase. Cell *69*, 1181-1189.

Yarnell, W.S., and Roberts, J.W. (1999). Mechanism of intrinsic transcription termination and antitermination. Science *284*, 611-615.

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 A resolution. Cell *98*, 811-824.

Zhang, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. Nat Struct Mol Biol *17*, 99-104.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. Science *338*, 1076-1080.