REGULATION OF TRANSCRIPTION ELONGATION

VIA INTERACTIONS OF RNA POLYMERASE WITH

SEQUENCE ELEMENTS AND ACCESSORY FACTORS

Erika L. Pearson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the
Department of Chemistry.

Chapel Hill

2008

Approved by:

Advisor: Dr. Dorothy Erie

Reader: Dr. Thomas Kunkel

Reader: Dr. Kevin Weeks

Chair: Dr. Gary Pielak

Dr. Linda Spremulli

ABSTRACT

ERIKA L. PEARSON

Regulation of Transcription Elongation Via Interactions of RNA
Polymerase with Sequence Elements and Accessory Factors
(Under the direction of Dr. Dorothy A. Erie)

RNA polymerase (RNAP) is the enzyme responsible for catalyzing the first step in a cascade of events leading towards gene expression. Since the dissemination of genetic information plays such a critical and integral role within the context of cellular processes, RNAP must uphold the stringent guidelines of catalyzing processive DNA-directed synthesis of RNA transcripts at a reasonable rate and with high fidelity. To ensure that these requirements are met and sustained, RNAP serves as a major target of rigorous regulation and is consequently subject to extrinsic regulatory factors, such as transcription accessory proteins, and to regulatory sequence elements, such as pause and termination sites located throughout the genome.

Chapter 2 explores the ability of RNAP to recognize a terminally misincorporated base in the nascent RNA transcript and to cleave that base from the transcript. Additionally, the effects of GreA and GreB, transcript cleavage factors, on RNAP containing terminally misincorporated transcript are also investigated. My results indicate that RNAP does not preferentially cleave a misincorporated base. In fact, the

GreB-mediated cleavage of terminally correct and terminally incorrect bases is indistinguishable. I conclude that, in the event of misincorporation, decay into the activated pathway serves as a signal to recruit GreA and GreB. Through Gre-mediated action, RNAP may backtrack along the nascent transcript and cleave the segment of RNA containing the error.

Chapter 3 studies the role of downstream elements in RNAP in regulating transcription elongation. RNAPs harboring mutations in two regions (Walker B motif and fork loop 2) composing the allosteric NTP binding site are studied for their ability to recognize pause and termination sites. I demonstrate that the Walker B motif is important for the recognition of pause and termination sites. The recognition of termination signals by the Walker B motif is also important *in vivo*. Additionally, I propose that the formation of two classes of pause states does not proceed through a common slow intermediate, as previously thought.

Chapter 4 presents the development of method by which the interaction of a transcription accessory factor with a stalled elongation complex can be probed. This method utilizes the technique of through-prism total internal reflection fluorescence (TIRF) microscopy. I present single-molecule data that permit me to estimate the lifetime and the dissociation constant of NusG.

For Ashley and Freya

ACKNOWLEDGEMENTS

Throughout these past six years, there has been a variety of people who have made this work possible, and as such I wish to acknowledge them here.

First, I wish to offer many thanks to my graduate research advisor, Dr. Dorothy Erie. Dorothy has allowed me to work in an untethered fashion in her lab for the majority of my tenure as a graduate student. In my interactions with several of my peers, it is quite clear that my natural inclination towards independent thought and exploration may not have been as supported and as cultivated under the direction of most other advisors. While this opportunity for independence permitted an extraordinary level of personal and scientific development, I readily admit that Dorothy, on a number of occasions, has had to reel me back in (so as to prevent me from reinventing the wheel a little too much), a point in which we derive much humor. Additionally, I would like to thank the members of my doctoral committee, Dr. Gary Pielak, Dr. Thomas Kunkel, Dr. Kevin Weeks, and Dr. Linda Spremulli. Tom was especially helpful in the review and the revision of this work, and for that I am particularly grateful. And of course, it is only right to give credit to Dr. Barry Lentz of the Biophysics Program for giving me tough skin and a dynamite ability to deliver a presentation. Finally, I wish to give special thanks to my undergraduate research advisor, the late Dr. Donald Nelson of Clark University. Don has had an enormously influential hand in shaping the scientist that I am today.

My scientific peers have provided me with much needed support, perhaps unsung until now. The members of the Erie Laboratory have each contributed to my growth in his/her own way. Dr. Scott Kennedy provided me with mutant RNA polymerase plasmids, which made possible the research presented in Chapter 3. Dan Burke introduced much needed levity and comic relief with swing dancing, bad jokes, and "Target Women" episodes. And of course, who could forget his famous chocolate mousse? Cherie Lanyi provided input in helpful discussions. Dr. Junghoon In gave a presence that made me happy to be in lab and around her (and of course driving me home on several occasions, helping me move, and keeping me sane throughout the dissertation writing experience). Vanessa DeRocco was extremely instrumental, not only scientifically but also personally. In addition to fielding rudimentary questions on fluorescence microscopy (which were also posed to the incredible Dr. Lauryn Sass!), Vanessa touched me deeply with her heartfelt friendship in both critical moments and in everyday situations. And of course, there's Dr. Liz Sacho, whom I first met when I moved to Chapel Hill back in 2002. Liz became my first true friend here, and our friendship has continued to grow every since.

Another one of my peers from day one is Dr. Denise Teotico, formerly of the Redinbo Laboratory. In my time here, Denise has been a good friend, room mate, and fellow scientist. And last (but certainly not least!), Caia Duncan of the Weeks Laboratory is a scientist, for whom I have enormous respect, and a close friend who has celebrated my "curiousness" and has educated me on the finer points of metal and euro-techno.

Friends peripheral to the scientific community have provided me with much stability, amusement and care. In particular, Lee Daub and Casey Nordell have gallivanted around the Triangle with me and have enjoyed many adventures, of which the memories will always remain fresh in my mind. I also wish to sincerely thank the members of the Choshinkan Dojo of Durham. I don't think I could possibly convey to this community, and in particular Sensei Danieley and Sensei Hazelrigg, all that they have done for me by acting as a surrogate family and unknowingly helping me to overcome one of the greatest challenges of my life thus far.

I also wish to give special thanks to Dr. Colleen Duffy and Dr. Amy Rountree for their continued support and professional expertise. I could not have made this journey without them.

And finally, my acknowledgements would not be complete without referencing my family, the greatest collection of unsung heroes whom I love dearly. Aunt Liz, Aunt Karen, and Uncle Don have all kept this lonely girl company from 800 miles away with supportive phone calls and loving and encouraging letters. Perhaps the male version of me, Erik has had an amazing sensitivity to all things nuanced and a deep understanding of many things unspoken. Rob has proven to me in the most crucial moments that there is more to this world than what meets the eye. My late stepfather, Dr. Herb Richmond, would have wanted to celebrate this moment with me, but celebration through spirit will do just fine! My dearest late Freya, there are some things that only she knows and I love her from the other side of existence. My sister Ashley, my worst tormentor and my fiercest protector, is the yang to my yin and knows me better than I think I know myself. And finally, I would not be here today if it weren't for my mother, Linda, who worked several jobs as a single parent and went without so that my sister and I might have everything we could ever need.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Transcription: A Matter of Perspective

The last sixty years have witnessed an unprecedented burst in knowledge derived from the study of biological sciences. Arguably one of the greatest cornerstones that lay the foundation for such a burst in achievement was Rosalind Franklin's illuminating study of the architecture of DNA by means of x-ray crystallography. Franklin's work undoubtedly influenced Watson and Crick to report the elucidated structure of double helical DNA in their seminal paper (Watson and Crick, 1953). A critical discovery that had once remained elusive to scientists, the solved structure of the double helix revealed an elegant, yet simple blueprint for life and served to provoke questions regarding the maintenance and dissemination of molecular information encoded within the DNA sequence.

How is this information transmitted and processed to direct the trajectory of life? Among the many cellular processes that hold the answers to this question, the process known as transcription is one that describes the initial step in a series of cascading events leading towards the expression of genes. As the name suggests, transcription is the act of directly transcribing information encoded within DNA via the generation of a sister

biomolecule, RNA, a replica of the coding strand of DNA. The information contained in this newly synthesized transcript can be further disseminated through additional cellular machinery that serves to translate the RNA in to a final genetic product, a polypeptide. These polypeptides may ultimately function as the structural framework of the cell, the basic unit of life, or they may act as enzymes, catalyzing the fundamental cellular reactions that promote and sustain all life.

Naturally, the thought follows that if such a cellular process is so critical and integral to life, then this process must be tightly governed to ensure proper continuity in a cell's life cycle. Indeed, cellular processes are governed by a myriad of regulatory networks, and transcription is no exception. Catalyzing transcript synthesis, RNA polymerase (RNAP) serves as a major point of regulation and is consequently subject to extrinsic regulatory factors, such as transcription accessory proteins, and to regulatory sequence elements, such as pause and termination sites located throughout the genome (Uptain, 1997; Erie, 2002). It is the study of such regulatory events that serves to illuminate the fundamental mechanisms of an indispensable cellular process, and it is to this topic that I have dedicated my dissertation research.

Overview of Transcription

The process of transcription is defined by three stages: initiation, elongation, and termination. A cartoon model of the transcription cycle in prokaryotic organisms outlined below is presented in Figure 1.1. During initiation, the RNAP holoenzyme, comprised of six subunits ($\sigma\alpha_2\beta\beta'\omega$), recognizes and binds to a promoter upstream of a

gene to be transcribed. The holoenzyme proceeds to melt and unwind approximately 17

base pairs (bp) of double stranded DNA, forming the open promoter complex (Yager and

von Hippel, 1991). Upon open complex formation, RNAP begins to catalyze the DNA-

directed synthesis of a nascent RNA transcript by binding nucleoside triphosphates

(NTPs), catalyzing a phosphotransfer reaction, releasing pyrophosphate, and

translocating its active site in preparation for the next round of nucleotide addition

(Chamberlin, 1974). Once the transcript reaches 8-9 nucleotides (nt) in length, RNAP

changes its conformation, releasing the σ subunit in promoter clearance, and enters into

the elongation stage as the core enzyme. From this point until termination, the RNAP

core enzyme continues to add nucleoside monophosphates (NMPs) in a processive

manner to the growing RNA chain, while maintaining a RNA-DNA hybrid length of ~8

nt and a DNA bubble length of ~12-15 bp (Yager and von Hippel, 1991; Nudler *et al*.,

1997; Komissarova and Kashlev, 1997; Wilson *et al*., 1999; Gnatt *et al*., 2001). During

elongation, RNAP does not dissociate from the ternary (RNAP-DNA-RNA) elongation

complex (TEC) until RNAP reaches a specific termination signal, at which point the

newly synthesized transcript and core enzyme are released and the two DNA strands

reanneal (Yager and von Hippel, 1991).

The Architecture of RNAP

Playing a central and essential role in the metabolism of nucleic acids, the

fundamental components of the core RNAP are evolutionarily conserved from bacteria to

humans (Archambault and Friesen, 1993). Recent advances in the crystallization and

structural determination of bacterial RNAP have provided basic insight into the

architecture of the simplest form of multi-subunit RNAPs. These solved crystal

structures include core and holoenzyme from *Thermus aquaticus* (*T. aquaticus*) and

holoenzyme from *Thermus thermophilus* (*T. thermophilus*) (Zhang *et al.*, 1999;

Murakami *et al.*, 2002; Vassylyev *et al.*, 2002). Greater structural detail of the elongation

complex is provided by the crystal structure of eukaryotic core RNAP (RNAP II) and of

T. thermophilus core RNAP involved in a ternary elongation complex (Gnatt *et al.*, 2001;

Vassylyev *et al.*, 2007). In conjunction with these core enzyme structures, crosslinking

experiments have elucidated a structural model detailing the interactions between RNAP

and the nucleic acids comprising the TEC, illustrated in Figure 1.2 (Korzheva *et al.*,

2000).

With a molecular mass of 400 kDa, the core enzyme is composed of several subunits

($\alpha_2\beta\beta'\omega$) and adopts an overall conformation that resembles a crab claw (Figure 1.2).

The α dimer creates a scaffold upon which the β and the β' subunits assemble, while the

ω subunit is postulated to serve as a chaperonin by way of completely wrapping around

the β' C-terminal tail.

As the largest subunits in RNAP, the β and the β' subunits define the global shape of

the crab claw, with one subunit comprising one arm of the crab claw and the other

subunit comprising the opposite arm of the claw (Figure 1.2). Along the full length of

each claw, there is an internal groove known as the main channel that is 150 Å long and

27 Å wide. This main channel is the major site of nucleic acid binding and

accommodates simultaneously both RNA and DNA. The non-template strand of the

DNA bubble makes contacts only with β, while the RNA-DNA hybrid interacts with the β'F helix, an important structural feature implicated in RNAP translocation (Zhang *et al.*, 1999; Nudler *et al.*, 1997).

A major interface between β and β' exists at the hinge of the crab claw conformation, where a highly conserved catalytic aspartate triad coordinates a Mg^{2+} ion in the active site of the enzyme. In addition to this tightly bound Mg^{2+} ion, another Mg^{2+} ion is recruited *ad hoc* during polymerization by an incoming NTP and is coordinated by the β,γ-phosphates of this NTP (Steitz, 1998; Sosunov *et al.*, 2003). This symmetric pair of Mg^{2+} ions stabilizes the structure and charge of the pentacovalent transition state generated by the SN2 attack of the nascent transcript's 3' O^- on the α-phosphate of the NTP to be incorporated.

As revealed by the model of the TEC, the DNA and the RNA follow distinct paths within RNAP. Downstream of the DNA bubble, dsDNA is situated between β' and β. The template and non-template strands in the transcription bubble make contacts to RNAP as previously described (Figure 1.2). The RNA-DNA hybrid is flanked by the catalytic Mg^{2+} ion and the β'F helix at the downstream edge and by a structure called the β'C rudder at the upstream edge. In the immediate vicinity of the β'F helix is a highly conserved loop in the β subunit known as βD-Loop II. This loop is responsible for closing over the incoming NTP to be incorporated and its cognate DNA template base during active synthesis. As RNAP translocates, the hybridized RNA passes by regions of β, including βD-Loop II, in which mutations confer rifampicin resistance to RNAP (Jin and Gross, 1988). When confronted by the β'C rudder 8 nt from the active site, the

transcript is separated from the template DNA strand and is extruded through a groove underneath the βG flap. Approximately 6 nt of RNA 5' of the RNA-DNA hybrid are held in this RNA binding site just prior to exiting RNAP via translocation. Once reannealed, the dsDNA upsteam of the transcription bubble interacts with an α-helical coiled-coil domain in β'. The entering downstream dsDNA and the exiting upstream dsDNA form a bend angle of approximately 90º.

In addition to the main channel, RNAP also contains a secondary channel that is 45 Å long and 10-12 Å wide. This channel provides a direct route from the external environment to the active center at the base of the main channel. As a result, it is believed that single NTP molecules may enter RNAP via the secondary channel and gain access to the active site.

The Dynamic Flexibility of RNAP

Despite the fact that RNAP's architecture is defined by salient structural features essential for function, RNAP has displayed an unprecedented ability to adopt a variety of conformations (Erie *et al*., 1993). Owning to its dynamic flexibility, RNAP can exist at multiple conformational states in response to various regulatory events that govern the rate and the fidelity of RNA synthesis. The observation of the multiple conformational states has led to a mechanistic model of RNA synthesis (Foster *et al*., 2001; Holmes and Erie, 2003). These states, as defined by the model, are depicted in flow-chart format in Figure 1.3.

During "normal" RNA synthesis, RNAP assumes an activated conformation that is cocked for rapid catalysis. In this conformation, the transcript's 3' hydroxyl moiety is optimally poised in the active site for subsequent nucleotide additions. Since this state is a high-energy state, the activated conformation must be supported by successive NTP binding and nucleotide incorporation events. When these requirements are met, the activated conformation is generally long-lived (Erie, 2002). In the instance where rapid nucleotide incorporation is suspended, the activated state decays into an unactivated, lower energy state. RNAPs adopting an unactivated conformation are still competent for RNA synthesis, however the rate of synthesis is significantly reduced by approximately 10-fold or more (Foster *et al*., 2001; Erie, 2002; Holmes and Erie, 2003). This reduced rate is due in part to the repositioning of the transcript's 3' hydroxyl moiety into what is postulated to be a suboptimal alignment within the active site. In response to the reduced rate of synthesis, the unactivated conformation of RNAP appears to exhibit increased fidelity in comparison to the activated conformation and may serve a pivotal role in maintaining transcriptional fidelity (Erie *et al*., 1993).

The unactivated conformation serves as a gateway to a variety of other conformational states that RNAP can assume. For example, during NTP deprivation or when confronted by pause signals in the template DNA strand, RNAP can experience additional rearrangements in the active site and enter an elemental pause state that is marked by the absence of translocation (Landick, 2006). This elemental pause is an obligate intermediate, from which further prolonged paused states are derived by one or more subsequent rearrangements. These rearrangements generally include translocation of RNAP's active site relative to the 3' end of the transcript. One mode of translocation is

reverse translocation, or backtracking, by RNAP along the RNA-DNA hybrid. During

reverse translocation, the active site is repositioned over an internal site on the transcript.

Experimental evidence has shown that RNAP is slow to recover from such a backtracked

state (Kashlev et al., 1993). Another mode of translocation is hypertranslocation by

RNAP. In hypertranslocation, the active site is slipped forward relative to the 3' end of

the transcript. These various decayed states are recognized and acted upon by different

transcription accessory proteins (Erie, 2002). The intrinsic and accessory protein-

mediated fates of the different decayed RNAP conformations are described below in

greater detail.

The Gre Family and Transcript Cleavage

As discussed above, RNAP has been shown to adopt several conformations in

response to various regulatory events. Through a paused intermediate conformation,

RNAP can experience additional rearrangements and translocate backwards along the

nascent transcript and the DNA template (Wang and Hawley, 1993; Rudd *et al*, 1994;

Markovtsov *et al*., 1996; Reeder and Hawley, 1996; Komissarova and Kashlev, 1997;

Nudler *et al*., 1997). In this backtracked state, RNAP dislodges the 3' end of the

transcript from the active site and shifts the transcription bubble upstream by unwinding

the dsDNA upstream of the initial bubble and concomitantly rewinding the template and

non-template DNA strands that previously formed the bubble. The translocated bubble

alters the original RNA-DNA interactions and consequently shifts the RNA-DNA hybrid

upstream as well. Increasing the strength of the RNA-DNA hybrid has been correlated

with decreasing the likelihood of backtracking (Nudler *et al*., 1997; Wilson *et al*., 1999; Palangat and Landick, 2001).

Once backtracked, RNAP has been shown to enter one of two possible pathways. The first of these involves RNAP's intrinsic ability to cleave the nascent transcript (Surratt *et al*, 1991; Wang and Hawley, 1993; Erie *et al*., 1993; Rudd *et al*., 1994). There is strong evidence to suggest that the repositioned active site, which catalyzes the canonical phosphotransfer reaction, also catalyzes the hydrolysis of the transcript, releasing the 3' end from the TEC. Therefore, the length of the RNA cleavage product may be diagnostic of the extent of backtracking. Upon cleavage, RNAP's active site holds a new 3' end with a free hydroxyl moiety and, in the presence of NTPs, can successfully resume transcript extension. The second of the pathways involves further structural rearrangements in RNAP to yield an arrested complex, also known as a dead-end complex (Erie *et al*., 1993; Wang and Hawley, 1993; Rudd *et al*., 1994; Reeder and Hawley, 1996; Artsimovitch and Landick, 2000). Even in the presence of high NTP concentrations, arrested complexes are kinetically trapped and are thus unable to cleave the transcript or resume elongation.

Two transcript cleavage factors, GreA and GreB, have been identified and characterized in *E. coli*. Much information regarding the function of these factors comes from *in vitro* assays (Borukhov *et al*., 1992; Borukhov *et al*., 1993; Erie *et al*., 1993; Koulich et al., 1997; Loizos and Darst; 1999). While activated RNAPs are resistant to the action of Gre proteins, unactivated RNAPs appear to be an obligate intermediate for Gre action (Erie *at al*., 1993). Both Gre proteins preferentially recognize unactivated states and are able to suppress or prevent transcriptional pausing by enhancing RNAP's

endogenous ability to cleave the transcript (Erie *et al*., 1993; Borukhov *et al*., 1993).

GreA induces cleavage of RNA products 2-3 nt long, whereas GreB can induce cleavage

of products up to 9 nt long. This difference in cleavage product length is associated with

the observation that GreB can recognize and rescue arrested complexes, while GreA

cannot. In addition to assisting unactivated or arrested complexes, the Gre proteins have

been implicated in reducing abortive initiation events and in maintaining fidelity (Erie *et*

al., 1993; Hsu *et al*., 1995). The roles of the Gre proteins exhibited *in vitro* have been

corroborated with *in vivo* studies demonstrating that transcription complexes are less

likely to recover from backtracked states in *greA*/*greB* knock-out strains (Toulme *et al*.,

2000; Marr and Roberts, 2000).

X-ray crystallography data (GreA) and computer modeling techniques (GreB) have

provided structural insights into the mechanism of Gre action (Stebbins *et al*., 1995;

Koulich *et al*., 1997; Polyakov *et al*., 1998). These structures are presented in Figure 1.4.

Displaying similar salient features, both GreA and GreB possess a coiled-coil N-terminal

domain (NTD) and a globular C-terminal domain (CTD). Additionally, both Gre proteins

exhibit asymmetric charge distribution on their surfaces. One face of the protein is

strongly acidic, while the other face is hydrophobic and contains an evolutionarily

conserved basic patch. This characteristic conserved basic patch in GreB is much larger

than that in GreA. The basic patch is located near the tip of the coiled-coil (NTD), which

is responsible for the induced cleavage activity specific to each Gre protein. Both the

NTD and the globular CTD contribute to specific binding to RNAP. The globular CTD

binds to the rim of the secondary channel, while the coiled-coil NTD protrudes the

secondary channel, extending all the way to the active site. Two invariant acidic residues

located at the tip of the NTD are responsible for coordinating the second Mg^{2+} ion, thereby directly facilitating hydrolysis of the nascent transcript (Laptenko *et al.*, 2003; Opalka *et al.*, 2003).

The Nus Family and Transcription Pausing and Termination

As alluded to earlier, RNAP's kinetic and conformational decay into the unactivated state has the potential to give rise to a variety of other subsequent lower energy states. Similar to the backtracked state that is delineated from an unactivated precursor, the pausing and termination states are generated by off-line pathways from a canonically transcribing RNAP via an unactivated obligate intermediate.

The structural rearrangements that trigger pausing are induced by the interaction of RNAP in the elongation complex with specific sequences in the DNA, RNA, or both (Landick, 2006). These sequence elements are generally known as pause signals and are crucial for synchronizing transcription with translation (Landick *et al.*, 1985; Squires and Zaporojets, 2000). The momentary halting of a transcribing RNAP permits a translating ribosome to catch up to the elongation complex and to subsequently induce escape from the paused state. In addition to coupling RNA synthesis with the movements of the ribosome, pauses also allow the recruitment of auxillary protein regulators and the proper folding of nascently synthesized RNA (Artsimovitch and Landick, 2002; Pan and Sosnick, 2006). Finally, pauses serve as a precursor to the termination of transcription with the concomitant release of RNAP and RNA and the collapse of the DNA bubble.

11

In *E. coli*, members of the Nus family are involved in transcription pausing and termination. Of this family, NusA and NusG play indispensably key roles in the general regulation of transcription elongation, as these proteins are both essential to cell viability, unlike GreA and GreB (Schmidt and Chamberlin, 1984; Sullivan and Gottesman, 1992). Each of these regulatory proteins is described in detail below.

NusA is a multifaceted protein, owing to the variety of roles in which this protein serves. Depending on the sequence context of the DNA or the RNA that confronts RNAP in an elongating complex, NusA can wield seemingly opposite effects on transcription. NusA enhances the stability, and hence the dwell time, of paused complexes, particularly of those halted at *his* and *trp* pause signals (Chan and Landick, 1993; Artsimovtich and Landick, 2000). A common feature of these two signals is the resultant generation of RNA hairpins that are stabilized by interaction with NusA. In relation to this pause-stimulating effect, NusA acts as a general elongation factor by reducing the overall rate of transcription, in some cases by 20%-35% *in vitro* (Burns *et al*., 1998; Rees et al., 1997). NusA also facilitates transcription termination at ρ-independent terminators, which are also responsible for generating RNA hairpins (Schmidt and Chamberlin, 1984; Schmidt and Chamberlin, 1987). Contrastingly, NusA can exhibit anti-termination activity at both ρ-dependent and ρ-independent terminators, depending on the presence of other transcription accessory proteins (Nudler and Gottesman, 2002).

Solved crystal structures of NusA from *Thermotoga maritima* and *Mycobacterium tuberculosis* have provided a physical basis for relating distinct structural features with specific functionality (Gopal *et al*., 2001; Worbs *et al*., 2001; Shin *et al*., 2003). As

shown in Figure 1.5, NusA is composed of two distinct domains, the NTD and the CTD. The NTD is composed of the RNAP-binding domain, while the CTD is composed of three integrated domains responsible for interaction with the nascent transcript. The RNAP-binding domain anchors NusA to RNAP's β' coiled-coil region located near the upstream opening of the primary channel. This site of interaction on the RNAP is the same site to which σ^{70} subunit binds (Traviglia et al., 1999). Comprising the RNA-binding CTD, the S1 domain and the two KH (KH1 and KH2) domains are globular structures that exhibit homology to the RNA-binding motifs of ribosomal protein S1 and pre-mRNA-binding protein K, respectively. When NusA is bound to RNAP, these RNA-binding domains extend beneath the RNAP βG flap. Highlighted earlier, the βG flap is the site on RNAP underneath which the nascent transcript is extruded. It is thought that the RNA-binding domains of NusA stabilize hairpin formation in a transcript emerging from RNAP (Pan *et al*., 1999; Toulokhonov *et al*., 2001).

In contrast to the pause-stimulating effects of NusA, NusG facilitates anti-pausing and anti-termination activity in RNAP (Sullivan and Gottesman, 1992; Li *et al*., 1993). Coupled with these effects is the observation that NusG serves as a general elongation factor by accelerating the rate of transcription by 25%-30% *in vivo* and *in vitro* (Burova *et al*., 1995; Burns *et al*., 1998). By accelerating the rate of transcription, NusG is thought to shift the conformational equilibrium of the elongation complex from a backtracked state to an elongation-competent state (Pasman and von Hippel, 2000). While not entirely conserved, NusG homologs have been identified in all prokaryotes, and orthologs have been identified in some eukaryotes, including yeast and humans (Hartzog *et al*., 1998).

An x-ray crystal structure of NusG from *Aquifex aeolicus* has been solved (Steiner *et al*., 2002; Knowlton *et al*., 2003) and is displayed in Figure 1.6. Based upon this structure, NusG is composed of three discrete domains that are loosely tethered together. The first and third domains, D1 and D3, are conserved regions found in all NusG proteins. In contrast, the second domain, D2, is defined as a variable-insertion region, of which only a small portion is found in *E. coli* NusG. Located near the flexible N-terminus, D1 is characterized by an RNP-like domain. It is speculated that NusG interacts weakly with RNAP via this RNP-like domain. At the C-terminus, D3 contains a KOW sequence motif, characteristically found in the ribosomal protein L24 (Kyrpides *et al*., 1996). Although this KOW motif is implicated in nucleic acid binding, direct binding to RNA and to DNA has been observed only in NusG from *Aquifex aeolicus* and *Thermotoga maritima* (Liao *et al*., 1996; Steiner *et al*., 2002). Direct interaction between NusG and nucleic acids has yet to be observed in *E. coli*.
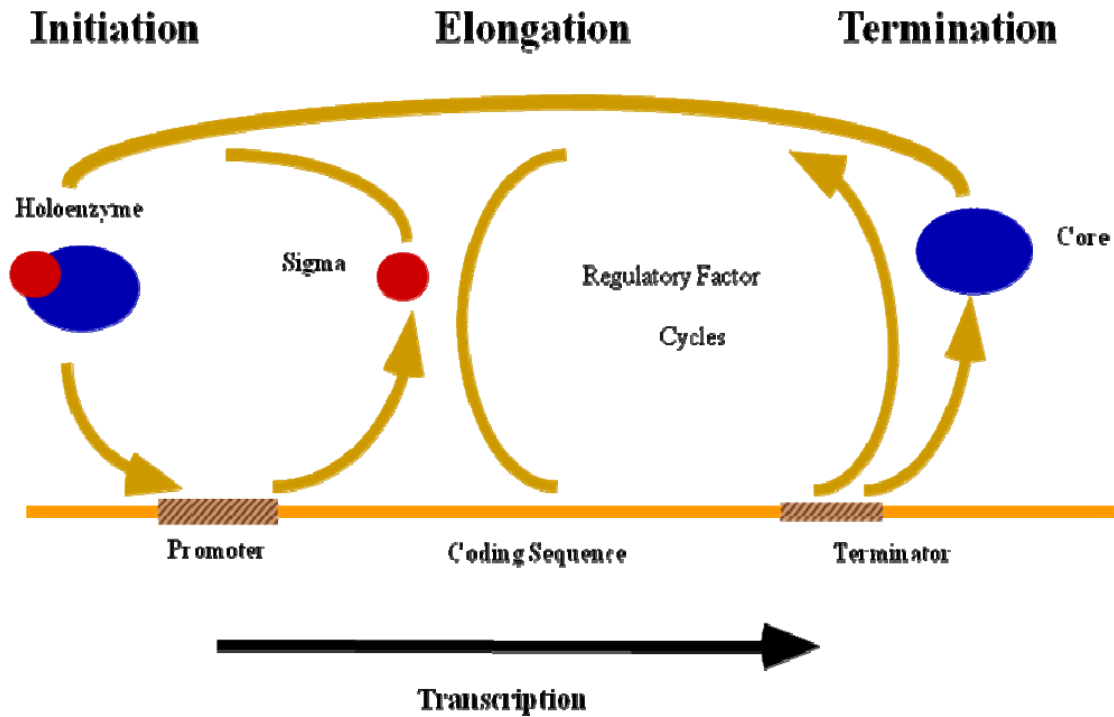
Figure 1.1 The transcription cycle in *Escherichia coli*. The transcription cycle is defined by three stages: initiation, elongation, and termination. During initiation, the holoenzyme (core RNAP and sigma) bind to a promoter upstream of a gene to be transcribed. RNAP melts the DNA within this region, generating an open promoter complex (OPC). The OPC binds NTPs and begins synthesizing a nascent RNA chain. Once the transcript reaches 8-9 nucleotides in length, sigma is released, and RNAP enters the elongation stage where catalysis is rapid and processive. Upon reaching a terminator signal, RNAP dissociates from the DNA and releases the nascent transcript. Core RNAP may then rebind sigma to begin the process anew.
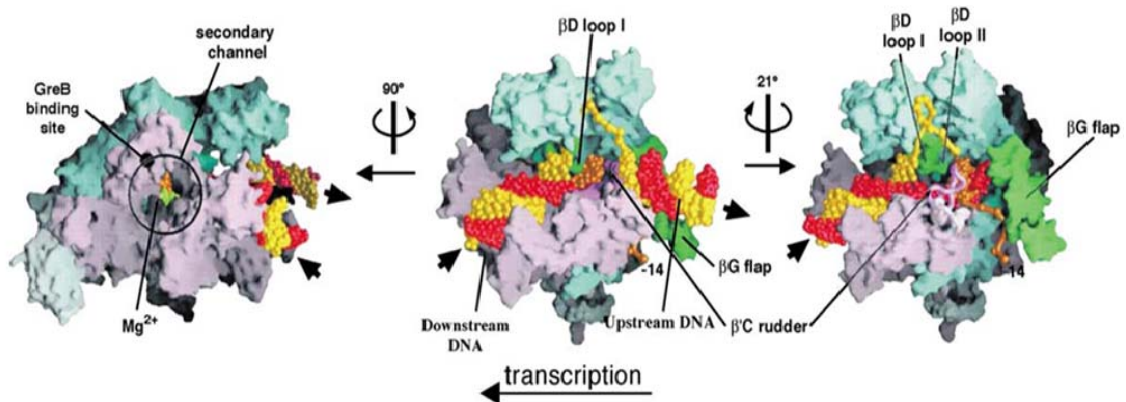
Figure 1.2 Model of the elongation complex (adapted from Korzheva et al., 2000).

The two largest subunits of RNAP, beta (cyan) and beta-prime (purple), are shown bound

to dsDNA. The templated strand is shown in red, while the non-templated strand is

shown in yellow. The nascent RNA is shown in orange. Rotating the central panel by 90

degrees, one can note the secondary channel, the small pore by which bulk solvent may

access the active site. A 21 degree rotation of the central panel displays the flap domain,

which makes contacts to a RNA chain emerging from the main channel.
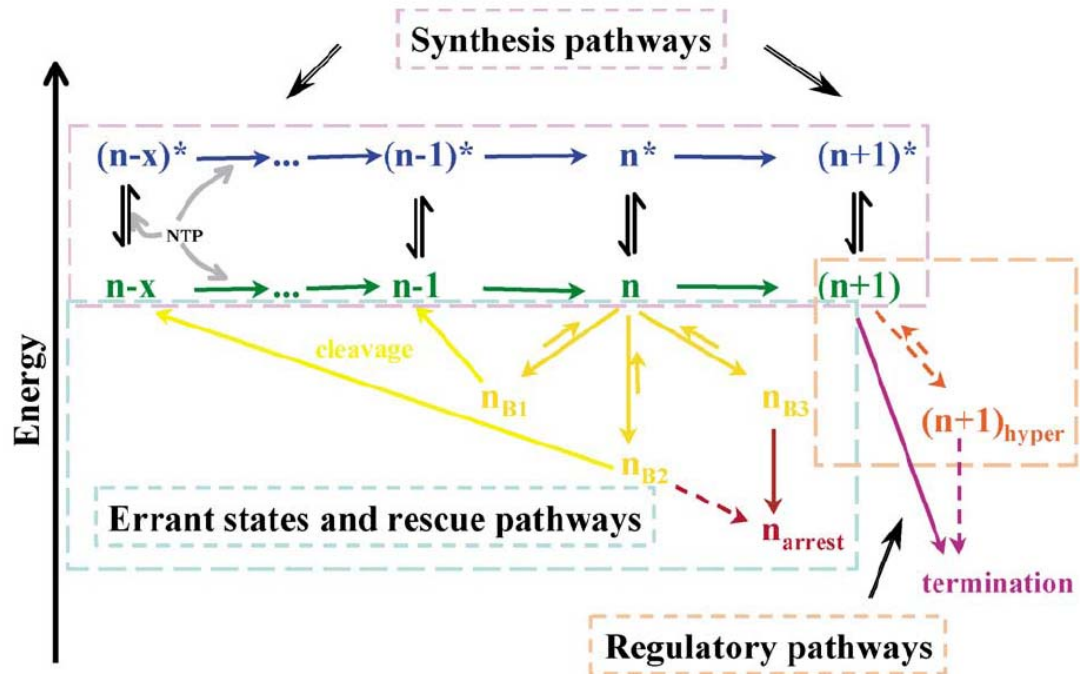
Figure 1.3 Multiple paths of synthesis by RNAP. During normal elongation, RNAP catalyzes rapid synthesis of RNA along the activated pathway (blue). When confronted by pause signals or during moments of NTP deprivation, RNAP can decay onto the unactivated pathway (green), where synthesis is continued at a much slower rate. It is thought that the unactivated pathway is a higher fidelity state and serves as a gateway to further decayed transcription states. Among these decayed states, are backtracked states (yellow) that can be induced to cleave their transcripts. If the backtracked state is prolonged, RNAP may become arrested (red), or kinetically trapped. Also, RNAP may hypertranslocate (orange) when confronted by pause or termination signals (purple).
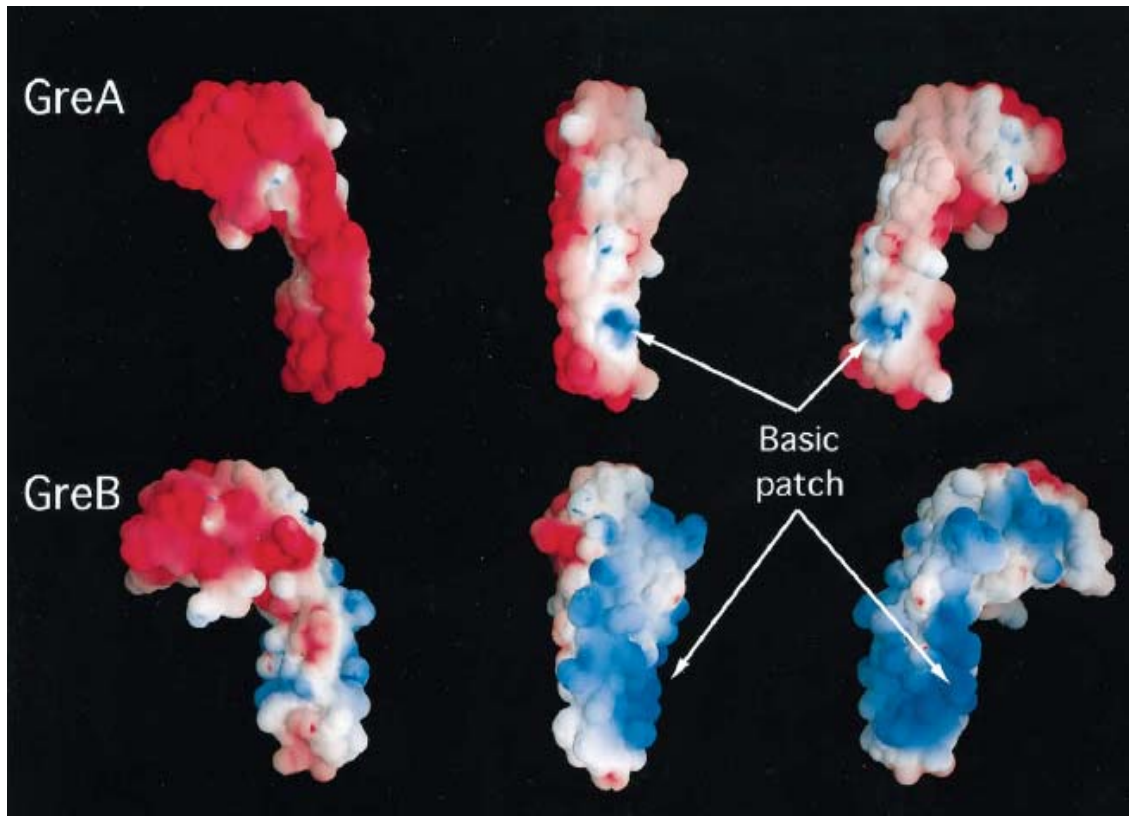
Figure 1.4 Structures of *Escherichia coli* GreA and GreB (adapted from Koulich et al., 1997). The structures of GreA and GreB reveal a globular C-terminal domain and a coiled-coil N-terminal domain. The C-terminal domain is responsible for binding to the surface of RNAP, while the coiled-coil protrudes into the secondary channel of RNAP and reaches the active site to promote transcript cleavage. The asymmetric charge distribution is indicate (red, acidic residues; blue, basic residues).
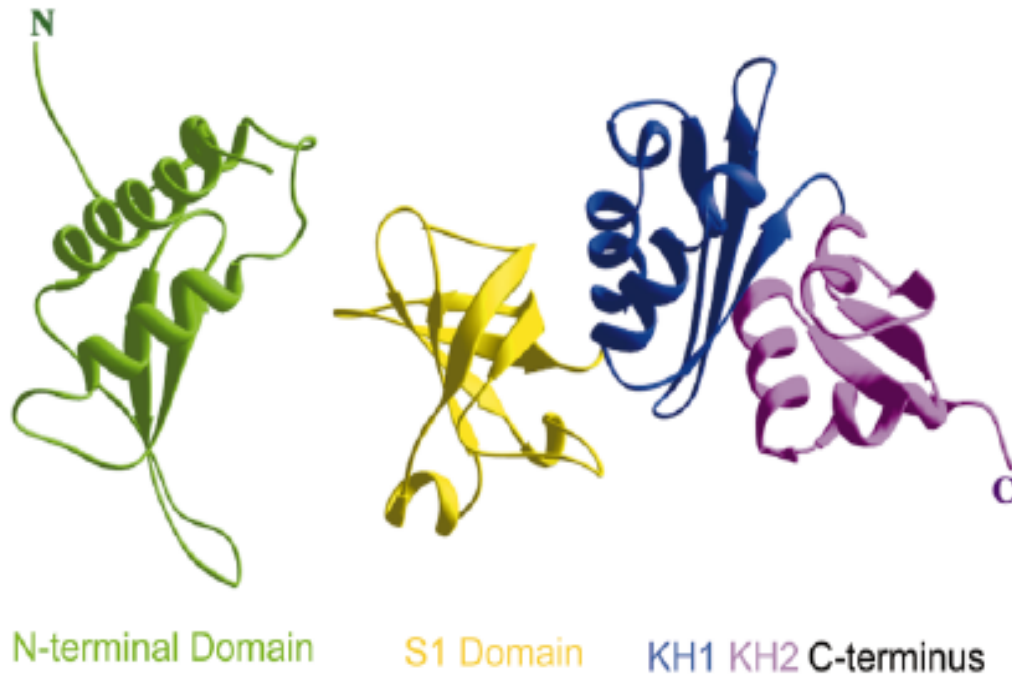
Figure 1.5 Structure of *Mycobacterium tuberculosis* NusA (adapted from Gopal et al., 2001). The structure of NusA reveals three distinct domains: the N-terminal domain, the S1 domain, and the KH domain. The N-terminal domain is responsible for binding to RNAP. Comprising the RNA-binding CTD, the S1 domain and the two KH (KH1 and KH2) domains are globular structures that exhibit homology to the RNA-binding motifs of ribosomal protein S1 and pre-mRNA-binding protein K, respectively.
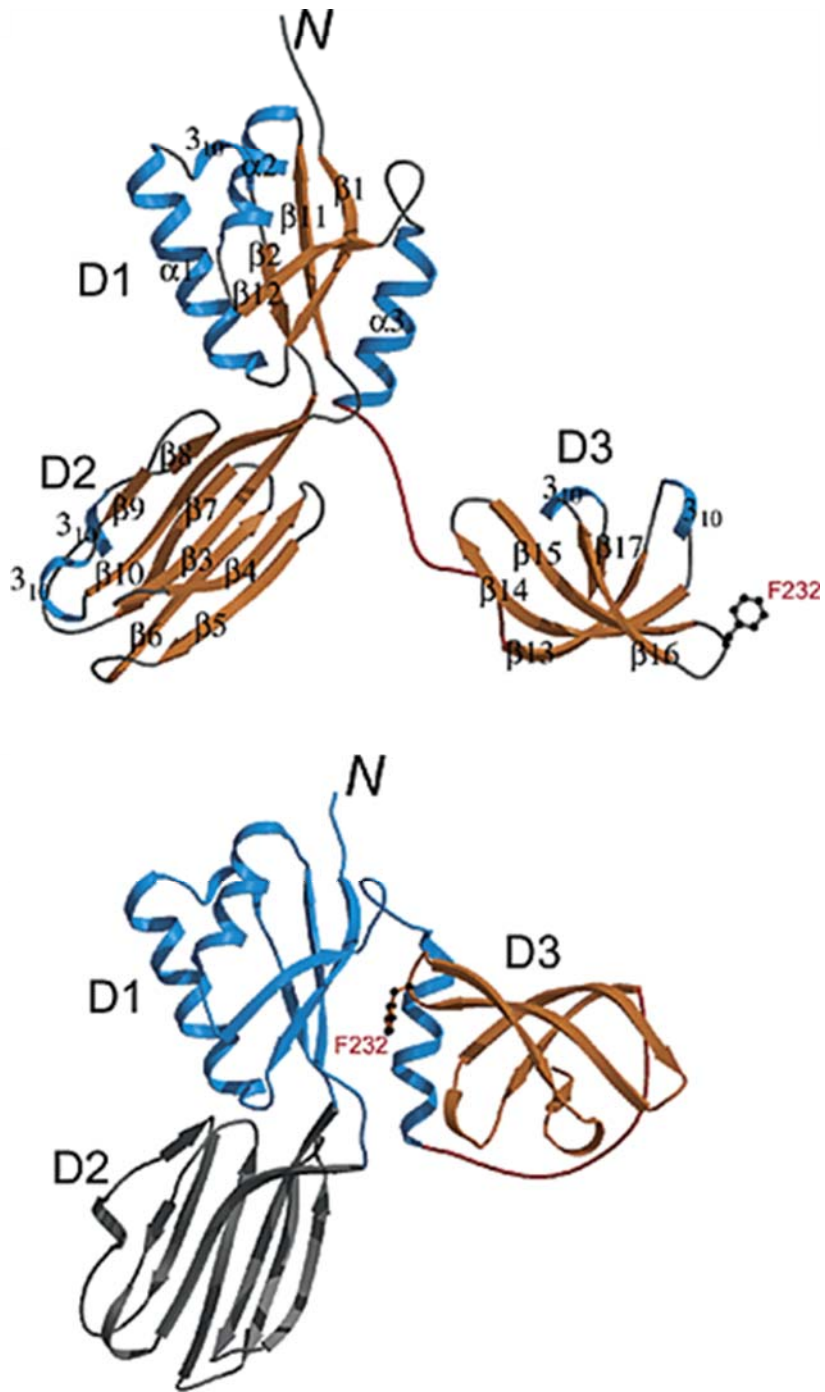
Figure 1.6 Structure of *Aquifex aeolicus* NusG (adapted from Knowlton et al., 2003).

Figure 1.6 Structure of *Aquifex aeolicus* NusG (adapted from Knowlton et al., 2003).

The structure of NusG reveals three distinct domains, with domains D1 and D3 being

highly conserved. D1 is characterized as a RNP-like domain, while D3 contains a KOW

motif. It is thought that NusG interacts with RNAP through the RNP-like domain.

BIBLIOGRAPHY

Artsimovitch, I., and Landick, R. (2000). Pausing by Bacterial RNA Polymerase is Mediated by Mechanistically Distinct Classes of Signals. Proc. Nat. Acad. Sci. *97*, 7090-7095.

Burova, E. et al., (1995). Escherichia coli NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. J. Bact. *177*(5), 1388-1392.

Chan, C.L and Landick R. (1993). Dissection of the his leader pause site by base substitution reveals a multipartite signal that includes a pause RNA hairpin. J. Mol. Biol. *233*(1), 25-42.

Erie, D.A. (2002). The Many Conformational States of RNA Polymerase Elongation Complexes and Their Roles in the Regulation of Transcription. Biochim. Biophys. Acta. *1577*(2), 224-39.

Erie, D.A. et al., (1993). Multiple RNA Polymerase Conformations and GreA: Control of the Fidelity of Transcription. Science *262*, 867-873.

Foster, J.E. et al., (2001). Allosteric Binding of Nucleoside Triphosphates to RNA Polymerase Regulates Transcription Elongation. Cell *106*, 243-252.

Gnatt, A.L. et al., (2001). Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. Science *292*, 1876-1882.

Gong, X.Q. et al., (2005). Dynamic Error Correction and Regulation of Downstream Bubble Opening by Human RNA Polymerase II. Molecular Cell *18*, 461-470.

Gopal, B. et al., (2001). Crystal Structure of the Transcription Elongation/Antitermination Factor NusA from Mycobacterium tuberculosis at 1.7 A Resolution. J. Mol. Biol. *314*, 1087-1095.

Hartzog, G.A. et al., (1998). Evidence that Spt4, Spt5, and Spt6 Control Transcription Elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes and Dev. *12*(3), 357-69.

Holmes, S.F., and Erie, D.A. (2003). Downstream DNA Sequence Effects on Transcription Elongation: Allosteric Binding of Nucleoside Triphosphates Facilitates Translocation Via a Ratchet Motion. J. Bio. Chem. *278*, 35597-35608.

Hsu, L.M. et al., (1995). Escherichia coli Transcript Cleavage Factors GreA and GreB Stimulate Promoter Escape and Gene Expression In Vivo and In Vitro. Proc. Natl. Acad. Sci. USA *92*, 11588-11592.

Knowlton, J.R. et al., (2003). A Spring-Loaded State of NusG in Its Functional Cycle is Suggested by X-ray Crystallography and Supported by Site-Directed Mutants. Biochemistry *42*, 2275-2281.

Komissarova, N., and Kashlev, M. (1997). RNA Polymerase Switches Between Inactivated and Activated States By Translocating Back and Forth along the DNA and the RNA. J. Bio. Chem. *272*, 15329-15338.

Korzheva N. et al., (2000). A Structural Model of Ttranscription Elongation. Science 289, 619-625.

Koulich, D. et al., (1997). Domain Organization of Escherichia coli Transcript Cleavage Factors GreA and GreB. J. Bio. Chem *272*(11), 7201-7210.

Landick, R. (2006). The Regulatory Roles and Mechanism of Transcriptional Pausing. Biochemical Society Transactions *34*(6), 1062-1066.

Landick, R. et al., (1985). Translation activates the paused transcription complex and restores transcription of the trp operon leader region. Proc. Natl. Acad. Sci. USA *82*, 4663-4667.

Laptenko, O. et al., (2003). Transcript Cleavage Factors GreA and GreB act as Transient Catalytic Components of RNA Polymerase. EMBO *22*(23), 6322-6334.

Li, J. et al., (1992). NusG, a New Escherichia coli Elongation Factor Involved in Transcriptional Antitermination by the N Protein of Phage Lamda. J. Bio. Chem. *267*(9), 6012-6019.

Li, J. et al., (1993). Elongation Factor NusG Interacts with Termination Factor Rho to Regulate Termination and Antitermination of Transcription. Genes and Dev. *7*, 161-172.

Liao, D. et al., (1996). A NusG-Like Protein from Thermotoga maritima Binds to DNA and RNA. J. Bact. *178*(14), 4089-4098.

Loizos, N. and Darst, S. A. (1999). Mapping Interactions of Escherichia coli GreB with RNA Polymerase and Ternary Elongation Complexes. J. Biol. Chem. *274*, 23378-23386.

Murakami, K. et al., (2002). Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 Å Resolution. Science *296*, 1280-1284.

Markovtsov, V. et al., (1996). Protein-RNA Interactions in the Active Center of Transcription Elongation Complex. Proc. Natl. Acad. Sci. USA. *93* (3221-3226).

Marr, M.T. and Roberts, J.W. (2000). Function of Transcript Cleavage Factors GreA and GreB at a Regulatory Pause Site. Mol. Cell. *6*, 1275-1285.

Mustaev, A. et al., (1993). Active Center Rearrangement in RNA Polymerase Initiation Complex. J. Bio, Chem. *268*, 19185-19187.

Nudler, E. and Gottesman, M. (2002). Transcription Termination and Antitermination in E. coli. Genes Cells *7*, 755-768.

Nudler, E. et al., (1997). The RNA-DNA Hybrid Maintains the Register of Transcription by Preventing Backtracking of RNA Polymerase. Cell 89(1), 33-41.

Opalka, N. et al., (2003). Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase. Cell *114*, 335-345.

Palangat, M., and Landick, R. (2001) Roles of RNA:DNA Hybrid Stability, RNA Structure, and Active Site Conformation in Pausing by Human RNA Polymerase II. J. Mol. Biol. *311*, 265-282.

Pan, T. and Sosnick, T. (2006). RNA folding during transcription. Annu. Rev. Biophys. Biomol. Struct. 35, 161-175.

Pasman, Z and von Hippel, P.H. (2000). Regulation of Rho-Dependent Transcription Termination by NusG is Specific to the Escherichia coli Elongation Complex. Biochemistry *39*, 5573-5585.

Polyakov, A. et al. (1998). Visualization of the Binding Site for the Transcript Cleavage Factor GreB on Escherichia coli RNA Polymerase. J. Mol. Biol. *281*, 465-473.

Reeder, T.C. and Hawley, D. K. (1996). Promoter Proximal Sequences Modulate RNA Polymerase II by a Novel Mechanism. Cell *87*, 767-777.

Rudd, M.D. et al., (1994). The Active Site of RNA Polymerase II Participates in Transcript Cleavage with Arrested Ternary Complexes. Proc. Natl. Acad. Sci. USA *91*, 8057-8061.

Schmidt, M.C. and Chamberlin, M.J. (1987). nusA Protein of Escherichia coli is an Efficient Transcription Termination Factor for Certain Terminator Sites. J. Mol. Biol. *195*, 809-818.

Schmidt, M.C. and Chamberlin, M.J. (1984). Amplification and Isolation of Escherichia coli nusA Protein and Studies of Its Effects on In Vitro RNA Chain Elongation. Biochemistry *23*, 197-203.

Shin, D.H. et al., (2003). Crystal Structure of NusA from Thermotoga maritime and Functional Implication of the N-Terminal Domain. Biochemistry *42*, 13429-13437.

Squires, C.L. and Zaporojets, D. (2000). Proteins Shared by the Transcription and Translation Machines. Annu. Rev. Microbiol. *54*, 775-798.

Stebbins, C.E. et al., (1995). Crystal Structure of the GreA Transcript Cleavage Factor from Escherichia coli. Nature *373*, 636-640.

Steiner T. et al., (2002). Crystal Structures of Transcription Factor NusG in Light of its Nucleic Acid- and Protein-Binding Activities. EMBO *21*(17), 4641-4653.

Sullivan, S.L. and Gottesman, M.E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell *68*(5), 989-994.

Surratt, C.K. et al., (1991). Spontaneous Cleavage of RNA in Ternary Complexes of Escherichia coli RNA Polymerase and its Significance for the Mechanism of Transcription. Proc. Natl. Acad. Sci. USA *88,* 7983-7987.

Toulme, F. et al., (2000). GreA and GreB Proteins Revive Backtracked RNA Polymerase In Vivo by Promoting Transcript Trimming. EMBO *19*, 6853-6859.

Toulokhonov, I. et al., (2001). Allosteric Control of RNA Polymerase by a Site that Contacts Nascent RNA Hairpins. Science *292*, 730-733.

Traviglia, S.L. et al., (1999). Targeted Protein Footprinting: Where Different Transcription Factors Bind to RNA Polymerase. Biochemistry *38*, 15774-15778.

Uptain, S.M., and Chamberlin, M.J. (1997). *Escherichia coli* RNA Polymerase Terminates Transcription Efficiently at Rho-independent Terminators on Single-Stranded DNA Templates. Proc. Nat. Acad. Sci. *94*, 13548-13553.

Vassylyev, D.G. et al., (2007a). Structural Basis for Substrate Loading in Bacterial RNA Polymerase. Nature *448*, 163-168.

Vassylyev, D.G. et al., (2007b). Structural Basis for Transcription Elongation by Bacterial RNA Polymerase. Nature *448*, 157-164.

von Hippel, P.H., and Yager, T.D. (1991). Transcript Elongation and Termination are Competitive Kinetic Processes. Proc. Natl. Acad. Sci. *88*, 2307-2311.

Wang, D. et al., (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. Cell *127*, 941-954.

Wang, D. and Hawley, D.K. (1993). Identification of a 3'-to-5' Exonuclease Activity Associated with Human RNA Polymerase II. Proc. Natl. Acad. Sci. USA *90*, 843-847.

Watson, J.D. and Crick, F.H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature *171*(4356), 737-8.

Westover, K.D. et al., (2004). Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. Cell *119*, 481-489.

Wilson, K.S. et al., (1999). Determinants of the stability of transcription elongation complexes: interactions of the nascent RNA with the DNA template and the RNA polymerase. J. Mol. Biol. *289*(5), 1179-94.

Worbs, M. et al., (2001). An Extended RNA Binding Surface Through Arrayed S1 and KH Domains in Transcription Factor NusA. Mol. Cell *7*, 1177-1189.

Zhang, G. et al., (1999). Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 Å Resolution. Cell *98*(6), 811-24.

CHAPTER TWO

PATTERNS OF ERROR RECOGNITION AND TRANSCRIPT CLEAVAGE BY ESCHERICHIA COLI RNA POLYMERASE

INTRODUCTION

Catalyzing the first step in gene expression, RNA polymerase (RNAP) must synthesize a nascent RNA transcript in a processive manner with high fidelity and at a reasonable rate. Due to these two stringent cellular requirements, a myriad of factors, such as accessory proteins and sequence elements, subject RNAP to rigorous regulation. In particular, during the elongation stage of transcription the ternary (RNAP-DNA-RNA) elongation complex (TEC) serves as a major point of such regulation. The nucleic acid-nucleic acid and protein-nucleic acid interactions that define and stabilize the TEC influence the mode of regulation (Erie, 2002).

RNAP has exhibited multiple conformations during transcription and displays an ability to catalyze several reactions (Chapter 1, Figure 1.3) (Erie *et al*., 1993; Erie, 2002). In conjunction with the canonical phosphoryl transfer reaction, RNAP has been shown to intrinsically catalyze a nucleolytic reaction (Surratt *et al*., 1991; Wang and Hawley, 1993;

Erie *et al*., 1993; Rudd *et al*., 1994). The same active site that extends a growing RNA

chain is also responsible for the cleavage of the nascent transcript. Upon reverse

translocation of RNAP along the transcript, the active site is repositioned over an internal

site in the transcript where a phosphodiester bond is hydrolyzed, releasing the transcript's

3' end from the TEC. The active site holds a regenerated 3' end with a free hydroxyl

moiety, and, in the presence of NTPs, can successfully resume transcript extension. In

this manner, the intrinsic cleavage activity of RNAP is reminiscent of the 3'→5'

exonucleolytic activity of DNA polymerases (DNAP).

While the nucleolytic activity is performed by RNAP's active center, the accessory

protein factors, GreA and GreB, greatly stimulate transcript cleavage (Borukhov *et al*.,

1992; Borukhov *et al*., 1993; Erie *et al*., 1993; Koulich *et al*., 1997; Loizos and Darst,

1999). These Gre factors are endowed with an evolutionarily conserved basic patch

located near the tip of the N-terminal extended coiled-coil domain (NTD) and two

conserved aspartic acid residues positioned immediately at the NTD tip (Stebbins *et al*.,

1995; Koulich *et al*., 1997; Polyakov *et al*., 1998). When bound to RNAP in the TEC,

the Gre factors insert the coiled-coil domain down the length of RNAP's secondary

channel, anchoring the transcript with the basic patch and placing the two conserved

aspartic acid residues in the vicinity of the active site. It is believed that these two

residues assist RNAP in recruiting and in chelating a Mg^{2+} ion required for transcript

cleavage (Laptenko *et al*., 2003; Opalka *et al*., 2003). In this fashion, GreA induces

cleavage of RNA products 2-3 nt long; whereas, GreB induces cleavage of products up to

9 nt long. By way of transcript cleavage, the Gre factors have been implicated in

suppressing transcriptional pausing, reducing abortive initiation events and maintaining

transcriptional fidelity (Erie *et al.*, 1993; Hsu *et al.*, 1995). Additionally, the prokaryotic Gre factors have an eukaryotic ortholog known as TFIIS (Wang and Hawley, 1993). Similar to the Gre factors, TFIIS is responsible for stimulating the transcript cleavage activity of the active site of RNA polymerase II (pol II).

The present study seeks to explore the potential endogenous proofreading capabilities of RNAP and the extent by which the Gre factors enhance this capacity. To this end, TECs bearing transcripts with either a terminally correct or terminally incorrect base were generated from a promoter and were subsequently subjected to intrinsic and Gre-mediated cleavage reactions.

MATERIALS AND METHODS

Source of RNA Polymerase and DNA

His-tagged *E. coli* RNA polymerase was purified from *E. coli* strain RL916 (gift of R. Landick) as previously described (Burgess and Jendrisak, 1975; Uptain *et al.*, 1997). Amplification of pDE13 (Erie *et al.*, 1993) by PCR generated a 5'-biotinylated 540 nt DNA template containing the λP_R promoter and encoding a transcript in which the first cytosine to be incorporated is at position +25. The mutated U20a template was derived from pDE13 using the QuikChange site-directed mutagenesis kit (Stratagene), was confirmed by analysis at an automated DNA sequencing facility at the University of North Carolina, and was amplified as above.

31

Expression and Purification of GreA and GreB

The plasmids pIA577 and pIA578 (gifts of I. Artsimovitch), encoding His-tagged GreB and GreA, respectively, were transformed into BL21(DE3) competent cells (Stratagene). The transformed cells were grown to mid-log phase and induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside. After having grown for three hours, the cells were pelleted at 2,500 rpm for 30 min and resuspended in lysis buffer (50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, 8M urea). The resuspended cells were then lysed by gently stirring and vortexing at room temperature for 45 min, followed by centrifugation at 15,000 rpm for 30 min. The cleared lysate was combined with 5 mL of pre-equilibrated Ni-NTA agarose beads (Qiagen) and gently rocked on a rotary shaker for 1 hr at room temperature. Upon spinning at 1,500 rpm for 1 min, the beads adsorbed with Gre protein were washed with 20 mL renaturation buffer I (50 mM NaH_2PO_4, 500 mM NaCl, 10 mM imidazole, 10 mM β -mercaptoethanol, 2M urea), rocked for 30 min at room temperature, and pelleted as above. To complete the renaturation of Gre protein, the pelleted beads were washed twice with 20 mL renaturation buffer II (same as renaturation buffer I without urea), rocked at 4 °C for 15 min, and pelleted. The beads were then washed with 20 mL wash buffer (renaturation buffer II supplemented with 20 mM imidazole), rocked at 4 °C for 5 min, and pelleted. Gre protein was eluted with 2 mL elution buffer (renaturation buffer II supplemented with 250 mM imidazole), analyzed by 10% SDS-PAGE, and dialyzed overnight against 1 L storage buffer (50 mM NaH_2PO_4, 500 mM NaCl, 0.1 mM EDTA, 1 mM dithiothreitol, 50% glycerol). The Gre protein was purified to 10 mg/mL at approximately 95% purity and stored at –20 °C.

Formation of Stalled Elongation Complexes (SECs)

Open promoter complexes (OPCs) were generated by incubating 100 nM RNAP and 50 nM 5'-biotinylated DNA template bound to streptavidin-coated magnetic beads at 37 $^{\circ}$C in 1X transcription buffer (30 mM HEPES, pH 8.0, 10 mM magnesium glutamate, 200 mM potassium glutamate, 25 µg/mL bovine serum albumin, 1 mM dithiothreitol). After 10 min of incubation, 10 µM ATP, UTP, [α-^{32}P]GTP (800 Ci/mmol), each, were added to the OPCs and allowed to react for 1.5 min at 37 $^{\circ}$C to generate complexes stalled at position +24 (A24). The resultant stalled elongation complexes (SECs) were placed next to a strong magnet and washed five times with ice-cold 1X transcription buffer. The SECs were then resuspended in 1X ice-cold transcription buffer and stored on ice. In most experiments, the SECs were "walked" to position +25 by incubation at room temperature with either 1 µM CTP (C25) or 100 µM UTP (U25) to prepare SECs bearing terminally correct and incorrect transcripts, respectively. These SECs were purified as above before use in subsequent reactions.

Transcript Cleavage Assays

At t=0, the SECs were combined with stoichiometric quantities of either GreA or GreB at 37 $^{\circ}$C. Reaction aliquots were removed at designated times and quenched with 95% formamide. Also at certain times, reaction aliquots were chased with 1 mM of all four NTPs to ensure that the complexes were active throughout the course of the experiment. The transcriptional products were resolved and analyzed by electrophoresis on 8 M, 20% acrylamide gels.

Kinetic Analysis of the Cleavage Reaction

The total amount of radioactivity in each lane of the gel was measured on a Biosciences PhosphorImager (Amersham) and analyzed with ImageQuant software. The percentage of complexes at each position on the template was calculated by dividing the amount of radioactivity in the indicated band by the total amount of radioactivity in all the bands. The disappearance of A24, C25, and U25 complexes was plotted as a function of time. Each data set was fit to a single exponential ($y = A\exp(-kt)$), and the pseudo-first-order rate constant for each reaction was determined. In addition, the appearance of distinct cleavage products was plotted as a function of time.

RESULTS

Intrinsic cleavage of correctly incorporated and misincorporated transcripts

We sought to explore error correction mechanisms in *E. coli* RNAP. The first step in this study was to generate TECs on the DE13 template and to stall the TECs at position +24 in the absence of CTP. These stalled complexes were designated as A24 complexes. From position +24, subsets of the TECs were "walked" with the addition of a specific NTP to extend the transcript by one nucleotide to position +25. One subset of the A24 complexes was correctly extended to position +25 in the presence of the templated CTP, and these complexes were designated as C25 complexes. The other

subset of the A24 complexes was incorrectly extended to position +25 in the presence of the non-cognate UTP, and these misincorporated complexes were designated as U25 complexes. Thus, in this manner of controlled transcript extension, we generated TECs harboring either a terminally correct or terminally incorrect base at the 3' end of the nascent transcript.

To test RNAP's capability of sensing a terminally misincorporated nucleotide at the 3' end of a nascent transcript, we performed intrinsic transcript cleavage reactions using the various complexes described above. A representative gel presented in Figure 2.1 displays the time course for the cleavage of A24, U25, and C25 transcripts. Inspection of the gel reveals a decrease in the intensities of bands corresponding to positions +24 and +25 as a function of time. Since the relative intensities of the bands in each lane of the gel correspond to the percentage of transcripts at each position, the decrease in the intensities indicate that the relative amounts of +24 and +25 complexes are decreasing with time as a result of cleavage by RNAP. The slight decrease in intensity, however, suggests that there is little intrinsic cleavage over time. Although the cleavage is slight, we can still follow the disappearance of +24 and +25 transcripts and quantify the rate of cleavage.

In addition to the cleavage reactions, the functional integrity of each complex was tested by subjecting reaction aliquots to all four NTPs at various points along the time trajectory. As shown in Figure 2.1, each complex successfully extends its transcript when challenged with the four NTPs, indicating that the complexes remain catalytically competent during the course of the cleavage reaction. Additionally, the 5' cleavage products, which are presumably held as part of the TEC upon cleavage, are also fully

extended. We must note, however, that these transcripts are not entirely a result of cleavage and that many of these transcripts are a direct result of prematurely stalled elongation complexes, which failed to extend to position +24 before the complexes were purified. This point will be discussed later in more detail. The extension of the 5' cleavage products supports the notion that RNAP can successfully resume elongation following cleavage of its transcript, lending credence to the cleave-and-restart mechanism.

Measuring the decrease in the intensity of the bands corresponding to the initial 24-mers and 25-mers yields the rate of disappearance of these transcripts over time. Since we are measuring the rate of disappearance of the parent transcripts, these rates correspond to the absolute rates of transcript cleavage. Figure 2.2 plots the disappearance of the C25 and U25 transcripts as a function of time. The decay of each 25-mer is characterized by a single exponential decay, from which a pseudo-first-order rate constant can be extracted. Both rate constants describing the decay of C25 and U25 are on the order of 10^{-2} events per minute, indicating a slow rate of intrinsic cleavage. The slow rate calculated in our study is consistent with slow rates of intrinsic cleavage observed in other studies (Surratt *et al*., 1991; Erie *et al*., 1993). The more striking result, however, is the comparatively faster rate of cleavage of C25 to that of U25. The cleavage is C25 is on the order of three times faster than that of U25. Cleaving a 3' terminally correct transcript, properly aligned with its cognate DNA base in the RNAP active site, faster than cleaving a terminally erroneous transcript, presumably contributing to de-optimized RNA-DNA base pairing geometry in the active site, is contradictory to other published results (Jeon and Agarwal, 1996; Zenkin *et al*., 2006). In contrast to our

experiments, the transcript cleavage experiments performed in these other studies have relied on TECs constructed from artificial scaffolds, as opposed to promoter-initiated TECs. These authors conclude from experiments performed on artificial scaffolds that 3' misincorporated nucleotides are preferentially removed via intrinsic cleavage. Alternatively, our results suggest that TECs synthesized *de novo* from a promoter cleave transcripts with a slight preferential cleavage of correct over incorrect nucleotides.

Gre-mediated cleavage of correctly incorporated and misincorporated transcripts

To explore the role of GreA and GreB in facilitating the removal of a terminal error in a nascent transcript, we performed the same cleavage reactions, as detailed above, but in the presence of each Gre factor. Each complex (A24, C25, U25) was coincidently combined with either GreA or GreB and placed at 37 °C to permit factor-mediated transcript cleavage. The consequent RNA products were electrophoresed, with representative gels corresponding to GreA- and GreB-mediated cleavage reactions presented in Figures 2.3 and 2.4, respectively. As demonstrated by the banding pattern in these figures, the signal intensities corresponding to the parent transcripts (the 24- and 25-mers) markedly wane over time when compared to those in Figure 2.1. It appears, therefore, that the disappearance of the parent transcripts is faster when GreA and GreB are present in the reaction. Expectedly, GreA and GreB significantly enhance the endogenous hydrolytic activity of RNAP.

The decay of the C25 and U25 transcripts were plotted as a function of time to extract the pseudo-first-order rate constants of the Gre-mediated cleavage reactions

(Figure 2.5). Corroborating the initial observation obtained from the electrophoretic decay pattern, the rate constants of the Gre-mediated cleavage reactions are an order of magnitude greater than those of the intrinsic cleavage reactions. Similar to the intrinsic cleavage reactions, the cleavage reactions involving GreA result in an order of magnitude slower rate of cleavage of U25 complexes compared to that of C25 complexes. Unlike the intrinsic and GreA-mediated cleavage reactions, however, the cleavage reactions involving GreB yield a slightly (2.5-fold) faster rate of cleavage of U25 complexes compared to that of C25 complexes. These results suggest that Gre-mediated transcript cleavage exhibits a differential preference for terminally correct or terminally misincorporated nucleotides. While GreA cleaves terminally correct nucleotides with greater efficiency, GreB cleaves terminal nucleotides without regard to the identity of the nucleotide.

Transcript cleavage patterns

While the RNA products resulting from intrinsic cleavage are not entirely discernible due to low cleavage rates (Figure 2.1), the products of the Gre-mediated cleavage reaction can be characterized. Gre-mediated cleavage produces a distinguishing pattern of bands, which are indicative of the appearance of 5' cleavage products (Figures 2.3 and 2.4). The identities of these cleavage products, including the number of bases in the transcript and the identity of the base positioned in the active site, are indicated to the left of each gel. Notably, both GreA- and GreB-mediated cleavage of terminally correct transcripts (A24 and C25) render the same banding pattern. In other words, the cleavage

of transcripts bearing terminally correct nucleotides at positions +24 or +25 yields the same cleavage products. Complexes bearing a perfectly matched DNA-RNA hybrid, namely A24 and C25, exhibit the same preferences for cleavage sites along the transcript. These preferential sites are cleaved in RNAP's active site to yield the 5' cleavage products A19 and G17. Both A24 and C25 transcripts are preferentially cleaved to yield a regenerated transcript possessing a 3' purine base poised in the active site. Alternatively, complexes bearing terminally misincorporated transcripts, namely U25, are stimulated to preferentially cleave their transcripts so as to yield a regenerated transcript with a 3' pyrimidine base positioned in the active site. As a result, the cleavage activity of U25 complexes gives rise to the 5' cleavage products U18 and U17. This observation suggests that RNAP, upon having misincorporated a pyrimidine at position +25, may systematically halt at the site of each successive pyrimidine as it reverse translocates, or backtracks, along the nascent transcript.

While every attempt is made to synchronize TECs at a given position, there is inevitable heterogeneity among the complexes during transcript extension to position +24. Therefore, in the synthesis of A24 complexes, a subset of complexes will stall before reaching A24 and will remain at a given position once purified. As a result, there is concern as to whether or not the bonafide cleavage products can be distinguished from those transcripts that originally failed to be extended to position +24. To ensure that the 5' cleavage products of U25, which yield regenerated 3' pyrimidines, are not a result of extension of transcripts less than 24 nucleotides with UTP, we conducted experiments with 3' end-labeled transcripts. In these experiments, A24 complexes were synthesized in the presence of non-radioactive NTPs. These unlabeled complexes were then walked to

position +25 with alpha-labeled UTP and subsequently subjected to GreB-mediated cleavage (Figure 2.5). Figure 2.6 displays the resultant RNA products of the cleaved, end-labeled parent transcript (U25). Only a single row of bands is observed on the gel, and these bands correspond to the parent U25 transcripts. As expected, no bands indicative of 5' cleavage products appear overtime. Likewise, the absence of bands that would electrophorese faster than the U25 bands suggests that radio-labeled UTP is not incorporated at other positions prior to position +25. Hence, the banding patterns observed in the body-labeled transcript experiments are due to the appearance of 5' cleavage products and not the extension of transcripts less than 24 nucleotides.

DISCUSSION

Our results of the intrinsic cleavage reaction lead us to the surprising conclusion that a misincorporated transcript is cleaved more slowly than correctly incorporated transcripts. In fact, the striking conclusion is precisely that misincorporated bases are not cleaved faster, as was expected. Previous studies have demonstrated that RNAP selectively and preferentially cleaves misincorporated bases from its transcripts (Jeon and Agarwal, 1996; Zenkin *et al*., 2006). These results are in direct contradiction with the ones that we have presented. An explanation to this discrepancy may involve the substrates used in the cleavage experiments. These previous studies utilized artificial scaffolds as subtrates, whereas our transcription complexes were promoter-initiated and contained longer (>100 nt) DNA substrates and RNA transcripts. The length of the DNA

substrate is an important detail, given that the artificial scaffolds tend to lack upstream

segments of templated and non-templated DNA,in addition to downstream segments of

DNA.   Since reverse translocation involves the shifting of the transcription bubble along

with the concomitant unwinding of the upstream dsDNA and the rewinding of the

downstream DNA, the contacts involved in the scaffold may not be sufficiently satisfied

upon reverse translocation due to an insufficient segment of upstream DNA. Satisfying

these regenerated contacts is absolutely crucial, in light of the energy required to disrupt

the original contacts in order to reverse translocate the entire transcription complex.   In

contrast, transcripts synthesized *de novo* do not share this feature of missing nucleic acid

segments. Protein-nucleic acid and nucleic acid-nucleic acid contacts can be fulfilled

upon shifting the transcription bubble in the backwards direction. Additionally, the

scaffolds do not permit the length of cleavage products that we are able to observe in our

experiments involving GreA and GreB. This result is most likely due to the

complications encountered upon reverse translocating a complex that is missing key

nucleic acid features, as detailed above. If a complex can not properly translocate

backwards, then it would be logical to assume that the extent of "trimming" of the

nascent transcript would be dramatically reduced, in light of evidence suggesting that the

length of a RNA cleavage product is diagnostic of the extent of complex backtracking.

It has been suggested that RNAP, dynamically capable of sampling a variety of

conformations, isomerizes into an "unactivated" conformation during NTP deprivation or

when confronted by pause signals (Erie, 2002). Transcribing at a much slower rate than

"normal" RNA synthesis, the unactivated state exhibits a higher fidelity than the rapidly

catalyzing state (Foster *et al*., 2001). As such, it is possible that, when RNAP has makes

an error of incorporating a non-cognate base into the growing RNA chain, RNAP decays

onto an unactivated pathway. In other words, the presence of an error at the site of

catalysis may drive the equilibrium of the TEC population towards the unactivated

pathway. A misincorporation event may, therefore, facilitate a quicker transition to the

unactivated state. Once on the unactivated pathway, RNAP catalyzes elongation at a

much slower rate. Previous studies have demonstrated that a misincorporation event

reduces the rate of incorporation of the next incoming templated nucleotide (Erie *et al.*,

1993; Foster *et al.*, 2001). If RNAP is slow to extend the transcript beyond the

misincorporated base, then RNAP may indeed have a slow rate of recovery to a state of

resumed elongation. In such an instance, RNAP may experience a longer dwell time at

the site of a misincorporated base in relation to that experienced at the site of a correctly

incorporated base. Therefore, the increased dwell time at a given position on the

transcript may serve as a signal for the recruitment of transcription accessory factors,

such as GreA and GreB, to assist in rescuing RNAP and promoting resumed elongation.

The Gre factors can promote backtracking of the transcription complex and stimulate

RNAP's endogenous nucleolytic activity to remove the segment of RNA containing the

misincorporated base. Supporting this notion is the observation made in a previous study

that terminally misincorporated transcripts were cleaved rapidly by GreA, but the

transcripts that were extended beyond the mismatch were not readily cleaved (Erie *et al.*,

1993). In this manner, transcriptional fidelity may be maintained, despite the momentary

occurrence of an error in the nascent transcript.

Taken together, our observations suggest that intrinsic differences in the rates of

cleavage of matched versus mismatched transcripts is likely to not play a significant role

in maintaining fidelity. A misincorporation event, however, does trigger RNAP to possess a higher probability of decaying into the unactivated pathway. It is this unactivated conformation adopted by RNAPs bearing terminally misincorporated transcripts that is recognized by Gre factors, particularly GreB. Hence, it is the fall from the activated pathway of rapid transcript extension, and not the immediate intrinsic removal of the misincorporated base, that results in higher transcriptional fidelity. If the unactivated pathway behaves like a checkpoint in transcription regulation, then the Gre factors, most likely GreB, may possess a greater affinity for unactivated complexes to facilitate the removal of an error, which originally resulted in the decay from the activated pathway. As a result of Gre action, errors may be removed from a nascent transcript, thereby increasing the fidelity of transcript synthesis by RNAP. In this manner, the Gre factors may serve as general rescue and fidelity factors in the process of transcription.
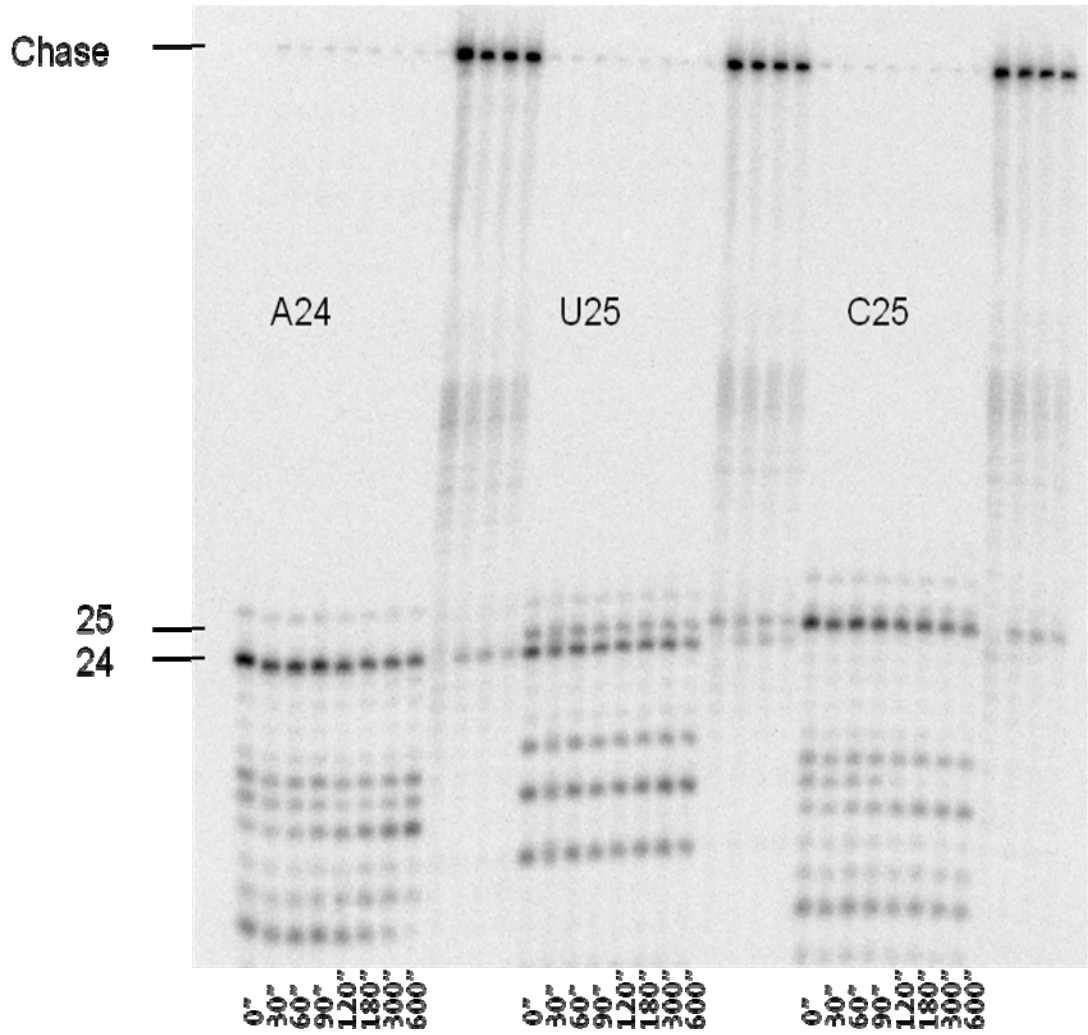
Figure 2.1 Intrinsic cleavage of terminally correct (A24, C25) and terminally incorrect (U25) transcripts.

Figure 2.1 Intrinsic cleavage of terminally correct (A24, C25) and terminally incorrect (U25) transcripts. Ternary elongation complexes were generated from the DE13 template and were stalled at position +24 (A24) with CTP deprivation. Subsets of these complexes were incubated with either the next templated nucleotide (CTP) or with UTP to induce terminal misincorporation. The resulting complexes, C25 and U25, in addition to the parent A24 complex, were monitored for decay due to transcript cleavage as a function of time. Reaction aliquots were removed and quenched at 0", 30", 1', 2', 5', 10', 20', and 40'. Additional aliquots were removed and chased with all four NTPs to ensure that the complexes were catalytically competent.
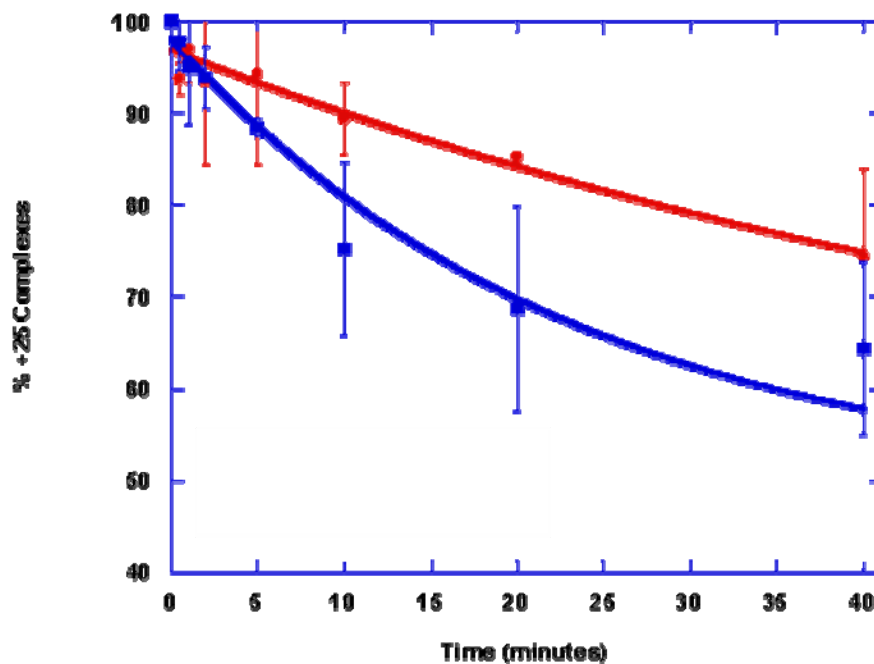
Figure 2.2 Intrinsic cleavage of complexes stalled at position +25. The decay of +25 transcripts, U25 (red) and C25 (blue), are plotted as a function of time. The decay curves are fit to single exponentials, and the resultant psuedo-first-order rate constants are elucidated. The intrinsic cleavage rate constants for U25 and C25 are 1.5 (+/- 0.2) * 10^{-2} min^{-1} and 4.3 (+/- 0.7) * 10^{-2} min^{-1}, respectively. The error obtained at each data point was calculated from the standard deviation of at least three independent trials of the intrinsic transcript cleavage experiment. The error in each measurement is the calculated standard deviation of at least three independent experimental trials.

Figure 2.3 GreA-mediated cleavage of terminally correct (A24, C25) and terminally incorrect (U25) transcripts. Ternary elongation complexes were generated from the DE13 template and were stalled at position +24 (A24) with CTP deprivation. Subsets of these complexes were either incubated with the next templated nucleotide (CTP) or with UTP to induce terminal misincorporation. The resulting complexes, C25 and U25, in addition to the parent A24 complex, were monitored for decay due to GreA-mediated transcript cleavage as a function of time. Reaction aliquots were removed and quenched at 0", 10", 20", 30", 40", 50", 1', 2', 5', and 10'. Additional aliquots were removed and chased with all four NTPs to ensure that the complexes were catalytically competent. The ladder to the left of the gel indicates the length of the 5' transcript cleavage products, as well as the identity of the regenerated 3' base held within RNAP's active site.
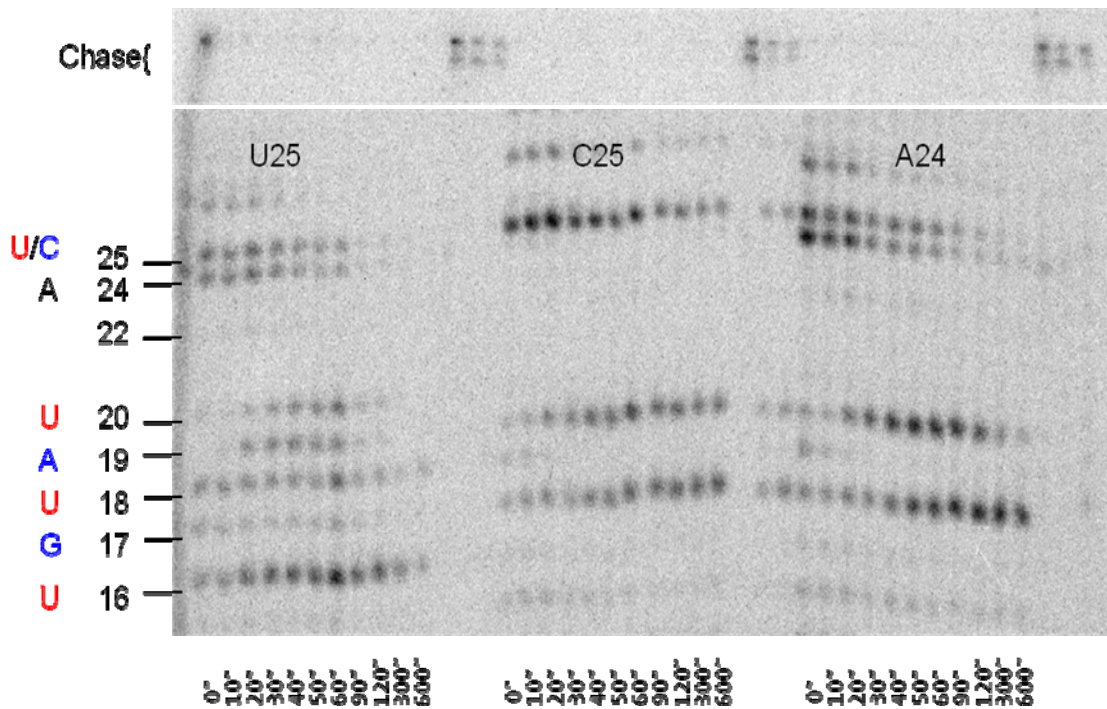
Figure 2.4 GreB-mediated cleavage of terminally correct (A24, C25) and terminally incorrect (U25) transcripts. Ternary elongation complexes were generated from the DE13 template and were stalled at position +24 (A24) with CTP deprivation. Subsets of these complexes were either incubated with the next templated nucleotide (CTP) or with UTP to induce terminal misincorporation. The resulting complexes, C25 and U25, in addition to the parent A24 complex, were monitored for decay due to GreB-mediated transcript cleavage as a function of time. Reaction aliquots were removed and quenched at 0", 10", 20", 30", 40", 50", 1', 2', 5', and 10'. Additional aliquots were removed and chased with all four NTPs to ensure that the complexes were catalytically competent. The ladder to the left of the gel indicates the length of the 5' transcript cleavage products, as well as the identity of the regenerated 3' base held within RNAP's active site.
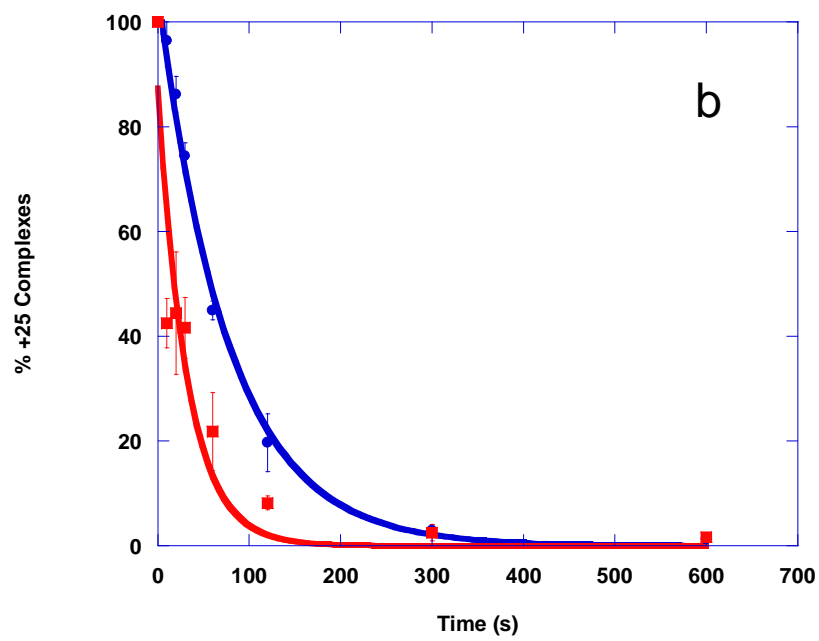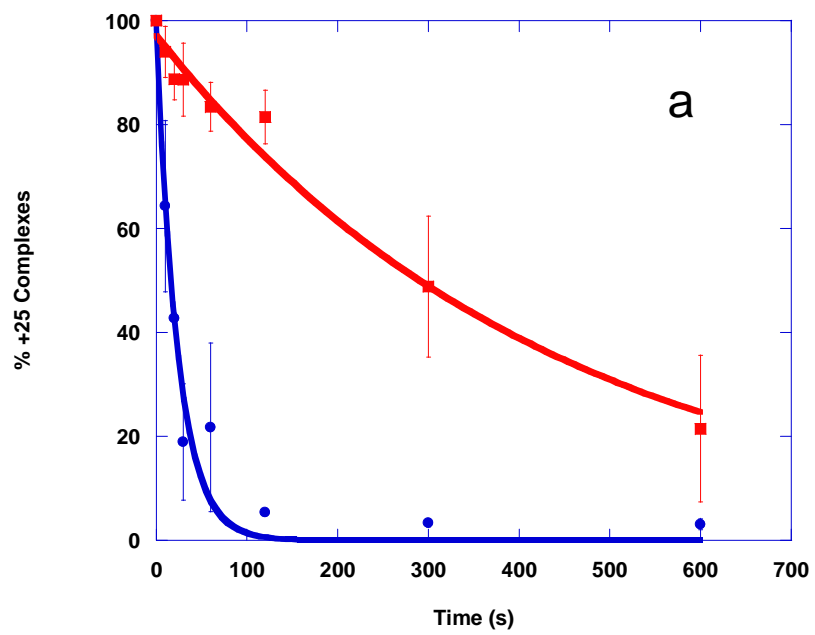
Figure 2.5a-b Gre-mediated cleavage of complexes stalled at position +25.

Figure 2.5a-b Gre-mediated cleavage of complexes stalled at position +25. The decay of +25 transcripts, U25 (red) and C25 (blue), in the presence of GreA (a) or GreB (b) are plotted as a function of time. The decay curves are fit to single exponentials, and the resultant psuedo-first-order rate constants are elucidated. The GreA-mediated cleavage rate constants for U25 and C25 are 0.23 (+/- 0.02) * 10^{-2} s^{-1} and 4.3 (+/- 0.6) * 10^{-2} s^{-1}, respectively. The GreB-mediated cleavage rate constants for U25 and C25 are 3.1 (+/- 0.8) * 10^{-2} s^{-1} and 1.3 (+/- 0.1) * 10^{-2} s^{-1}, respectively. The error in each measurement is the calculated standard deviation of at least three independent experimental trials.

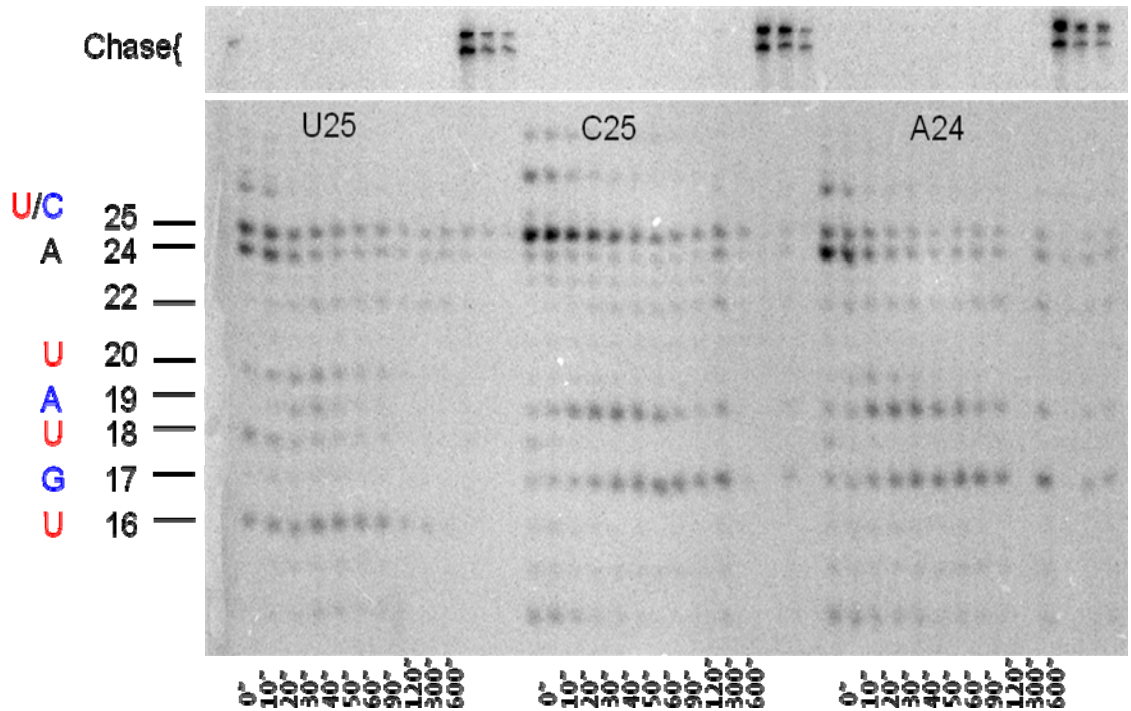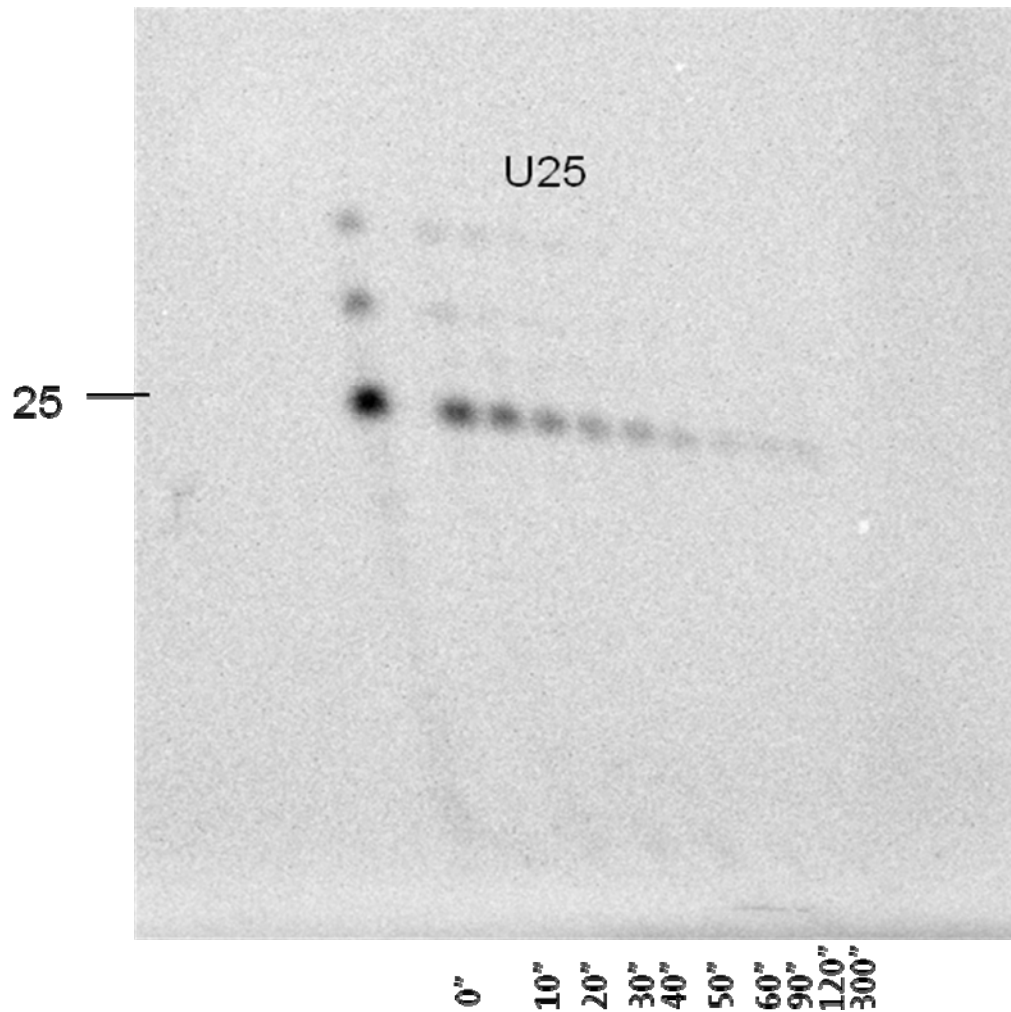**Figure 2.6 Gre-mediated cleavage of end-labeled U25 complexes.** Ternary elongation complexes were generated from the DE13 template and were stalled at position +24 (A24) with CTP deprivation. These complexes were then incubated with radiolabeled UTP to induce terminal misincorporation. The U25 complexes were monitored for decay due to Gre-mediated transcript cleavage as a function of time. Reaction aliquots were removed and quenched at 0", 10", 20", 30", 40", 50", 1', 2', 5.

BIBLIOGRAPHY

Burova, E. et al., (1995). Escherichia coli NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. J. Bact. *177*(5), 1388-1392.

Burgess, R.R. and Jendrisak, J.J. (1975). A Procedure for the Rapid, Large-Scall Purification of Escherichia coli DNA-Dependent RNA Polymerase involving Polymin P Precipitation and DNA-Cellulose Chromatography.

Borukhov, S. et al., (1993). Transcript Cleavage Factors from E. coli. Cell *72*, 459-466.

Borukhov, S. et al., (1992). GreA Protein: A Transcription Elongation Factor from Escherichia coli. Proc. Natl. Acad. Sci. USA *89*, 8899-8902.

Erie, D.A. et al., (1993). Multiple RNA Polymerase Conformations and GreA: Control of the Fidelity of Transcription. Science *262*, 867-873.

Foster, J.E. et al., (2001). Allosteric Binding of Nucleoside Triphosphates to RNA Polymerase Regulates Transcription Elongation. Cell *106*, 243-252.

Holmes, S.F., and Erie, D.A. (2003). Downstream DNA Sequence Effects on Transcription Elongation: Allosteric Binding of Nucleoside Triphosphates Facilitates Translocation Via a Ratchet Motion. J. Bio. Chem. *278*, 35597-35608.

Hsu, L.M et al., (1995). Escherichia coli Transcript Cleavage Factors GreA and GreB Stimulate Promoter Escape and Gene Expression In Vivo and In Vitro. Proc. Natl. Acad. Sci. USA *92*, 11588-11592.

Jeon, C and Agarwal K. (1996). Fidelity of RNA Polymerase II Transcription Controlled by Elongation Factor TFIIS. Proc. Natl. Acad. Sci. USA *93*, 13677-13682.

Komissarova, N., and Kashlev, M. (1997). RNA Polymerase Switches Between Inactivated and Activated States By Translocating Back and Forth along the DNA and the RNA. J. Bio. Chem. *272*, 15329-15338.

Koulich, D. et al., (1997). Domain Organization of Escherichia coli Transcript Cleavage Factors GreA and GreB. J. Bio. Chem *272*(11), 7201-7210.

Laptenko, O. et al., (2003). Transcript Cleavage Factors GreA and GreB act as Transient Catalytic Components of RNA Polymerase. EMBO *22*(23), 6322-6334.

Loizos, N. and Darst, S. A. (1999). Mapping Interactions of Escherichia coli GreB with RNA Polymerase and Ternary Elongation Complexes. J. Biol. Chem. *274*, 23378-23386.

Markovtsov, V. et al., (1996). Protein-RNA Interactions in the Active Center of Transcription Elongation Complex. Proc. Natl. Acad. Sci. USA. *93* (3221-3226).

Marr, M.T. and Roberts, J.W. (2000). Function of Transcript Cleavage Factors GreA and GreB at a Regulatory Pause Site. Mol. Cell. *6*, 1275-1285.

Mustaev, A. et al., (1993). Active Center Rearrangement in RNA Polymerase Initiation Complex. J. Bio, Chem. *268*, 19185-19187.

Opalka, N. et al., (2003). Structure and Function of the Transcription Elongation Factor GreB Bound to Bacterial RNA Polymerase. Cell *114*, 335-345.

Polyakov, A. et al. (1998). Visualization of the Binding Site for the Transcript Cleavage Factor GreB on Escherichia coli RNA Polymerase. J. Mol. Biol. *281*, 465-473.

Reeder, T.C. and Hawley, D. K. (1996). Promoter Proximal Sequences Modulate RNA Polymerase II by a Novel Mechanism. Cell *87*, 767-777.

Rudd, M.D. et al., (1994). The Active Site of RNA Polymerase II Participates in Transcript Cleavage with Arrested Ternary Complexes. Proc. Natl. Acad. Sci. USA *91*, 8057-8061.

Stebbins, C.E. et al., (1995). Crystal Structure of the GreA Transcript Cleavage Factor from Escherichia coli. Nature *373*, 636-640.

Surratt, C.K. et al., (1991). Spontaneous Cleavage of RNA in Ternary Complexes of Escherichia coli RNA Polymerase and its Significance for the Mechanism of Transcription. Proc. Natl. Acad. Sci. USA *88,* 7983-7987.

Uptain, S.M., and Chamberlin, M.J. (1997). *Escherichia coli* RNA Polymerase Terminates Transcription Efficiently at Rho-independent Terminators on Single-Stranded DNA Templates. Proc. Nat. Acad. Sci. *94*, 13548-13553.

Zenkin, N et al., (2006). Transcript-Assisted Transcriptional Proofreading. Science *313*, 518-520.

CHAPTER THREE

TO ARREST OR NOT TO ARREST:
A STUDY OF THE RESPONSE TO PAUSING AND TERMINATION SIGNALS

INTRODUCTION

RNA polymerase (RNAP) is the enzyme responsible for catalyzing the first step in a cascade of events leading towards gene expression. Since the dissemination of genetic information plays such a critical and integral role within the context of cellular processes, RNAP must uphold the stringent guidelines of catalyzing processive DNA-directed synthesis of RNA transcripts at a reasonable rate and with high fidelity. To ensure that these requirements are met and sustained, RNAP serves as a major target of rigorous regulation and is consequently subject to extrinsic regulatory factors, such as transcription accessory proteins, and to regulatory sequence elements, such as pause and termination sites located throughout the genome (Erie et al., 1992; Uptain and Chamberlain, 1997; Erie, 2002).

One of the many ways that RNAP is thought to be regulated is through the various conformations that it samples. Indeed, RNAP has demonstrated unprecedented dynamic and conformational flexibility in catalyzing multiple reactions along a branched pathway (Erie *et al*., 1993; Erie, 2002). A later study documented that RNAP is capable of incorporating nucleotides into a growing transcript along a fast and a slow pathway by adopting an "activated" and an "unactivated" conformation, respectively (Foster *et al*., 2001). Additional evidence revealed a quadratic dependence of nucleotide incorporation on the concentration of a downstream templated NTP (Foster *et al*., 2001; Holmes and Erie, 2003). This observed quadratic dependence suggested that transcription elongation is mechanistically governed by allosteric control and thus pointed strongly to the existence of an allosteric NTP binding site in addition to the canonical catalytic NTP binding site. In an attempt to locate a candidate allosteric site, Holmes and Erie studied the existing crystal structures of *T. aquaticus* and *T. thermophilus* RNAPs and subsequently built a working structural model of *E. coli* RNAP, shown in Figure 3.1 (Zhang *et al*., 1999; Murakami *et al*., 2002; Vassylyev *et al*., 2002). By studying the crystal structures, the investigators noted a structural region found within the β-subunit side of RNAP's main channel that may serve as a putative allosteric site. Located in the vicinity of the rifampicin binding region, this putative NTP binding site is comprised of two major structural components, fork loop 2 (βD-loop I) and a Walker B motif. Sequence alignments of these two regions are presented in Figure 3.2 (Holmes and Erie, 2003). Fork loop 2 is a highly conserved, flexible loop that lies across from the downstream DNA and contributes to an overall fold, including the β-pincer, which also interacts with the downstream DNA and resembles a nucleotide binding site (Holmes and

Erie, 2003). The Walker B motif, on the other hand, is composed of a highly conserved

consensus sequence, DHXG, and is located at the back of fork loop 2 (Figure 3.1). In

NTPases, the DXXG consensus sequence provides an aspartate residue to chelate,

indirectly through a water molecule, the Mg^{2+} of the Mg^{2+}-NTP complex, and a glycine

residue, whose backbone hydrogen bonds to the γ-phosphate of the NTP (Pai *et al*., 1990;

Bourne *et al*., 1991; Vassylyev *et al*., 2002). It is suggested that this sequence plays a

similar role in RNAP for NTPs that may bind to the putative allosteric site (Holmes and

Erie, 2003). It is thought that binding of a templated NTP to this putative site enhances

the rate of nucleotide incorporation in an allosteric manner by facilitating translocation of

the ternary (RNAP-DNA-RNA) elongation complex (Holmes and Erie, 2003).

To probe this putative site for the ability to bind templated NTPs and to affect the

incorporation of nucleotides, mutations were generated in fork loop 2 and in the Walker

B motif (Kennedy and Erie, submitted; Kennedy, 2007). The fork loop 2 mutant, named

Δ-loop RNAP, contains a deletion of four key residues, R542-F545. In the Walker B

motif mutant, named Walker RNAP, the conserved aspartate and glycine residues of the

DXXG motif are replaced with alanine residues (D446A/G449A). Transient-state kinetic

analysis of multiple nucleotide incorporations revealed that both mutant RNAPs exhibit

altered nucleotide incorporation kinetics within the context of template-dependent NTP

binding. Specifically, Δ-loop RNAP maintains a slow phase of synthesis

indistinguishable from that of wild-type (wt) RNAP, while the fast phase, a hallmark of

allostery, is dramatically reduced (Kennedy and Erie, submitted). In contrast, Walker

RNAP maintains a fast phase of synthesis indistinguishable from that of wtRNAP, but

demonstrates a four-fold increase in the rate of synthesis in the slow phase (Kennedy,

2007). These results strongly suggest that the proposed site is indeed a NTP binding site and that binding of templated NTPs to this site facilitates translocation and the shuttling of NTPs to the active site.

In light of these altered nucleotide incorporation kinetics, we wished to examine the role of this allosteric binding site in transcriptional pausing and termination. To this end, we tested Δ-loop and Walker RNAP for the ability to detect and to escape two different classes of pause sites: one which is dependent of a hairpin in the RNA transcript (*his* pause) and one that has no secondary structure in the RNA transcript (*ops* pause) (Figure 3.3). Additionally, we investigated the termination efficiency of these mutant RNAPs at a simple ρ-independent terminator. We also explored the degree to which NusA and NusG may modulate the pausing and termination properties of the mutants, in light of the fact that NusA increases dwell time at the *his* pause site and termination efficiency at ρ-independent terminators, while NusG decreases dwell time at the *ops* pause site (Schmidt and Chamberlain 1987; Artsimovitch and Landick, 2000). Finally, we studied the *in vivo* effects on cell growth and transcription termination efficiency of cells expressing Δ-loop and Walker RNAP. As discussed below, our results strongly suggest that the allosteric site plays important roles in both pausing and termination.

MATERIALS AND METHODS

Source of Protein

His-tagged *E. coli* wild type (wt) RNA polymerase was purified from mid-log phase cells of strain RL916 (gift of R. Landick), as previously described (Burgess and Jendrisak, 1975; Uptain *et al*., 1997). The plasmid pIA509 (gift of I. Artsimovitch), which encodes the β, β', and both α subunits of RNAP, was used as a template in site-directed mutagenesis reactions to generate the Walker B motif variant of RNAP (Walker RNAP) (Kennedy, 2007). The Walker RNAP mutant was purified from late-log phase cells of strain HMS174(DE3)pLysS (Novagen), as described previously (Kennedy, 2007). The plasmid pRL706 (gift of R. Landick), which encodes the β subunit of RNAP, was used as a template in site-directed mutagenesis reactions to generate the βD-Loop variant of RNAP (Δ-loop RNAP) (Kennedy, 2007). The Δ-loop RNAP was purified from late-log phase cells of strain TOM100 (gift of T. Santangelo), as described previously (Kennedy, 2007). Purified *E. coli* NusA and NusG were gifts of I. Artsimovitch.

Source of Oligonucleotides

Linear DNA templates for *in vitro* transcription reactions were amplified via the polymerase chain reaction (PCR) from the plasmids pDE13 (Erie *et al*., 1993) and pIA349 (Artsimovitch, 2007). Amplification of pDE13 generated a 5'-biotinylated 540 nt DNA template containing the λP$_R$ promoter and encoding a transcript in which the first

cytosine to be incorporated is at position +25. Additionally, the DE13 template encodes

the λ tR2 terminator at position +264. Amplification of pIA349 generated a 5'-

biotinylated 500 nt DNA template containing the T7A1 promoter and encoding a

transcript in which the first uridine to be incorporated beyond the transcription start site is

at position +38. The IA349 template also encodes two pause sites, the *ops* pause at

position +45 and the *his* pause at position +145, and a terminator sequence, the *his*

terminator at position +225.

In Vitro Transcription Termination Assay

Transcription termination assays were performed primarily on the DE13 template,

which contains the tR2 terminator, with wtRNAP, Δ-loop RNAP, and Walker RNAP.

Open promoter complexes (OPCs) were formed by incubation of 100 nM RNAP with 50

nM 5'-biotinylated DE13 template in transcription buffer (30 mM HEPES, pH 8.0, 10

mM magnesium glutamate, 200 mM potassium glutamate, 25 μg/mL bovine serum

albumin (BSA), 1 mM dithiothreitol (DTT)) at 37 °C. Upon a 10 minute incubation

period, the OPCs were brought to ambient temperature and were supplemented with 20

μM ATP, UTP, and [α-^{32}P]GTP (800 Ci/mmol) (MP Biomedicals) to initiate transcription

in a 15 μL reaction volume. OPCs generated by wtRNAP were allowed to react for 30

seconds in the presence of NTPs, whereas those generated by either the Δ-loop RNAP or

the Walker RNAP were allowed to react for a total of 2 minutes. Transcription reactions

initiated in the absence of CTP yielded elongation complexes stalled at position +24

(A24), otherwise known as stalled elongation complexes (SECs). Newly formed SECs

were spiked with 50 μM of all four NTPs to begin the termination assay, and 1 μL

reaction aliquots were removed at specified times over a course of 10 minutes and were

quenched in 4 μL quench dye (95% formamide, 0.05% bromophenol blue, and 0.05%

xylene cyanol). Termination assays were performed in the presence and absence of 100

nM NusA to measure both NusA-mediated and intrinsic transcription termination at the

tR2 terminator. RNA products resulting from the termination assay were resolved and

analyzed by electrophoresis on 10 % polyacrylamide/8 M ureas gels.

In Vitro Transcription Pausing Assay

Transcription pausing assays were performed on the IA349 template with wtRNAP,

Δ-loop RNAP, and Walker RNAP. Open promoter complexes (OPCs) were preformed

by incubation of 75 nM RNAP with 50 nM 5'-biotinylated IA349 template bound to

streptavidin-coated magnetic beads (Promega) in transcription buffer (30 mM HEPES,

pH 8.0, 10 mM magnesium glutamate, 200 mM potassium glutamate, 25 μg/mL bovine

serum albumin (BSA), 1 mM dithiothreitol (DTT)) at 37 °C. Upon a 10 to 15 minute

incubation period, the OPCs were brought to ambient temperature and were

supplemented with 100 μM pApU (Midland Oligos) and 5 μM ATP, CTP, and [α-

^{32}P]GTP (800 Ci/mmol) to initiate transcription in a 20 μL reaction volume. OPCs were

allowed to react for approximately 5-7 minutes in the absence of UTP to generate SECs

halted at position +37 (G37). The resultant SECs were placed beside a strong magnet,

and the immobilized SECs were washed three times with ice-cold transcription buffer.

The purified SECs were resuspended in 45 μL transcription buffer and kept on ice until

the pausing assays were initiated. To begin the pausing assay, purified SECs were

brought to ambient temperature and were spiked with 100 μg/mL heparin and 100 μM of

all four NTPs in the absence or in the presence of either 100 nM NusA or NusG in a 15

μL reaction volume. At designated times, 1-2 μL reaction aliquots were quenched in 4

μL quench dye over a course of 10 minutes. RNA products resulting from the pausing

assay were resolved and analyzed by electrophoresis on 10% acrylamide/8 M urea gels.

Analysis of Pausing and Termination Assays

The radioactively body-labeled RNA products resulting from the above *in vitro*

assays were electrophoresed on denaturing poly-acrylamide gels. The gels were exposed

overnight against a phosphor screen, and these screens were scanned for radiation

exposure on a Biosciences PhosphorImager (Amersham). The patterned radioactive

bands detected via scanning were analyzed with ImageQuant software. Each band

corresponded to a RNA of a particular length, and measuring the intensity of each band in

each lane could be related to the relative amounts of different RNA products present over

time. Therefore, the percentage of transcribing complexes at any given position along the

template was calculated by dividing the amount of radioactivity in the corresponding

RNA band by the total amount of radioactivity in all the bands of equal in the same lane.

In the case of the pausing and termination assays, the percentage of complexes at a given

template position, whether at a pause or at a termination site, was measured with respect

to only those complexes at positions downstream of these complexes. In other words,

paused or terminated complexes were analyzed in relation to those complexes that "read through" the pause or termination sites.

In Vivo Growth Phenotype Studies

E. coli strain HMS174(DE3)pLysS transformed with either wt pIA509 or Walker mutant pIA509 were grown overnight at 37 °C, with shaking, from a single colony to a freshly saturated culture in super broth supplemented with 100 μg/mL carbenicillin. The saturated cultures were diluted 50-fold into fresh super broth supplemented with 100 μg/mL carbenicillin and were grown at 37 °C, with shaking, to an optical density (OD) of ~0.3 at 600 nm. Five serial 10-fold dilutions of the cultures were made in super broth (25 g tryptone, 15 g yeast extract, 5 g NaCl per liter of broth), and 5 μL of each of these serial dilutions were plated on LB-agar-ampicillin media supplemented with either 0 mM, 0.05 mM, 0.1 mM, 0.5 mM, or 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). In a second set of experiments, 10 μL of the serial dilutions were grown in the presence of 0.2 mM, 0.3 mM, 0.4 mM, and 0.5 mM IPTG. The plates were incubated at 37 °C for approximately 18 hours, and were monitored for growth of colonies expressing either wtRNAP or Walker RNAP.

In Vivo Termination Assays

E. coli strain RL211 (gift of R. Landick; Landick *et al.*, 1990) was transformed with pIA509 and pRL706 derivatives and were plated in the presence of 100 μg/mL ampicillin

on LB-agar plates. Resulting colonies were transferred in regular arrays to LB-agar-ampicillin plates supplemented with 0.5 mM IPTG. These master plates were incubated at 37 °C for approximately 18 hours and were replica-plated by velvet transfer to a series of different media. The replica was made onto LB-agar-ampicillin-IPTG plates supplemented with either 10 µg/mL or 20 µg/mL chloramphenicol. The LB-agar plates were incubated at 37 °C for approximately 18 hours.

RESULTS

Intrinsic Termination at the tR2 Terminator

Previous studies have demonstrated that both the Δ-loop RNAP and the Walker RNAP exhibit altered kinetics with respect to the binding of incoming NTPs and to the incorporation of templated NMPs (Kennedy and Erie, submitted; Kennedy, 2007). To determine whether or not either of these mutants exhibits any additional aberrant elongation properties, the transcription termination efficiency of wt and mutant RNAPs were examined. In this study, SECs were generated from both wt and mutant RNAPs on the DE13 template encoding the tR2 intrinsic terminator. The SECs were chased with 50 µM NTPs in the absence of any protein co-factor, and the ensuing transcript extension was monitored as a function of time. The electrophoresed RNA products resulting from this reaction are displayed in Figure 3.4. As can be seen in Figure 3.4, a distinct banding pattern develops over time during the resumption of transcript synthesis. Most notably, a characteristic band with marked intensity corresponds to transcripts that have been

64

extended to the tR2 terminator and have been subsequently terminated. In addition to this characteristic band, other bands of higher molecular weight transcripts are observed above the terminated transcript band. The transcripts that correspond to these specific bands are ones that have reached the tR2 terminator and have been extended past this intrinsic termination signal. As such, these transcripts are considered to be "read-through" transcripts. Termination efficiency is therefore defined as the ratio of the percentage of terminated transcripts to the summed percentages of terminated and read-through transcripts.

Figure 3.6 displays the plot of the termination efficiencies of wt-, Δ-loop-, and Walker RNAP at the tR2 terminator, as measured ten minutes after the resumption of elongation. At ten minutes, most RNAPs have reached, or in some cases have transcribed past, the tR2 terminator. The wtRNAP exhibits 60% intrinsic termination efficiency, the highest efficiency of all the RNAPs examined in this study. This measured termination efficiency of wtRNAP is within error of that determined in a previous study (Schmidt and Chamberlin, 1987). The Δ-loop RNAP exhibits similar, yet slightly lower, termination efficiency with respect to wtRNAP. In contrast, the Walker RNAP exhibits significantly reduced termination efficiency of 30%. These results indicate that the termination efficiency of Walker RNAP is only half that of wt- and Δ-loop RNAP.

NusA-Mediated Termination at the tR2 Terminator

In conjunction with exploring any potential aberrant intrinsic termination efficiencies of the mutant RNAPs, we wished to determine whether or not these mutant RNAPs would display altered sensitivities to the effects of NusA, a transcription termination accessory protein known to enhance termination at the tR2 terminator *in vitro* (Schmidt and Chamberlin, 1987). If the RNAP mutants are relatively insensitive to the presence of a termination signal, then perhaps these mutants would also exhibit resistance to the fundamental action of NusA.

To test whether or not NusA could modulate the termination efficiency of these mutants, termination assays were performed in the presence of 100 nM NusA. Given that NusA exhibits ~ 30 nM binding constant (von Hippel), we can comfortably assume that we are using saturating levels of NusA in this experiment. The electrophoresed RNA products resulting from the NusA-mediated termination reaction are displayed in Figure 3.5. Similar to Figure 3.4, Figure 3.5 reveals a distinct banding pattern with two sets of emerging prominent bands. Again, these sets of bands correspond to transcripts that have been either terminated at the tR2 terminator or extended past the tR2 terminator. In addition to these bands that have been observed also in Figure 3.4, intense bands unique to the set of date presented in Figure 3.5 are observed below the terminated transcript bands. These additional dark bands are discernible, however, only in the reactions with wtRNAP and with Δ-loop RNAP. The presence of these bands, corresponding to transcripts of various lengths, indicates that populations of elongating complexes are collecting for a period of time at particular positions along the DNA template and are subsequently resuming synthesis. This observation is consistent with previous studies

demonstrating that NusA, while an extrinsic termination factor, is also responsible for increasing pausing during *in vitro* transcription elongation (Kassavetis and Chamberlin, 1981; Schmidt and Chamberlin, 1984). Figure 3.5 demonstrates that NusA enhances pausing at distinct positions along the DNA template upstream of the termination signal. Although wt- and Δ-loop RNAP are sensitive to this sequence-dependent pausing mediated by NusA, Walker RNAP appears to be insensitive to the general enhanced pausing effects of NusA.

In addition to intrinsic termination efficiencies, Figure 3.6 depicts the NusA-mediated termination efficiencies of the various RNAPs at the tR2 terminator. As illustrated in Figure 3.6, wtRNAP demonstrates marked enhancement of transcription termination in the presence of NusA, as the termination efficiency is strengthened from 60% to 90%. This observed increase in termination efficiency at the tR2 terminator due to NusA action is supported by results obtained in earlier studies (Schmidt and Chamberlin, 1987). Also exhibiting a termination efficiency of 90%, the Δ-loop RNAP is clearly responsive to NusA-mediated termination, as this mutant achieves the same level of efficiency as that of wtRNAP in the presence of NusA. Lastly, the Walker RNAP exhibits the greatest enhancement of termination, with an intrinsic efficiency of 30% increasing to a NusA-mediated efficiency of 60%. Despite the Walker RNAP's noticeable intrinsic termination deficiency when compared to that of wt and of Δ-loop RNAPs, it appears that NusA is able to partially restore the Walker RNAP's termination efficiency to levels of *intrinsic* wt termination efficiency.

Since transcription termination arises from an RNAP in a paused transcriptional state (Erie, 2002; Landick, 2006), it follows logically to explore effects on pausing by Δ-loop and Walker RNAPs. To this end, we employed a DNA template amplified from pIA349 that encodes several sequence elements, including an *ops* pause site at position +45, a *his* pause site at position +145, and a *his* terminator at position +225. Using the IA349 template, SECs were generated from wt and mutant RNAPs and stalled at position +37 with UTP deprivation. SECs were purified from free enzyme and NTPs and were subsequently chased with 100 μM of all four NTPs. These chase reactions were designed to explore intrinsic RNAP effects on pausing. In a subset of reactions, 100 nM of either NusA or NusG were included to examine Nus-dependent effects on pausing. The transcripts extended by wt and mutant RNAPs in the presence and absence of NusA and NusG were resolved on denaturing acrylamide gels as a function of time. The resultant RNA products synthesized by wt-, Δ-loop-, and Walker RNAP are presented in Figures 3.7, 3.8, and 3.9, respectively.

By inspection of the gels shown in Figures 3.7-3.9, an immediate observation is the emergence of distinct banding patterns as a function to time. In each of the Figures, one can note the initially stalled, purified complex at position +37 at the beginning of each reaction. With the introduction of NTPs, the transcripts are briefly extended before pausing at the *ops* pause, as denoted by the presence of bands around position +45. As transcript extension is resumed, bands begin to appear around position +145, signaling the arrival of elongation complexes at the *his* pause. Finally, a subset of transcripts are

terminated at the *his* terminator (+225), while the rest are "read through" this terminator, as reflected by the emergence of higher molecular weight bands.

The emergence of these characteristic bands, particularly ones representing the pausing at the *ops* and *his* pause sites, as a function of time were quantified and are presented in Figures 3.10 and 3.11. Intrinsic and NusG-mediated pausing at the *ops* pause site are shown in Figure 3.10a-b (respectively), while intrinsic and NusA-mediated pausing at the *his* pause site are shown in Figure 3.11a-b (respectively). In both Figures, the percentages of paused complexes are plotted as a function of time. The plots yield simple exponential decay curves, and from these curves, two characteristic variables that define a pausing event, pause efficiency and the rate of pause escape, can be determined. Pause efficiency measures the ratio of complexes that pause at a defined template position to the total number of complexes present in the reaction. On the other hand, the rate of pause escape measures the rate at which those complexes that have paused at a site resume elongation of their transcripts. The rate of pause escape can also be expressed as the dwell time, or the inverse of the rate of escape, of complexes at a particular pause site. Table 3.1 catalogs the pause efficiencies and the dwell times of the mutant and wtRNAPs at each pause site.

As shown in Figure 3.10a, wt, Δ-loop, and Walker RNAPs display differing intrinsic pause efficiencies at the *ops* pause. The measured pause efficiency of wtRNAP is approximately 70%, a value similar to that of wtRNAP published in a previous study (Artsimovitch and Landick, 2000). Compared to that of wtRNAP, the pause efficiency of Walker RNAP is reduced by ~50%, while the pause efficiency of Δ-loop RNAP is reduced by ~30%. Similarly, the calculated dwell times at the *ops* pause are slightly

69

different among the RNAPs. Walker and wtRNAPs exhibit dwell times on average of 21 s, approximately 20% greater than that of Δ-loop RNAP. When NusG is present, the pause efficiencies of wt and Walker RNAPs undergo a modest decrease, but pause efficiency of Δ-loop RNAP remains unchanged. Interestingly, however, both mutant RNAPs have characteristically unaltered dwell times at the *ops* pause in the presence of NusG. NusG is expected to reduce dwell times experienced by RNAP specifically at the *ops* pause, as is the case with wtRNAP in this study (Figure 3.10b; Artsimovitch and Landick, 2000).

In Figure 3.11a, wt and Δ-loop RNAPs exhibit similar pause efficiencies at the *his* pause, with wtRNAP efficiency at 75% and Δ-loop RNAP efficiency just under 70%. This measured wtRNAP pause efficiency at the *his* pause is similar to previously reported results (Artsimovitch and Landick, 2000). Comparatively, Walker RNAP exhibits reduced pause efficiency, measured at approximately 45%. Although Walker RNAP is characterized with having a weaker pause efficiency, Walker and wtRNAPs display comparable dwell times at the *his* pause. Likewise, while Δ–loop and wtRNAPs have similar pause efficiencies, Δ-loop RNAP exhibits a slightly longer dwell time than wtRNAP. These trends in pause efficiency and dwell time are maintained when NusA is present, as shown in Figure 3.11b. Expectedly, NusA increases pause efficiencies and dwell times for all three RNAPs at the *his* pause. The pause efficiency of Walker RNAP is increased with NusA (approximately 65%), yet not fully restored to the level of intrinsic pause efficiency of wtRNAP. The dwell times of Walker and wtRNAPs are lengthened by NusA action to a comparable level of intrinsic dwell time observed for Δ-loop RNAP.

In addition to these pausing experiments involving a 100 μM NTP spike, we also performed some preliminary experiments involving reaction conditions used in a previous study (Artsimovitch and Landick, 2000). These reaction conditions include a limiting concentration of GTP (10 μM) mixed with 150 μM of the other three NTPs. The limiting-GTP spike supports an increased propensity to pause, as well as prolonged dwell times, at the *ops* pause site. These two features permit not only a greater ease of measuring lifetimes at the *ops* pause, but also better detection of pausing at two distinct sites, which include the *ops1* and the *ops2* sites, that comprise the *ops* element (Herbert et al., 2006). A representative gel showing the results of the pausing experiment performed with wt- and Δ-loop RNAPs under the GTP-limiting conditions is presented in Figure 3.12. Inspection of this gel reveals that pausing at the *ops1* and the *ops2* sites is quite discernible, compared to the results of the higher concentration experiments shown in Figures 3.7-3.9.

In Vivo Growth Phenotypes

In light of the deficiencies in transcription pausing and termination exhibited by the Walker RNAP in the *in vitro* experiments described above, we were interested in determining what effect, if any, this mutant would exert within a living *E. coli* cell. One simple way to probe for *in vivo* effects is to monitor the growth of *E. coli* that are expressing various levels of Walker RNAP and to compare this growth to that of *E. coli* that are expressing various levels of wtRNAP.

To perform such a test, HMS174(DE3)pLysS were grown to early log phase and were induced on LB-agar-ampicillin plates with varying concentrations of IPTG (0-1 mM) to facilitate expression of corresponding levels of either wtRNAP or Walker RNAP. Induction of these cells was implemented in early log phase (OD ~ 0.3 at 600 nm) in an attempt to reduce the amount of background wtRNAP expressed from genomic DNA. After 18 hours of incubation at 37 °C, the induced *E. coli* were examined for their extent of growth as judged by the number of colonies present on the solid media. Representative LB-agar-IPTG plates displaying growth of induced colonies are displayed in Figure 3.13a-d.

By inspecting the control 0 mM IPTG plate, one may note that cells harboring plasmid encoding either wtRNAP or Walker RNAP exhibit similar growth patterns. Cells plated at a concentration of 10^7 cells/mL display the greatest density of growth, when compared to the growth of the other plated concentrations. As expected, this density of cellular growth decreases concomitantly with decreasing plated concentrations. This observed growth pattern remains relatively unchanged on the plates supplemented with 0.05 mM and 0.1 mM IPTG. It is only in the range of 10^4-10^5 cells/mL that subtle differences in the amount of colonies on these plates are discernible, reflected equally in cells expressing either wtRNAP or Walker RNAP. Contrastingly, the plates supplemented with either 0.5 mM or 1 mM IPTG display marked changes in growth density. Not surprisingly, the cells expressing higher levels of wtRNAP fail to grow as densely as those subjected to lower concentrations of IPTG. The cells expressing the mutant RNAP, however, demonstrate even less density in cell growth, with only a few detectable colonies in the 10^7 cells/mL plated concentration.

To better resolve the difference in growth density between cells expressing either wtRNAP or Walker RNAP, plated cultures were induced with 0.2 mM, 0.3 mM, 0.4 mM, and 0.5 mM IPTG, and these plates are presented in Figure 3.14a-e. As can be seen from inspection of these plates, there is a marked reduction in the number of colonies expressing Walker RNAP in relation to those expressing wtRNAP at IPTG concentrations ranging from 0.2-0.4 mM.

In Vivo Termination Efficiencies

While the growth phenotype experiments demonstrated a general *in vivo* effect experienced by cells expressing Walker RNAP, we sought to advance our understanding of this mutant's properties by examining termination efficiency *in vivo*. To accomplish this end, we used DNA constructs that allow us to examine decreased termination efficiencies. To screen for alterations in *in vivo* termination efficiencies, derivatives of both pIA509 and pRL706 were used to transform strain RL211 (Landick *et al.*, 1990). RL211 is a W3110 derivative constructed with λRL12 that carries a 2-kb DNA fragment encoding a transcriptional fusion of *S. marcescens* Δ*trpL430trpE* (Stroynowski and Yanofsky, 1982) to *cat* from plasmid pACYC184 (Chang and Cohen, 1978). Deficient in *recA*, RL211 permits expression of altered termination efficiencies, as this strain minimizes recombination between recombinant RNAP and genomic wtRNAP. Decreased termination efficiencies are detected indirectly by monitoring termination at a simple ρ-independent terminator preceding a *cat* gene. If mutant RNAPs are termination deficient at this terminator, cells expressing these mutants will be insensitive to

chloramphenicol. RL211 transformed with pIA509 and pRL706 derivatives were transferred via velvet replica plating techniques to agar plates in the presence of chloramphenicol. The plates were monitored for growth and are presented in Figure 3.16a-b. Upon incubation, only a few colonies were present on the LB-agar-chloramphenicol plates. These colonies were largely those expressing a pIA509 derivative, namely Walker RNAP. A few colonies expressing wtRNAP, both pIA509 and pRL706 derivatives, were observed to grow in the presence of chloramphenicol, while no colonies expressing Δ-loop RNAP, a pRL706 derivative, survived. Based on these observations of chloramphenicol insensitivity, it appears that Walker RNAP is deficient in terminating at a ρ-independent terminator within the context of a living cell. On the contrary, Δ-loop RNAP is efficient at terminating *in vivo* transcription at the intrinsic terminator, as cells expressing this mutant are exceptionally sensitive to chloramphenicol. Not unexpectedly, a handful of colonies reflect a background level of wtRNAP-expressing cells that are insensitive to chloramphenicol.

DISCUSSION

Examining the various structures of RNAP, one can note two distinct channels, the main channel and the secondary channel (Chapter 1, Figure 1.2). Since the secondary channel is a funnel-shaped pore that permits direct access of bulk solvent to the active

site, many research teams suspect that NTPs bind to the active site via preferential entry through the secondary channel (Zhang *et al*., 1999; Adelman *et al*., 2004; Mukhopadhyay, *et al*., 2004; Batada, *et al*., 2004). Still, other evidence suggested that NTPs could bind inside the main channel (Holmes and Erie, 2003; Gong *et al*., 2005; Xiong and Burton, 2007), although there does not appear to be an "open" path that would allow NTPs bound in the main channel to be shuttled into the active site due to the presence of downstream DNA bound within this channel (Gnatt *et al*., 2001; Kettenberger *et al*., 2004; Westover *et al*., Wang *et al*., Vassylyev *et al*., 2007a; Vassylyev *et al*., 2007b). Combining these two different proposals with the studies of Δ-loop RNAP incorporation kinetics, Kennedy and Erie proposed a mechanistic model detailing three potential pathways to nucleotide incorporation (Kennedy and Erie, submitted). The first and ultimately the main pathway is through the secondary channel when RNAP's active site is located in the post-translocated register. Entry through the secondary channel is considered to be non-allosteric, providing that there is not a NTP bound in the allosteric site, since NTPs can directly access the active site along this pathway. Conversely, the other two pathways involve binding to the allosteric site, which is comprised partially of fork loop 2 and Walker B motif (Holmes and Erie, 2003; Kennedy and Erie, submitted). In the first of these two pathways, a templated NTP binds to the allosteric site, followed by rapid binding of a NTP to the active site via the secondary channel. Rapid binding of a NTP from bulk solvent is possible along this pathway since the binding of a NTP to the allosteric site shifts the equilibrium of the active site from the pre- to the post-translocated register, thereby opening the active site. Fork loop 2 affects this rapid NTP-facilitated pathway, in light of the fact that kinetics of

the fast phase of synthesis are reduced for Δ-loop RNAP. The other allosteric pathway, known as the "shuttling" pathway, involves binding of a templated NTP to the the allosteric site and subsequent shuttling of this NTP from the main channel to the active site. Since this latter pathway is the slower of the two allosteric pathways, shuttling of NTPs from the allosteric to the active site is favored when NTPs can not regularly access the secondary channel. The Walker B motif appears to affect this shuttling pathway, in that the rate of the slower phase is fast for Walker RNAP relative to wtRNAP.

As outlined in Figure 1.2 (Chapter 1), unactivated transcription complexes catalyzing "slow" synthesis of RNA are a precursor to further decayed states, including paused states. Pausing may be induced when a transcribing RNAP encounters genetic elements encoded in the DNA template or expressed in the nascent RNA (Erie, 2002). Two such pausing elements have been documented, the Class I and the Class II pause sites (Artsimovitch and Landick, 2000). Like r-independent terminators, Class I pauses require the formation of a hairpin RNA, such as those encountered at the *his* pause. Class II pauses, on the other hand, produce transcripts without any secondary structure, such as those produced at the *ops* pause, and these complexes are thought to be prone to backtracking after pause recognition. These two classes of pauses respond differently to transcription accessory factors. While Class I pauses are enhanced by NusA action, Class II pauses are reduced by NusG action and are also sensitive to Gre factors as a consequence of induced backtracking. A key feature to these pauses is the observation that two RNAPs, expressing different mutations within the rifampicin-binding domain, exhibit the same recognition of the two pause classes as wt RNAP, suggesting that recognition of the different pause classes proceed through a common "slow" intermediate

76

(Artsimovitch and Landick, 2000). As discussed below, the data presented here yield additional insight into the mechanism of pausing and termination and further suggest that Class I and Class II pauses may not proceed through a common intermediate.

Processes Involving RNA Hairpin Formation are Differentially Affected by the Fork Loop 2 Mutation: Recognition is Unaltered, but Dwell Time is Enhanced

Our results indicate that Δ-loop RNAP exhibits the same level of pausing efficiency at the *his* pause site when compared to that of wtRNAP. Additionally, the termination efficiencies of wt- and Δ-loop RNAP at the tR2 terminator are essentially equivalent. An important, common feature shared by the *his* pause, a Class I pause, and the tR2 terminator is the generation of RNA hairpins when RNAP transcribes these sequence elements. Therefore, since the efficiencies of Δ-loop RNAP at these particular sites remain unaltered, it appears that fork loop 2 is perhaps not involved in the recognition of hairpin induced pausing and termination. Although the pausing efficiency is preserved, the dwell time at the *his* pause is almost 50% greater for Δ-loop RNAP. This result suggests that, while fork loop 2 may not be involved in *his* pause detection, fork loop 2 participates in escape from this pause site. Therefore, the role of fork loop 2 is confined to the post-recognition stage of pause site sensing.

Previous studies of transcriptional pausing have yielded a model outlining the process of forming decayed transcriptional states from a precursory actively transcribing complex (Artsimovitch and Landick, 2000; Landick, 2006). In this model (Figure 3.16a), an actively transcribing complex encounters a pause signal, and upon recognition of this

signal, the active complex decays into a conformational state known as the "slow intermediate". It has been suggested that the conformation of the slow intermediate is parallel to the conformation adopted by complexes transcribing along the unactivated kinetic pathway (Foster *et al.*, 2001; Artsimovitch and Landick, 2000; Landick, 2006). Similar to the unactivated conformation, the slow intermediate is itself a precursor to further decayed, errant states, such as the prolonged paused, arrested, or terminated states, and as such, is considered to be an "elemental pause". As such, the formation of the slow intermediate controls pause recognition. Within the context of this model, fork loop 2 does not affect the formation of the slow intermediate, since fork loop 2 is not involved in pause site recognition. Rather, fork loop 2 affects states following the elemental pause, including the formation of the characteristic RNA hairpin.

The observation that pause recognition remains unaltered is surprising given that mutations in fork loop, as noted earlier, alter the fast phase of transcript synthesis (Kennedy and Erie, submitted). The model proposed by Kennedy and Erie predicts that the fork loop 2 mutation would preferentially partition complexes into the slow phase of synthesis, thereby favoring the recognition of pause sites. Such a prediction, however, is not what we observe. Instead, we find that the fork loop 2 mutation has the same probability of recognizing a pause and subsequently decaying into a paused state when compared to the wild-type fork loop 2. Under some nucleotide concentrations, however, it is possible that there may be an effect of partitioning complexes into different states, since the four residue deletion in fork loop 2 has not completely abolished the loop's functioning. On the other hand, the observation that pause lifetimes are enhanced by the fork loop 2 mutation is precisely what the model, proposed by Kennedy and Erie, would

predict. The fork loop 2 mutation would not efficiently facilitate NTP-driven translocation and would therefore cause RNAP to be slow to recover from a paused state.

The Walker B Motif Mutation Affects Recognition of Pausing and Termination Signals

Unlike the fork loop 2 mutation, for the Walker B mutation the rate of the fast phase of synthesis is unaffected, but there is a slightly reduced amount of complexes in the fast phase. The slow phase of synthesis, however, is increased by a factor of four (Kennedy, 2007). It is further suggested that the increased rate of the slow phase results in an increased rate of shuttling of the NTP from the allosteric site to the active site, because a lowered affinity to the allosteric site for NTPs would allow the allosteric site to release bound NTPs with greater ease. These results led to the suggestion that the mutation of the Walker B motif reduces the affinity of NTPs for the allosteric site. In light of the observation that Walker RNAP exhibits reduced pausing, both at the *his* (Class I) and the *ops* (Class II) elements, and termination efficiency, we propose that the shuttling mechanism is involved in the recognition of pausing and termination signals. The observation that Walker RNAP exhibits reduced recognition of pausing and termination signals indicates that the Walker mutation affects the formation of the slow intermediate, the precursor that controls the decay into paused and terminated states (Figure 3.17a). Our results suggest that the formation of the slow intermediate may be associated with the shuttle path. Perhaps the signal to pause or to terminate, in part, induces changes in the ability to shuttle NTPs from the allosteric site to the active site.

Class I and Class II Pauses may not Share a Common Intermediate

Comparing the pause efficiency at the *his* and the *ops* pause elements, it appears that Δ-loop RNAP exhibits different pausing efficiencies at these sites. Although wild-type for pause recognition at the *his* site, Δ-loop RNAP demonstrates reduced pausing efficiency at the *ops* site. These results suggest that, at least in the case of the fork loop 2 mutation, the *his* and the *ops* elements produce different initial recognition states. This suggestion is bolstered by the observation that both the *his* and the *ops* elements share a common characteristic of promoting longer dwell times, an event downstream of the initial pause recognition, for the fork loop 2 mutation. The diverging responses to these two pause sites indicate that Class I and Class II pauses may not share a common intermediate, as proposed previously (Artsimovitch and Landick, 2000). Such a proposal is an alternate one to the previously proposed model, whose central concept is the claim that Class I and Class II pauses are produced from the same precursory elemental pause state (Figure 3.16a) (Artsimovitch and Landick, 2000; Landick, 2006). The suggestion of a common intermediate arose from the observation that two RNAP mutants, altered in the rifampicin-binding region, demonstrate similar pause efficiencies in relation to wtRNAP at both pause classes (Artsimovtich and Landick, 2000). Therefore, we revise this model to include an additional component to suggest chiefly that responses to Class I and Class II pause elements yield two distinct slow intermediate states (Figure 3.16b). The Class I slow intermediate can decay further into *his* paused and ρ-independent terminated states, while the Class II slow intermediate can decay further into *ops* paused state.

While the components of pauses are fairly well understood, researches in the field of transcriptional pausing have not identified, with certainty, the initial signal that induces

pausing by RNAP. As noted above, experimental results have suggested that all pause states arise from a common elemental pause state (Artsimovitch and Landick, 2000). However, we have a RNAP mutant that suggests a separation of function. Unlike the rifampicin-binding mutants, our fork loop 2 mutation does exhibit different pause recognition capabilities at the different pause sites. The fork loop 2 mutation only recognizes Class I pauses well, while the rifampicin-binding mutants are capable of recognizing both Class I and Class II pauses.

Downstream Elements of RNAP and the Regulation of Transcription

As our results demonstrate, deficiencies in fork loop 2 and in Walker B motif are responsive to NusA effects. NusA enhances pause dwell time and termination efficiency of both Δ-loop and wtRNAPs by similar proportions. In the case of Walker RNAP, NusA enhances pause dwell time in proportion to that of wild-type, but interestingly, NusA rescues the termination efficiency to intrinsic levels of that of both wild-type and Δ-loop RNAP. Evidently, these mutations in the vicinity of the active site do not hinder the ability of NusA to anchor an RNA hairpin emerging from the exit channel to the β-flap of RNAP, as a previous model suggested for wtRNAP (Toulokhonov *et al*., 2001). The β-flap is quite distal from the active site, with an approximate 65 Å separation between these two sites. Despite the remote location of the β-flap relative to the active site, the authors provide evidence that the interaction between the flap domain and an emergent RNA hairpin is communicated allosterically via a two-stranded, antiparallel β sheet, dubbed the connector, to the active site. The opening of the RNA exit channel due

to the hairpin-flap interaction transmits a signal through the connector that potentially constrains the active site residues, thereby reducing the affinity of NTPs for the active site. Addition of NusA results in a tripartite interaction that serves to increase the stability of the hairpin-flap domain interaction and potentially further constrain active site residues. It is not too surprising, however, that the action of NusA is independent of Walker B and fork loop 2, given the distance between the site of NusA interaction and the allosteric site.

While these RNAP variants are subject to regulation by NusA, NusG seemingly has no *in vitro* regulatory effect on these mutants. NusG is capable of exerting regulatory effects on wtRNAP, as evidenced by the decrease in dwell time at the *ops* pause site. Interestingly, the lifetime of both the his and the ops pauses are unaffected by NusG for both Δ-loop and Walker RNAPs. Therefore, it appears that both fork loop 2 and Walker B motif are essential for effective NusG action on RNAP at the *ops* pause. Consequently, disruption of either fork loop 2 or the Walker B motif confers NusG resistance to RNAP. Although a direct interaction between NusG and RNAP has been documented (Li *et al*., 1992), the specific site of NusG binding on RNAP is currently unknown. Biochemical studies have demonstrated that NusA and NusG do not compete with each other for sites of action on RNAP (Burns *et al*., 1998). Therefore, one may argue that NusG does not bind to the flap domain, the specific site of NusA action. Alternatively, perhaps NusG binds to RNAP in such a way that a signal can be transmitted to the downstream edge of the transcription bubble, where fork loop 2 and Walker B motif are located. If either of these two sites is disrupted, then it is possible that NusG-mediated communication from a relatively distal point on RNAP to these sites would prove to be ineffective.

Interestingly, studies performed in *B. subtilis*, one of the few members of eubacteria in which NusG is non-essential, demonstrate that mutation of a conserved glutamine residue (Q513 using the E. coli numbering system) to a basic residue confers hypersensitivity to NusG with respect to autoregulation of *nusG* (Ingham and Furneaux, 2000). It is this same glutamine residue that, when mutated in *E. coli* RNAP to form the variant *rpoB8*, is responsible for increasing both ρ-dependent and ρ-independent termination and altering transcription elongation (Jin and Gross, 1991). More studies need to be performed to better resolve the site of NusG binding on RNAP.

Deficiencies in In Vitro Transcription Termination are Translated to Phenotypic Alterations within Living E. coli Cells

In addition to the marked deficiency *in vitro* ρ-independent termination, the Walker B mutation imparts a dramatic effect on live *E. coli* cells. Our results suggest that expression of Walker RNAP does produce a negative global *in vivo* effect with regard to cell growth. Based upon our results, it appears that a properly functioning the Walker B motif plays a vital role in maintaining healthy cells. How might a transcription termination deficiency affect cell growth? One line of evidence suggests that increased amounts of RNA may result in cell toxicity and may therefore be deleterious to cell growth (Taguchi et al., 1987; Brosius, 1984). Another possibility is that transcription and translation may be uncoupled and may subsequently affect the tightly regulated cell cycle check points. Alternatively, if RNAP is not properly signaled to terminate, RNAP may continue to remain bound to DNA. Since there is a discreet amount of RNAP expressed

in the cell, if RNAP ignores signals to terminate transcription and release itself from the

DNA in a timely fashion, then there may not be enough available RNAP to rebind

promoters and begin a new round of transcription. This unchecked, constant transcription

would also serve to disrupt important cell cycle signals. Therefore, it is clear that the

allosteric site plays a fundamental role in regulating transcription and, ultimately, in

properly coordinating transcription with other biological processes in the cell.
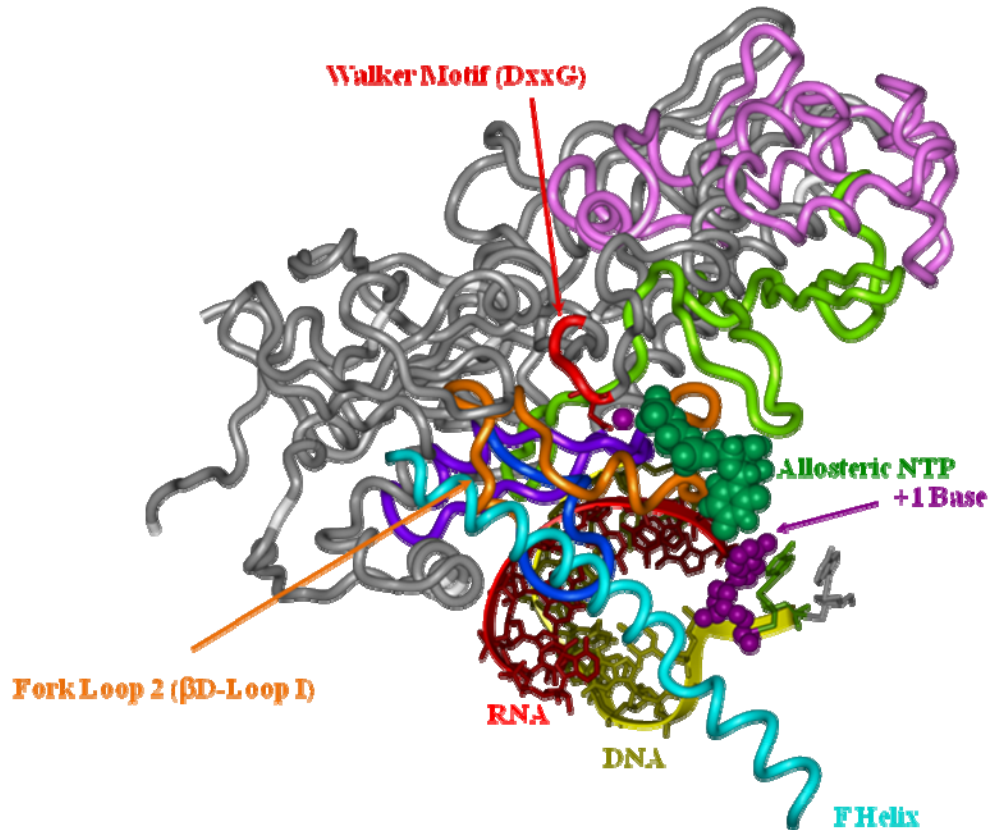
Figure 3.1a Model of the *Escherichia coli* (*E. coli*) transcription elongation complex
(adapted from Holmes and Erie, 2003). The essential components of the proposed
allosteric site are highlighted. These components include fork loop 2 (orange) and a
Walker B motif (red). An allosteric NTP bound to the allosteric site is rendered in space
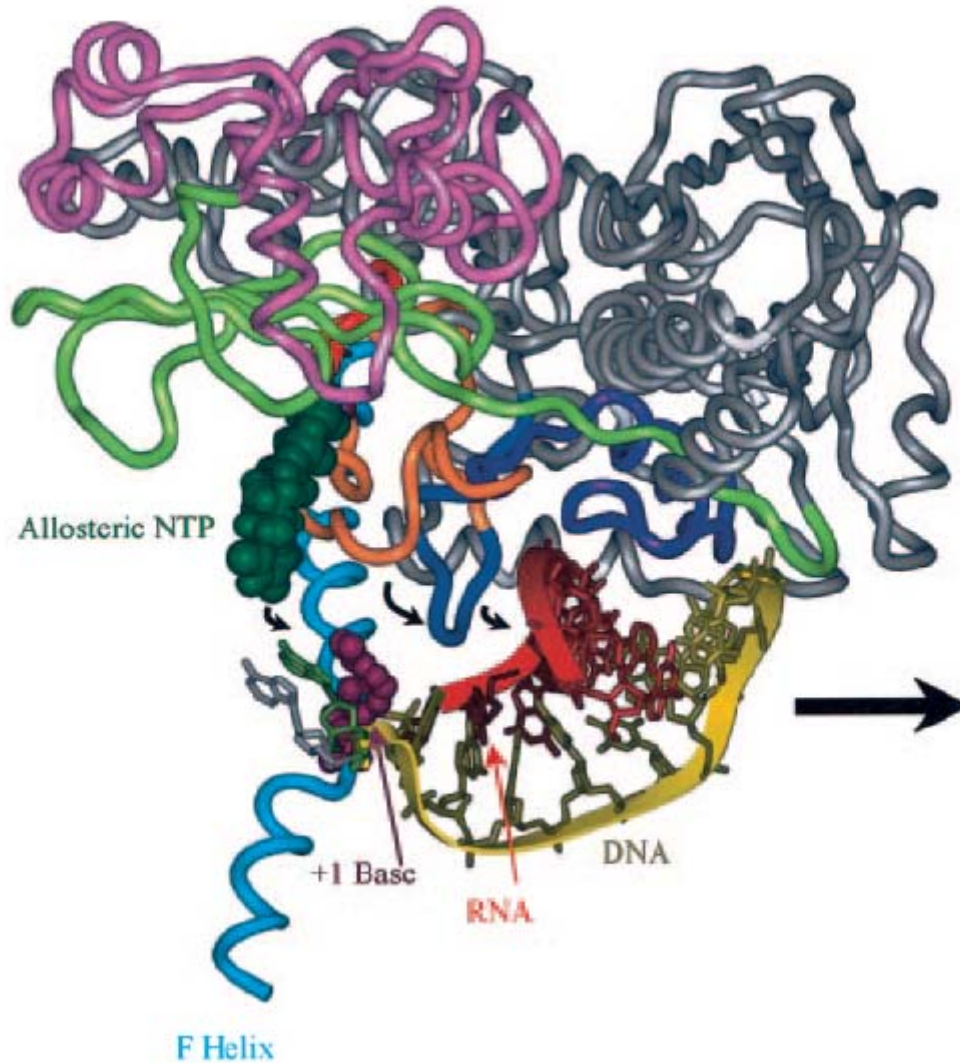filling (green).

Figure 3.1b Model of the *Escherichia coli* (*E. coli*) transcription elongation complex (adapted from Holmes and Erie, 2003). Another orientation of the essential components of the proposed allosteric site is highlighted. These components include fork loop 2 (orange) and a Walker B motif (red). An allosteric NTP bound to the allosteric site is rendered in space filling (green). The arrows indicate the direction of translocation.
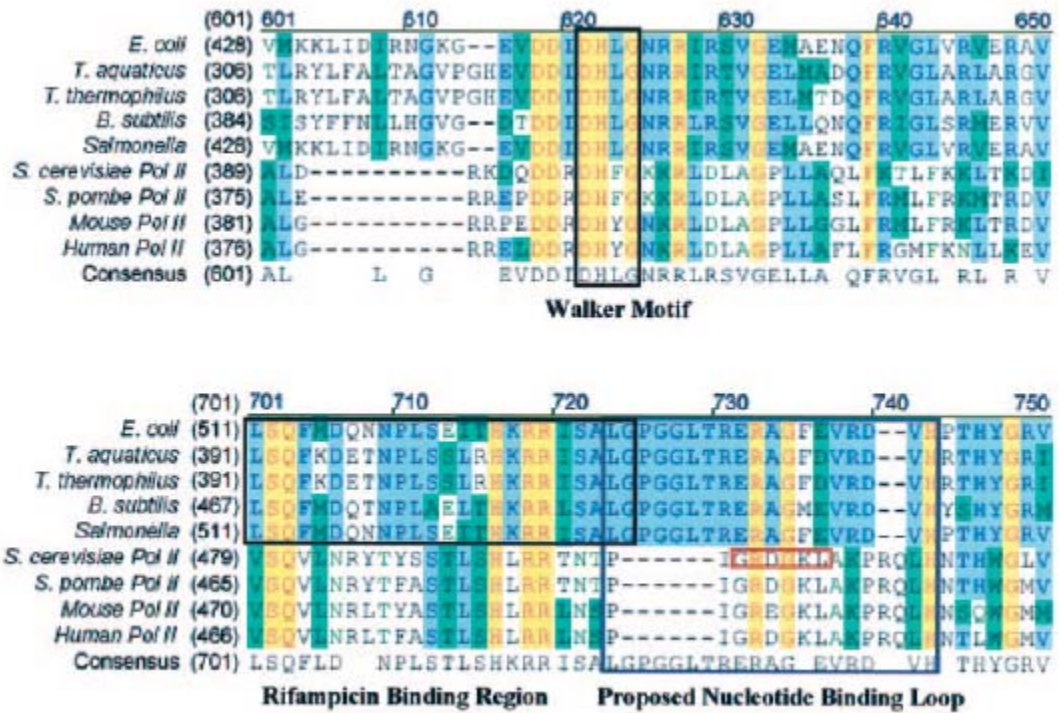
Figure 3.2 Sequence alignments of the β subunit of RNA polymerase (adapted from Holmes and Erie, 2003). The alignment presents the sequences of key regions in the β subunit of RNA polymerase from a variety of organisms. Both the Walker B motif (top panel) and fork loop 2 (bottom panel) are indicated. The putative allosteric site is composed of the Walker B motif and fork loop 2.
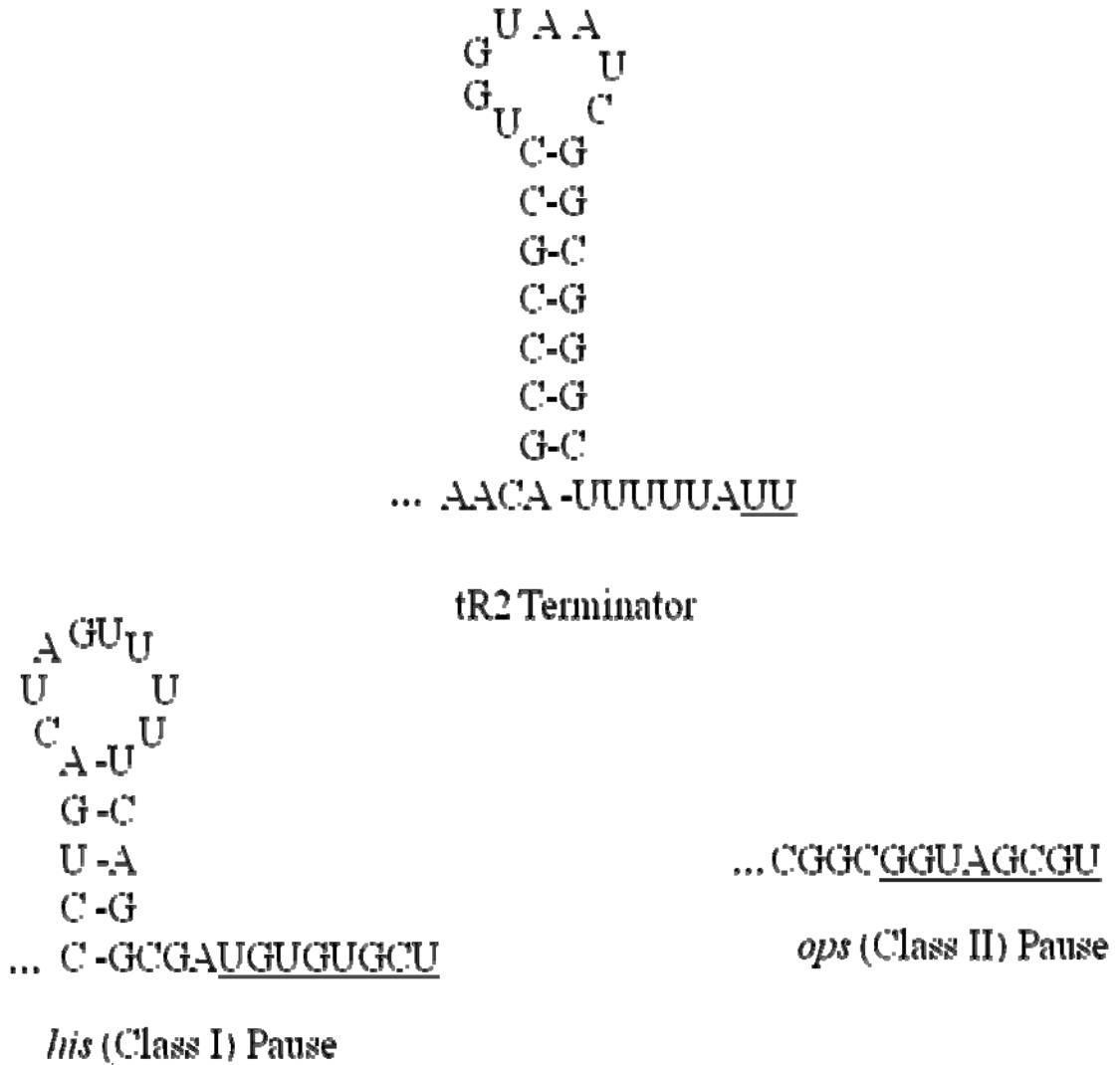
```
        U A A
     G           U
   G
    U           C
       C-G
       C-G
       G-C
       C-G
       C-G
       C-G
       G-C
  ... AACA -UUUUUAUU

       tR2 Terminator
```

```
   A GU U
  U         U
   C      U
    A-U
    G -C
    U -A
    C -G
... C -GCGAUGUGUGCU
```

his (Class I) Pause

```
... CGGCGGUAGCGU
```

ops (Class II) Pause

Figure 3.3 Primary and secondary structures of the tR2 terminator and the *his* and *ops* pause elements. The bases underlined are the site of termination/pausing.
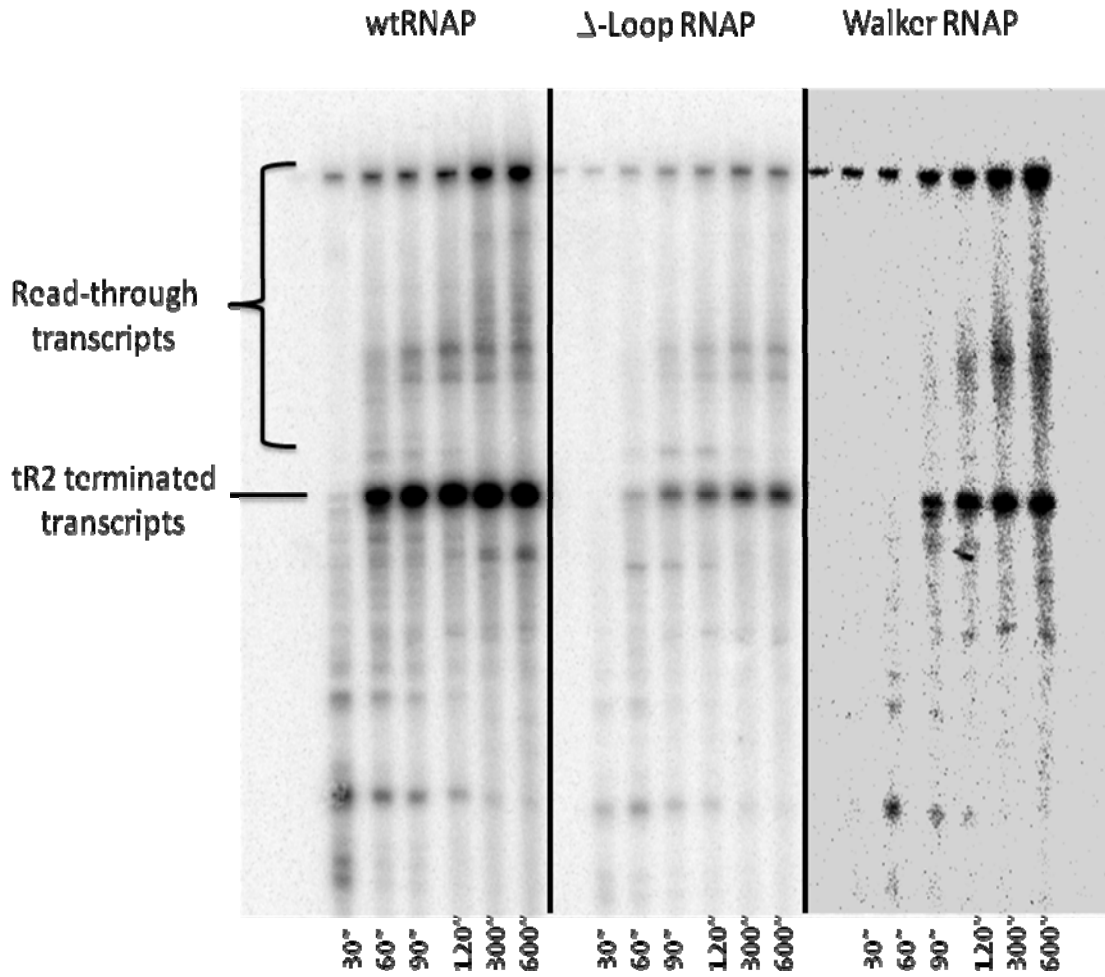
Figure 3.4 Intrinsic termination at the tR2 terminator. Stalled elongation complexes composed of wt- (right panel), Δ-loop- (middle panel), and Walker RNAP (left panel) were generated from the DE13 template under conditions of CTP deprivation. Elongation was resumed with the addition of 50 μM NTP mix, and reaction aliquots were removed and quenched at 30'', 1', 1'30, 2', 5', and 10'. Read-through transcripts and transcripts terminated at the tR2 terminator are indicated.
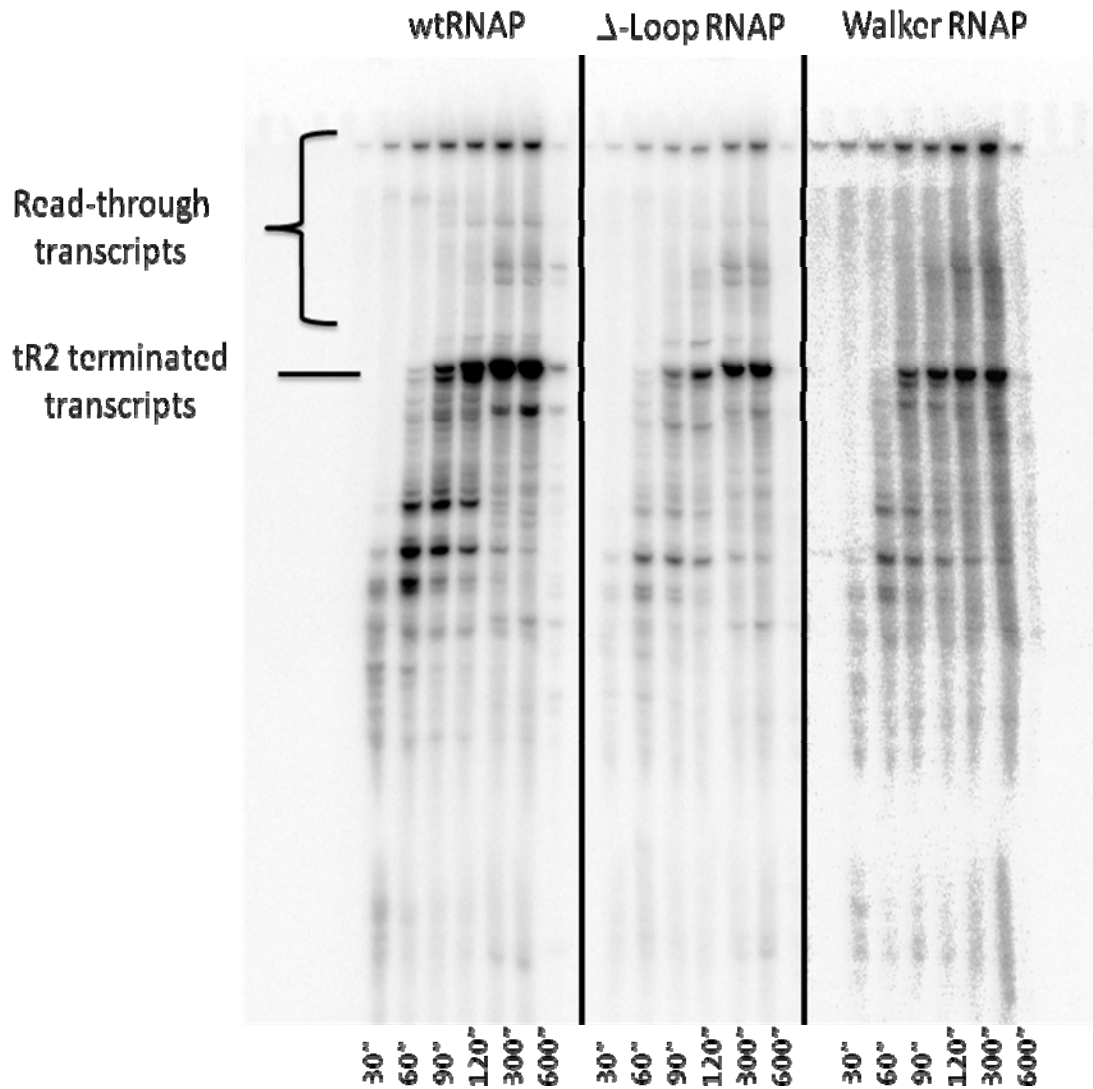
Figure 3.5 NusA-mediated termination at the tR2 terminator. Stalled elongation complexes composed of wt- (right panel), Δ-loop- (middle panel), and Walker RNAP (left panel) were generated from the DE13 template under conditions of CTP deprivation. Elongation was resumed with the addition of 50 μM NTP mix and 100 nM NusA, and reaction aliquots were removed and quenched at 30'', 1', 1'30, 2', 5', and 10'. Read-through transcripts and transcripts terminated at the tR2 terminator are indicated.
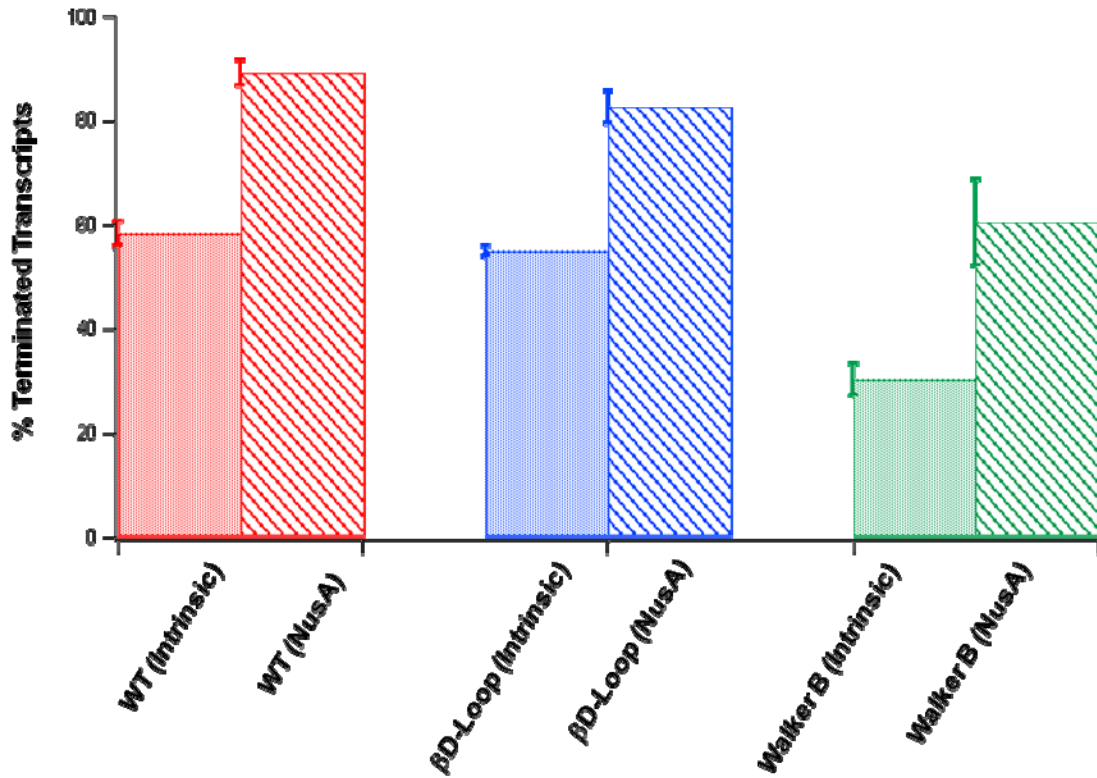
Figure 3.6 Percent terminated transcripts at the tR2 terminator. The percentage of transcripts terminated by wt- (red), Δ-loop- (blue), and Walker RNAP (green) in the absence (solid) and presence (hatch-mark) of NusA are shown. The termination efficiency of each RNAP was calculated by measuring the intensity of the bands corresponding to terminated transcripts and dividing this value by the sum of measured intensities of terminated and read-through bands in each lane of the gel (Figures 3.4 and 3.5). The error in each measurement is the calculated standard deviation of at least three independent experimental trials.
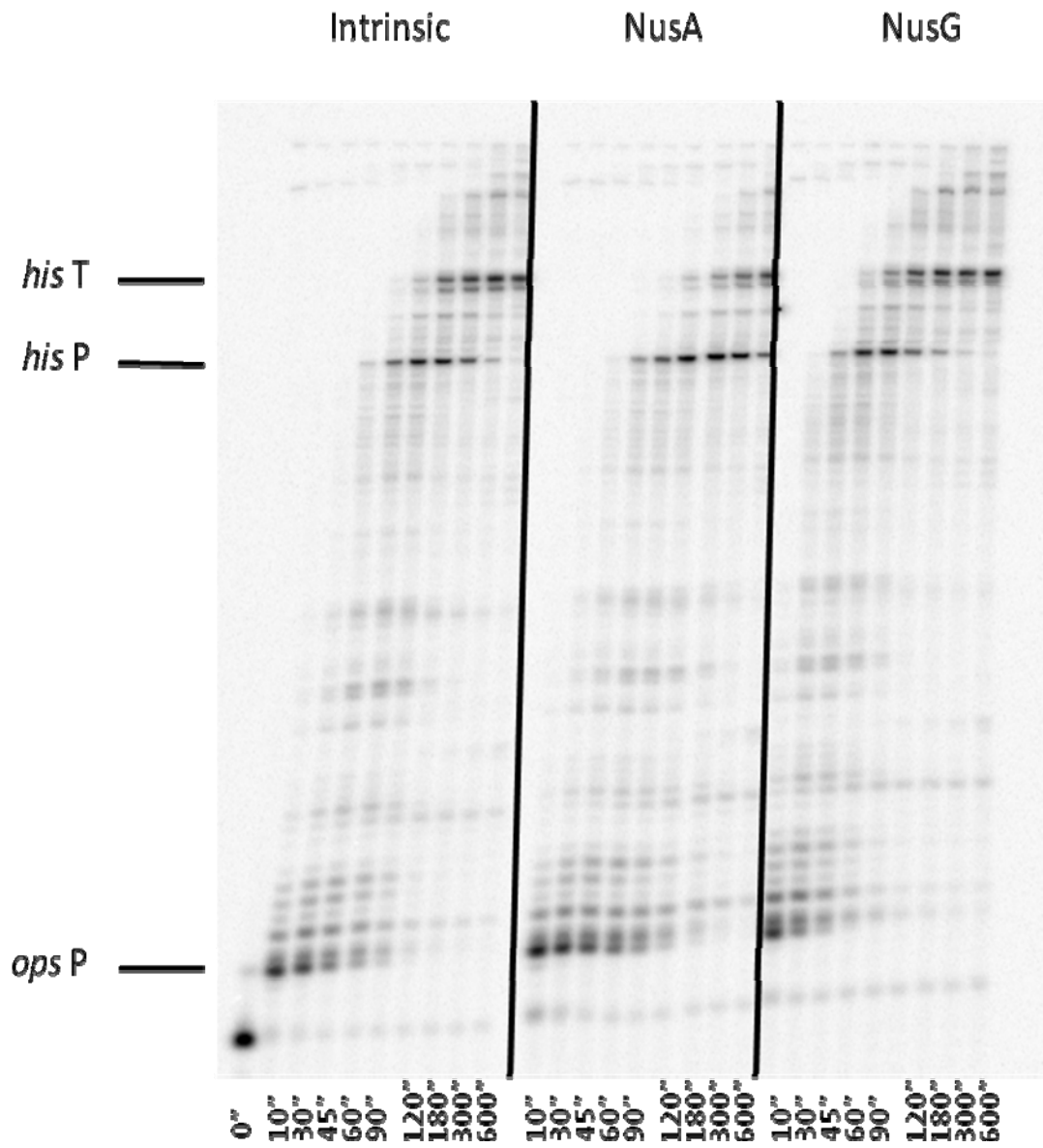
Figure 3.7 Transcriptional pausing of wtRNAP at the *ops* and the *his* pause elements.

Figure 3.7 Transcriptional pausing of wtRNAP at the *ops* and the *his* pause elements. Stalled elongation complexes composed of wtRNAP were generated from the IA349 template under conditions of UTP deprivation. Elongation was resumed with the addition of 100 μM NTP mix in either the absence of Nus-factors (left panel) or the presence of 100 nM Nus factors (NusA-middle panel, NusG-right panel). Reaction aliquots were removed and quenched at 10'', 20'', 30'', 45'', 1', 1'30, 2', 3', 5', and 10'. Elongation complexes paused at the *ops* and the *his* pause sites are indicated. Transcripts terminated at the *his* terminator are also shown.
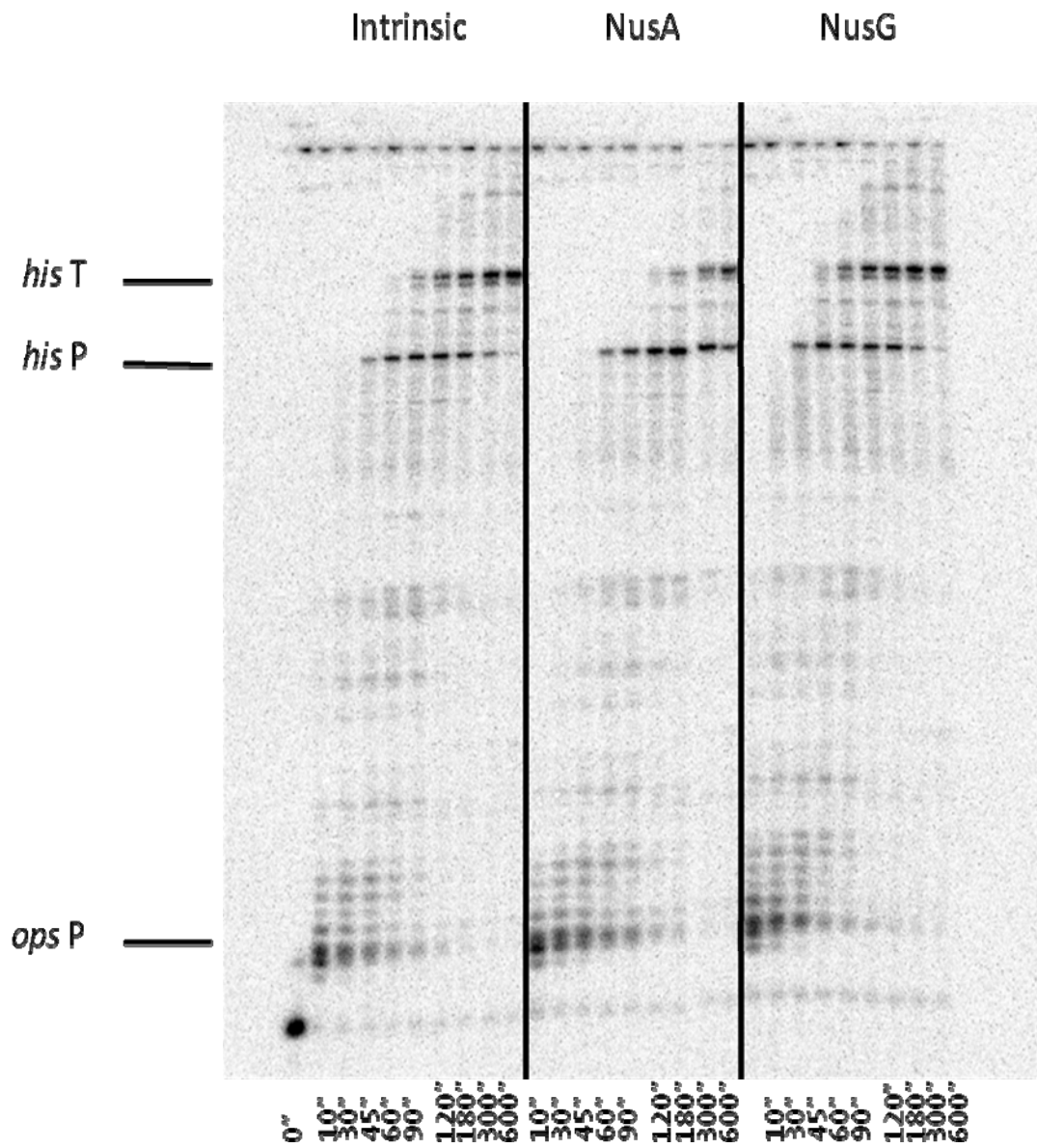
Figure 3.8 Transcriptional pausing of Δ-loop RNAP at the *ops* and the *his* pause elements.

Figure 3.8 Transcriptional pausing of Δ-loop RNAP at the *ops* and the *his* pause elements. Stalled elongation complexes composed of Δ-loop RNAP were generated from the IA349 template under conditions of UTP deprivation. Elongation was resumed with the addition of 100 μM NTP mix in either the absence of Nus-factors (left panel) or the presence of 100 nM Nus factors (NusA-middle panel, NusG-right panel). Reaction aliquots were removed and quenched at 10'', 20'', 30'', 45'', 1', 1'30, 2', 3', 5', and 10'. Elongation complexes paused at the *ops* and the *his* pause sites are indicated. Transcripts terminated at the *his* terminator are also shown.
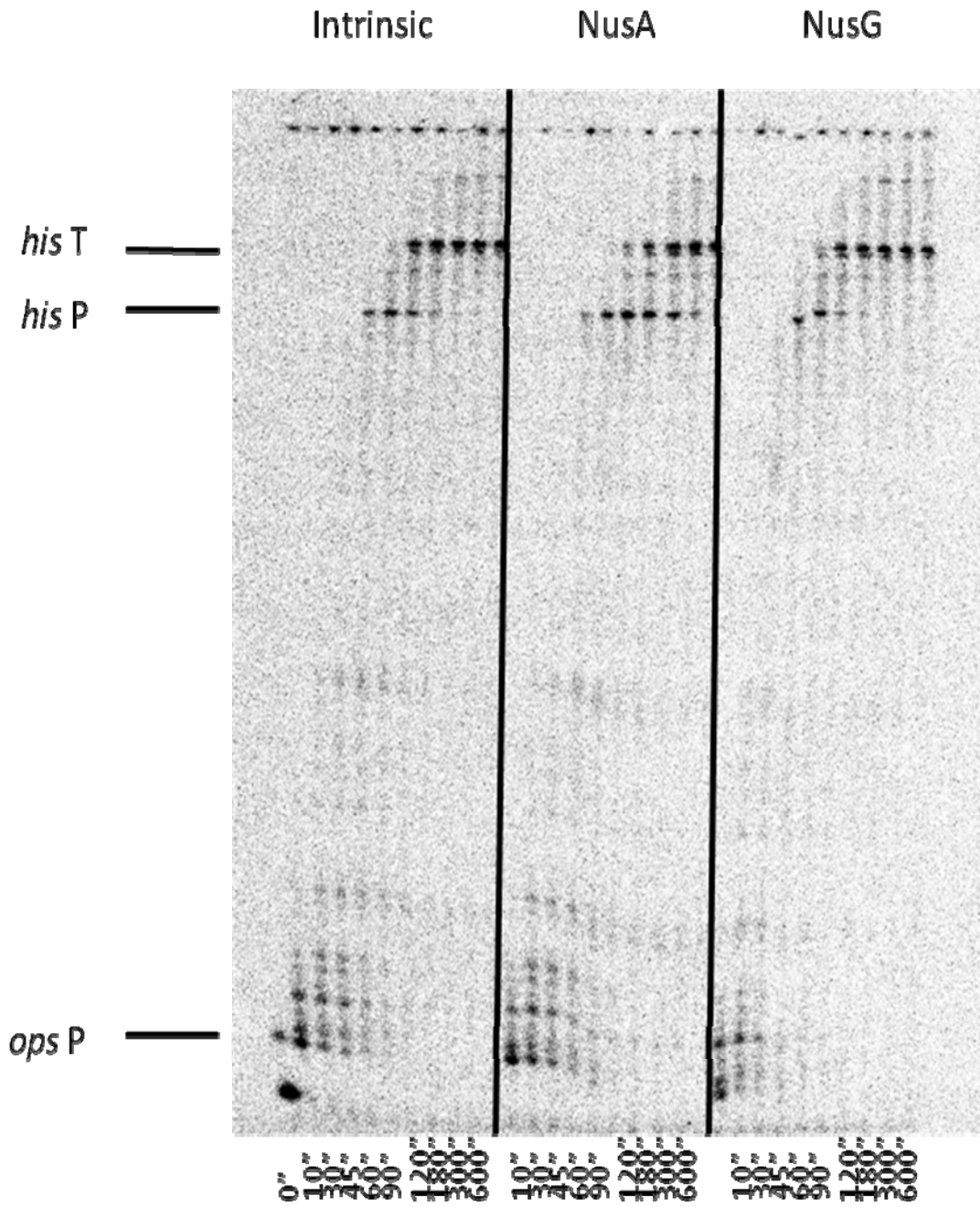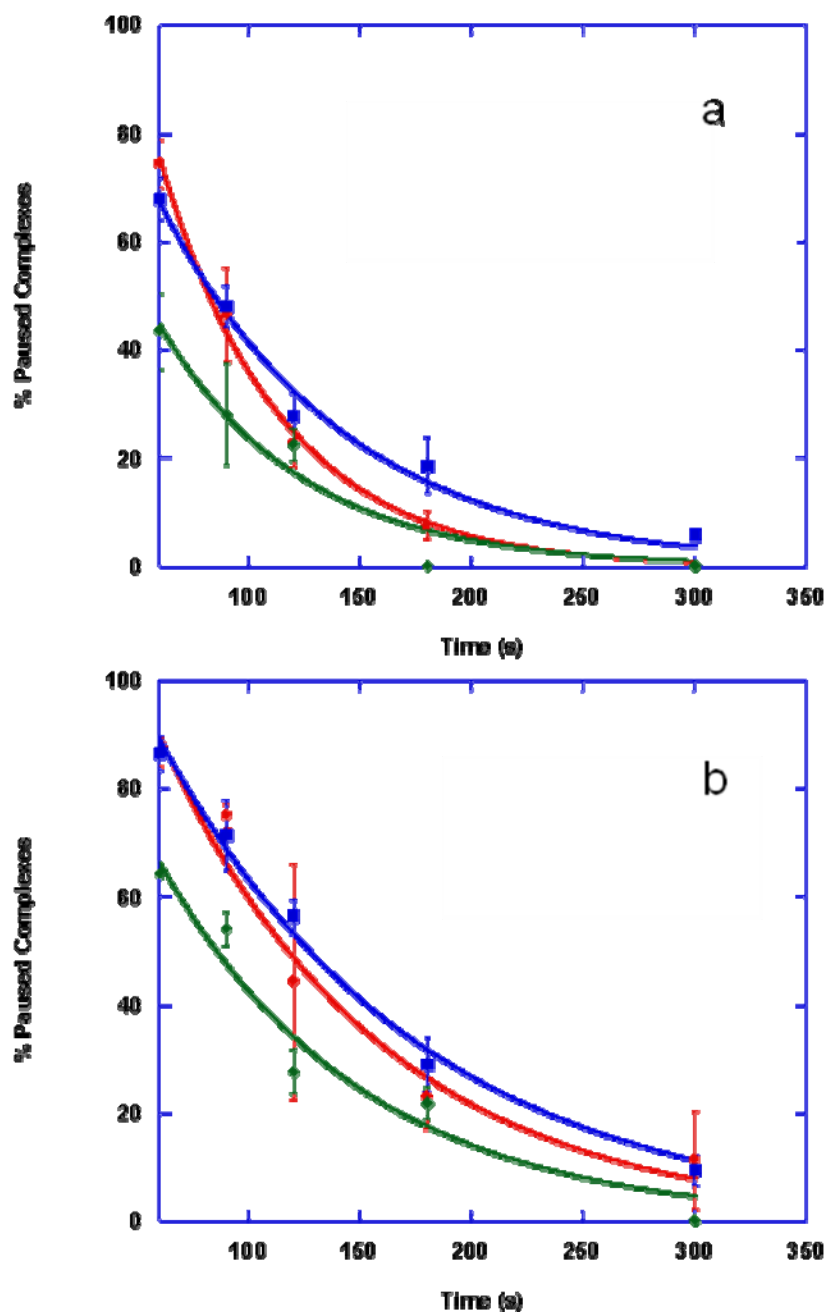
Figure 3.9 Transcriptional pausing of Walker RNAP at the *ops* and the *his* pause elements.

Figure 3.9 Transcriptional pausing of Walker RNAP at the *ops* and the *his* pause elements. Stalled elongation complexes composed of Walker RNAP were generated from the IA349 template under conditions of UTP deprivation. Elongation was resumed with the addition of 100 μM NTP mix in either the absence of Nus-factors (left panel) or the presence of 100 nM Nus factors (NusA-middle panel, NusG-right panel). Reaction aliquots were removed and quenched at 10'', 20'', 30'', 45'', 1', 1'30, 2', 3', 5', and 10'. Elongation complexes paused at the *ops* and the *his* pause sites are indicated. Transcripts terminated at the *his* terminator are also shown.
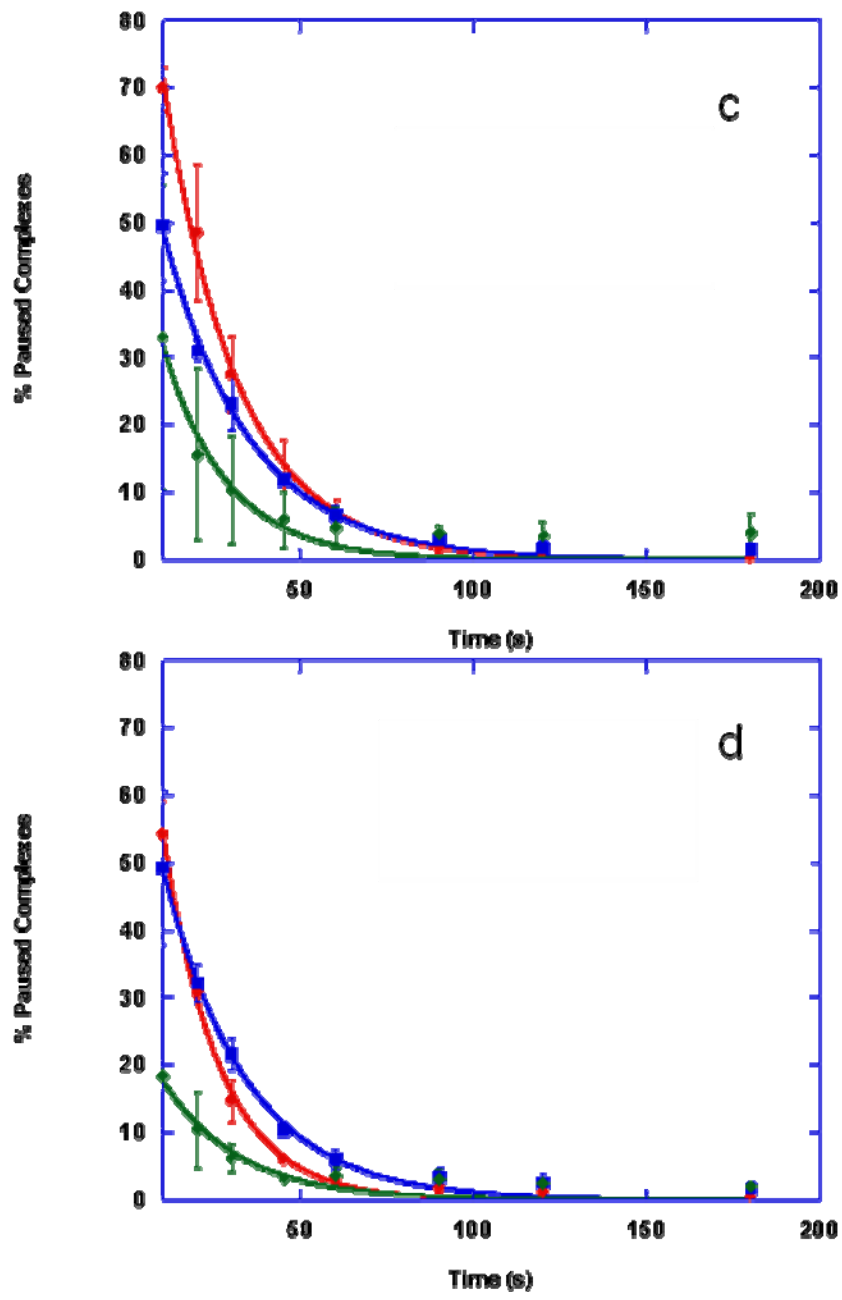
Figures 3.10a-b Percentage of complexes paused at the *his* (Class I) element as a function of time.

Figures 3.10a-b Percentage of complexes paused at the *his* (Class I) element as a function of time. The percentage of wt- (red), Δ-loop- (blue), and Walker RNAPs (green) paused in the absence (a) and presence (b) of NusA are presented. The pausing efficiency of each RNAP was calculated by measuring the intensity of the bands corresponding to complexes paused at the *his* element and dividing this value by the sum of measured intensities of paused and read-through complexes in each lane of the gel (Figures 3.7-3.9). The pause escapes were fit to simple exponential decays, and the pause lifetimes were calculated from the curves. The error in each measurement is the calculated standard deviation of at least three independent experimental trials.

Figures 3.11a-b Percentage of complexes paused at the *ops* (Class II) element as a function of time.

Figures 3.11a-b Percentage of complexes paused at the *ops* (Class II) element as a function of time. The percentage of wt- (red), Δ-loop- (blue), and Walker RNAPs (green) paused in the absence (a) and presence (b) of NusG are presented. The pausing efficiency of each RNAP was calculated by measuring the intensity of the bands corresponding to complexes paused at the *ops* element and dividing this value by the sum of measured intensities of paused and read-through complexes in each lane of the gel (Figures 3.7-3.9). The pause escapes were fit to simple exponential decays, and the pause lifetimes were calculated from the curves. The error in each measurement is the calculated standard deviation of at least three independent experimental trials.
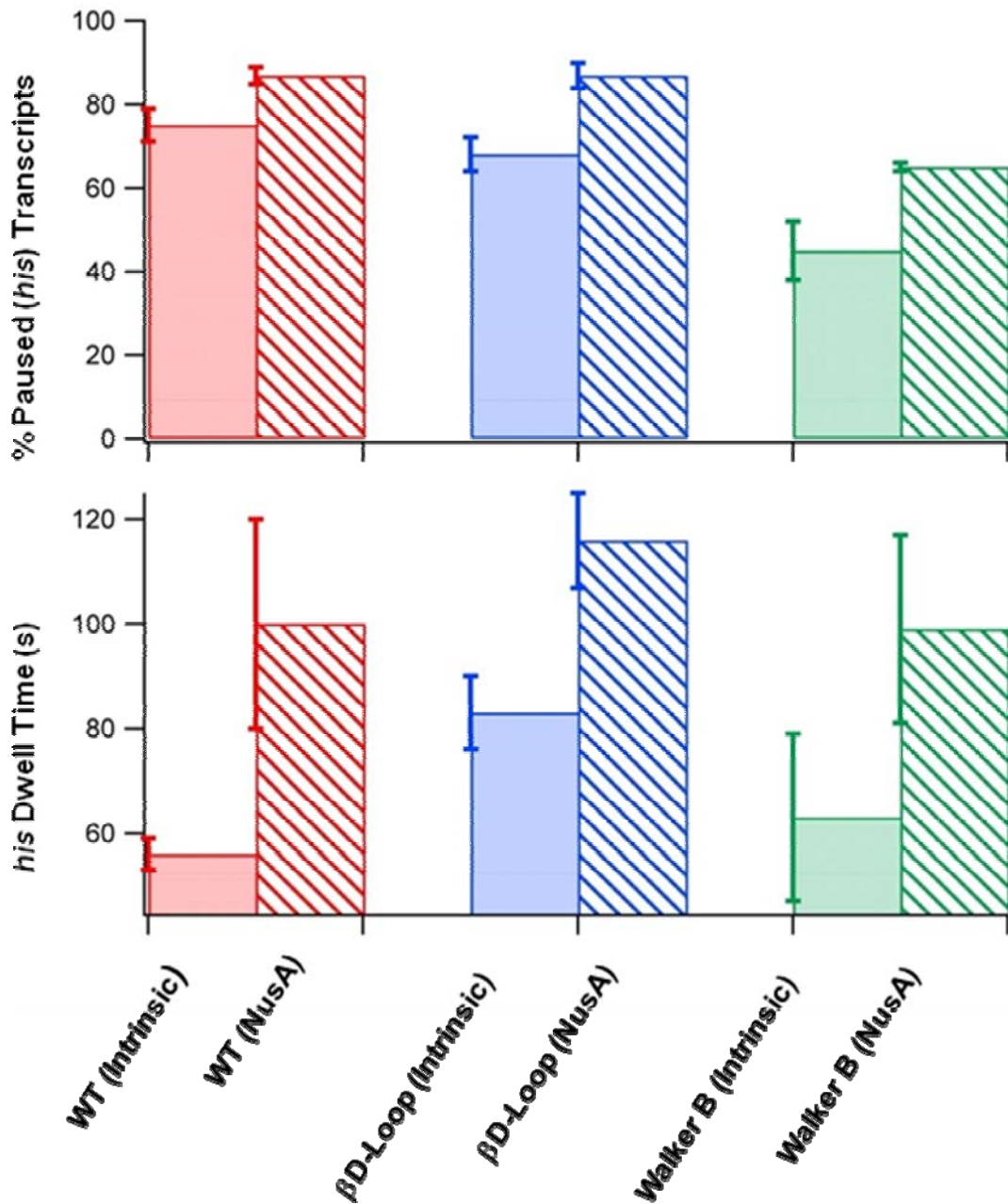
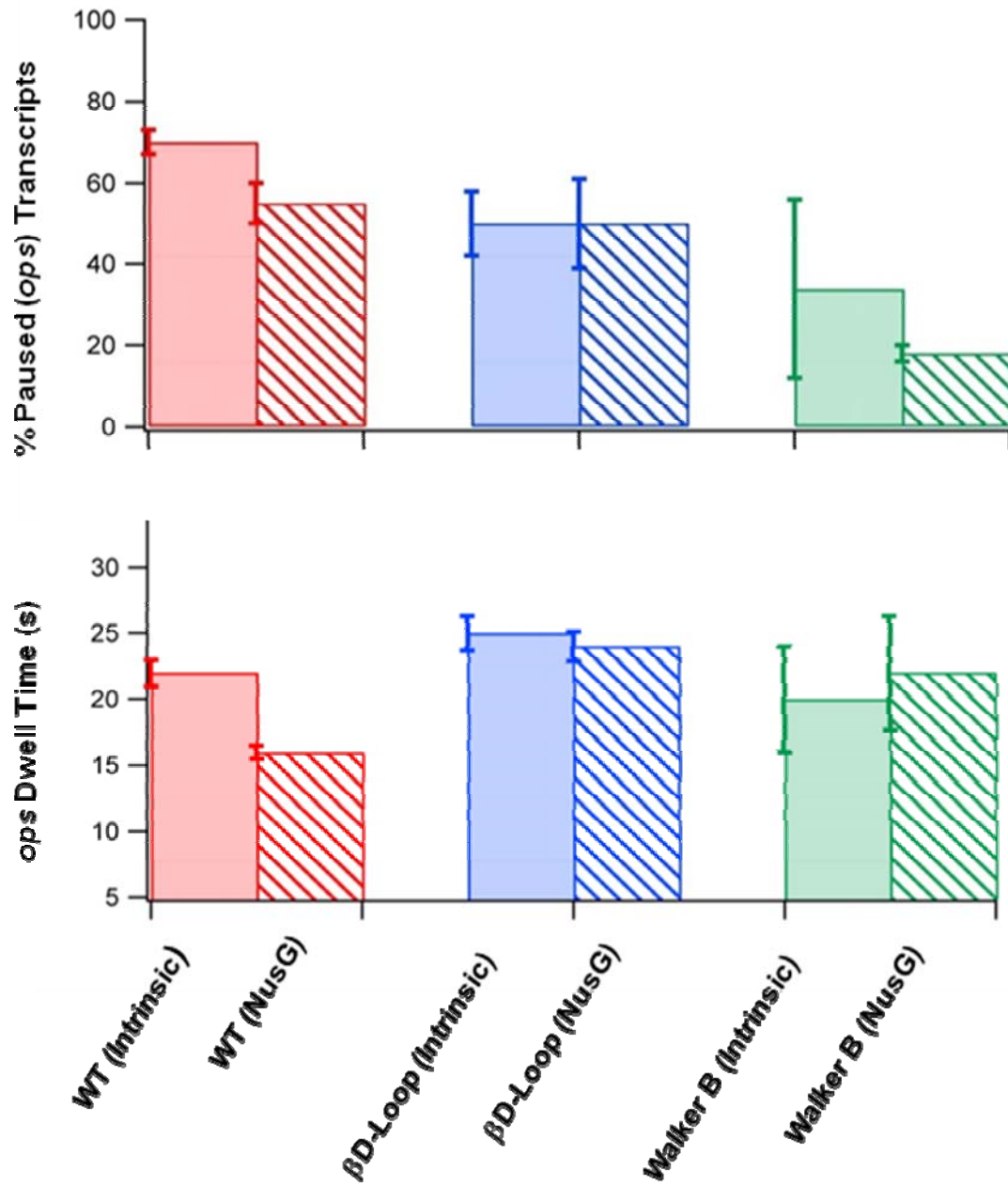**Figure 3.12 Pause efficiencies and dwell times at a Class I pause element.**

Figure 3.13 Pause efficiencies and dwell times at a Class II pause element.
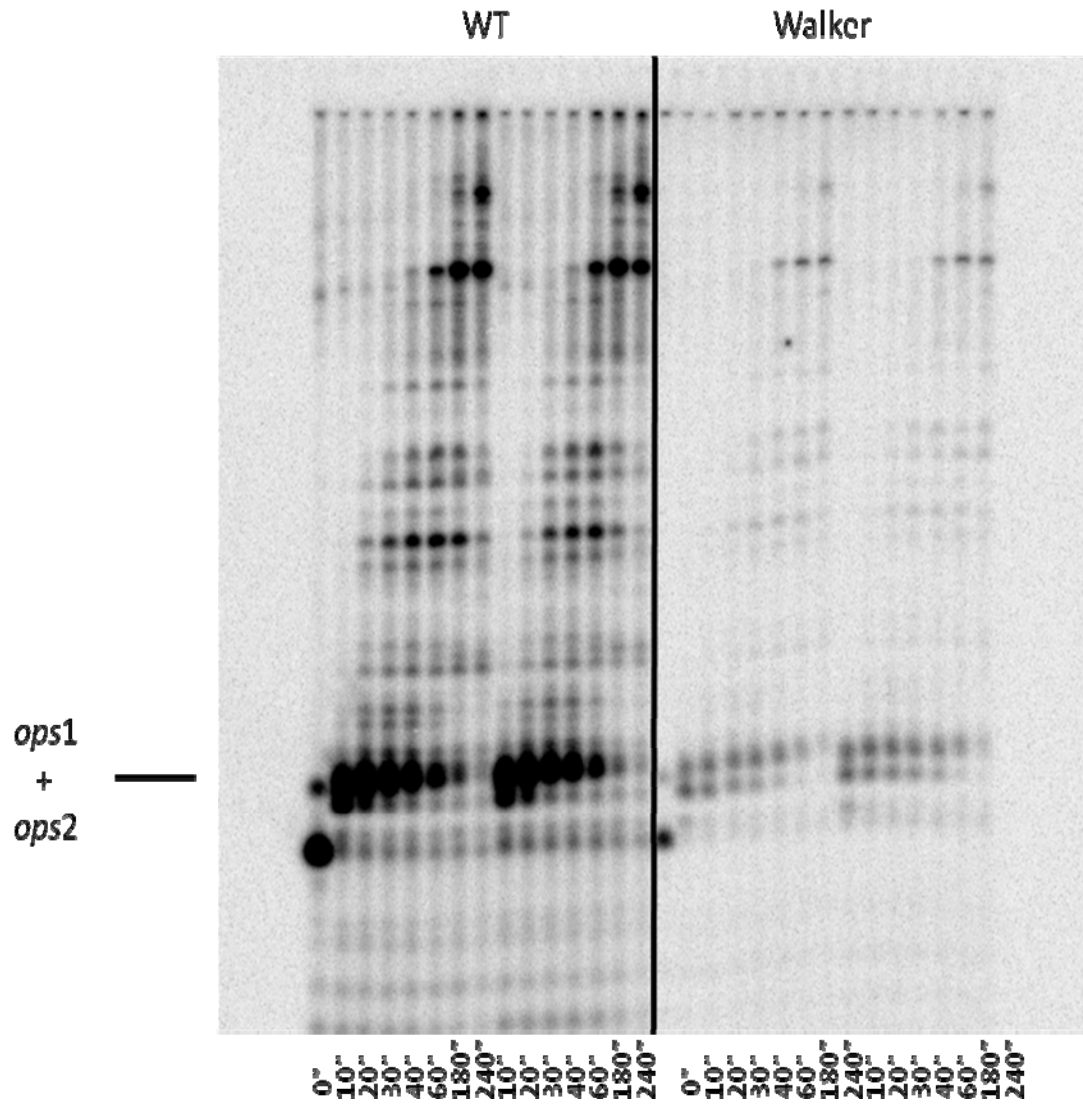
Figure 3.14 Transcriptional pausing of Δ-loop- and wtRNAPs at the *ops* pause element.

Figure 3.14 Transcriptional pausing of Δ-loop- and wtRNAPs at the *ops* pause

elements. Pausing by wt (left panel) and Δ-loop (right panel) RNAPs at both *ops*1 and

*ops*2 elements is resolved under limiting GTP concentration and is indicated to the left of

the gel. The experiment was conducted in the absence (left side of each panel) and in the

presence (right side of each panel) of 100 nM NusG. The results of this experiment are

only preliminary and are, therefore, not yet quantified.
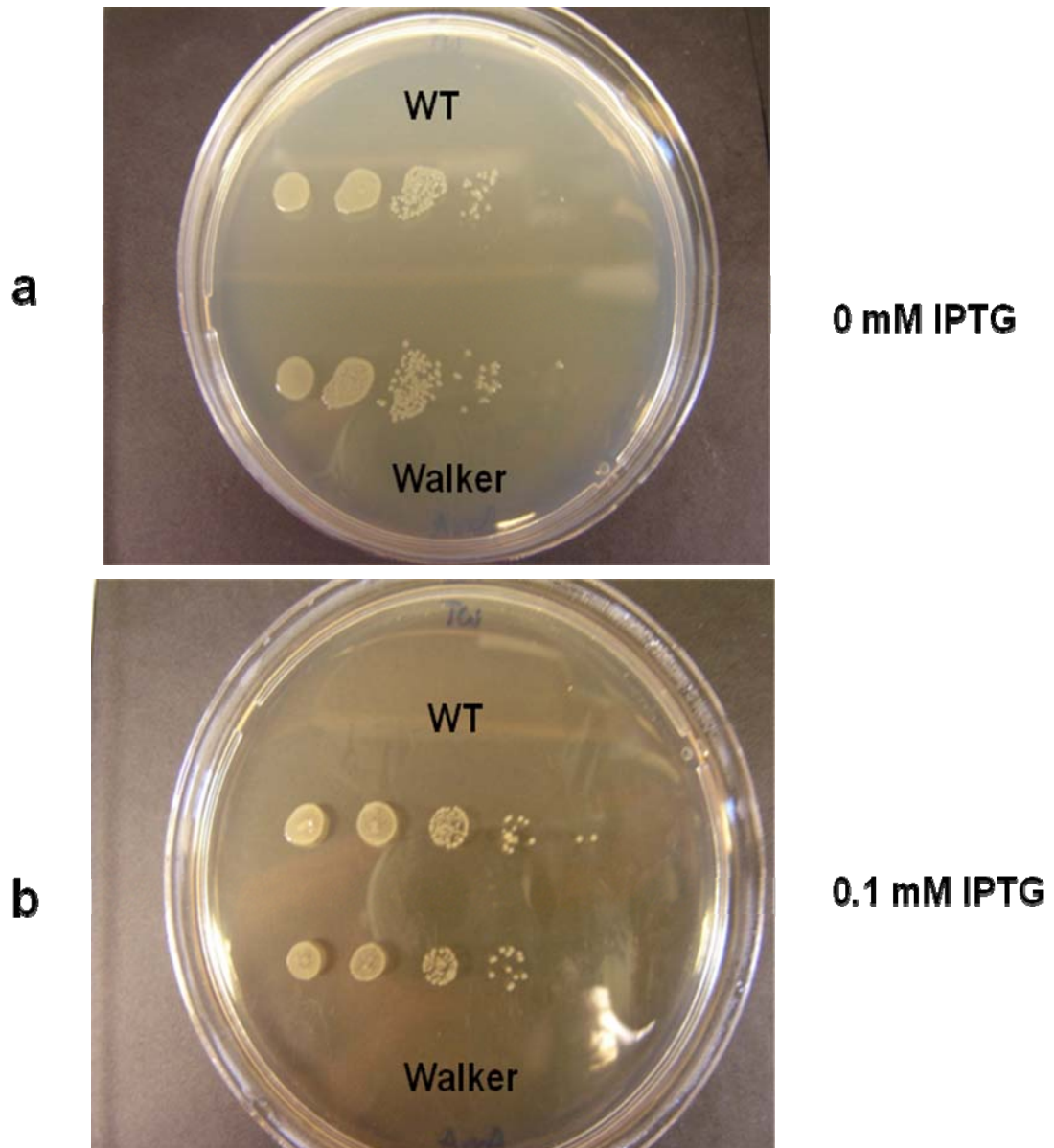
Figures 3.15a-b Growth phenotypes of *E. coli* expressing wt- and Walker RNAPs.
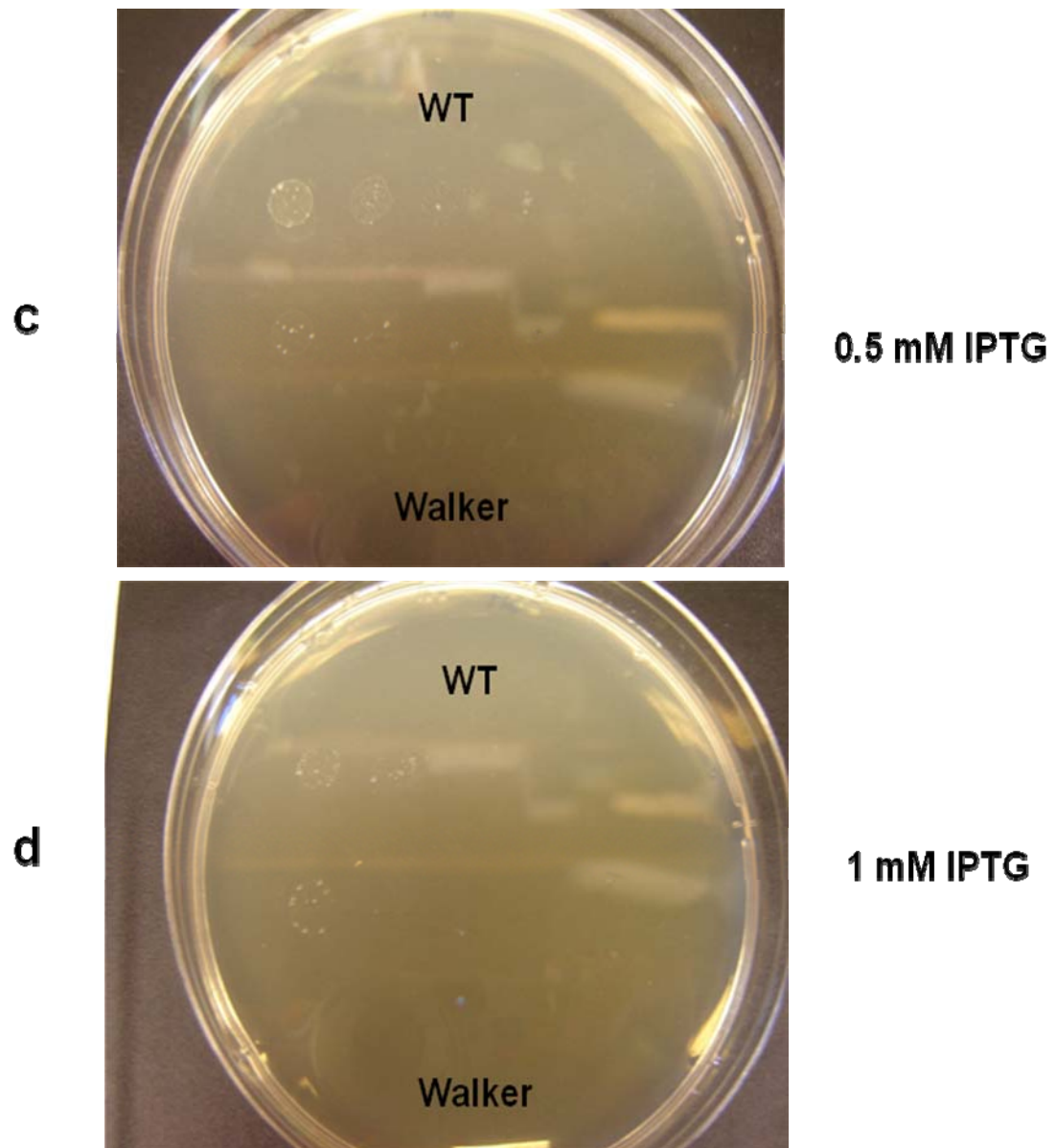
Figures 3.15c-d Growth phenotypes of *E. coli* expressing wt- and Walker RNAPs.

Figures 3.15a-d Growth phenotypes of *E. coli* expressing wt- and Walker RNAPs.

Cells were grown to early log phase and were serially diluted five times to yield 10^7-10^3

cells/mL. Five microliters of each dilution were plated in descending order (left-to-right)

against 0 mM (a), 0.1 mM (b), 0.5 mM (c), and 1 mM (d) IPTG. Cells expressing

wtRNAP were plated along the top row, while cells expressing Walker RNAP were

plated along the bottom row.
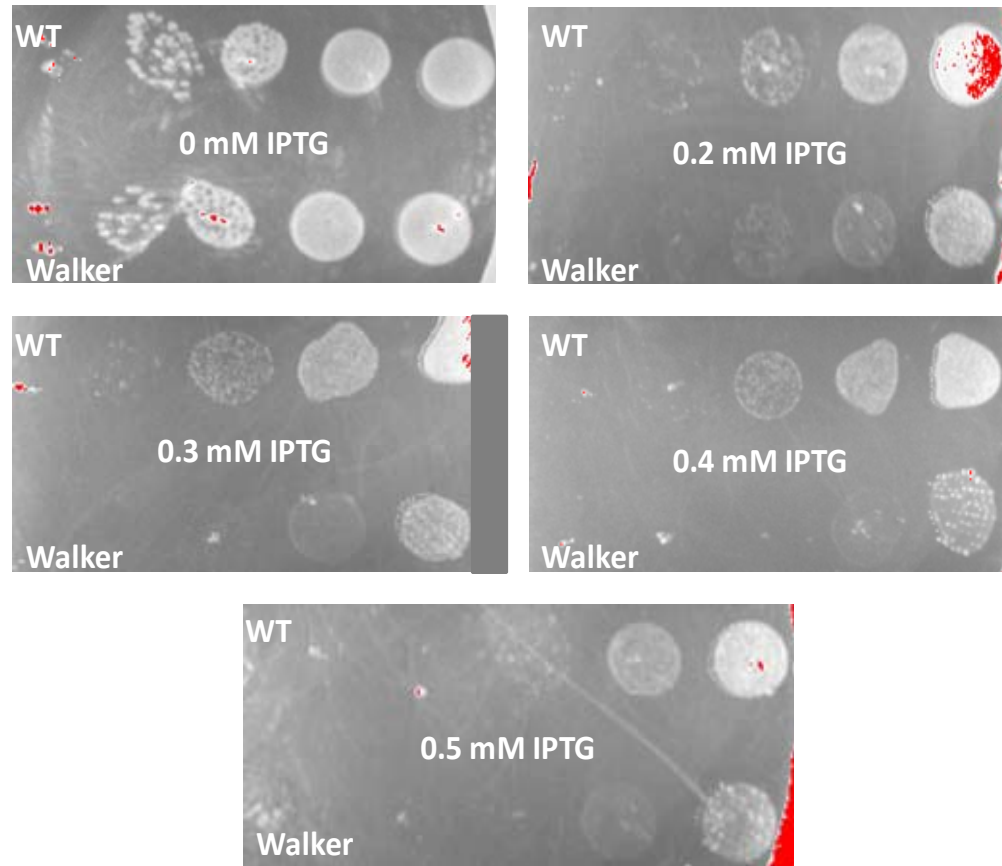
Figures 3.16 Growth phenotypes of *E. coli* expressing wt- and Walker RNAPs. Cells were grown to early log phase and were serially diluted five times to yield 10^7-10^3 cells/mL. Ten microliters of each dilution were plated in ascending order (left-to-right) against 0 mM, 0.2 mM, 0.3 mM, 0.4 mM, and 0.5 mM IPTG. Cells expressing wtRNAP were plated along the bottom row, while cells expressing Walker RNAP were plated along the top row.
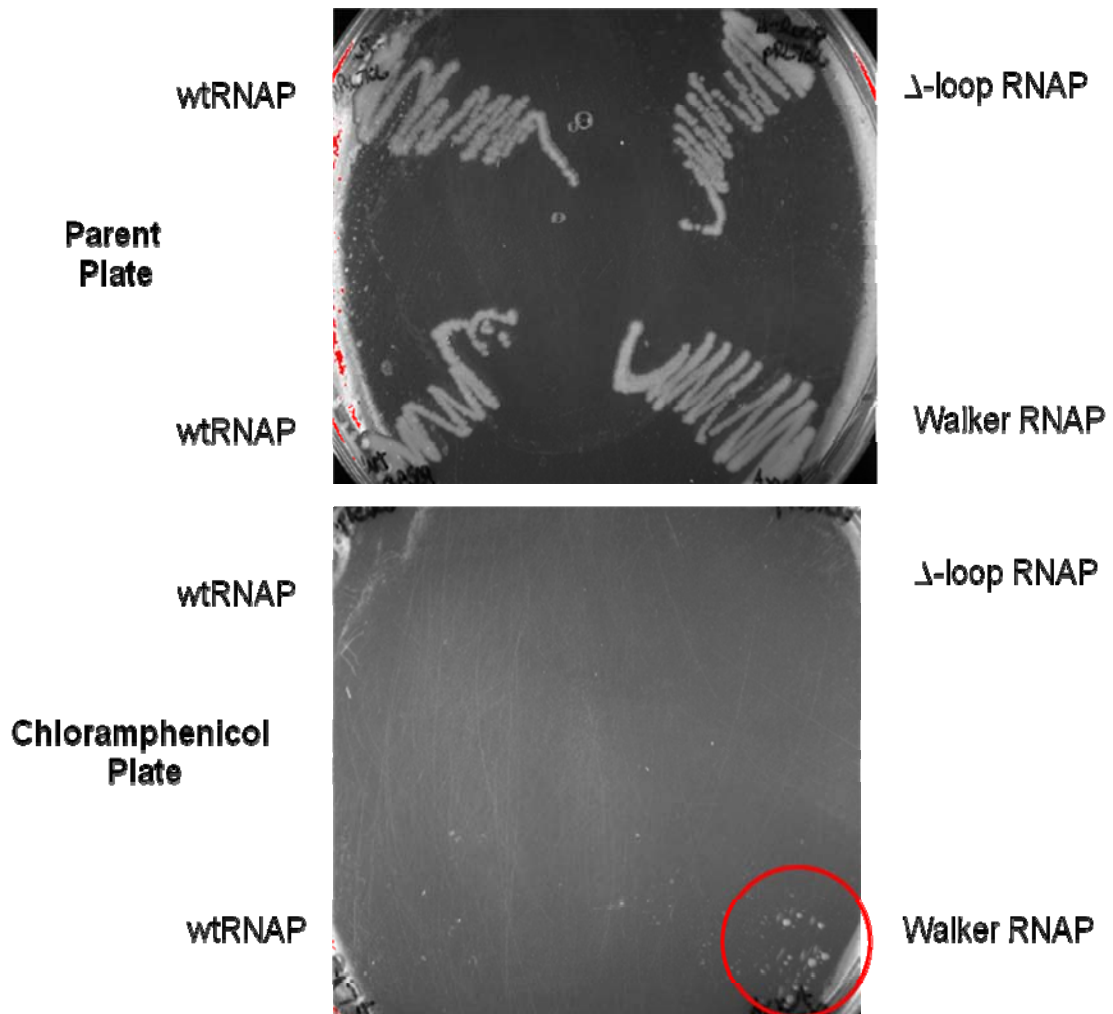
Figure 3.17 *In vivo* transcription termination phenotypes. Cells were plated on a parent plate supplemented with 0.5 mM IPTG. A velvet replica was made of the parent plate to transfer colonies to a plate supplemented with 10 mg/mL chloramphenicol. The colonies circled in red on the chloramphenicol plate are the most densely populated on this plate. The colonies are expressing Walker RNAP. Since these colonies are able to survive in the presence of chloramphenicol, it appears that the expressed Walker RNAP is termination deficient *in vivo*.

Figure 3.18a Model of the decay into paused states. This model is based upon the one originally proposed by Artsimovitch and Landick which stipulates that Class I and Class II pause states proceed through a common intermediate known as the elemental pause.

Figure 3.18b Modified model of the decay into paused states. This model is a modification of the one originally proposed by Artsimovitch and Landick which stipulates that Class I and Class II pause st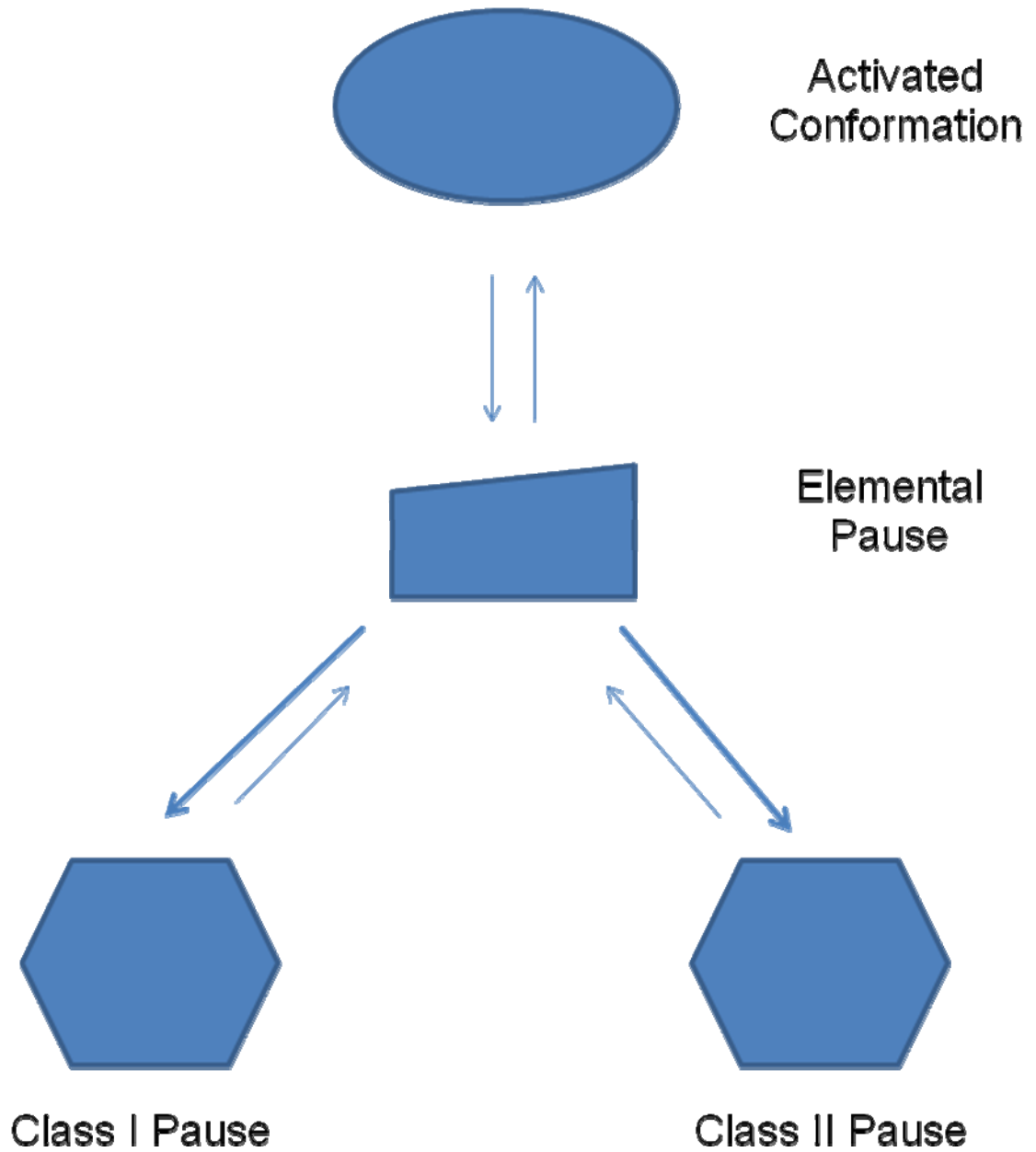ates proceed through a common interm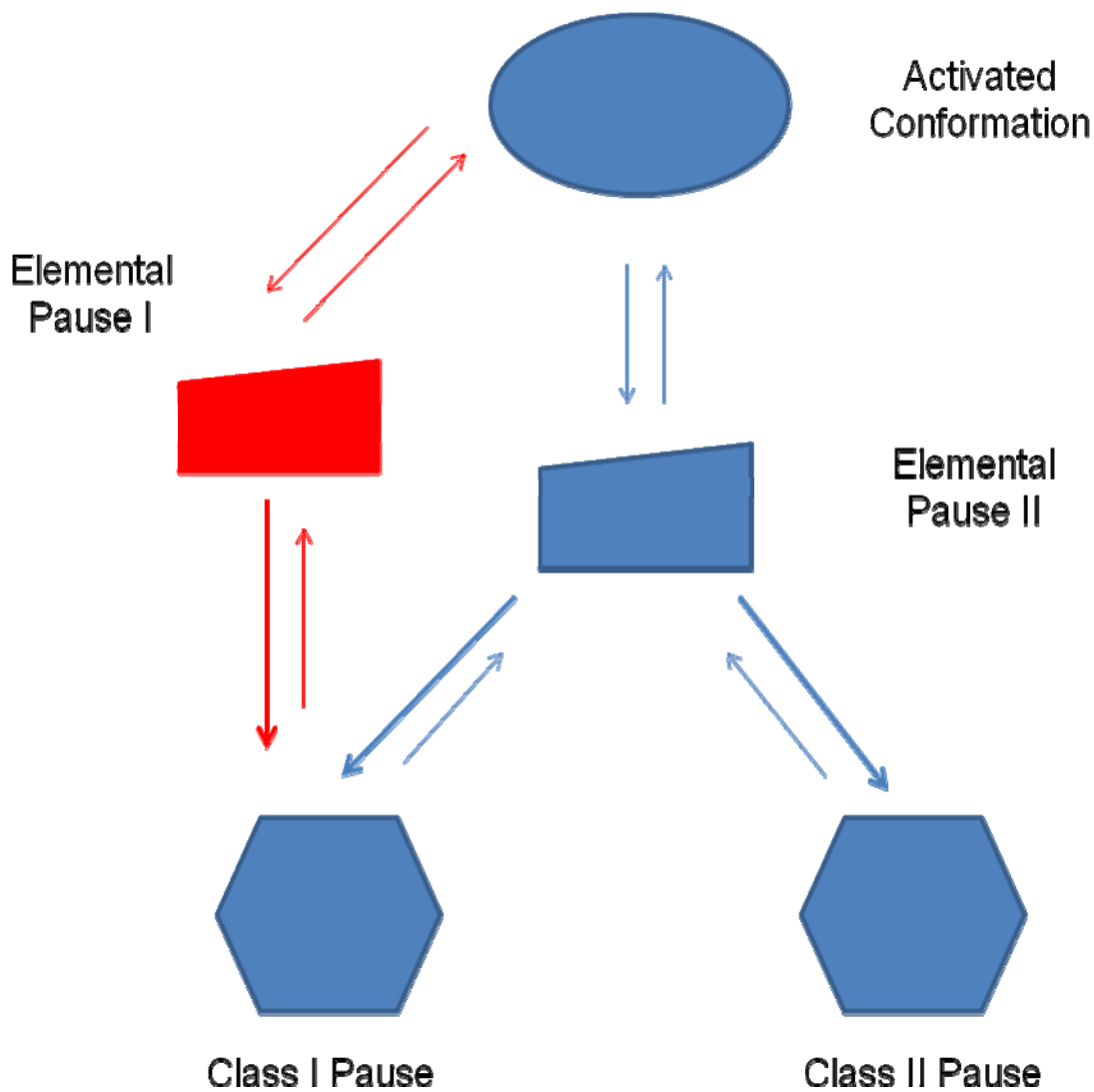ediate known as the elemental pause. This modified model indicates that Class I and Class II pause states proceed through different slow intermediates states.

BIBLIOGRAPHY

Artsimovitch, I., and Landick, R. (2000). Pausing by Bacterial RNA Polymerase is Mediated by Mechanistically Distinct Classes of Signals. Proc. Nat. Acad. Sci. *97*, 7090-7095.

Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A Ratchet Mechanism of Transcription Elongation and Its Control. Cell *120*, 183-193.

Batada, N.N., Westover, K.D., Bushnell, D.A., Levitt, M., and Kornberg, R.D. (2004). Diffusion of Nucleoside Triphosphates and Role of the Entry Site to the RNA Polymerase II Active Center. Proc. Nat. Acad. Sci. *101*, 17361-17364.

Burova, E. et al., (1995). Escherichia coli NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. J. Bact. *177*(5), 1388-1392.

Burton, Z.F., Feig, M., Gong, X.Q., Zhang, C., Nedialkov, Y.A., and Xiong, Y. (2005). NTP-Driven Translocation and Regulation of Downstream Template Opening by Multi-Subunit RNA Polymerases. Biochem. Cell Bio. *83*, 486-496.

Burgess, R.R. and Jendrisak, J.J. (1975). A Procedure for the Rapid, Large-Scall Purification of Escherichia coli DNA-Dependent RNA Polymerase involving Polymin P Precipitation and DNA-Cellulose Chromatography.

Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science *292*, 1863-1876.

Erie, D.A. (2002). The Many Conformational States Of RNA Polymerase Elongation Complexes And Their Roles In The Regulation Of Transcription. Biochim. Biophys. Acta *1577*, 224-239.

Erie, D.A. et al., (1993). Multiple RNA Polymerase Conformations and GreA: Control of the Fidelity of Transcription. Science *262*, 867-873.

Foster, J.E. et al., (2001). Allosteric Binding of Nucleoside Triphosphates to RNA Polymerase Regulates Transcription Elongation. Cell *106*, 243-252.

Herbert, K.M., Porta, A.L., 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.

Holmes, S.F., and Erie, D.A. (2003). Downstream DNA Sequence Effects on Transcription Elongation: Allosteric Binding of Nucleoside Triphosphates Facilitates Translocation Via a Ratchet Motion. J. Bio. Chem. *278*, 35597-35608.

Gnatt, A.L., et al., (2001). Structural Basis of Transcription: An RNA Polymerase II Elongation Complex at 3.3 Å Resolution. Science *292*, 1876-1882.

Gong, X.Q. et al., (2005). Dynamic Error Correction and Regulation of Downstream Bubble Opening by Human RNA Polymerase II. Molecular Cell *18*, 461-470.

Kettenberger, H., Armache, K-J., and Cramer, P. (2004). Complete RNA Polymerase II Elongation Complex Structure and Its Interactions with NTP and TFIIS. Mol. Cell *16*, 955-965.

Kireeva, M.L., Nedialkov, Y.A., Cremona, G.H., Purtov, Y.A., Kubkowska, L., Malagon, F., Burton, Z.F., Strathern, J.N., and Kashlev, M. (2008). Transient Reversal of RNA

Polymerase II Active Site Closing Controls Fidelity of Transcription Elongation. Mol. Cell *30*, 557-566.

Komissarova, N., and Kashlev, M. (1997). RNA Polymerase Switches Between Inactivated and Activated States By Translocating Back and Forth along the DNA and the RNA. J. Bio. Chem. *272*, 15329-15338.

Landick, R. (2006). The Regulatory Roles and Mechanism of Transcriptional Pausing. Biochemical Society Transactions *34*(6), 1062-1066.

Leipe, D.D., Wolf, Y.I., Koonin, E.V., and Aravind, L. (2002). Classification and Evolution of P-loop GTPases and Related ATPases. J. Mol. Bio. *317*, 41-72.

Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R.M., and Ebright, R.H. (2004). Antibacterial Peptide Microcin J25 Inhibits Transcription by Binding within and Obstructing the RNA Polymerase Secondary Channel. Mol. Cell *14*, 739-751.

Murakami, K. et al., (2002). Structural Basis of Transcription Initiation: RNA Polymerase Holoenzyme at 4 A Resolution. Science *296*, 1280-1284.

Palangat, M., and Landick, R. (2001) Roles of RNA:DNA Hybrid Stability, RNA Structure, and Active Site Conformation in Pausing by Human RNA Polymerase II. J. Mol. Biol. *311*, 265-282.

Pan, T. and Sosnick, T. (2006). RNA folding during transcription. Annu. Rev. Biophys. Biomol. Struct. 35, 161-175.

Pasman, Z and von Hippel, P.H. (2000). Regulation of Rho-Dependent Transcription Termination by NusG is Specific to the Escherichia coli Elongation Complex. Biochemistry *39*, 5573-5585.

Schmidt, M.C. and Chamberlin, M.J. (1987). nusA Protein of Escherichia coli is an Efficient Transcription Termination Factor for Certain Terminator Sites. J. Mol. Biol. *195*, 809-818.

Sullivan, S.L. and Gottesman, M.E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell *68*(5), 989-994.

Toulokhonov, I. et al., (2001). Allosteric Control of RNA Polymerase by a Site that Contacts Nascent RNA Hairpins. Science *292*, 730-733.

Toulokhonov, I. et al., (2007). A Central Role of the RNA Polymerase Trigger Loop in Active-Site Rearrangement During Transcriptional Pausing. Mol. Cell *27*, 406-419.

Uptain, S.M., and Chamberlin, M.J. (1997). *Escherichia coli* RNA Polymerase Terminates Transcription Efficiently at Rho-independent Terminators on Single-Stranded DNA Templates. Proc. Nat. Acad. Sci. *94*, 13548-13553.

Vassylyev, D.G. et al., (2007a). Structural Basis for Substrate Loading in Bacterial RNA Polymerase. Nature *448*, 163-168.

Vassylyev, D.G. et al., (2007b). Structural Basis for Transcription Elongation by Bacterial RNA Polymerase. Nature *448*, 157-164.

Wang, D. et al., (2006). Structural Basis of Transcription: Role of the Trigger Loop in Substrate Specificity and Catalysis. Cell *127*, 941-954.

Westover, K.D. et al., (2004). Structural Basis of Transcription: Nucleotide Selection by Rotation in the RNA Polymerase II Active Center. Cell *119*, 481-489.

Zhang, G. et al., (1999). Crystal Structure of Thermus aquaticus Core RNA Polymerase at 3.3 A Resolution. Cell *98*(6), 811-24.

CHAPTER FOUR

MOLECULAR BEACONS: DETECTION OF NUSG INTERACTION WITH A STALLED ELONGATION COMPLEX

INTRODUCTION

During the process of transcription elongation, the ternary (RNAP-DNA-RNA) elongation complex (TEC) is upheld to the stringent standards of synthesizing a nascent transcript at reasonable rates and with high fidelity. As such, the TEC serves as a point of major regulation and is targeted by a myriad of factors, including transcriptional accessory proteins (Erie, 2002). These transcriptional accessory proteins include members of the Nus family, which are involved in transcriptional pausing and termination. Of this family, NusG plays indispensably key roles in the general regulation of transcription, as this protein is essential to cell viability (Sullivan and Gottesman, 1992). NusG functions to enhance anti-pausing and anti-termination activity in RNAP (Sullivan and Gottesman, 1992; Li *et al*., 1993). Coupled with these effects is the observation that NusG serves as a general elongation factor by accelerating the rate of transcription by 25%-30% *in vivo* and *in vitro* (Burova *et al*., 1995; Burns *et al*., 1998).

By accelerating the rate of transcription, NusG is thought to shift the conformational equilibrium of the elongation complex from a backtracked state to an elongation-competent state (Pasman and von Hippel, 2000). While not entirely conserved, NusG homologs have been identified in all prokaryotes, and orthologs have been identified in some eukaryotes, including yeast and humans (Hartzog *et al*., 1998).

In addition to functional information, recent studies have advanced our understanding of the NusG architecture. An x-ray crystal structure of NusG from *Aquifex aeolicus* has been solved (Chapter 1, Figure 1.6) (Steiner *et al*., 2002; Knowlton *et al*., 2003). Based upon this structure, NusG is composed of three discrete domains that are loosely tethered together. The first and third domains, D1 and D3, are conserved regions found in all NusG proteins. In contrast, the second domain, D2, is defined as a variable-insertion region, of which only a small portion is found in *E. coli* NusG. Located near the flexible N-terminus, D1 is characterized by an RNP-like domain. It is speculated that NusG interacts weakly with RNAP via this RNP-like domain. At the C-terminus, D3 contains a KOW sequence motif, characteristically found in the ribosomal protein L24 (Krypides et al., 1996). Although this KOW motif is implicated in nucleic acid binding, direct binding to RNA and to DNA has been observed only in NusG from *Aquifex aeolicus* and *Thermotoga maritima* (Liao et al., 1996; Steiner et al., 2002).

Despite recent advancements illuminating the structural components of NusG, we still do not know how NusG interacts with RNAPs involved in transcription elongation complexes. In light of evidence that NusG is absolutely essential to cell viability (Li *et al*., 1992), we assume that the interaction between NusG and an elongation complex is highly important. Therefore, we sought to explore methods to study this presumably vital

119

interaction. An earlier study used bulk fluorescence techniques to study binding of transcription factors to core RNAP (Stanley *et al.*, 1991). In this study, the authors were able to estimate the binding constants of NusA and of σ^{70} to core RNAP using fluorescence anisotropy techniques. While bulk studies permit the estimation of binding constants, recent technological advances in fluorescence spectroscopy have allowed us to explore beyond ensemble averages and to detect the conformational distributions and time trajectories of individual molecules within a heterogeneous population (Weiss, 1999). Specifically, the single molecule technique of total internal reflection fluorescence (TIRF) microscopy has permitted the detection of single molecule binding events and conformational dynamics (Weninger *et al.*, 2003; Bowen *et al.*, 2005). In our present study, we have chosen to develop a system by which we may utilize through-prism TIRF techniques to observe the interaction of individual NusG molecules with stalled elongation complexes. Using a single molecule approach, we can not only estimate the binding constant of NusG to an elongation complex, but also explore the dynamics of this interaction and obtain a glimpse of transient events that may be missed using bulk techniques.

MATERIALS AND METHODS

Cloning of nusG

Genomic DNA was purified from DH5α using standard biochemical techniques of phenol / chloroform / isoamyl alcohol extraction and isopropanol precipitation. The

purified genomic DNA served as the template for the polymerase chain reaction (PCR) amplification of *nusG*. For specific amplification of *nusG*, the following forward and reverse DNA primers, synthesized at the UNC Oligonucleotide Synthesis Facility, were used:

(forward) 5'-CATGCGGCATATGTCTGAAGCTCCTAAAAAG-3',

(reverse) 5'-GACGGATCCTTAGGCTTTTTCAACCTGGCTG-3'.

In a 50 μL reaction, approximately 200 ng of purified DH5α genomic DNA were combined with 200 μM dNTP mix, 0.3 μM forward and reverse primers, 3 mM MgSO$_4$, and 0.5 μL Vent DNA polymerase (NEB) in 1x Thermo Pol buffer (NEB). This reaction was subjected to the following thermocycling program: 30 iterations of 1) 30 seconds at 95 °C, 2) 30 seconds at 60 °C, 3) 45 seconds at 72 °C. Once the specificity of the amplification was verified, *nusG* was purified using standard biochemical techniques of phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation.

In a 50 μL double restriction digest reaction, approximately 3 μg of *nusG* and 3 μg of the vector pET16b (Novagen) were digested separately with 20 U NdeI (NEB) and 50 U BamHI (NEB) in 1x NEB Buffer #3 for 1.5 hours at 37 °C. The doubly restricted DNA was purified by phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. In a 10 μL ligation reaction, approximately 120 fmol of *nusG* sticky ends were ligated to 40 fmol of pET16b sticky ends with 1 mM ATP and 1 μL T4 DNA ligase (Invitrogen) in 1x ligation buffer (50 mM Tris-Cl pH 7.6, 5 mM MgCl$_2$, 5 mM DTT, 50 μg/mL BSA). Upon ligation at 15 °C for 9 hours, 100 μL of ultra-competent XL1-Blue

(Novagen) were transformed with 2 μL of the ligation reaction using heat shock at 42 °C

for 45 seconds. The cells were recovered in SOC broth, plated on LB-agar-ampicillin

plates, and incubated overnight at 37 °C.

The colonies that grew successfully against ampicillin were grown to saturation in

terrific broth (12 g tryptone, 24 g yeast extract, 4 mL glycerol, 17 mM dibasic potassium

phosphate, and 72 mM monobasic potassium phosphate per liter broth) and were

subsequently harvested. The recombinant plasmids that were harbored in these colonies

were extracted and purified by the PEG precipitation method. Upon the preliminary

identification of positive clones through a diagnostic NdeI/BamHI restriction digest, the

proper cloning of *nusG* was confirmed by sequencing analysis through the UNC

Sequencing Facility. The resultant recombinant plasmid encoding NusG was named

pFB1.

Engineering of a Non-Native Cys in NusG

The *nusG*-pET16b recombinant plasmid encodes an N-terminal nine amino acid

linker between the histidine tag and the native *E. coli* NusG. Located within this linker is

a serine (Ser) residue, which was selected for a genetically conservative point mutation to

a cysteine (Cys) residue. To construct such a site-specific Ser→Cys point mutation, the

following primers were used:

(forward) 5'-CATCACAGCTGCGGCCATATC-3',

(reverse) 5'-GATATGGCCGCAGCTGTGATG-3'.

In a 50 μL reaction, 50 ng pFB1, 200 ng of both forward and reverse mutagenesis primers, 1 μL dNTP mix and 1 μL Turbo DNA polymerase (Stratagene) were combined in 1x mutagenesis buffer (Stratagene). This reaction was subjected to the following thermocycling program: 12 iterations of 1) 30 seconds at 95 °C, 2) 1 minute at 60 °C, 3) 6 minutes at 68 °C. The parental (non-mutagenized) strands of DNA were digested with DpnI (Stratagene), and the remaining intact plasmids were subsequently used to transform DH5α.

The DH5α colonies that grew successfully in the presence of ampicillin were grown to saturation in terrific broth and were subsequently harvested. The recombinant plasmids that were harbored in these colonies were extracted and purified by the PEG precipitation method. The successful point-mutagenesis of *nusG* was confirmed by sequencing analysis through the UNC Sequencing Facility. The sequencing analysis also confirmed that the mutagenesis reaction did not introduce any changes to *nusG* other than the targeted cysteine residue. The recombinant plasmid containing the non-native Cys variant of *nusG* (NusG-Cys) was named pFB2.

Expression and Purification of NusG-Cys

The expression strain BL21(DE3) was transformed with pFB2. Transformants were grown to mid-log phase in LB supplemented with 100 μg/mL ampicillin and were induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG). Expression of NusG-Cys was allowed to continue for 3.5 hours before the cells were harvested. One-half liter of culture yielded approximately 1 g cell pellet.

The cell pellet was resuspended in 5 mL lysis buffer [20 mM sodium phosphate, pH 7.4, 0.5 M NaCl, 10 mM imidazole, and 7 mM beta-mercaptoethanol (βME)] supplemented with 0.1 mg/mL lysozyme and 0.1 mM phenylmethylsulfonyl fluoride (PMSF). The resuspended cells were incubated on ice for 30 minutes prior to sonication at 10% pulse with maximum output for 10 seconds. Cellular debris was pelleted at 10,000x g for 25 minutes. The resultant cleared lysate (~5 mL) was added to 1 mL Ni^{2+}-NTA resin pre-equilibrated with lysis buffer, and the lysate resin mixture was incubated with gentle rocking at 4 °C for 1 hour. Upon gently pelleting the resin, the resin was gently resuspended in 10 mL 20 mM imidazole wash buffer and rocked for 5 minutes. The subsequent pelleting and resuspension of the resin was cycled using a step gradient with wash buffers containing 30 mM, 40 mM, 50 mM, and 100 mM imidazole. NusG-Cys was eluted from the resin with 300 mM imidazole and was shown via SDS-PAGE analysis to be greater than 90% pure.

Fluorescent Modification of NusG-Cys

Purified NusG-Cys was dialyzed against conjugation buffer (20 mM Tris-Cl, pH 7.4, 250 mM NaCl, 10% glycerol, and 0.1 mM Tris(2-carboxyethyl)phosphine (TCEP)) at 4°C. A five-fold molar excess of the maleimide Alexa Fluor 546 (AF 546, Molecular Probes) dissolved in anhydrous DMSO was added dropwise to NusG-Cys (60 μM). The conjugation reaction of AF 546 to the single Cys residue of NusG-Cys was protected in aluminum foil and was allowed to proceed overnight at 4 °C with gentle rotation. To terminate the reaction and purify modified NusG-Cys (NusG-AF546) from unconjugated

AF 546, NusG-AF546 was dialyzed against buffer with 0.1 mM dithiothreitol (DTT) covered in foil at 4 °C.

To determine the resultant molar ratio of NusG-Cys to AF 546, the optical density of a sample of NusG-AF546 was measured at 280 nm and at 556 nm, the wavelength of maximum absorption for AF 546. Using Beer's Law and the corresponding ε values ($\varepsilon_{280, \text{NusG}}$ = 15,470 $\text{M}^{-1}\text{cm}^{-1}$, $\varepsilon_{556, \text{AF 546}}$ = 104,000 $\text{M}^{-1}\text{cm}^{-1}$), the individual molar components of NusG-Cys and of AF 546 were calculated (Molecular Probes). The calculation of the NusG-Cys molar concentration was corrected for the contribution of AF 546 absorbance at 280 nm by multiplying the correction factor of 0.12 with the absorbance of AF 546 at 556 nm and subtracting this value from the absorbance of NusG-Cys at 280 nm (Molecular Probes). The resulting molar ratio of NusG-Cys to AF 546 was determined to be 1:0.6.

Generation of a Fluorescently Modified DNA Template

To generate a fluorescently modified variant of the DE13 template (see Chapters 2 and 3), a primer (primer 2-Cy5) was designed and constructed with a 5'-Cy5 fluorophore (IDT). Amplification of pDE13 with primer 1b (5'-biotin-GCGACGTGCGTCCTCAAGC) and primer 2-Cy5 (5'-Cy5-GCGGTTGCGTTCCTGAATGG) produced a doubly functional DE13 template, with a 5'-biotin on one strand and a 5'-Cy5 on the other strand. This modified DE13 template is displayed in Figure 4.1. Modified DNA was purified from excess primers using Microcon spin columns (Amicon).

Quartz Slide and Sample Preparation

Quartz microscope slides (3" x 1" x 1 mm, G. Finkenbeiner, Inc.) were drilled with 220G diamond point drill bits (part 115005, Starlite) to generate a series of four holes, which would serve as channel openings, along the 3" sides of each slide. Prior to use, the drilled slides underwent an exhaustive cleaning regiment that included 15 min incubations in a bath sonicator while submerged in successive rounds of solvents: alconox, acetone, 95% ethanol, 1 M KOH (potassium hydroxide), 95% ethanol, 1 M KOH (Weninger *et al.*, 2003). Upon cleaning, the slides were stored in deionized water and were dried over a propane torch immediately before use. Strips of double-sided Scotch tape were gently pressed across the dried slide on either side of the channel holes to define the individual channels. A disposable 24 mm x 40 mm No. 1.5 glass coverslip (VWR) was gently torched to evaporate the waxy residue on the surface of the coverslip and was subsequently placed on top of the slide lined with tape. Excess fringes of tape were excised with a razor blade, and epoxy was lathered along the adjoining sides of the slide and coverslip to create a seal for the channels.

The channels were flushed with syringe-filtered 1x transcription buffer (30 mM HEPES, pH 8.0, 200 mM potassium glutamate, 10 mM magnesium glutamate, 14 mM beta-mercaptoethanol (βME)). The quartz surface of the channels was then coated with 1 mg/mL biotinylated bovine serum albumin (BSA, Sigma) during a five-minute incubation. Excess biotinylated BSA was flushed from the channel with 1x transcription buffer. Streptavidin at 0.1 mg/mL (Molecular Probes) was flowed into the channel and allowed to bind to the biotinylated BSA during a five-minute incubation. Excess streptavidin was flushed from the channel with 1x transcription buffer.

126

Open promoter complexes (OPCs) were formed upon incubation of 50 nM biotinylated, Cy5-DE13 with 75 nM RNA polymerase (RNAP) in 1x transcription buffer at 37 °C for 10 minutes. OPCs were then brought to ambient temperature and were allowed to react with 10 μM ATP, UTP, and GTP for 45 seconds to form stalled elongation complexes (SECs). SECs were diluted to 10 pM and flowed into the channel over the biotinylated BSA-streptavidin coated surface. During a five-minute incubation, the SECs were immobilized to the surface of the quartz slide via the interaction between the bound streptavidin and the biotinylated DNA of the SEC. A schematic of the immobilized SECs is presented in Figure 4.1. Unbound complexes were flushed from the channel with ice-cold 1x transcription buffer.

Immediately before imaging the sample at room temperature, channels with immobilized SECs were flushed with ice-cold 1x transcription buffer supplemented with enzymatic oxygen scavenging components (2% glucose (Sigma), 0.1 mg/mL glucose oxidase (Sigma), 0.025 mg/mL catalase (Sigma)) and a triplet state quencher (50 μM cyclooctatetraene (Aldrich)). The oxygen scavengers enhanced fluorophore lifetime, while the triplet state quencher reduced fluorophore blinking. All buffers with these components were freshly made and passed through a 0.2 μm filter just prior to use. To locate the SECs in the presence of NusG, NusG was diluted in ice-cold oxygen scavenging buffer and flown directly into the channel without additional flushing.

TIR Fluorescence Microscopy

Samples were imaged and data were collected using a through-prism total internal reflection fluorescence (TIRF) microscope, as previously described (Weninger *et al*., 2003; Bowen *et al*., 2005). In this optical setup, two diode lasers were utilized by directing them onto the prism. The "red" laser emitted light at 635 nm and directly excited the Cy5 fluorophore, while the "green" laser emitted light at 532 nm and directly exited the AF 546 fluorophore at the quartz-solution interface. Fluorescence emission was collected through a 60x 1.2 numerical aperature water immersion objective and split by a 645 nm dichroic mirror into short and long wavelength paths (Optical Insights Dual-View beam-splitter). These paths were filtered for Cy5 and AF 546 emission using HQ 700/75 and HQ 585/70 bandpass filters, respectively. The respective spectrally-resolved emissions were relayed as dual images on an intensified charge-coupled device camera (Roper Scientific Cascade II 512B), as shown in Figure 4.2. Images were exposed at 10 frames per second and collected using software written in-house (Keith Weninger, NCSU).

To perform colocalization experiments, a prepared sample was exposed to red laser excitation for 1.6 seconds to permit location of the immobilized, Cy5-labelled stalled elongation complexes excited by red light. After briefly locating these molecules, the sample was then exposed to green laser excitation for 50 seconds to excite AF 546-labelled NusG. Colocalization of NusG to the immobilized stalled complexes was calculated only during green excitation. At the end of the experiment, the sample was once again exposed to red laser excitation to permit appropriate background corrections.

Data Analysis

The emission intensities detected from excited single molecules were integrated with software written in-house (Keith Weninger, NCSU) to generate individual fluorescent emission time traces, as previously described (Bowen *et al*., 2005). Background corrections were made for all emission time traces. An edge detection algorithm, written previously to analyze dynamic fluorescence resonance energy transfer (FRET) efficiency traces (Computer Science Department, UNC), was modified by a peer (Cherie Lanyi, UNC) to analyze colocalization of individual fluorescent molecules. This modified edge analysis allowed us to separate emission intensity transitions of NusG molecules colocalized to transcription complexes. Once the transitions were sorted, the lifetimes of each observed NusG emission intensity were calculated. False transition edges were eliminated by altering the scale spaces with varying thresholds.

RESULTS AND DISCUSSION

Colocalization of NusG to a Stalled Elongation Complex

We sought to advance our understanding of NusG interaction with a transcription complex. To better explore this interaction, we conducted single-molecule fluorescence experiments designed to detect the colocalization of fluorescently-labeled NusG to stalled elongation complexes fluorescently labeled at one end of the DNA substrate (Figure 4.1). In these experiments, 1 nM NusG was flowed over 50 pM immobilized transcription

complexes, and the data obtained from this system are preliminary. Representative

sample time traces generated from these preliminary experiments are presented in Figures

4.2a-d. These time traces show the red fluorescent emissions of single transcription

complexes at the beginning and the end of the experiment. In sorting through the traces,

we kept and subsequently analyzed red fluorescent emission data that exhibited values

between 500 and 2000 units. Intensities greater than 2000 units are indicative of multiple

molecules. In addition to the complexes, we are able to detect individual NusG

molecules in our field of view due to the presence of green fluorescent emissions in the

time traces (Figures 4.2a-d). These NusG molecules appear to be colocalized to the

elongation complexes tethered to the surface of the quartz slide. Inspection of the traces

reveals that the NusG-colocalization is quite dynamic. Figures 4.2a-b show examples of

colocalization events characterized by longer dwell-times when NusG remains bound to

the complexes for a significant amount of time. Conversely, Figures 4.2c-d show

examples of colocalization events characterized by much shorter dwell-times and

frequent transitions from bound to unbound states of NusG. It is important to point out

that control reactions were performed involving the flowing of NusG molecules over

immobilized naked DNA templates. Analysis of these control reactions reveal that

colocalization of individual NusG molecules to the immobilized DNA template is not

observed.

Lifetime Analysis of NusG Colocalization

The individual time traces are subsequently analyzed using an edge analysis
algorithm. Figure 4.3 illustrates how this edge analysis works. As shown in Figure 4.3,
the data are sorted by FRET efficiency, or in the case of the present study, by fluorescent
emission intensities. The time elapsed during the emission of each intensity observed, as
well as the number of transitions observed, are determined for each trace. Once sorted, a
histogram (Figure 4.4) of the dwell-time can be constructed. As can be seen, the
histogram in Figure 4.4 can be fitted well to a single exponential decay. The single
exponential decay curve yields an off rate (k_{off}) of 0.2 s^{-1}. Therefore, NusG exhibits a
lifetime of 5 seconds in association with a transcription complex. Assuming a diffusion-
limited binding event of NusG to a transcription complex, we can estimate k_{on} to a value
of ~ 10^8 M^{-1} s^{-1}. This estimation had been previously made for the association rate
constant of the lac repressor to specific operator site on DNA (von Hippel and Berg,
1989). To estimate k_{on} for the lac repressor, the authors used the Smoluchowski equation
for the association rate constant ($k_{assoc} = 4\pi\kappa af(D_A + D_B)N_0/1000$) (Barkley, 1981) and set
the diffusion constant of the repressor (D_A) to 5 x 10^{-7} cm^2/s, the diffusion constant of the
DNA (D_B) to 0 cm^2/s, the interaction distance (*a*) to 50 x 10^{-8} cm, the electrostatic factor
(*f*) to unity, and the fraction of reactive particles (κ) to 0.05. Using the mathematical
definition of the dissociation constant, K_D, ($K_D = k_{off}/k_{on}$), we can estimate the
dissociation constant of NusG to be ~ 200 nM. While this dissociation constant is not a
particularly strong one, it is in agreement with other previous experimental evidence
suggesting that NusG interaction with core RNAP is rather weak, given the observation
that only approximately 10% of recovered core RNAPs, using affinity chromatography

techniques, had an associated NusG protein (Li *et al*., 1992). These authors found that other accessory proteins, including N and NusA, are required for the stable association of NusG with core RNAP.

Transition Analysis of NusG Binding Events

As indicated above (Figure 4.3), the edge analysis algorithm can be used to detect transitions in the fluorescent emission intensities of NusG molecules colocalized to stalled elongation complexes. The edge analysis allows us to gauge the "start" intensity and the "stop" intensity of each transition. A histogram of these start and stop intensities can be constructed to generate a three-dimensional topographical image, as shown in Figure 4.5. This image is color coded such that highly populated transitional states are shaded in red, while the less densely populated transitional states are shaded in indigo. In Figure 4.5, two sets of salient peaks emerge and are indicated by red arrows. The one set of major peaks that corresponds to a start intensity of approximately 200 units and a stop intensity of approximately 300-400 units indicates a binding event. Since two major peaks are observed in this subset, the binding event appears to be governed by two distinct transitions. Conversely, the other set of major peaks that corresponds to a start intensity of 300-400 units and a stop intensity of 200 units indicates a dissociation event. Again, this event appears to be governed by two distinct transitions.

As stated earlier, the results of the single molecule fluorescence study presented here are only preliminary. While the results are promising, more work is needed to fully develop this system by which individual transcription accessory proteins can be detected

in association with a stalled elongation complex. Once the system involving stalled

elongation complexes is sufficiently developed, the next step would be to utilize a single

flow channel apparatus, which would allow us to flow NTPs, in addition to accessory

factors, over stalled elongation complexes and to monitor the association of accessory

factors with an actively transcribing elongation complex. Additionally, other DNA

templates may be used, including ones that encode such elements as the *ops* pause, in

generating the stalled complexes. These complexes can then resume elongation with the

reintroduction of NTPs, and we can observe NusG interaction, for example, with an

elongation complex transcribing through the *ops* pause. Experiments such as the ones

described here may aide our understanding in the protein-protein interactions that govern
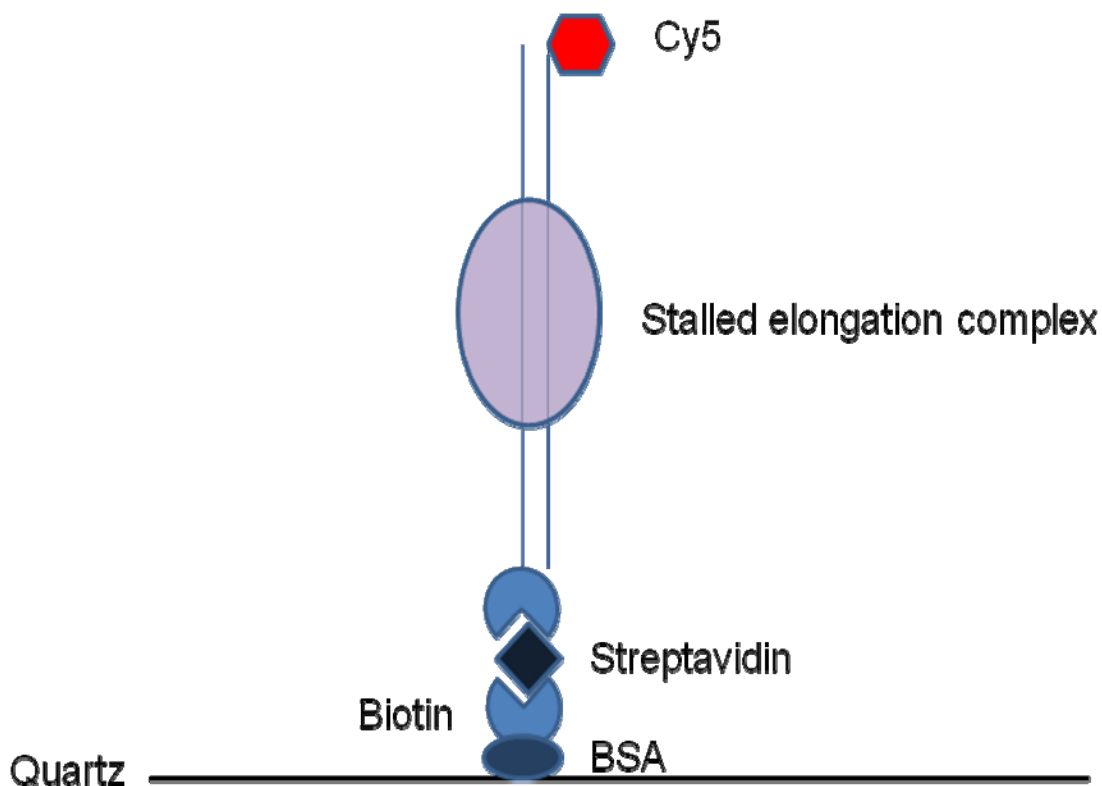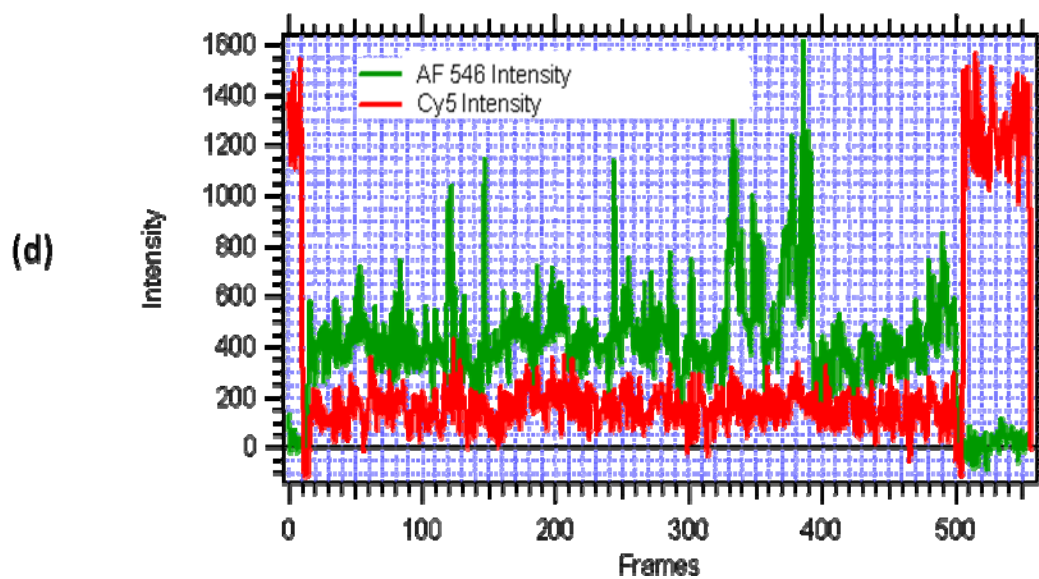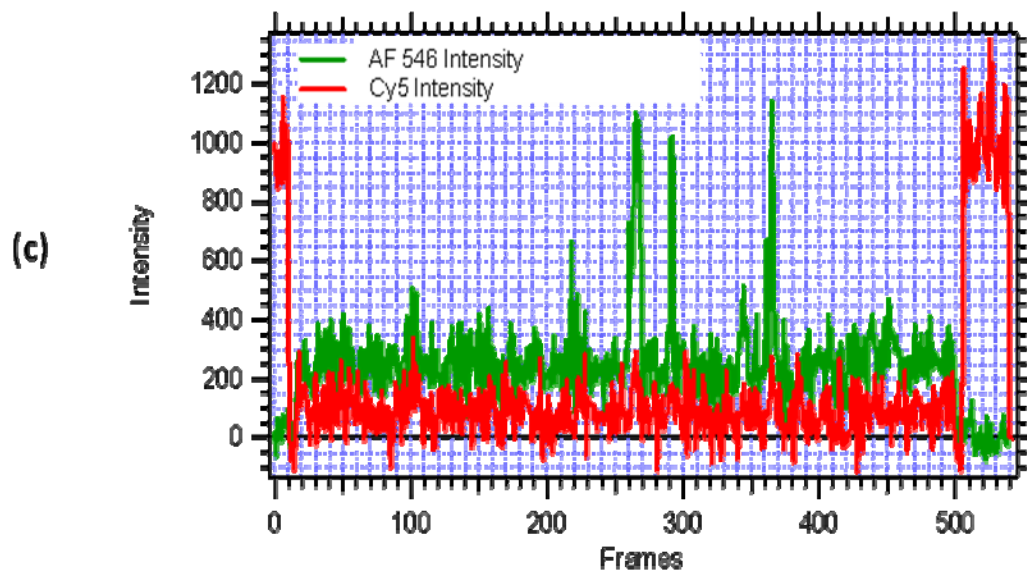
the regulation of transcription.

Figure 4.1 Schematic of a stalled elongation complex immobilized to the surface of a quartz microscope slide. The stalled elongation complex was generated from the DE13 template containing a biotin molecule at one 5' end and a Cy5 fluorophore at the other 5' end. The quartz surface was coated with 1 mg/mL biotinylated bovine serum albumin (BSA). Streptavidin at 0.1 mg/mL was then bound to the BSA-coated surface. Stalled elongation complexes (10 pM) were tethered to the coated surface via interaction between the bound streptavidin and the 5'-biotinylated DNA template.
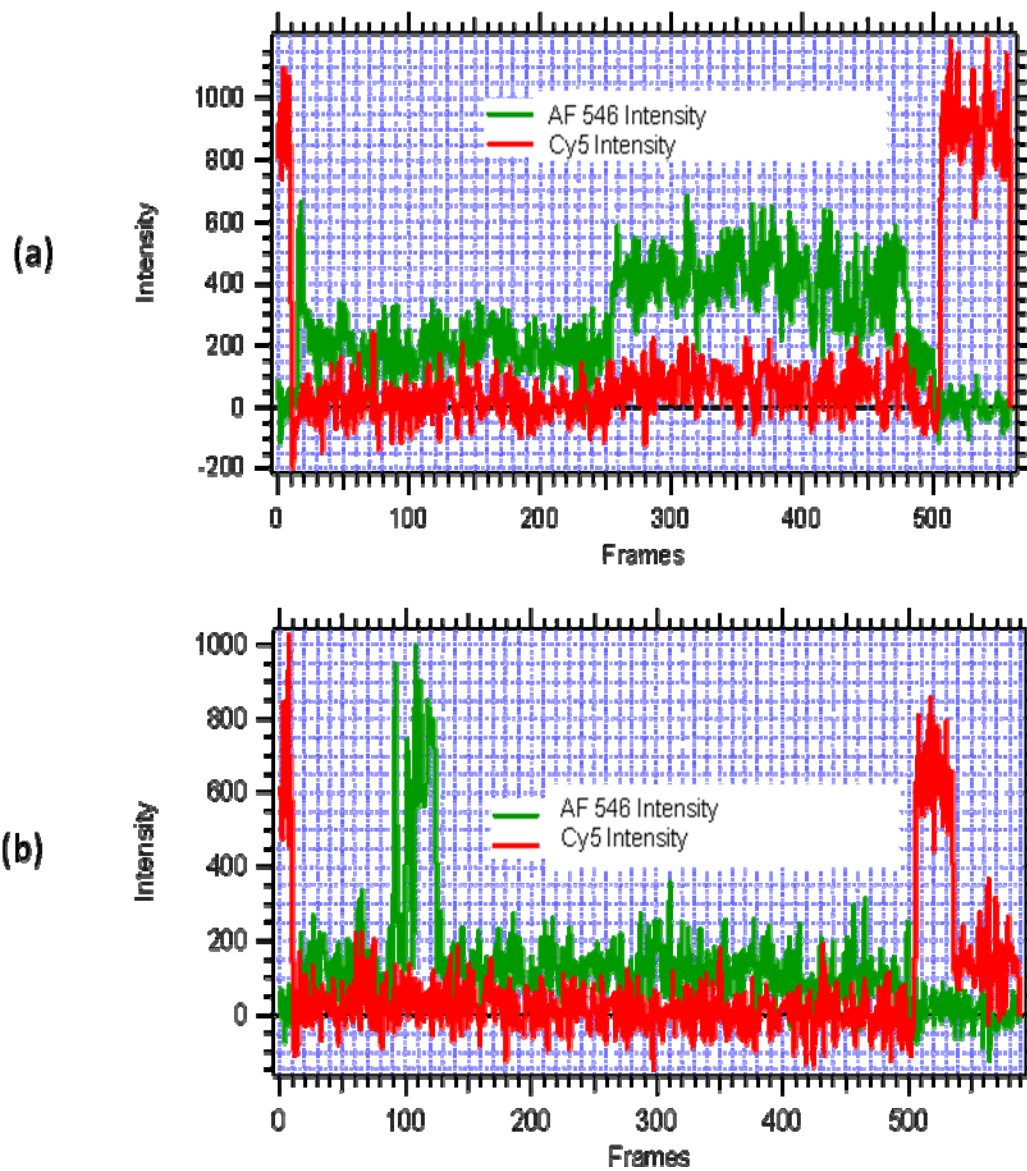
Figures 4.2a-b Representative fluorescence emission traces as a function of time.

Figures 4.2a-b Representative fluorescence emission traces as a function of time.

Data were collected at 10 frames per second. Shown are the intensities of AF 546 (green)

emission, from single NusG molecules, and of Cy5 (red) emission, from single stalled

elongation complexes tethered to the surface of the microscope. Using a "red-alt-green"

schematic, the sample was subjected to red laser excitation for 1.6 s, then to green laser

excitation for 50s, followed by brief red laser excitation.

(c)

(d)

Figures 4.2c-d Representative fluorescence emission traces as a function of time.

Figures 4.2c-d Representative fluorescence emission traces as a function of time.

Data were collected at 10 frames per second. Shown are the intensities of AF 546 (green)

emission, from single NusG molecules, and of Cy5 (red) emission, from single stalled

elongation complexes tethered to the surface of the microscope. Using a "red-alt-green"

schematic, the sample was subjected to red laser excitation for 1.6 s, then to green laser

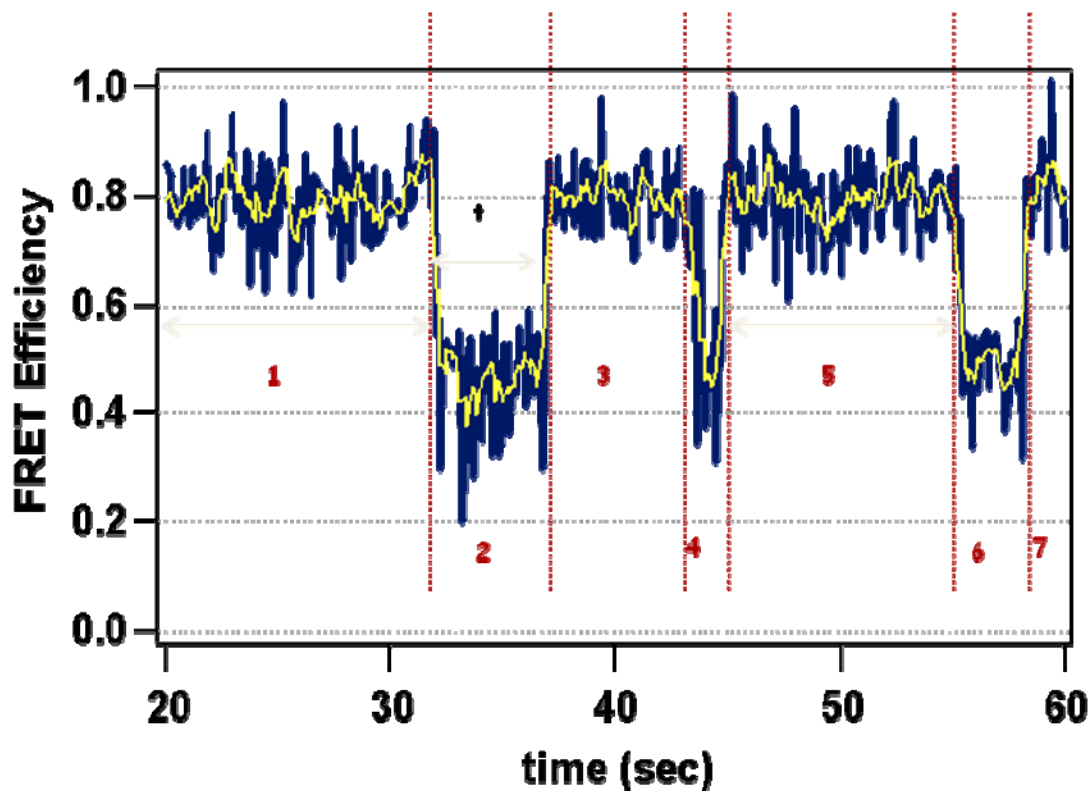excitation for 50s, followed by brief red laser excitation.

Figure 4.3 Schematic of the edge analysis algorithm. The edge analysis algorithm was used to analyze the fluorescent intensity traces as a function of time. In the image displayed above, FRET efficiency is analyzed. The original algorithm was specific for FRET efficiency values and was thus modified (Cherie Lanyi, UNC) to analyze the intensity of fluorescent emissions. The important features of this analysis are the sorting of intensities based on value, the time elapsed during emission at each of these intensities, and the number of transitions observed between the different intensities.
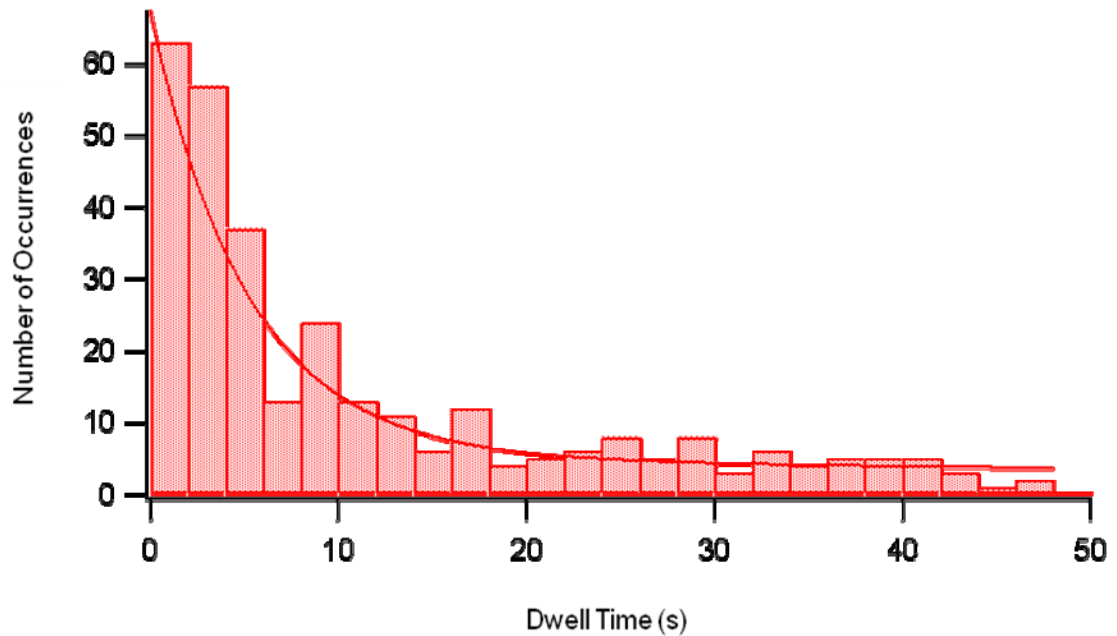
Figure 4.4 Lifetime analysis of NusG colocalized to stalled elongation complexes.
The distribution of colocalized NusG molecules obtained from 357 data points is fitted to a single exponential decay. This fit yields an off rate (k_{off}) of 0.2 s^{-1}. Estimating a value of $\sim 10^8$ M^{-1} s^{-1} for the on rate (k_{on}), we estimate that the dissociation constant (K_D) of NusG is $\sim$ 200 nM.
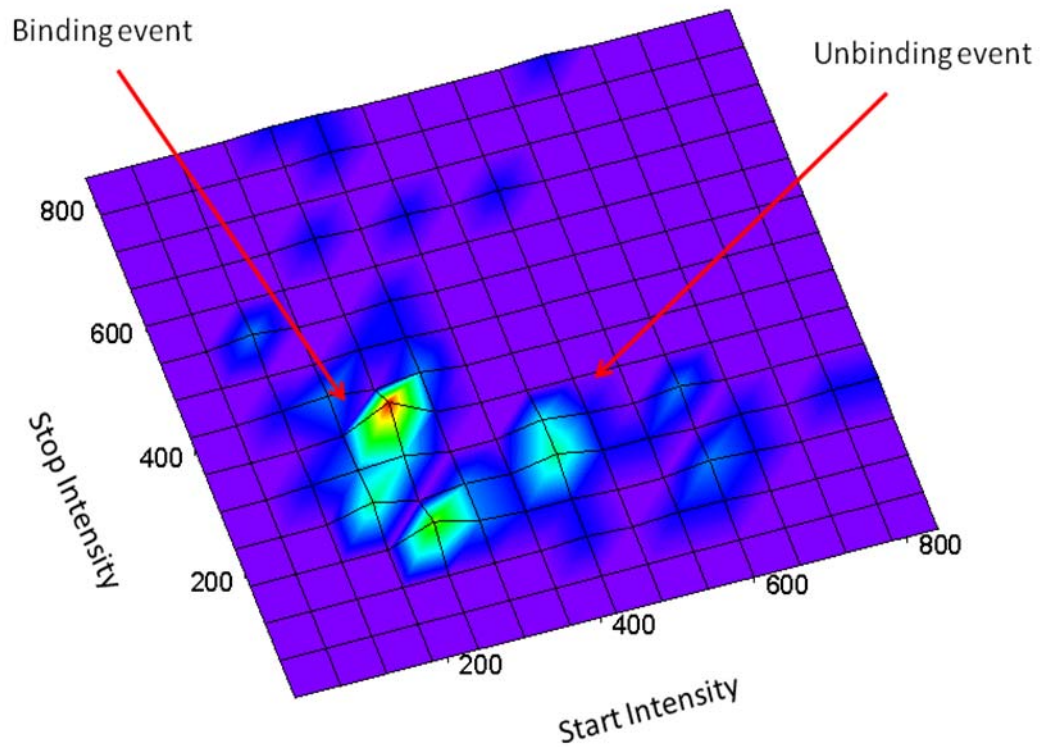
Figure 4.5 Transition analysis of NusG binding events. A histogram of the start and stop intensities of each observed transition was used to generate a three-dimensional topographical image. The image is color coded such that highly populated transitional state are shaded in red, while the less densely populated transitional states are shaded in indigo. Indicated are the binding and unbinding events observed in the colocalization of NusG with stalled elongation complexes.

BIBLIOGRAPHY

Barkley, M.D. (1981). Salt Dependence of the Kinetics of the Lac Repressor-Operator Interaction: Role of Nonoperator Deoxyribonucleic Acid in the Association Reaction. Biochemistry *20*, 3833-3842.

Bowen, M.E. et al., (2005). Single-Molecule Studies of Synaptotagmin and Complexin Binding to the SNARE Complex. Biophys. J. *89*(1), 690-702.

Burns, C.M. et al., (1998). Combinatorial Effects of NusA and NusG on Transcription Elongation and Rho-dependent Termination in Escherichia coli. J. Mol. Biol. 278, 307-316.

Burova, E. et al., (1995). Escherichia coli NusG Protein Stimulates Transcription Elongation Rates In Vivo and In Vitro. J. Bact. *177*(5), 1388-1392.

Erie, D.A. (2002). The Many Conformational States of RNA Polymerase Elongation Complexes and Their Roles in the Regulation of Transcription. Biochim. Biophys. Acta. *1577*(2), 224-39.

Gill, S.C. et al., (1991). Escherichia coli Sigma 70 and NusA Proteins. I. Binding Interactions with Core RNA Polymerase in Solution and within the Transcription Complex. J. Mol. Bio. *220*(2), 307-24.

Hartzog, G.A. et al., (1998). Evidence that Spt4, Spt5, and Spt6 Control Transcription Elongation by RNA polymerase II in Saccharomyces cerevisiae. Genes and Dev. *12*(3), 357-69.

Knowlton, J.R. et al., (2003). A Spring-Loaded State of NusG in Its Functional Cycle is Suggested by X-ray Crystallography and Supported by Site-Directed Mutants. Biochemistry *42*, 2275-2281.

Li, J. et al., (1992). NusG, a New Escherichia coli Elongation Factor Involved in Transcriptional Antitermination by the N Protein of Phage Lamda. J. Bio. Chem. *267*(9), 6012-6019.

Li, J. et al., (1993). Elongation Factor NusG Interacts with Termination Factor Rho to Regulate Termination and Antitermination of Transcription. Genes and Dev. *7*, 161-172.

Liao, D. et al., (1996). A NusG-Like Protein from Thermotoga maritima Binds to DNA and RNA. J. Bact. *178*(14), 4089-4098.

Pasman, Z and von Hippel, P.H. (2000). Regulation of Rho-Dependent Transcription Termination by NusG is Specific to the Escherichia coli Elongation Complex. Biochemistry *39*, 5573-5585.

Steiner T. et al., (2002). Crystal Structures of Transcription Factor NusG in Light of its Nucleic Acid- and Protein-Binding Activities. EMBO *21*(17), 4641-4653.

Sullivan, S.L. and Gottesman, M.E. (1992). Requirement for E. coli NusG protein in factor-dependent transcription termination. Cell *68*(5), 989-994.

von Hippel, P.H. and Berg, O.G. (1989). Facilitated Target Location in Biological Systems J. Bio. Chem. *264*(2), 675-678.

Weiss, S. (1999). Fluorescence Spectroscopy of Single Biomolecules. Science *283*, 1676-1683.

Weninger, K.M. et al., (2003). Single-Molecule Studies of SNARE Complex Assembly Reveal Parallel and Antiparallel Configurations. Proc. Natl. Acad. Sci. USA *100*(25), 14800-5.