

I give permission for public access to my thesis and any copying to be done at the discretion of the archives' librarian and/or the College library.

Mya Thandar

Signature

05/10/2012

Date

EVIDENCE SUPPORTING A NOVEL MECHANISM OF
ABORTIVE INITIATION BY *E. COLI* RNA
POLYMERASE

BY:

MYA THANDAR

A Paper Presented to the
Faculty of Mount Holyoke College in
Partial Fulfillment of the Requirements for
the Degree of Bachelors of Arts with
Honor

Department of Biochemistry
South Hadley, MA 01075

May 2012

This thesis was prepared under the direction of:

Dr. Lilian M. Hsu
Professor of Biochemistry
for ten credits.

For Mama & Papa

ACKNOWLEDGMENTS

First and foremost, I would like to thank Professor Lilian Hsu for being the most understanding and supportive mentor. Words could not express my gratitude to her. Since the very first day, her wisdom has guided me and her encouragement has motivated me to achieve more than I could have wished for. Completing this thesis would have been impossible without her patience, knowledge and expertise. It has been a privilege having her as my research and academic adviser.

Secondly, I would like to thank Tenaya Vallery for showing me the tips and tricks of different techniques, from putting in scratching time during time-course reactions to checking frozen reagents. She has shared with me her passion for Biochemistry and we have had many productive discussions/arguments. I would not have been able to get this far in the project without her guidance and substantial contributions.

Third, I would like to thank Ahri Lee for her joint effort in preparing the DNA for footprinting experiments and for ice-cream time-outs. I would also like to thank Rebekah Wieland and Ellen Shamansky for all the good times together in lab and at conferences. In addition, I would like to thank all my friends, especially San Theingi, Macy Akalu, Saranya Ramakrishnan and Anastasia Hyrina, for bringing balance to my college life. These individuals and other Mount Holyoke College Faculty and Staff have made this place a second home

away from home.

Next, I would like to thank NSF grant RUI (MCB0841452) awarded to L. M. Hsu for funding my research.

Finally, I would like to thank my grandma and parents. They have loved me for who I am and their unwavering kindness and affection have provided me with the energy to make it through all the difficult times.

TABLE OF CONTENTS

	PAGE
List of Figures	vii
List of Tables	ix
Abstract	x
Introduction	1
Materials and Methods	
Materials	
DNA primers and plasmids	29
Proteins and enzymes	33
Reagents	34
Methods	
<i>Roadblock transcription reactions</i>	
EQ-series DNA construction	35
Purification of EQ-series DNA	36
Purification of roadblock protein (E ₁₁₁ Q EcoRI)	37
Transcription assays	40
<i>Permanganate footprinting reactions</i>	
Different approaches to prepare single end-labeled promoter templates for footprinting	42
Cloning/modification of pSAN1 and pSAN2 for N25 and N25 _{anti} promoters	44
Plasmid labeling: Generating single end-labeled fragment from plasmid DNAs	47
Maxam-Gilbert sequencing ladder	50
Permanganate footprinting	51
Results	
Roadblock transcription experiments	
Rationale	55
Results	57
Summary	66
Permanganate footprinting experiments	
Rationale	71
Optimization	71
Results on template strand	79
Results on non-template strand	82
Summary	89
Discussion	93
References	100

LIST OF FIGURES

FIGURES	PAGE
1. The schematic of transcription initiation	3
2. The breakdown of promoter elements recognized by $E\sigma^{70}$	5
3. Models of closed complex (RP_c) and open complex (RP_o)	10
4. The role of DNA scrunching in abortive initiation and promoter escape	13
5. The GreB rescue of backtracked RNAP	16
6. The GreB-resistant nature of <u>v</u> ery <u>l</u> ong <u>a</u> abortive <u>t</u> ranscripts (VLATs)	19
7. The postulated model of RNAP hyperforward translocation	22
8. The reaction mechanism and procedure of permanganate footprinting	26
9. Comparison of the four footprinting promoters, N25, N25 _{anti} , DG203 and SP _{fullcon}	28
10. Gel image of <i>in vitro</i> transcription experiments with and without GreB on EQ-series DNA	59
11. Gel image of <i>in vitro</i> transcription experiments without EcoRI and GreB, with EcoRI alone, and with both EcoRI and GreB on EQ-series DNA	62
12. Gel image of <i>in vitro</i> transcription experiments without EcoRI and GreB, with EcoRI alone, with both EcoRI and GreB, and with GreB alone on selected EQ-series DNA	65
13. The quantification of the results from roadblock transcription experiments	68

14. Gel image of <i>in vitro</i> transcription experiments on the footprinting promoter templates N25, N25 _{anti} , DG203 and SP _{fullcon} : comparison of buffer conditions	74
15. Gel image of <i>in vitro</i> transcription experiments on the footprinting promoter templates N25, N25 _{anti} , DG203 and SP _{fullcon} : KCl titration	76
16. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled DG203 and SP _{fullcon} promoter fragments prepared by primer extension	78
17. Gel image of permanganate footprinting experiments on 5'-template strand end-labeled promoter fragments derived from plasmid DNA	81
18. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled promoter fragments derived from plasmid DNA	84
19. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled N25 _{anti} promoter fragment prepared from plasmid DNA: KMnO ₄ titration	86
20. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled N25 _{anti} promoter fragment prepared from plasmid DNA: NTP titration	88
21. Summary of permanganate footprinting experiments	92

LIST OF TABLES

TABLES	PAGE
1. The upstream (u) and downstream (d) EQ-primer sequences used in generating EQ-series DNA, reading 5' to 3'	30
2. The upstream (u) and downstream (d) sequences of VL-primers used in generating DNA for single end-labeling, reading 5' to 3'	31
3. The non-template (NT) and template (T) sequences of SAN-primers, reading 5' to 3'	32
4. Summary of the roadblock transcription analysis	70

ABSTRACT

Transcription is one of the highly regulated steps in gene expression and comprises three phases – initiation, elongation and termination. In prokaryotes, the initiation phase is further divided into two steps: (1) promoter binding and activation; and (2) abortive initiation and promoter escape. Our wild-type promoter of interest, T5 phage N25, is a prototypical strong *E. coli* RNA polymerase σ^{70} promoter that produces abundant transcripts. However, due to its consensus-like promoter elements, the transcription initiation is rate-limited at the promoter escape step, leading to extensive abortive initiation. Studies have shown that the position of promoter escape depends not only on the promoter-polymerase affinity but also on the initial transcribed sequence (ITS) (Kammerer et al., 1986; Hsu et al., 2006). While N25 aborts to +11 and escapes at +12, its ITS variants, such as N25_{anti} and DG203, abort to +15 and +19 and escape at +16 and +20, respectively (Hsu et al., 2006).

Studies with the transcription factor GreB showed that all abortive transcripts ≤ 15 nucleotides (nt) arise from RNA polymerase (RNAP) backtracking and those longer than 4 nt can be rescued (Hsu et al., 1995). “Rescue” involves GreB, bound in the RNAP secondary channel, stimulating the hydrolytic activity of RNAP to cleave the backtracked nascent RNA at the active site, aligning a new 3'-OH group such that the 5'-nascent RNA can be further elongated (Opalka et al., 2003). However, on the DG203 promoter, it was

discovered that a fraction of the very long aabortive transcripts (VLATs) of 16-19 nt in length that are GreB-resistant. Consequently, it was proposed that the GreB-resistant VLATs emerge from a different mechanism called RNAP hyper-forward translocation: Upon reaching the promoter escape transition stage, a fraction of the RNAP is propelled forward by more than one nucleotide, causing the 3'-end of nascent RNA to move upstream past the active center and become lodged in the RNA exit channel instead. Thus inactivated, the RNA polymerase complex subsequently releases the nascent RNA as abortive transcripts (Chander et al., 2007).

In the current study, the proposition of GreB-resistant VLAT formation by RNAP hyper-forward translocation was tested by placing a mutant EcoRI protein roadblock at various locations in the downstream stretches of DG203 promoter and performing *in vitro* transcription to examine the progress of RNAP. The results of “roadblock transcription” analysis gave the following conclusions: 1. VLATs are indeed products of RNAP forward movement, and they are produced during the promoter escape transition on the DG203 promoter; 2. We observed a spatial requirement of at least 2 bp, but optimally 4-5 bp, for RNAP to undergo hyper forward translocation to produce the GreB-resistant VLATs.

Moreover, it was postulated that the energy that propels RNAP forward by more than one nucleotide comes from DNA scrunching (Revyakin et al., 2006; Chander et al., 2007). If true, the initial transcribing complex of DG203 on the cusp of escape must have scrunched in 19 bp of DNA, greatly expanding the

transcription bubble from its normal 13-14 bp size. To test this prediction, permanganate footprinting approach was used to characterize the sizes of the initial transcribing complex (ITC) bubbles. For comparison, a variant of the DG203 promoter called SP_{fullcon} was further created. It differs in sequence from DG203 only in the 17-bp spacer region connecting the -35 and -10 elements. SP_{fullcon} shows the same position of escape as DG203, but its level of full-length transcripts is greatly diminished (Chander and Hsu, manuscript in submission), allowing us to exclusively capture the melted DNA region during initiation and scrunching. Using four different promoters—N25, N25_{anti}, DG203 and SP_{fullcon}—the sizes of the open complex and transcribing complex bubbles were characterized and compared. The results indicated that the VLAT-producing ITC bubbles were much larger; bubble expansion to the +20 position could be detected. This observation suggested a higher degree of DNA scrunching in the DG203 and SP_{fullcon} promoters which, in turn, propelled the RNAP forward by more than one nucleotide during the escape transition, giving rise to the formation of GreB-resistant VLATs.

INTRODUCTION

Transcription is a process in which cellular RNA is synthesized by evolutionarily conserved DNA-dependent RNA polymerase (RNAP). In prokaryotes, it is the first step in the regulation of gene expression. Like other macromolecular synthesis processes, transcription can be divided into three phases: initiation, elongation and termination. The initiation phase can further be divided into steps of promoter binding, open complex formation, RNA synthesis, abortive initiation and promoter escape (Figure 1, adapted from Hsu, 2002b). During initiation of house-keeping genes in *Escherichia coli* (*E. coli*), the RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma^{70}$ or $E\sigma^{70}$) first recognizes and binds to the upstream promoter sequences with the help of the specific initiation factor σ^{70} to form a closed complex (RP_c) (Murakami et al., 2002a; Mooney et al., 2005). The promoter DNA elements that the holoenzyme $E\sigma^{70}$ recognizes are the -35 and -10 elements (and the extended -10 nucleotides), the spacer sequence connecting the two (Murakami et al., 2002a), and an upstream or UP element in some promoters (Ross et al., 1993). The double-stranded DNA (dsDNA) is then melted apart from the upstream boundary of the -10 element to about +2 (relative to the +1 start site of transcription) by the interaction between domain 2 of σ -subunit (σ_2) and the -10 element (Feklistov and Darst, 2011). The breakdown of the promoter DNA elements and their interactions with multiple subunits of RNAP is shown in Figure 2.

Figure 1. The schematic of transcription initiation. Transcription commences when the RNAP (R) recognizes promoter DNA (P), forming a closed complex (RP_c). The closed complex then isomerizes to form the open complex (RP_o). When ribonucleoside triphosphates (NTPs) are present, the RNAP begins the *de novo* synthesis of RNA. At this initial transcribing complex (ITC) stage, the RNAP repeatedly synthesizes and releases nascent RNAs (abortive initiation) until the RNAP-promoter contacts are disrupted (promoter escape) for the RNAP to move on to the elongation phase (formation of ternary elongation complex (TEC)). The branched pathway model of abortive and productive initiation showed the presence of RP_o' and unproductive ITC' that abortively initiate without undergoing escape (Hsu, 2002b).

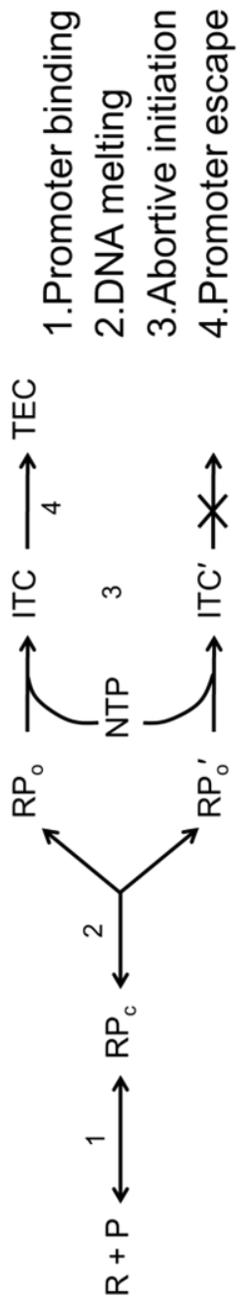
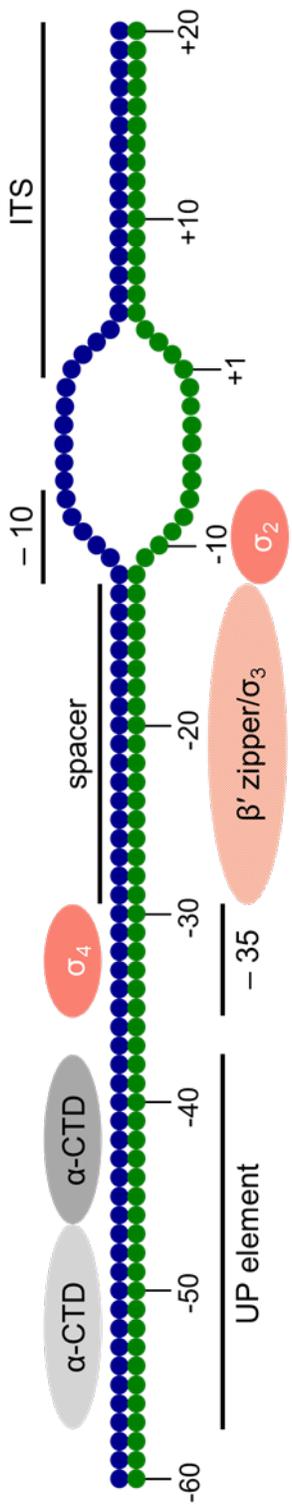


Figure 2. The breakdown of promoter elements recognized by $E\sigma^{70}$. The subunits of *E. coli* RNAP holoenzyme $E\sigma^{70}$ are colored: light gray, C-terminal domain (CTD) of αI ; dark gray, CTD of αII ; salmon, σ_2 and σ_4 ; and light salmon, β' zipper and σ_3 . The non-template strand is colored blue and the template strand green, and each bead represents a nucleotide. Different promoter regions are denoted above and below the sequences. Each α -CTD interacts with the distal and proximal UP element. The domains 4 and 2 of σ -subunit bind to the -35 and -10 elements, respectively. The 17-bp spacer is optimal for the placement of σ_4 onto -35 and σ_2 onto -10. The β' zipper region and σ_3 also form contacts with the spacer element. The double-stranded DNA is melted apart inside the active site channel formed by the subunits β and β' (not shown). Although the multiple RNAP-promoter interactions stabilize the open complex, they also ‘immobilize’ the RNAP.



In the open complex (RP_o), the +1 start site is positioned at the RNAP's active center which has two bound Mg^{2+} ions (Murakami et al., 2002a, 2002b). In the presence of ribonucleoside triphosphates (NTPs), the RNAP performs *de novo* synthesis of RNA transcripts by a general two-metal-ion mechanism (Steitz and Steitz, 1993; Vassylyev et al., 2002). At this stage, the RNAP carries out abortive initiation where it repeatedly synthesizes and releases short nascent RNAs of varying lengths (Carpousis and Gralla, 1980). Only after the nascent transcript is sufficiently long does the RNAP achieve promoter escape and transition into the elongation phase (Borukhov and Nudler, 2008). However, on some promoters that have high binding affinity to the RNAP, the promoter escape step is shown to be rate-limiting and the RNAP undergoes extensive abortive initiation repeatedly prior to escape (Vo et al., 2003).

At the crux of the abortive initiation-promoter escape problem is a promoter that has maximized its ability to bind RNA polymerase and form a highly stable open complex. Such a promoter contains sequences and features very similar to the consensus $E\sigma^{70}$ promoter made up of two highly conserved hexanucleotides—**TTGACA** (at the -35 region) and **TATAAT** (at the -10 region)—separated by a 17-bp spacer DNA (Hawley and McClure, 1983), the bold nucleotides being the most conserved (Vo et al., 2003). Although the spacer region does not exhibit sequence conservation like -35 and -10 elements, its length was shown to be critical for promoter recognition and binding. That an RNAP binds more tightly to those promoters with sequences closer to the

consensus has been confirmed with *in vitro* binding-abortive initiation assays (Mulligan et al., 1984; Brunner and Bujard, 1987) and genetic mutational analysis (reviewed in McClure, 1985). In fact, X-ray crystallographic investigations have given us a high resolution structure of an open complex containing thermophilic bacterial RNAP holoenzyme bound to a consensus promoter DNA (Murakami et al., 2002b). The open complex is stabilized through multiple σ subunit-promoter DNA interactions: σ_4 domain with the -35 element, σ_2 domain with the -10 element, and σ_3 domain with the extended -10 bases at -14/-15 in the spacer DNA. The 17-bp spacer is crucial for the optimal positioning of σ_4 /-35 and σ_2 /-10 interactions (Murakami et al., 2002a, 2002b; Vassylyev et al., 2002, Hsu, 2002a). Beyond spacer length, other more recent studies indicate that σ^{70} domain 3, the short loop between domains 2 and 3, and the β' zipper or zinc-binding domain (β' -ZBD) all interact with different parts of spacer DNA in a sequence-specific manner (Barne et al., 1997; Singh et al., 2011; Yuzenkova et al., 2011).

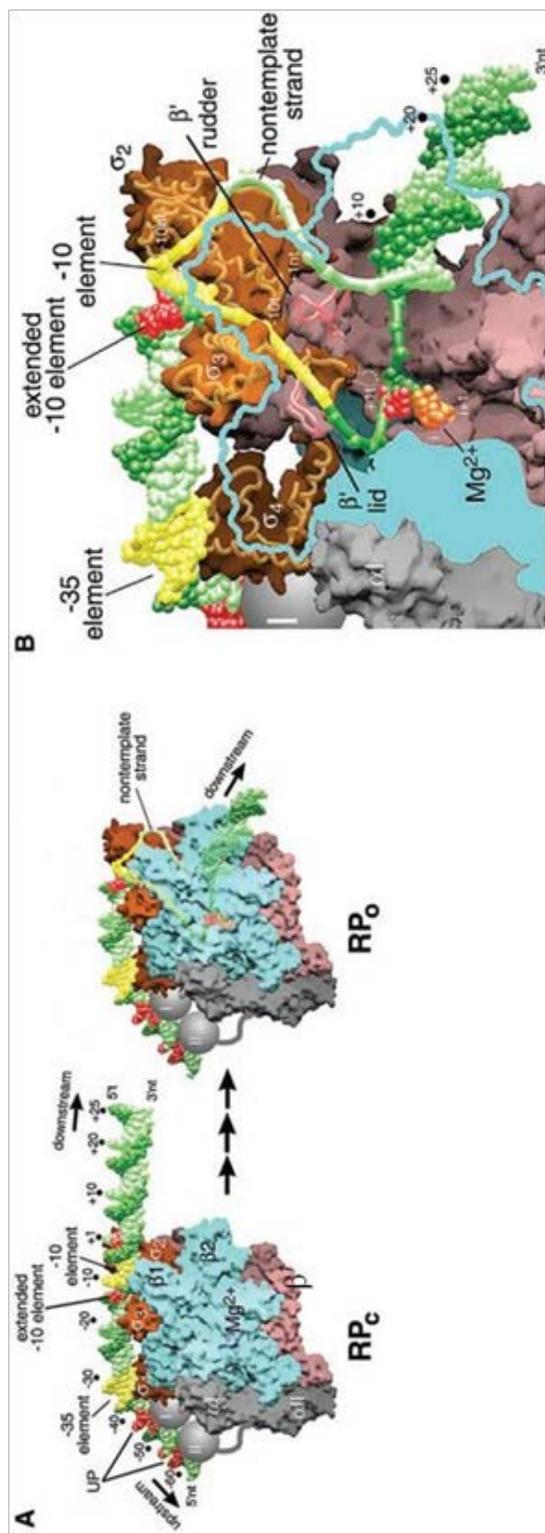
Another DNA element recognized and bound by the RNAP is the UP element (sequences further upstream of the -35 element). Studies have shown that the C-terminal domains of RNAP α -subunits (α -CTDs) interact with the AT-rich UP element in *rrnB* P1 promoter to enhance transcription (Ross et al., 1993). Since the UP element is not as ubiquitous as the -35 or -10 elements (i.e. not every promoter contains an UP element), *in vitro* selection and *in vivo* screening approaches were utilized to identify the full UP element consensus sequence: -59 nnAAAWWTWTTTTnnAAAAnnnn -38 (W = A or T and n = any nucleotide)

(Estrem et al., 1998). The full UP consensus sequence was later shown to be divisible into the distal and proximal subsites, each with its respective consensus sequence: the distal subsite, -57 AWWWWWT TTTT -47; and the proximal subsite, -46 AAAAAAR_nR -38 (R = A or G). Each is contacted by one α -CTD. In promoters containing only the proximal UP element, only one of the two α -CTDs is engaged in the open complex (Estrem et al., 1999).

After the RNAP has made these specific contacts with promoter elements or RP_c formation, σ_2 domain serves to melt apart the dsDNA at the -10 element. The recently published crystal structure of σ_2 bound to single-stranded -10 region reveals that the conserved A (-11) and T (-7) are flipped out into a protein pocket, thereby initiating open complex formation (Feklistov and Darst, 2011). The template strand is directed inside the RNAP active-site channel by the basic amino acids of $\sigma_{2,4}$ and σ_3 through the tunnel formed by parts of σ_2 , σ_3 , β_1 , β' lid, and β' rudder, forming a sharp 90° bend. Meanwhile the non-template strand is held between β_1 and β_2 , the two lobes of the β -subunit; thus the two strands of the DNA are stably melted apart for 12-13 bp inside the RNAP active site channel (Figure 3, adapted from Murakami et al., 2002a). The RP_o is catalytically active and the RNA synthesis commences if NTPs are present at this stage.

Due to the multiple RNAP-DNA interactions illustrated in Figure 3, upon initiation, the enzyme cannot extricate itself and move along the downstream DNA during transcription of the initial transcribed sequence (ITS) region. Instead, the RNAP undertakes a translocation process called DNA scrunching to transcribe

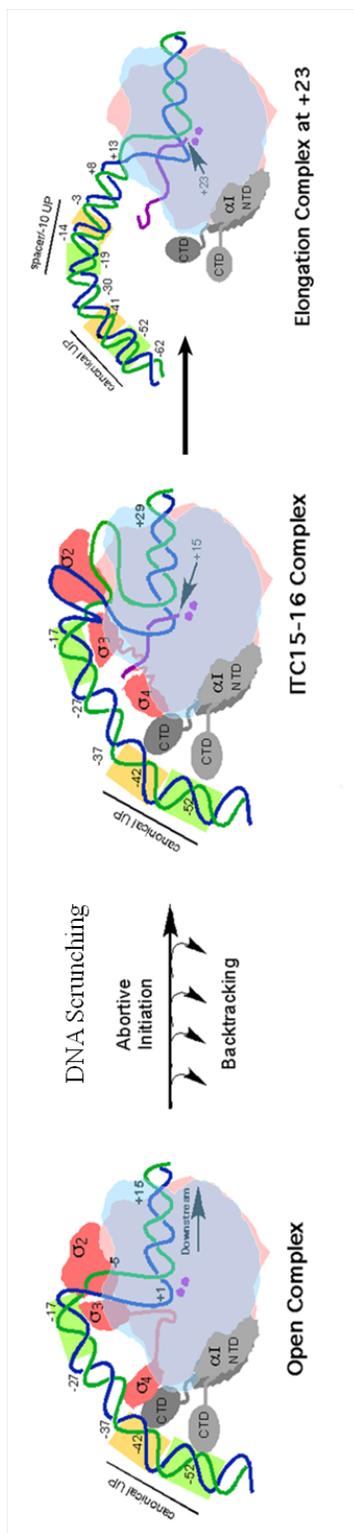
Figure 3. Models of closed complex (RP_c) and open complex (RP_o). The subunits of the RNAP were colored: gray, α I, α II and ω ; cyan, β ; pink, β' ; and orange, σ . The β -subunit was rendered transparent in (B) for visualization of the active center. The double-stranded DNA is shown as atoms and single-stranded as phosphate backbone worms with only phosphates visible as atoms. The non-template strand is light-green and template strand green. (A) After the formation of RP_c, the double-stranded DNA is melted apart at the -10 element. The template strand is directed towards the active center. (B) The close-up of the melted region of the DNA showed the non-template strand locked inside the tunnel formed by parts of σ_2 , σ_3 , β 1, β' lid, and β' rudder while the non-template is held between the two lobes of β -subunit, β 1 and β 2 (adapted from Murakami et al., 2002a). These models showed the RNAP-promoter interactions discussed in Figure 2 at the molecular level.



the downstream sequences until the promoter escape ensues (Figure 4, adapted from Hsu et al., 2006). This mechanism was elucidated by X-ray crystallography of a T7 RNAP initiating complex (Cheetham and Steitz, 1999), fluorescence resonance energy transfer (FRET) experiments (Kapanidis et al., 2006) and single-molecule DNA nanomanipulation experiments (Revyakin et al., 2006). It can also explain the observations from earlier DNA-footprinting results showing that the protected upstream boundary of the DNA in either open or transcribing complexes did not change (Carpousis and Gralla, 1985; Straney and Crothers, 1987; Krummel and Chamberlin, 1989). According to this model, the RNAP remains stationary by binding to the upstream promoter sequences while the downstream initial transcribed sequence DNA is unwound and pulled into the enzyme active site for template-dependent transcription. The stress generated from DNA-unwinding and DNA-compaction is implicated to be the force used by the RNAP to disrupt the upstream promoter contacts, subsequently the upstream edge of the open complex bubble rewinds, allowing the RNAP to transition into the elongation phase (Roberts, 2006). In general, the stronger the RNAP-promoter contacts, the more difficult it is for the RNAP to undergo promoter escape (McClure, 1980; Carpousis et al., 1982; Vo et al., 2003).

However, the accumulated energy is more often dispensed by the re-winding of the downstream DNA, causing backtracking. In this case, the nascent RNA is displaced from its RNA-DNA hybrid arrangement, the 3'-single-stranded RNA tail becomes extruded into the secondary (NTP entry) channel and the

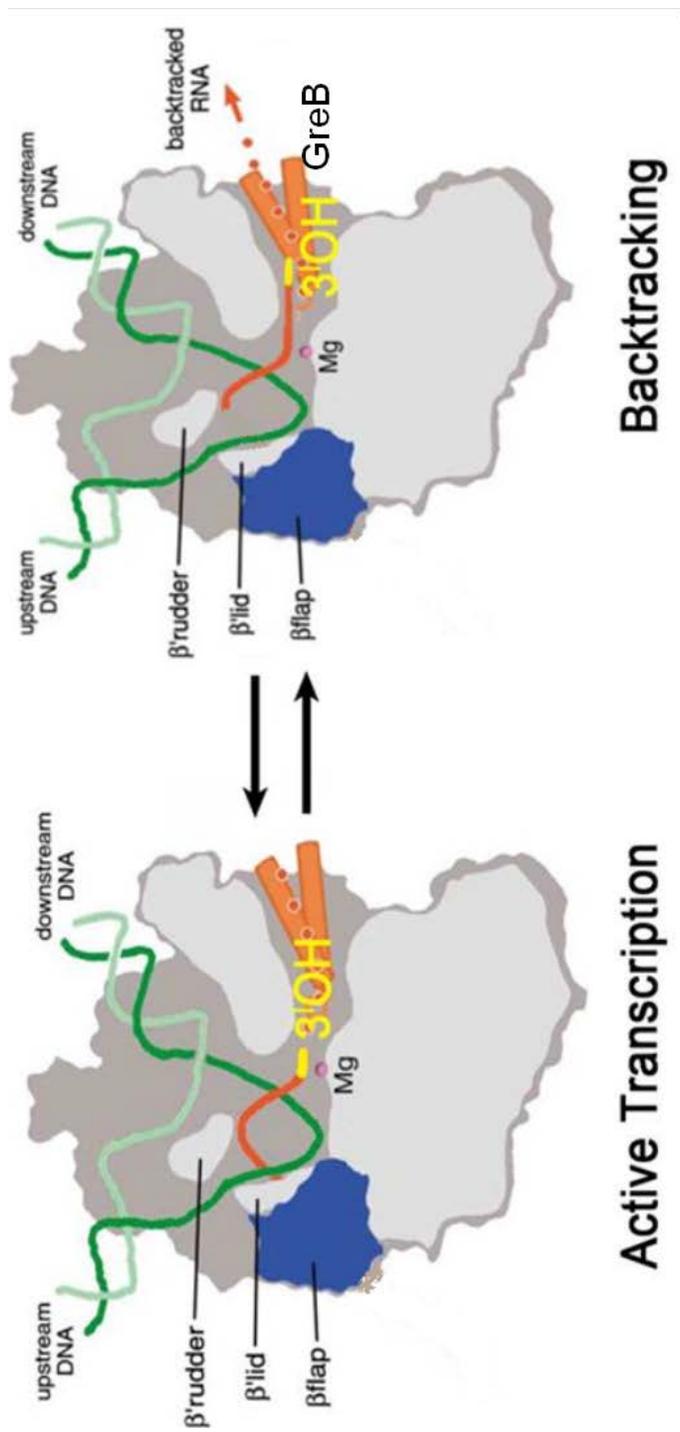
Figure 4. The role of DNA scrunching in abortive initiation and promoter escape. The ‘immobilized’ RNAP in stable open complex has to pull in the initial transcribed sequences (ITS) during the transcription initiation, which inevitably creates a stressed intermediate from DNA-unwinding and DNA-compaction. The stress generated from DNA scrunching in initial transcribing complex (ITC) can be dispensed by re-winding the upstream promoter DNA or the downstream initial transcribed sequence (ITS) DNA. In the former case, the RNAP successfully achieves promoter escape and transition into elongation phase. In the latter case, the RNAP backtracks and releases nascent RNA (abortive initiation) (adapted from Hsu et al., 2006).



RNAP will be arrested at the backtracked positions (Figure 5, adapted from Opalka et al., 2003). The half-life of a backtracked complex is dependent on the residual RNA-DNA hybrid interactions; the shorter the hybrid, the less stable is the backtracked complex, and the sooner will the backtracked transcript be released as an abortive RNA. It was shown that RNA-DNA hybrid of 4 bp represents the lower limit of complex stability (Hsu et al., 2006). Since the 3'-OH end of the nascent RNA is mis-aligned with the active center and cannot be further elongated, the release of the misaligned transcript allows the RNAP to transform back to an open complex and to catalyze RNA synthesis again (Komissarova and Kashlev, 1997; Hsu, 2002b). On the other hand, transcription factor such as GreA or GreB is shown to be able to rescue the abortive transcripts released from RNAP backtracking and increase the efficiency of promoter escape (Hsu et al., 1995). To bring about cleavage-rescue, GreB binds in the RNAP secondary channel and is positioned near the active center where it stimulates the hydrolytic activity of RNAP to cleave the backtracked RNA so that a new 3'-OH end is aligned at the active center for further elongation (Opalka et al., 2003). It has been shown that the level of abortive transcripts ≤ 15 nt diminishes and that of full-length RNAs increases when GreB is present (Chander et al., 2007).

Given these known facts about the RNAP backtracking and GreB rescue in abortive initiation, it is not surprising that the discovery of very long abortive transcripts (VLATs) whose levels increase in the presence of GreB, prompted the proposal of a novel mechanism called RNAP hyper-forward translocation

Figure 5. The GreB rescue of backtracked RNAP. When the nascent RNA is dislocated into the secondary (NTP entry) channel from RNAP backtracking, the transcription factor GreB (depicted as orange molecular scissor) stimulates the hydrolytic activity of the RNAP to cleave the nascent RNA. This cleavage positions a new 3'-OH end at the active site for the RNAP to resume the RNA synthesis. In the absence of GreB, the backtracked RNA is released as an abortive transcript (adapted from Opalka et al., 2003).



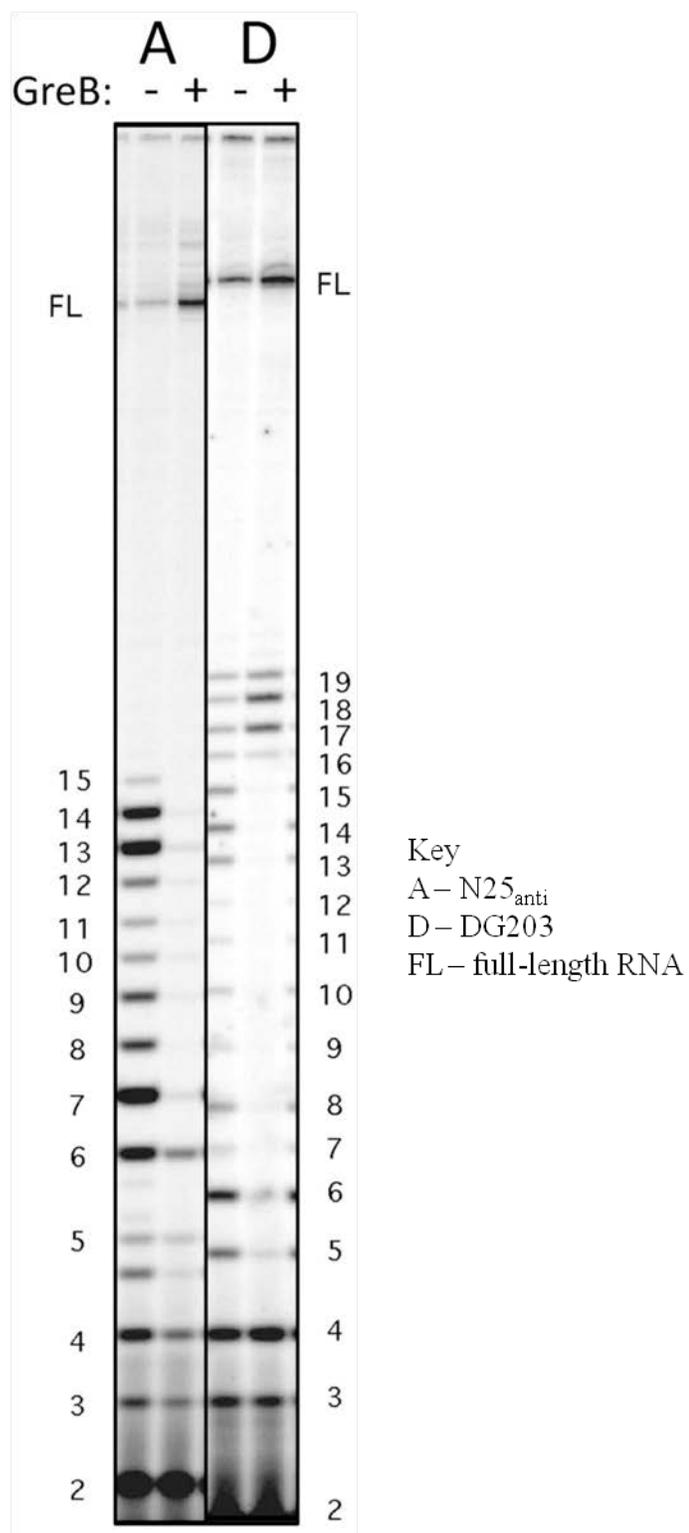
Backtracking

Active Transcription

(Chander et al., 2007). The VLATs are 16-19 nt long and are produced from a group of promoters derived from the random mutagenesis of the +3 to +10 region of the ITS of N25 promoter. The wild-type T5 phage N25 promoter is a classic example of the promoter with consensus-like promoter sequences and rate-limiting step at promoter escape – the RNAP releases 2-11 nt long abortive transcripts before escape occurs at +12. Previous studies conducted to investigate the influence of the ITS over abortive profile have shown that when the ITS from +3 to +20 of N25 is changed (A ↔ C and G ↔ T) to generate N25_{anti} promoter, the level of productive synthesis decreases 10 fold (Kammerer et al., 1986). The RNAP also undergoes a higher degree of abortive initiation and releases 2-15 nt long abortive transcripts (Hsu et al., 2003). The effect of ITS over abortive initiation is further explored by random mutagenesis of the ITS from +3 to +10 and these studies yielded the DG203 promoter on which the RNAP produces up to 19-nt long abortive transcripts (Hsu et al., 2006). On both N25_{anti} and DG203, 5-15 nt long abortive transcripts are derived from RNAP backtracking, because their levels decreased in the presence of GreB (Figure 6, adapted from Hsu, 2009). The mechanism proposed to explain the distinguishing feature of VLATs being resistant to GreB is as follows (Chander et al., 2007).

N25 and DG203 promoters share identical promoter recognition elements; thus, both promoters can bind RNAP and form open complexes equally well. However, they differ in their ITS region from +3 to +10, changing the highly AT-rich sequence in N25 to the highly GC-rich sequence in DG203. (The first 10

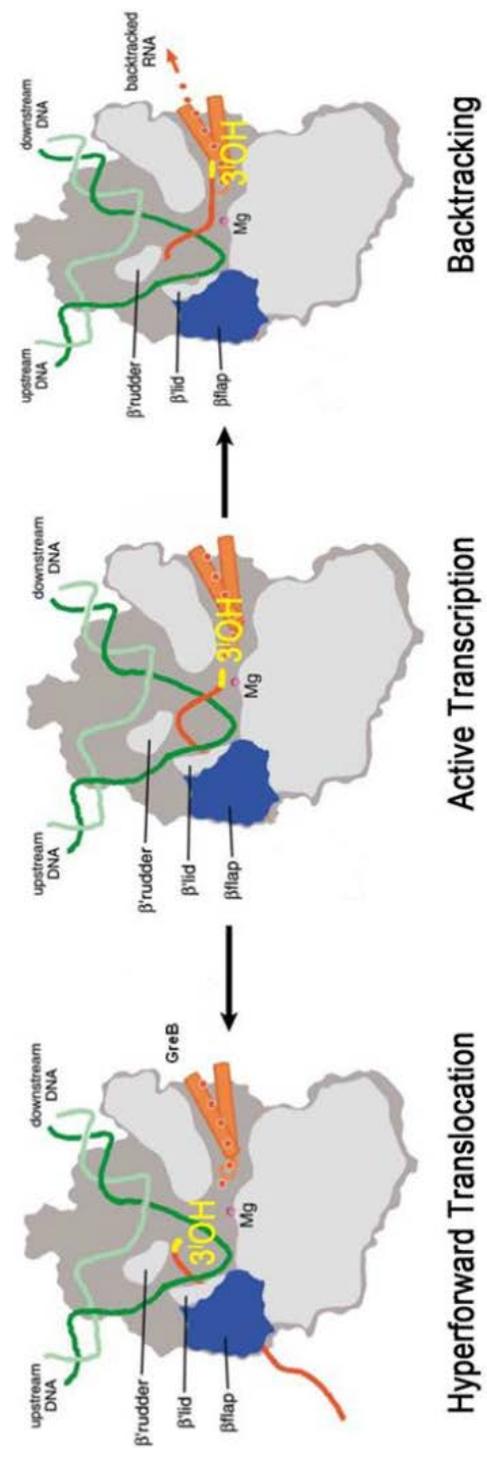
Figure 6. The GreB-resistant nature of very long abortive transcripts (VLATs). During transcription of N25_{anti} (A) and DG203 (D) promoters, the presence of GreB diminishes the level of ≤ 15 nt long abortive transcripts and increases the level of full-length transcripts. However, 16-19 nt long transcripts released from DG203 promoter are resistant to GreB rescue. These VLATs are proposed to be released from hyperforward translocation of RNAP (Chander et al., 2007) (adapted from Hsu, 2009).



bases in N25 ITS are ATAAATTTGA, and in DG203, ATGCGACCGG.) This sequence change dramatically shifts the position of escape to further downstream. For N25, escape occurs at +12, after the RNAP has abortively initiated and produced a ladder of abortive RNAs ranging from 2-11 nt. For DG203, escape is not completed until +20; prior to escape, the RNAP has generated an abortive ladder of 2-19 nt. This means that RNAP has to scrunch in nearly twice as many bp of DNA on the DG203 promoter than on the N25 promoter to reach the cusp of escape, generating that much more stress in the DG203 complex. Thus, when this stress energy is used to disrupt the polymerase-DNA interactions and allow RNAP to move downstream by forward translocation, the larger amount of stress energy released by the DG203 complex will bring about hyper forward translocation. That is, the polymerase will skip forward a few nucleotides. This creates an undesirable consequence where the 3'-OH end of the nascent RNA is shoved upstream past the active center and now situated in the RNA exit channel where it cannot be rescued by GreB-stimulated cleavage activity. As a result, these transcripts, 16-19 nt in length and GreB-resistant, are released out of the RNA exit channel (Figure 7, adapted from Opalka et al., 2003).

This research aims to provide evidence to support this mechanism using two approaches: roadblock transcription and permanganate footprinting. The first technique, roadblock transcription, was utilized to stall the forward (downstream) movement of RNAP, thereby examining whether the synthesis of GreB-resistant VLATs requires RNAP forward movement, and whether the forward

Figure 7. The postulated model of RNAP hyperforward translocation. To explain the GreB-resistant nature of VLATs produced on DG203 promoter, Chander et al. (2007) proposed that the RNAP was propelled forward by more than one nucleotide due to the excessive DNA scrunching required in DG203 to release the promoter contacts. On the verge of promoter escape, the upstream DNA re-wound with such force that the RNAP hyperforward translocated, thereby leaving the 3'-OH end of the nascent RNA upstream of the active center. Unlike in the backtracking complex, GreB could not rescue this mis-located RNA and thus it is released as an abortive transcript (adapted from Opalka et al., 2003).



Postulated model

Backtracking

Active Transcription

Hyperforward Translocation

translocation occurs by more than one nucleotide. The roadblock protein used was *EcoRI* protein containing a glutamate to glutamine mutation at the 111th residue position (E₁₁₁Q *EcoRI*). King et al. (1989) showed that the mutant protein can bind specifically at the *EcoRI* recognition site (GAATTC) without cleaving the DNA. Pavco and Steege (1990) later reported that the binding affinity is strong enough to halt the forward movement of an *E. coli* RNAP. This happens, during the elongation phase, when the front edge of the RNAP, which is ~14 bp downstream of the active center, bumps into the stationary *EcoRI* dimer (Pavco and Steege, 1990). The dimer covers two nucleotides beyond the first nucleotide of the *EcoRI* binding site (McClarín et al., 1986). To investigate the RNAP hyper-forward translocation mechanism, *EcoRI* binding sites were introduced at +23, +28, +32, +35, +38, +41, +43 and +48 (the numbering refers to the first base G of the *EcoRI* binding site) on DG203 promoter, monitoring the forward movement of RNAP as it releases VLATs (Jiang, 2009; Vallery, 2010). Based on their results, more binding sites were introduced at +33, +34, +36 and +37. The *EcoRI* binding site-containing DG203 templates are named EQ-xx templates. All of the EQ-series templates were used in this study to pinpoint the cusp when RNAP hyper-forward translocation occurs.

The second approach, permanganate footprinting, was used to investigate whether excessive DNA scrunching must occur to provide the stress energy necessary to propel the RNAP to jump/skip/slide forward by more than one nucleotide. Hayatsu and Iida (1969) showed that permanganate can be used to

specifically modify single-stranded thymine residues in DNA. Thus, using potassium permanganate as a footprinting reagent, the extent of DNA bubble expansion can be inferred from the thymine residues detected as being single-stranded during transcription (Figure 8). To better probe the initial transcribing complex bubbles, a derivative of DG203 promoter called SP_{fullcon} was also included. The only difference between DG203 and SP_{fullcon} is the spacer sequence. The 17-bp sequence in DG203 contains an UP-like element that is contacted by the α -CTD of RNAP during transcriptional progression; this interaction gives rise to more GreB-susceptible VLATs than GreB-resistant ones from DG203 (Chander and Hsu, manuscript in submission; Lee and Hsu, unpublished results). The SP_{fullcon} spacer sequence is one of many associated with the *in vitro* selected consensus promoter (Gaal et al., 2001); it contains 12 centrally located As and Ts and was the spacer sequence used for crystallizing the promoter DNA-holoenzyme open complex structure (Murakami et al., 2002b). This 17-bp substitution greatly diminished the production of full-length transcripts on SP_{fullcon} to an undetectable level (Chander and Hsu, manuscript in submission). Thus, the DNA bubbles captured by permanganate footprinting on this promoter would be mostly associated with initial transcribing complexes, not elongation complexes. The permanganate footprinting was carried out on four promoters, N25, N25_{anti}, DG203 and SP_{fullcon}, from both non-template and template strands by differential radiolabeling. For easy reference, the sequences of the promoters of interest (N25, N25_{anti}, DG203 and SP_{fullcon}) are compared in Figure 9.

Figure 8. The reaction mechanism and procedure of permanganate

footprinting. The permanganate's ability to differentially modify single-stranded thymine residues (T's) and spare those in double-stranded form is utilized to capture the expansion of the DNA melted region in transcribing complexes. The footprinting experiments were carried out on single-end labeled DNA. In this reaction scheme, the DNA was incubated with the RNAP to allow the formation of open complex bubble. The addition of NTP then prompted the RNAP to start transcribing and scrunching in the downstream DNA. KMnO_4 treatment oxidized the single-stranded T's and the DNA was cleaved at modified sites by treatment with weak base, piperidine. Analysis and comparison of the fragments of labeled DNA from open complex and transcribing complex on the sequencing gel allowed the visualization of the transcribing complex bubble expansion.

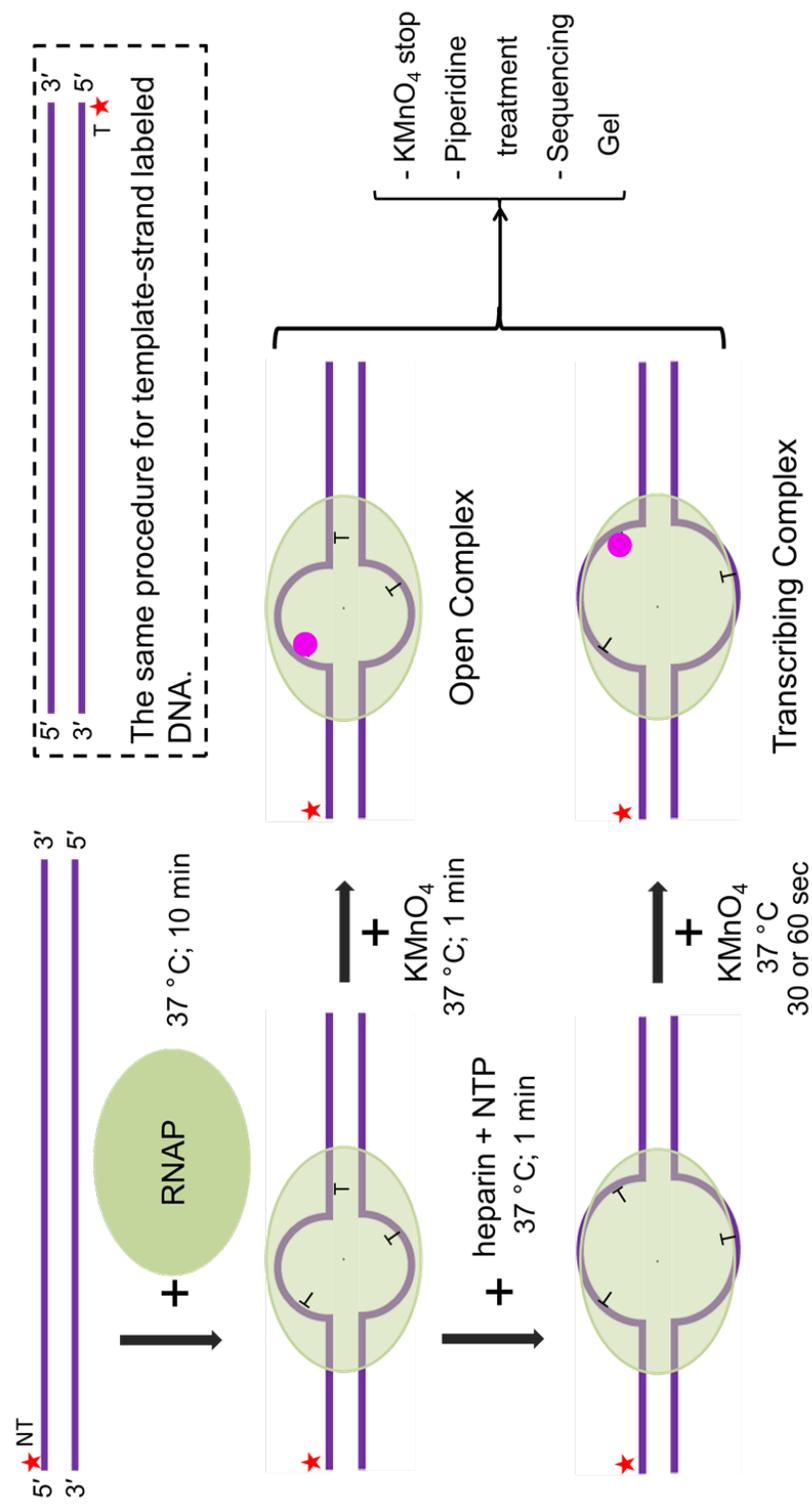


Figure 9. Comparison of the four footprinting promoters, N25, N25_{anti}, DG203 and SP_{fullcon}. The N25 promoter is one of the wild-type T5 phage promoters and the N25_{anti} and DG203 are initial transcribed sequence (ITS)-variants of N25. SP_{fullcon} is a derivative of DG203 with changes in spacer region. The -35 and -10 promoter elements are shown in blue and the differences in the sequences among the four promoters are shown in magenta. Due to these changes, the abortive and productive initiation patterns were altered (see Figure 14).

	-35		-10	+1	+20
N25 (N)	TTTAT	TTGCT	CTGTA	GATTC	TTGA
N25 _{anti} (A)	TTTAT	TTGCT	TAATA	GATTC	GAGAG
DG203 (D)	TTTAT	TTGCT	TAATA	GATTC	GAGAG
SPfullcon (S)	TTTAT	TTGCT	TAATA	GATTC	GAGAG

MATERIALS AND METHODS

MATERIALS

DNA primers and plasmids

The oligonucleotides used as DNA primers to prepare the DNA templates for roadblock transcription reactions (EQ-primers) and permanganate footprinting reactions (VL-primers and SAN-primers) were purchased from Integrated DNA Technologies, Inc. (IDT). Their sequences are listed in Tables 1-3.

The EQ-primers (Table 1) were utilized to construct the EQ-series of promoters, each containing an *EcoRI* site in the transcribed region, for delineating the mechanism of very long abortive transcript (VLAT) formation by including E₁₁₁Q *EcoRI* roadblock during transcription (Jiang, 2009; Vallery, 2010).

The VL-primers (Table 2) were used to prepare single end-labeled DNA templates by primer extension for footprinting the initial transcribing complexes of N25, N25_{anti}, DG203 and SP_{fullcon}. However, due to sub-optimal quality of the DNA prepared by this method, the VL templates were instead cloned into pSA508 vector (Choy and Adhya, 1993) so that the labeling procedure can be initiated from plasmid DNA (described later). Only the DG203 and SP_{fullcon} promoters were successfully cloned by T. K. Vallery '10.

The SAN-primers (Table 3) were designed for modifying earlier versions of the N25 and N25_{anti} promoters in pSA508 vector – pSAN1 and pSAN2

Table 1. The upstream (u) and downstream (d) EQ-primer sequences used in generating EQ-series DNA, reading 5' to 3'. The upstream and downstream primers had 16 bp overlapped from -5 to +11. The -35 and -10 promoter elements and the +1 start site are indicated in italics and bold, respectively. The *EcoRI* binding site is underlined. DG203 promoter served as a control without *EcoRI* binding site.

Primer	Sequence
DG203-u (XMB)	(-85) GGCTC GAGCG ATTCC CGGGG ATCCG TCGAG GGAAA TCATA AAAAA TTTAT TTGCT TTCAG GAAAA TTTTT CTGTA TAATA GATTC ATGCG ACCGG G (+11)
DG203-d (PH)	(+60) CCGTT CTGCA GAAGC TTTCT GCGAG AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-23-d	(+60) CCGTT CTGCA GAAGC TTTCT GCGAG AACCA GCGAA TTCTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-28-d	(+60) CCGTT CTGCA GAAGC TTTCT GCGAG AAGAA TTCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-32-d	(+60) CCGTT CTGCA GAAGC TTTCT GCGGA ATTCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-33-d	(+60) CCGTT CTGCA GAAGC TTTCT GCGAA TTCCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-34-d	(+60) CCGTT CTGCA GAAGC TTTCT GGAAT TCCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-35-d	(+60) CCGTT CTGCA GAAGC TTTCT GAATT CACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-36-d	(+60) CCGTT CTGCA GAAGC TTTCTG AATTC AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-37-d	(+60) CCGTT CTGCA GAAGC TTTGA ATTCT AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-38-d	(+60) CCGTT CTGCA GAAGC TTGAA TTCAG AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-41-d	(+60) CCGTT CTGCA GAAGG AATTC GCGAG AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)
EQ-43-d	(+60) CCGTT CTGCA GAGAA TTCCT GCGAG AACCA GCCAT ATTTA AACTC CTCTC CCGGT CGCAT GAATC (-5)

Table 2. The upstream (u) and downstream (d) sequences of VL-primers used in generating DNA for single end-labeling, reading 5' to 3'. The VL1-u served as the upstream primer for N25, N25_{anti} and DG203 promoters and VL2-u for SP_{fullcon}. The downstream primers VL3-d, VL4-d, VL5-d and VL6-d were for N25, N25_{anti}, DG203 and SP_{fullcon}, respectively. The primers had 26 bp overlapped from -1 to -26.

Primer	Sequence
VL1-u	(-100) GACTA GCTCG AGCCT CCGTC GAGGA ATTGC CGGG ATTTC TCGAG GGAAA TCATA AAAAA TTTAT TTGCT TTCAG GAAAA TTTT CTGTA TAAATA GATTTC (-1)
VL2-u	(-100) GACTA GCTCG AGCCT CCGTC GAGGA ATTGC CGGG ATTTC TCGAG GGAAA TCATA AAAAA TTTAT TTGCT TAAAG TGTTA AATTG TGCTA TAAATA GATTTC (-1)
VL3-d	(+74) GGTTT TAAGC TTGGT ACCCC TGCAC TTTCT TTCTG CGAGA ACCAG CCATA TTATA ACTCC TCTCT CAAAT TTATG AATCT ATTAT ACAGA AAAAT TTTCC (-26)
VL4-d	(+74) GGTTT TAAGC TTGGT ACCCC TGCAC TTTCT TTCTG CGAGA ACCAG CCATA TTATC CGGA AGAGG ATTCC GGATG AATCT ATTAT ACAGA AAAAT TTTCC (-26)
VL5-d	(+74) GGTTT TAAGC TTGGT ACCCC TGCAC TTTCT TTCTG CGAGA ACCAG CCATA TTATA ACTCC TCTCC CGGTC GCATG AATCT ATTAT ACAGA AAAAT TTTCC (-26)
VL6-d	(+74) GGTTT TAAGC TTGGT ACCCC TGCAC TTTCT TTCTG CGAGA ACCAG CCATA TTATA ACTCC TCTCC CGGTC GCATG AATCT ATTAT AGCAC AATTT AACAC (-26)

Table 3. The non-template (NT) and template (T) sequences of SAN-primers, reading 5' to 3'. The SAN-primers were used to modify pSAN1 and pSAN2 plasmids.

Primer	Sequence
SAN1a-NT	(-97) TCGAG CTCGC CTCCG TCGAG GAATT GCCGG GGATT CG (-61)
SAN1a-T	(-57) TCGAC GAAATC CCCGG CAAATT CCTCG ACGGA GGCGA GC (-93)
SAN1b-NT	(+55) CCCGG ATATC CTGCA (+69)
SAN1b-T	(+65) GGATA TCCGG GTGCA (+51)
SAN1b1	(-103) GGTAC CTCGA GCTCG CCTCC GTCGA GG (-77)
SAN1b2	(+49) GATTT CCTC GACGA ATCCC CGGCA ATTCC TCGAC GGAGG CGAGC (-93)
SAN1b3	(-67) GGATT CGTCG AGGGA AATCA TAAAA AATTT ATTTG CTTTC AGGAA AATTT TTCTG (-13)
SAN1b4	(+30) GCCAT ATTTA AACTC CTCTC TCAAA TTTAT GAATC TATTA TACAG AAAAA TTTTC CTG (-28)
SAN1b5	(+11) GAGAG GAGTT TAAAT ATGGC TGGTT CTCGC AGAAA GCTTC TGCAC CCGG (+59)
SAN1b6	(+73) TTGCT GCAGG ATATC CGGGT GCAGA AGA (+46)

plasmids – to create the same-length promoters (final 158 bp) as DG203 and SP_{fullcon} for footprinting comparison.

Proteins and Enzymes

The His₆-tagged RNAP was prepared by Dr. Lilian M. Hsu from *E. coli* RL721 strain (from Dr. Robert Landick, University of Wisconsin-Madison) grown at U. C. Berkeley fermentor facility using the procedure of Burgess and Jendrisak (1975) modified by Dr. Susan Uptain. The RNAP contained 50% active molecules at the time of use (Chamberlin et al., 1979). *E. coli* GreB protein was isolated by L. M. Hsu from IPTG-induced JM109 strain containing the plasmid pGF296 (Feng et al., 1994). Fresh E₁₁₁Q EcoRI protein was prepared from IPTG-induced T7 Express Competent *E. coli* (High Efficiency) cells (NEB) containing the plasmid pVS9 (a gift from Dr. Irina Artsimovitch, Ohio State University). The pVS9 plasmid was created by cloning the modified *EcoRI* gene (PCR amplified from the genomic clone from Modrich (King et al., 1989) into pET33 expression vector (Novagen) containing the kanamycin resistance gene. As expressed, the encoded E₁₁₁Q EcoRI protein is tagged N-terminally with His₆. All other enzymes used were purchased from New England BioLabs[®] Inc. (NEB); these include Klenow polymerase, T4 polynucleotide kinase (T4 PNK), Antarctic phosphatase (AP), EcoRV, Eco53kI, Acc65I, XhoI.

Reagents

The reagents were obtained from commercial sources. Ultrapure grade of common laboratory reagents obtained from US Biochemicals/Affymetrix include: isopropyl β -D-1-thiogalactopyranoside (IPTG), acrylamide, bis-acrylamide, ammonium persulfate, Trisma Base, boric acid, urea, ethylenediaminetetraacetate disodium salt (Na_2EDTA), 0.5 M EDTA (pH 8) solution, β -mercaptoethanol (β -ME), dithiothreitol (DTT), cacodylic acid, phenol-chloroform-isoamyl alcohol (25:24:1) mixture, ribonucleoside triphosphates (NTPs), deoxyribonucleoside triphosphates (dNTPs) and lysozyme (source: chicken egg white). Other molecular biology grade reagents purchased from Fisher Scientific include: ampicillin sodium salt, tetramethylethylenediamine (TEMED), sodium dodecyl sulfate (SDS), Elutip-D, formamide, xylene cyanol (XC) and Tween[®]-20. The reagents for Maxam-Gilbert sequencing and permanganate footprinting such as formic acid, potassium permanganate, heparin and piperidine were purchased from Sigma-Aldrich[®] (Sigma). Also obtained from Sigma were diethylpyrocarbonate (DEPC), kanamycin, agarose, bromophenol blue (BPB), amaranth, mixed bed resin beads, ethidium bromide and Triton[™] X-100. The reagents from NEB include: ColorPlus prestained protein ladder and bovine serum albumin or BSA (10 mg/mL stock). The Ni-NTA agarose resin was obtained from QIAGEN. The acetylated BSA was provided by S. Uptain. The cOmplete EDTA-free Protease Inhibitor Cocktail Tablets were obtained from Roche Applied Science. The glycogen obtained from Sigma was made nuclease-

free by repeated (~10x) phenol (H₂O-saturated; from USB/Affymetrix) extraction, followed by ethanol precipitation (prepared by L. M. Hsu).

METHODS

Roadblock Transcription Reactions

EQ-series DNA construction

The linear DG203 promoter DNA templates each containing an *EcoRI* binding site at distinct downstream positions were generated by primer extension. Briefly, the lyophilized upstream and downstream EQ-primers (Table 1) were re-suspended in 1x TE buffer (10 mM Tris-HCl, 1 mM Na₂EDTA, pH 8.0) to a working stock concentration of 10 μM. The overlapping pairs of upstream and downstream primers were annealed by incubating the mixture at 70 °C for 20 min, and 55 °C, 42 °C and 37 °C for 10 min each. The 50 μL mixture contained 2 μM upstream primer, 2 μM downstream primer and 1x NEBuffer 2 (50 mM NaCl, 10 mM Tris-HCl, pH 7.9, 10 mM MgCl₂, 1 mM DTT). After annealing, 50 μL of dNTP mixture (250 μM dNTP and 1x NEBuffer 2) was added to the previous mixture. Then 3 μL of DNA polymerase I Klenow fragment (NEB; 5 units/μL) was added to the reaction mixture which was incubated at 37 °C for 30 min. The enzyme was heat inactivated at 75 °C for 20 min. Each reaction was conducted in two sets which were then combined for later steps. The presence of the 145-bp

long DNA templates (spanning -85 to +60) was ascertained by agarose gel electrophoresis and UV detection.

Purification of EQ-series DNA

The primer-extended promoters were purified by two steps of ethanol (EtOH) precipitation, first from 0.3 M sodium acetate (NaAc) solution, and next from 5 M ammonium acetate (NH₄Ac) solution. In NaAc-EtOH precipitation, the samples were mixed thoroughly with one-tenth the reaction volume of 3 M NaAc. Then three times the total volume of 95% EtOH was added, mixed well and stored overnight at -20 °C. The solutions were centrifuged (13,400 rpm) for 30 min at 4 °C and the supernatants removed. The pellets dried in a Speed Vac for 15 min were re-dissolved in 100 µL TE into which the same volume of 10 M NH₄Ac was added and mixed thoroughly before three times the total volume of 95% EtOH was added and mixed well. The mixtures were incubated at room temperature for 30 min to 2 hr. They were then centrifuged (13,400 rpm) for 30 min at room temperature. The pellets were recovered by carefully removing the supernatant and drying in the Speed Vac. They were re-dissolved in 200 µL TE and the promoters were further purified by two rounds of phenol-chloroform-isoamyl alcohol (25:24:1) extraction followed by one round of chloroform-isoamyl alcohol (24:1) extraction. Each extraction involved mixing an equal volume of the DNA solution with the organic solvent, centrifuged briefly to separate the phases and the upper aqueous phase was recovered and transferred to

a new tube. The final aqueous phase was set up for another round of EtOH precipitation with 0.3 M NaAc. The final DNA pellets collected by microcentrifugation were washed with 500 μ L of 70% EtOH before being dried and re-dissolved in 100 μ L TE. The concentrations of the samples were determined by absorbance at 260 nm using NanoDrop™ ND-2000c Spectrophotometer (Thermo Scientific, Inc.) and diluted to 300 nM for use in transcription assays.

Purification of roadblock protein (E₁₁₁Q EcoRI)

To prepare fresh E₁₁₁Q EcoRI protein, pVS9 plasmid was transformed into T7 Express Competent *E. coli* (High Efficiency) cells from NEB following the recommended protocol. The transformed cells were spread on a 1.5% agar plate of L broth containing 30 μ g/mL kanamycin (LB/Kan plate) and incubated overnight at 37 °C. A single colony was picked the following day and inoculated into 5 mL LB with 30 μ g/mL kanamycin (LB/Kan) and incubated overnight in 37 °C shaker. In order to test isopropyl β -D-thiogalactopyranoside (IPTG) induction, 100 μ L of the overnight culture was inoculated into a fresh 10 mL LB/Kan and allowed to grow in 37 °C shaker until OD₆₀₀ reached ~0.6. From this mid-log culture, 1 mL of 40% glycerol stock was prepared, and another 1 mL of the culture was set aside as the uninduced control. To the rest of the culture, IPTG was added to 1 mM in final concentration, the cells were allowed to grow for another 3 hours under the same condition, and then 1 mL of the culture was

collected. The OD_{600} was measured at this point for calculating the amount of cracking buffer to be used to lyse the cells. Both pre- and post-induction cells were re-suspended to 25 OD_{600} /mL in cracking buffer (10 mM Tris-HCl, pH 6.8, 141 mM β -mercaptoethanol (β -ME), 1% SDS and 6 M urea). From each sample, 20 μ L was mixed with the same volume of 2x sample buffer (0.1 M Tris-HCl, pH 6.8, 2% SDS, 0.28 M β -ME, 20% glycerol and 0.5% bromophenol blue (BPB)) and heated at 95 °C for 5 min before being loaded into a 4% stacking-8% resolving SDS-polyacrylamide gel electrophoresis (PAGE) gel. (The compositions of 4% stacking are 4% acrylamide-bisacrylamide (37.5:1), 0.125 M Tris-HCl (pH 6.8), 0.1% SDS, and 8% resolving gel solutions 8% acrylamide-bisacrylamide (37.5:1), 0.375 M Tris-HCl (pH 8.8), 0.1% SDS, respectively. One-hundredth solution volume of 10% APS and one-thousandth solution volume of TEMED was added to polymerize the gel.) The induced protein expression was confirmed by the presence of an EcoRI-sized band around 28.5 kDa (referenced to ColorPlus Prestained Protein Ladder, Broad Range (10-230 kDa)) in post-induction sample.

The protein was then purified, according to the procedure provided by Dr. Artsimovitch, from a 500-mL culture induced with IPTG in the same manner. Briefly, the cell pellet was collected by centrifugation ($6,000 \times g$) at 4 °C for 10 min after the culture had been chilled on ice for 15 min. It was re-suspended in 25 mL lysis buffer (50 mM Tris-HCl, pH 7.5, 1.5 M NaCl, 5% glycerol and 1 mM β -ME) supplemented with 1 mg/mL lysozyme, 0.1% Tween-20 and Complete

EDTA-free Protease Inhibitors (Roche) per manufacturer's instructions. The suspension was incubated on ice for 30 min and the cells were homogenized with a blender at 4 °C. (Previous attempts at lysis by sonication did not work out well.) The supernatant was collected after centrifugation ($27,000 \times g$) of the homogenized cells at 4 °C for 15 min and its volume noted for further steps (i.e. reference volume ~31 mL). The supernatant was incubated with 0.1 volume of 50% Ni-NTA agarose (Qiagen) slurry in lysis buffer for 30 min at 8 °C with agitation. The suspension was poured into a disposable plastic column (a 5-mL syringe barrel plugged with glass wool) and drained by gravity. The column was washed with 0.5 volume of lysis buffer and 0.5 volume of wash buffer (50 mM Tris-HCl, pH 7.5, 0.5 M NaCl, 5% glycerol and 1 mM β -ME). The protein was eluted from the column with 1-mL aliquots of 0.1 volume each of elution buffer (50 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% glycerol, 1 mM β -ME) containing different concentrations of imidazole (10 mM, 20 mM, 50 mM, 100 mM and 200 mM). The flow rate was adjusted to be about a drop in 3 sec and 1-mL aliquots were collected manually. The protein concentrations of the aliquots were measured by absorbance at 280 nm using NanoDrop™ Spectrophotometer and the ones with high concentrations were checked using SDS-PAGE. Given the relative purity of the samples containing E₁₁₁Q EcoRI, they were not further purified using heparin-sepharose column, but were dialyzed in 500 volumes of storage buffer (20 mM Tris-HCl, pH 7.5, 300 mM KCl, 50% glycerol, 0.2 mM DTT, 1 mM EDTA). The samples were checked again for concentration and the one with the highest

concentration (22.4 μM) was stored at $-20\text{ }^{\circ}\text{C}$ and used in roadblock transcription assays. The rest were stored at $-80\text{ }^{\circ}\text{C}$.

Transcription assays

The *in vitro* transcription reactions were performed with the EQ-series DNA templates. The reaction conditions were devised by Jiang (2009) and Vallery (2010) to minimize residual EcoRI cleavage activity and achieve optimal binding of the roadblock protein (E₁₁₁Q EcoRI). Four sets of reaction were conducted for each promoter template – with RNAP only, with RNAP and EcoRI, with RNAP-GreB and EcoRI, and with RNAP-GreB. All the enzymes (RNAP, E₁₁₁Q EcoRI and RNAP-GreB) were diluted with the same enzyme diluent (10 mM Tris-HCl, pH 8.0, 10 mM β -ME, 10 mM KCl, 5% glycerol, 0.1 mM EDTA, 0.4 mg/mL bovine serum albumin (BSA), 0.1% v/v Triton X-100). The 10 μL reaction contained 30 nM DNA, 100 or 200 mM KCl, 1x new transcription buffer (50 mM Tris-HCl, pH 7.2, 5 mM MgCl_2 , 10 mM β -ME and 10 $\mu\text{g}/\text{mL}$ acetylated BSA), enzymes that were 30 nM or 60 nM RNAP with or without GreB (GreB:RNAP = 10:1) and with or without EcoRI (EcoRI:DNA = 30:1), 100 μM NTP mixture containing ~ 10 cpm/fmol of [γ -³²P] ATP (Perkin Elmer), all of which were mixed in diethylpyrocarbonate-treated water (DEPC-H₂O). Among the various conditions tested, the experimental setup described below yielded the most analyzable data shown in the results section.

The reaction was set up and proceeded in the following sequence: first, the 30 nM DNA/1x new transcription buffer/200 mM KCl and the 100 μ M NTP (labeled and unlabeled) mixtures were separately prepared; second, the enzymes were diluted; third, the RNAP (RNAP:DNA = 1:1) or RNAP-GreB (GreB:RNAP = 10:1) was added to the DNA mixture and incubated at 37 °C for 3 min to form the open complex; fourth, the EcoRI (EcoRI:DNA = 30:1) or DEPC-H₂O (for minus-EcoRI control) was added and incubated at 37 °C for another 3 min to allow binding, and finally, the NTP mixture was added to initiate transcription and the reaction proceeded for 10 min at 37 °C. The reaction was terminated by the addition of 100 μ L of glycogen-EDTA-sodium acetate or GES mix (3 M NaAc, 0.1 M EDTA and 10 mg/mL glycogen in DEPC-H₂O). Three times the total volume of 95% EtOH was added and the RNA was allowed to precipitate overnight at -20 °C. The pellets were collected by centrifugation (13,400 rpm) at 4 °C for 15 min, thorough removal of the supernatant and drying in the Speed Vac for 10 min. They were re-suspended in 10 μ L of formamide loading buffer or FLB (1x TBE (89 mM Tris, 89 mM boric acid, 2.5 mM EDTA, pH 8.3), 10 mM EDTA (pH 8.0), 0.08% xylene cyanol and 0.08% amaranth in de-ionized formamide solution). To fractionate the RNA products by size, a 4- μ L aliquot of each sample was loaded to a 23% (10:1) polyacrylamide 7 M urea gel submerged in 1x TBE in the top reservoir and 1x TBE/0.3 M NaAc at the bottom. The electrophoresis was carried out at 35 W for ~4 hours or until the amaranth dye is ~1 cm from the bottom edge. The gel exposed overnight to Storage Phosphor

Screen was scanned by Storm 820 Phosphorimager (Amersham Biosciences) the following day. The results were quantified by the software ImageQuant 5.2. Each experiment was completed three times to ensure reproducibility. The control reactions without enzymes were also carried out.

Permanganate Footprinting Reactions

Different approaches to prepare single end-labeled promoter templates for footprinting

To obtain single 5'-end radiolabeled promoter templates (N25, N25_{anti}, DG203 and SP_{fullcon}), the non-template or template strand VL-primer was first phosphorylated with T4 polynucleotide kinase and [γ -³²P] ATP before the overlapping pair was used for constructing a DNA template by primer extension. The resulting ³²P-labeled DNA fragments were purified either by using Elutip®-d minicolumns (Whatman) or by gel extraction (described later). However, when DNA prepared by the above approaches was used in footprinting reactions, DNA damage or strand breakage was evident, making these DNA undesirable as starting material for footprinting analysis. At the suggestion of Dr. Wilma Ross (University of Wisconsin-Madison), these promoters were cloned into the pSA508 vector and radiolabeled by the plasmid labeling approach (described later) which finally yielded pure starting DNA for footprinting studies.

In terms of generating the pSA508 plasmid clones of the four promoter fragments, DG203 and SP_{fullcon} were cloned by T. K. Vallery '10 whereas direct cloning of N25 and N25_{anti} fragments was unsuccessful repeatedly. Therefore, pSAN1 and pSAN2 plasmids—respectively earlier versions of the N25 and N25_{anti} promoters cloned in pSA508 vector—were modified to create equivalent constructs such that plasmid labeling would yield single end-labeled DNA of the same length (158 bp) from all four templates. (The existing clones would yield labeled DNA of 106 bp, which was not ideal for direct comparison of the four promoters). First, the longer insert (37 bp), from primers SAN1a-NT and SAN1a-T annealing, was cloned in at *XhoI* restriction site (C/TCGAG) of each plasmid. Next, the shorter one (15 bp), from SAN1b-NT and SAN1b-T annealing, was cloned in at the *PstI* restriction site (CTGCA/G). The pSA508 plasmid containing N25_{anti} promoter was successfully generated by this approach (described later). After several repeated modification attempts for N25 promoter failed, it was decided instead to generate a vector from N25_{anti}-containing plasmid by *XhoI*+*PstI* digestion and clone in a single-piece N25 promoter which was first produced by one-pot PCR (Engler et al., 2009) of the primers (SAN1b1, 1b2, 1b3, 1b4, 1b5 and 1b6). However, due to the melting temperature (T_m) differences of each overlapping region, the attempt was unsuccessful. Two alternate approaches were pursued to solve the T_m problem: The same amount of the six primers were annealed and extended with DNA polymerase I Klenow fragment, followed by the addition of extra SAN1b1 and 1b6, and One-pot PCR; and, the primers VL1-u,

SAN1b1, 1b4, 1b5 and 1b6 were used instead in One-pot PCR. Both approaches yielded the desired N25 promoter product, but only the second one is described in details later.

Cloning/modification of pSAN1 and pSAN2 for N25 and N25_{anti} promoters

First, 37 bp of DNA (primers SAN1a-NT (non-template) and SAN1a-T (template)) was inserted at *XhoI* cloning site. Each 50 μ L plasmid DNA digestion reaction contained 3 μ g of respective vector plasmid, 1x NEBuffer 4 (20 mM Tris-acetate, pH 7.9, 50 mM potassium acetate, 10 mM magnesium acetate and 1 mM DTT), 0.1 mg/mL BSA and 1 μ L of *XhoI* (NEB; 20 units/ μ L). The reaction was incubated at 37 °C for 2 hours and then the enzyme was heat-inactivated at 65 °C for 20 min. The complete digestion was confirmed by agarose gel electrophoresis and UV detection. The reaction mixture was supplemented with 6 μ L of 10x Antarctic Phosphatase (AP) buffer (0.5 M Bis-Tris-Propane-HCl, pH 6.0, 10 mM MgCl₂ and 1 mM ZnCl₂) and 3 μ L of doubly-distilled water (ddH₂O). Then 1 μ L of AP (NEB; 5 units/ μ L) was added to the now 60 μ L reaction which was incubated at 37 °C for 30 min and the AP was heat-inactivated at 65 °C for 20 min. The dephosphorylated linearized plasmid DNA was concentrated by NaAc-EtOH precipitation. The pellet was re-suspended with 9 μ L of TE and 1 μ L of 10x glycerol dye (40 % glycerol, 0.1 M EDTA, 0.5 % xylene cyanol, 0.5% BPB) and fractionated by agarose gel electrophoresis at 100 V for 30 min. The dephosphorylated vector DNA band

was recovered from the gel using QIAquick Gel Extraction Kit (QIAGEN). The concentrations were measured by Nanodrop and 0.1 pmol of each vector plasmid was used in respective ligation reaction.

To prepare the phosphorylated annealed insert, each primer was treated with T4 polynucleotide kinase (T4 PNK). The 50 μ L phosphorylation reaction contained 300 pmol of primer, 1x T4 PNK buffer (70 mM Tris-HCl, pH 7.6, 10 mM MgCl₂ and 5 mM DTT), 1 mM ATP and 1 μ L of T4 PNK (NEB; 10 units/ μ L). The reaction was incubated at 37 °C for 30 min and T4 PNK was heat-inactivated at 65 °C for 20 min. The phosphorylated primers were used directly in ligation reaction which was set up to contain 0.1 pmol of vector, 0.3 pmol of each insert primers, 1x T4 DNA ligase buffer (50 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 1 mM ATP and 10 mM DTT), 1 mM ATP and 1 μ L of T4 DNA ligase (NEB; 400 cohesive end units/ μ L). The 17- μ L reaction mixture without ATP and the enzyme was first incubated at 70 °C, 55 °C and 42 °C for 5 min each. It was allowed to cool down to room temperature and then 2 μ L of 10 mM ATP and 1 μ L T4 DNA ligase was added before incubating overnight at 16-18 °C. Control ligation reactions were also set up with dephosphorylated linearized plasmid DNA alone. Heat-shock transformation was carried out with CaCl₂-treated competent XL1-Blue *E. coli* cells and the cells were spread on 1.5% agar LB plate containing 0.1 mg/mL ampicillin (LB/Amp plate). Quite a few colonies were picked and individually grown in 5 mL LB containing 0.1 mg/mL ampicillin (LB/Amp) overnight in 37 °C shaker. The plasmids were isolated using QIAprep

Spin Miniprep Kit (QIAGEN) and the presence of the insert was confirmed by double restriction digestion with *XhoI* and *PstI*. After several attempts, the 37-bp insertion into pSAN2 was successful and yielded the pSAN2a plasmid.

The same procedure except for the use of 1x NEBuffer 3 (50 mM potassium acetate, 20 mM Tris acetate, pH 7.9, 10 mM magnesium acetate and 1 mM DTT) for *PstI* (20 units/ μ L) digestion was utilized to insert 15 bp of DNA (annealed from primers SAN1b-NT and SAN1b-T) containing an *EcoRV* site into the *PstI* cloning site. The presence of the insert was confirmed by *SalI* and *EcoRV* double digest and the promising plasmids were sent to Elim Biopharmaceuticals, Inc. for sequencing. The 40% glycerol stocks of transformed cells containing the plasmid with the correct sequence, now designated pSAN2b, were stored at -80 °C and used as starter cultures for preparing fresh plasmid DNA for the plasmid labeling procedure.

Unfortunately, the above insertion-cloning procedure was not successful with the N25 vector pSAN1. After a few attempts as described above, it was decided to generate the N25 plasmid pSAN1b by replacing the promoter insert in pSAN2b. The vector was prepared from the double restriction enzyme digestion of pSAN2b DNA using *XhoI* and *PstI*. The 50 μ L reaction contained 3 μ g of N25_{anti} plasmid pSAN2b, 1x NEBuffer 3, 0.1 mg/mL BSA and 1 μ L each of *PstI* and *XhoI*. The linearized plasmid was de-phosphorylated and purified as described above (Antarctic Phosphatase and QIAquick Gel Extraction Kit). The insert was prepared by 100 μ L One-Pot PCR containing the following primers:

0.3 μM SAN1b1, 0.02 μM VL1-u, 0.02 μM each of SAN1b4 and SAN1b5, 0.3 μM SAN1b6, 1x Phusion® HF buffer (NEB; composition not available), 0.25 mM dNTP and 1 μL of Phusion® High-Fidelity DNA Polymerase (NEB; 2 units/ μL). The PCR cycle was set up to be 95 °C, 53 °C and 72 °C, 20 sec each for 40 cycles. The presence of the insert fragment was confirmed by agarose gel electrophoresis and UV detection. The PCR product was digested with *Xho*I and *Pst*I, and subsequently ligated into the pSAN2b/*Xho* I + *Pst* I/alkaline phosphatase-treated vector. Transformation of XL1-Blue cells was successful in yielding pSAN1b clones containing the N25 promoter. Accuracy of the clones was confirmed through sequencing.

Plasmid labeling: Generating single end-labeled fragment from plasmid DNAs

Each pSA508 plasmid containing the N25 (abbreviated N), N25_{anti} (A), DG203 (D) or SP_{fullcon} (S) promoter was purified from an overnight LB/Amp culture using either the Plasmid Midi Kit (Qiagen) or the Plasmid Plus Midi Kit (Qiagen). To use the former kit, cells from glycerol stock were inoculated into 100 mL of LB/Amp culture and incubated overnight at 37 °C shaker whereas for the latter, 50 mL of the overnight culture was sufficient. The plasmid labeling procedure was adapted from the protocol provided by Dr. Ross. Plasmid labeling required starting out with 11 μg (~5 pmol) of a purified plasmid. To label the 5' end of template strand, the N25 and N25_{anti} (NA) plasmids, and the DG203 and SP_{fullcon} (DS) plasmids, were individually digested with *Eco*RV and *Acc*65I,

respectively. Each 20 μL restriction enzyme reaction contained 11 μg plasmid DNA, 1x NEBuffer 3, 0.1 mg/mL BSA and 1 μL of *EcoRV* for NA or 1 μL of *Acc65I* for DS. The reactions were incubated at 37 $^{\circ}\text{C}$ for 2 hours after which 0.5 μL of the respective enzymes were added and the incubation was continued for another hour. The enzymes were heat-inactivated by incubating at 80 $^{\circ}\text{C}$ (*EcoRV*) or 65 $^{\circ}\text{C}$ (*Acc65I*) for 20 min. The linearized plasmid DNA was then dephosphorylated by expanding the reaction volume to 100 μL in 1x AP buffer and 3 μL of AP. The reactions were incubated at 37 $^{\circ}\text{C}$ for 30 min and the AP was heat-inactivated at 65 $^{\circ}\text{C}$ for 20 min. The dephosphorylated plasmid DNA was then recovered through NaAc-EtOH precipitation overnight at -20 $^{\circ}\text{C}$.

The following day, the pellets were collected by centrifugation (13,400 rpm) at 4 $^{\circ}\text{C}$ for 30 min; after supernatant removal, the pellets were washed 3 times with 70% EtOH and dried in the Speed Vac for 15 min. Each pellet was re-suspended in 20 μL ddH₂O and set up for radiolabeling in a 50- μL reaction volume containing, in addition, 1x T4 PNK buffer, 150 μCi of [γ -³²P] ATP (NEG035C, Perkin Elmer), 1.25 μM ATP and 2 μL of T4 PNK. The reactions were incubated at 37 $^{\circ}\text{C}$ for 30 min and T4 PNK was heat-inactivated at 65 $^{\circ}\text{C}$ for 20 min. Next, NaAc-EtOH precipitation was carried out in the presence of 100 μg of glycogen at -80 $^{\circ}\text{C}$ for 30 min to remove the unincorporated radioactive ATP. The pellets were collected as in previous step and were re-suspended in 17 μL TE to set up a 20- μL reaction for cutting with the second restriction enzyme. For the NA plasmids, the second restriction enzyme reaction contained 1x

NEBuffer 4, 0.1 mg/mL BSA and 1 μ L of *Xho*I; for the DS plasmids, the reaction contained 1x NEBuffer 4 and 1 μ L *Eco*53kI (NEB; 10 units/ μ L). The reactions were incubated at 37 °C for 2 hours, followed by the addition of 0.5 μ L of the respective enzyme and another hour of incubation. Afterwards, NaAc-EtOH precipitation was carried out with 40 μ g of glycogen carrier at -80 °C for 30 min. The pellets were collected as in previous step (except that the samples were washed with 70% EtOH only once), and re-suspended in 18 μ L TE and 2 μ L 10x glycerol dye solution. The samples were loaded onto a 6% (19:1) polyacrylamide native gel and electrophoresis was carried out at 200 V for 2.5 hours to separate the restriction digest fragments. The DNA fragments were located through its radioactivity by a 5-min exposure to the Storage Phosphor Screen and scanning by the Phosphorimager. The gel pieces containing the short labeled DNA (158 bp) were cut out, individually placed in 600 μ L low-salt buffer (20 mM Tris-HCl, pH 7.4, 0.2 M NaCl and 1 mM EDTA) and incubated in a 37 °C shaker for a day. The supernatant from each sample was collected, purified by phenol-chloroform extraction, and concentrated by NaAc-EtOH precipitation. The final pellets were collected by centrifugation (13,400 rpm) for 30 min and careful supernatant removal, washed with 70% EtOH three times, dried in the Speed Vac, and re-suspended in a total of 30 μ L TE. The cpm/fmol were calculated from Nanodrop and scintillation counting data. A successful round of labeling usually gave rise to a range of specific activity of 2000 – 5000 cpm/fmol.

Due to recent complications in the NA-plasmid labeling, the above procedure was slightly modified for all four DNA templates as follows. Instead of 20- μ L reaction volume, restriction enzyme digestions were carried out in 30- μ L reaction to avoid potential star activity of the enzymes. The centrifugation durations were also increased from 30 min to one hour for better recovery. To radiolabel the non-template strand, the order of the restriction enzyme cuts were simply reversed, e.g. for NA non-template labeling, *XhoI* was used first and *EcoRV* second. The purified radiolabeled DNA is then used in Maxam-Gilbert sequencing and permanganate footprinting.

Maxam-Gilbert sequencing ladder

To reference the sequences of the footprint, a Maxam-Gilbert ladder (G+A) was prepared for each promoter template as described (Maxam and Gilbert, 1980) and specific conditions were modified by A. Lee '10. Briefly, a 14- μ L (G+A) sequencing reaction contained 200,000 cpm of labeled DNA and 1 μ g of sonicated salmon sperm (sss) DNA brought to 10 μ L volume, to which 4 μ L of 10% formic acid was added. The reaction was incubated at 37 °C for 15 min to modify the purine bases (the protonated purine bases become depurinated). The reaction was terminated by adding 200 μ L of (G+A) stop solution (0.3 M NaAc and 50 μ g/mL RNA), followed by EtOH precipitation with 750 μ L of 95% EtOH in a dry ice/EtOH bath for 30 min. The pellets were collected by centrifugation (13,400 rpm) for 30 min, washed once with 70% EtOH, dried in the Speed Vac and re-

suspended in 100 μL of 10% piperidine. The cleavage reaction in a weak base (i.e. 10% piperidine) was allowed to proceed at 90 $^{\circ}\text{C}$ for 15 sec (with cap open) and for 30 min (with cap closed and weighed down). The samples were concentrated using NaAc-EtOH precipitation with 100 μg glycogen in dry ice/EtOH bath for 30 min. The pellets were collected by centrifugation (13,400 rpm) for 30 min, washed three times with 70% EtOH and dried in the Speed Vac. They were then re-suspended in 20 μL ddH₂O, and allowed to dry in the Speed Vac for another 45 min. The final pellets were re-suspended in 9 μL of FLB with NaOH (1x TBE, 10 mM EDTA, 0.06% amaranth, 0.06% xylene cyanol, 20 mM NaOH in de-ionized formamide solution). Aliquots of the Maxam-Gilbert ladder (3 μL) were loaded alongside the footprinting samples to index the position of the footprints.

Permanganate footprinting

In transcriptional investigations, permanganate is used to probe the single-stranded regions of a promoter-RNA polymerase complex by reacting with unpaired thymidines. A typical permanganate reaction contained labeled DNA, transcription buffer, KCl, RNAP, 0.1 μg of heparin, NTP, and potassium permanganate, but specific conditions as well as incubation time were fine-tuned to yield a snapshot of the initial transcribing complexes. First, the reaction conditions in which initial transcribing complexes were abundant for permanganate footprinting, such as different transcription buffers and salt concentrations were tested by regular transcription assays. The conditions tested

were the buffers: 1x old transcription buffer (50 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, 10 mM β-ME, 10 μg/mL acetylated BSA), 1x new transcription buffer (50 mM Tris-HCl, pH 7.2, 5 mM MgCl₂, 10 mM β-ME and 10 μg/mL BSA) and 1x cacodylate buffer (20 mM cacodylate, pH 7.0, 10 mM MgCl₂, 10 mM β-ME and 10 μg/mL BSA), and the salt concentrations: 50 mM, 100 mM and 200 mM KCl. From the abortive profiles of N25, N25_{anti}, DG203 and SP_{fullcon} in these tests, it was concluded that 50 mM KCl was optimal for capturing the initial transcribing complexes and that the buffer conditions did not have much effect. Therefore, the permanganate footprinting reactions were all carried out in 1x old transcription buffer and 50 mM KCl. For each set of footprinting reaction, the same amount and cpm of each promoter were used to achieve similar specific activity and, subsequently, the same RNAP:DNA ratio in open complex formation. To achieve this criterion, we first calculate the amount of promoter DNA required to give a total of 200,000 cpm from the cpm/fmol data. Then, non-radiolabeled promoters were supplemented to those with higher cpm/fmol values to match the one with the lowest cpm/fmol. Next, to ensure full occupancy of promoter by RNAP, different ratios of DNA to RNAP—1:3, 1:5 and 1:10—were tested; as a result, the ratio of DNA:RNAP of 1:10 was chosen. In addition, different permanganate concentrations (1 mM, 3 mM, 4.5 mM and 9 mM) for footprinting either the template strand (T) or the non-template strand (NT) were tested to ensure single-hit reaction conditions. The results so far indicated that 1 mM and 4.5 mM were optimal for NT- and T-labeled DNA, respectively. The amount of incubation

time for heparin and NTP (30 sec and 1 min) as well as that for permanganate (30 sec, 1, 2, 3, 5 and 10 min) were also tested. Furthermore, different concentrations of NTP such as 1 μ M, 10 μ M and 100 μ M were tested on NT-labeled DNA.

So far, the optimal condition of permanganate footprinting is 200,000 cpm DNA (with specific activity \sim 100-2000 cpm/fmol), RNAP:DNA (10:1), 1x old transcription buffer, 50 mM KCl, 0.1 mM NTP, 0.1 μ g heparin and 1 mM KMnO_4 (NT-labeled DNA) or 4.5 mM KMnO_4 (T-labeled DNA). Each reaction was proceeded in three steps. The 8- μ L DNA mixture containing the labeled DNA, transcription buffer and KCl in ddH₂O was made. The RNAP was diluted with enzyme diluents and was added to the DNA mixture, incubating at 37 °C for 10 min. Then, heparin only or heparin with NTP was added for open complex (RP_o) or transcribing complexes, respectively, and the reaction was incubated for another 30 sec at 37 °C. Finally, respective KMnO_4 solution was added and the reaction was incubated for 30 sec for RP_o and 30 sec and 2 min for transcribing complexes. After each time point, the reaction was terminated by KMnO_4 stop solution (0.3 M NaAc, 0.5 M β -ME, 0.8 μ g sssDNA in TE buffer). At this point, the same procedure was followed as in Maxam-Gilbert sequencing reaction after the termination by (G+A) stop solution. However, the final pellet was re-suspended in 7 μ L of FLB/NaOH solution and 4- μ L aliquots were loaded onto a 10% (19:1) polyacrylamide 7 M urea sequencing gel. The gel was submerged in 1x TBE buffer both at the top and the bottom reservoirs and the electrophoresis was carried out at 35 W for approximately 4 hours or until the xylene cyanol dye

was ~7 cm from the bottom edge. After the electrophoresis, the gel was transferred to a piece of moistened Whatman 3MM paper and dried under vacuum in a Bio-Rad gel dryer at 80 °C for an hour. The gel was exposed to storage phosphor screen overnight and later scanned with Phosphorimager. The results were analyzed with ImageQuant 5.2.

RESULTS

To support the postulated hyper-forward translocation mechanism for very long abortive transcript (VLAT) formation, two experimental methods were employed. They were *in vitro* transcription assays in the presence of E₁₁₁Q EcoRI (roadblock experiments) to investigate the RNAP forward translocation and permanganate footprinting to examine the sizes of the transcription bubbles. Accordingly, the results from these methods provided evidence that support the following required features of the proposed mechanism: 1) The VLATs are the products of RNAP forward movement, produced during the promoter escape transition on DG203 and SP_{fullcon} promoters, 2) The RNAP requires at least 2 bp (optimally 4-5 bp) to forward translocate and produce GreB-resistant VLATs, and 3) On VLAT-producing promoters, the initial transcribing complex bubble expands to at least +20 through DNA scrunching.

Roadblock Transcription Experiments

Rationale

The roadblock experiments were carried out on EQ-series DNA templates to test whether the RNAP forward movement was responsible for the production of VLATs. The specific and tight binding of the E₁₁₁Q EcoRI was reported to be able to stop the RNAP during elongation at the position where the RNAP active

center and the first nucleotide of the *Eco*RI recognition site (G) are 14 bp apart (Pavco and Steege, 1990). However, since the earlier roadblock protein studies were conducted for elongation complexes only, we investigated whether this distance relationship holds true for the initial transcribing complexes (EQ-23 and EQ-28); we included two elongation complexes formed on our DNA templates (EQ-41 and EQ-43) for comparison. Under our reaction conditions, it was observed that the RNAP was stopped when the aforementioned distance was 14-16 nt for initiation complexes and 15-16 nt for elongation complexes (Vallery, 2010). Based on the results of Jiang (2009) and Vallery (2010), we constructed additional roadblock transcription templates containing an *Eco*RI site bracketing the promoter escape positions (i.e. +16 to +19 on DG203-related templates). Thus, the EQ-series promoters that were examined in this project include DG203 (the control promoter without an *Eco*RI binding site), EQ-23, -28, -32, -33, -34, -35, -36, -37, -38, -41, -43; the number denotes the position of the first nucleotide (G) of the *Eco*RI binding site (GAATTC).

Previous transcription experiments with EQ-23, -28, -41, and -43 showed that a cluster of two or three bands was generated when RNAP encounters the roadblock (Jiang, 2009; Vallery, 2010; see Figure 11 below). This imprecise stoppage might be due to the heterogeneous population of RNAP molecules scrunching DNA with three different speeds (or forces). The fastest fraction could pull the DNA all the way until the downstream edge of the RNAP is at the *Eco*RI binding site, while the slowest fraction was halted at the outer edge of the

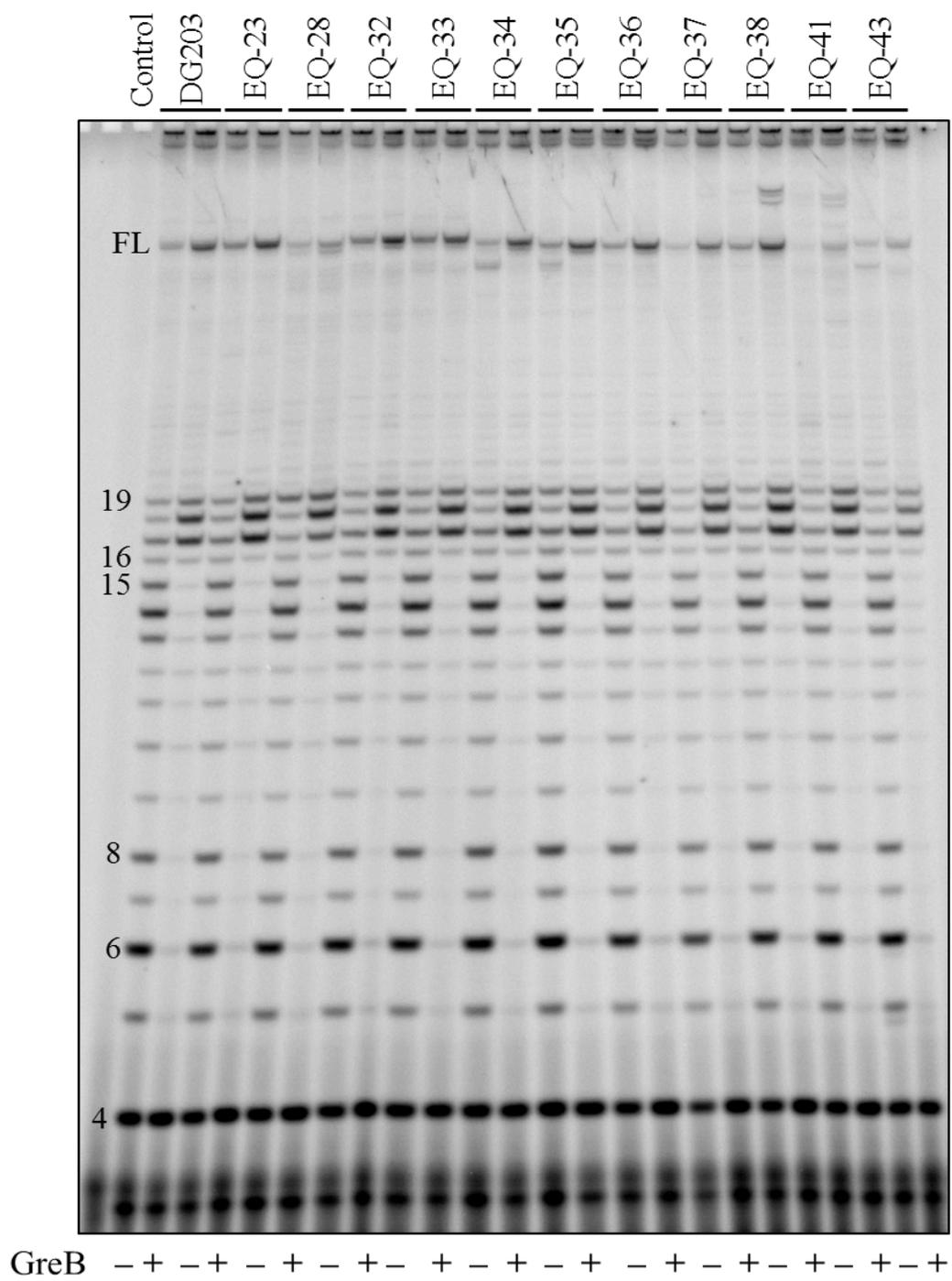
EcoRI-DNA complex 2 bp before the *EcoRI* site (McClarín et al., 1986). Then, there could be an intermediate fraction that stopped at 1 bp before the *EcoRI* site. Accordingly, the transcripts synthesized by RNAP in the presence of roadblock protein could be of three populations: those released due to backtracking, those that bump into the *EcoRI* protein and become stalled, and those that were released due to RNAP hyper forward translocation. To sort out the three groups, a set of roadblock transcription reactions in the presence of GreB was conducted. GreB decreases the level of backtracked transcripts, leaves the level of simply-halted transcript unaffected, and enhances the level of GreB-resistant VLATs.

Results

The *in vitro* transcription experiments with and without GreB were conducted to ensure that the presence of downstream *EcoRI* binding sites did not alter the production of GreB-resistant VLATs on DG203 promoter. As expected, the level of VLAT production without or with GreB was comparable across all EQ-series promoters (Figure 10). However, GreB enhanced escape and increased the level of full-length RNA. The increased amount of VLATs (17-19 nt) observed in (+GreB) lanes suggested that as more backtracked RNAPs were rescued, the chance for the RNAPs to come through the VLAT-producing stage increased.

Next, transcription experiments without *EcoRI* or GreB ($EcoRI^- GreB^-$), with *EcoRI* only ($EcoRI^+ GreB^-$), with both *EcoRI* and GreB ($EcoRI^+ GreB^+$) were

Figure 10. Gel image of *in vitro* transcription experiments with and without GreB on EQ-series DNA. The conditions were 30 nM DNA, RNAP:DNA (1:1), GreB:RNAP (10:1) and 200 mM KCl, 100 mM NTP with [γ - 32 P]-ATP label. The control reaction contained no enzymes. This set of reactions was carried out to confirm that GreB-resistant-VLAT formation was not altered on EQ-series DNA by the presence of downstream *Eco*RI binding sites. Similar to the results on DG203 template, the abortive transcripts of ≤ 15 nt disappeared in the presence of GreB and the level of full-length RNA (FL) increased. The VLATs (16-19 nt) persisted and even increased; the latter is due to the increased probability of RNAP coming to the VLAT-producing stage as the backtracked RNAs were rescued.



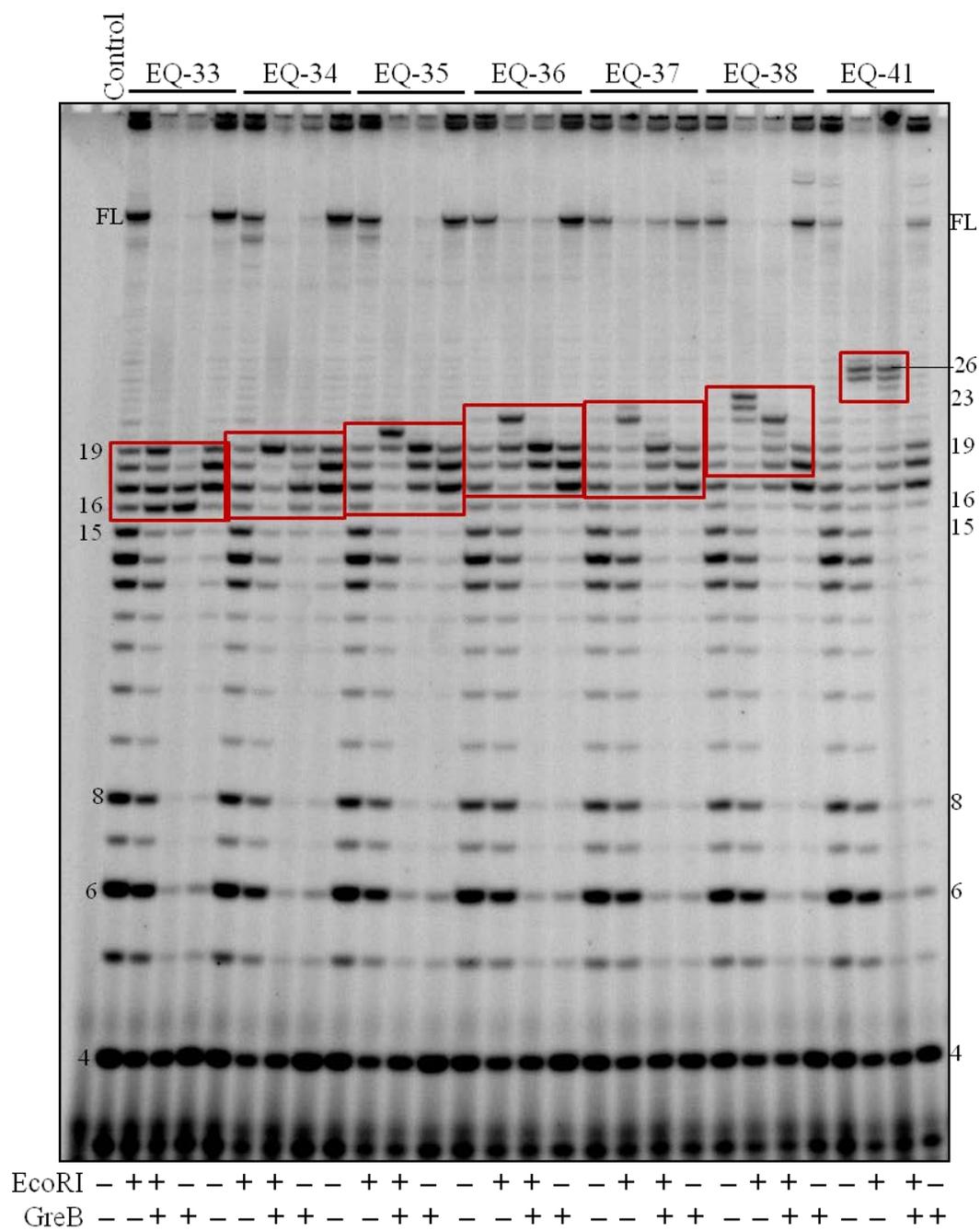
performed for the EQ-series templates and the RNAs recovered from each set were loaded side-by-side on the transcription gel for direct comparison (Figure 11). The data analysis indeed showed the existence of all three types of abortive transcripts described in the *Rationale*. For example, on EQ-23 promoter, EcoRI blockage of the RNAP led to the synthesis of 6-8 nt long abortive transcripts. Of these, the 6-nt long transcript was not affected whereas 7- and 8-nt long RNAs were rescued by GreB. This result indicated that 7- and 8-nt long abortive transcripts were the products of RNAP backtracking and 6-nt long RNA was from RNAP stalling at the roadblock. On this initial transcribing complex, stalling occurred when the active center and the *EcoRI* binding site were 17 bp apart. In contrast, on EQ-41 promoter, the RNAP was blocked by EcoRI when the active center was at +25, +26, or +27. The level of these RNAs did not change in the presence of GreB, indicating that these transcripts were produced neither from RNAP backtracking nor hyper forward translocation (Figure 11). Thus, on an elongation complex, RNAP was stalled when the active center and the *EcoRI* binding site were 14-16 bp apart. The different distances for bumping into the EcoRI protein between the two complexes suggests a conformational change in the RNAP complex between the initiation phase and the elongation phase.

Since our criterion for determining whether an abortive transcript is derived from RNAP hyper forward translocation was the resistance to GreB treatment, a set of four reactions ($EcoRI^- GreB^-$, $EcoRI^+ GreB^-$, $EcoRI^+ GreB^+$ and $EcoRI^- GreB^+$) was performed for selected EQ-series promoter templates and

Figure 11. Gel image of *in vitro* transcription experiments without EcoRI and GreB, with EcoRI alone, and with both EcoRI and GreB on EQ-series DNA. The conditions were 30 nM DNA, RNAP:DNA (1:1), GreB:RNAP (10:1) and 200 mM KCl. The control reaction contained no enzymes. The presence of EcoRI effectively halts RNAP transcription on EQ-series promoters, as evidenced by the lack of full-length transcripts and increased abundance of the blocked transcripts shown in red boxes. In (GreB + EcoRI) lanes, the level of backtracking-induced transcripts was greatly diminished. The levels of those stably stalled by EcoRI did not change whereas those from RNAP hyper forward translocation increased. Moreover, the increased level of GreB-resistant VLATs was detected only in reactions where the roadblock placement did not interfere with promoter escape and RNAP forward movement, e.g. EQ-33 (for 17-nt VLATs), EQ-34 (for 16-, 17- and 18-nt VLATs), EQ-35 (for 17-, 18- and 19-nt VLATs), EQ-36 (for 17-, 18- and 19-nt VLATs), EQ-37 (for 17-, 18-, 19-, and 20-nt VLATs) and EQ-38 (for 17-, 18-, 19-, 20- and 21-nt VLATs).

loaded side-by-side for direct comparison (Figure 12). On EQ-32 to EQ-38 promoters, the roadblock protein was bound at positions that affected promoter escape and the formation of GreB-resistant VLATs, indicating that GreB-resistant VLATs are produced during the promoter escape transition. The results in Figures 11 and 12 showed that the RNAP underwent the escape transition into elongation phase at multiple positions, a small fraction beginning at +16 (see EQ-32) and most completing the process at +19 (see EQ-37), with some lingering until +21 (see EQ-38). Of all the VLATs made at an EQ promoter, only those with sufficient spatial requirement for hyper forward translocation (i.e. RNAP translocates by more than 1 bp distance) became GreB-resistant VLATs whose level was also elevated. This is best illustrated with the example analysis of EQ-36 transcription (see Figure 12). The presence of EcoRI caused the accumulation of blocked transcripts that were 19 and 21 nt long. Of these, the 21-nt transcripts were rescued by the presence of GreB, resulting in greatly diminished levels, indicating they were the products mainly from RNAP backtracking. The 19-nt abortive transcript would be the stably stalled RNA. However, since the 19-nt RNA containing complex is on the verge of escape, it is unlikely that this complex can stay stably stalled; rather, the RNAP most likely hyper translocated to give rise to elevated level of GreB-resistant VLAT-19. By the same token, the 17- and 18-nt RNA are highly GreB-resistant, because the RNAP in these two complexes can hyper translocate 4-5 bp before clashing with the roadblock bound at the *EcoRI* binding site.

Figure 12. Gel image of *in vitro* transcription experiments without EcoRI and GreB, with EcoRI alone, with both EcoRI and GreB and with GreB alone on selected EQ-series DNA. This gel image allowed the direct comparison of GreB-resistant VLATs in the presence and absence of the roadblock protein on promoters where the RNAP was on the verge of escape. The conditions were 30 nM DNA, RNAP:DNA (1:1), GreB:RNAP (10:1) and 200 mM KCl. The control reaction contained no enzymes. The red box showed the gradual increase in the length of VLATs released by the RNAP as it undergoes the escape transition and immediately encountering a differently positioned downstream roadblock.

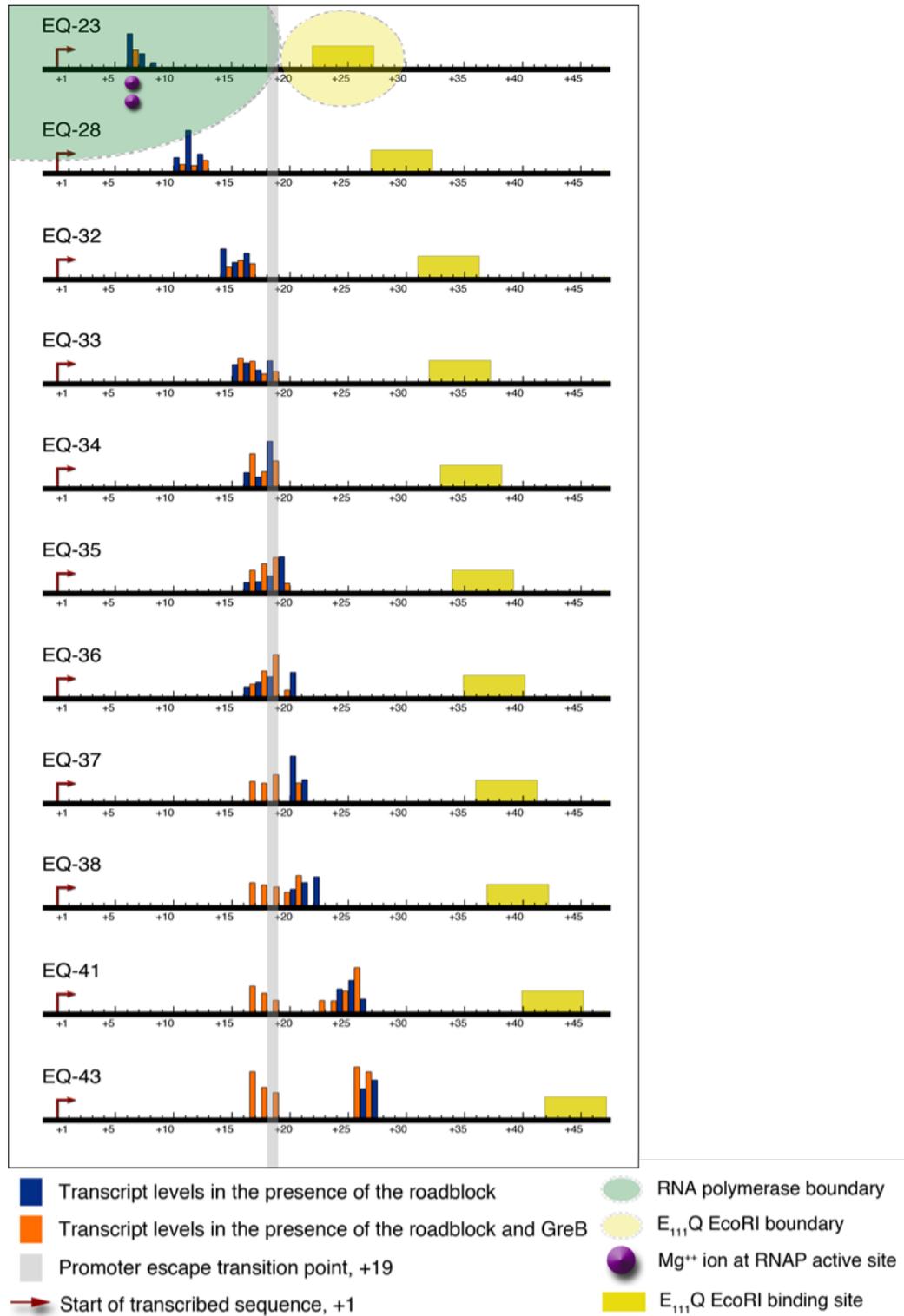


Summary

The data obtained from roadblock transcription experiments are summarized in Figure 13 and Table 4. Figure 13 was constructed by T. K. Vallery '10 to illustrate the changes in the amount of GreB-resistant abortive transcript production as the roadblock protein was placed further downstream. By comparing the transcript levels in the presence of the roadblock (blue) and those in the presence of both the roadblock and GreB (orange), it can be deciphered that GreB-resistant VLATs are produced on the verge of promoter escape by RNAP forward movement. The presence of the roadblock at the transition points could prevent the RNAP from producing VLATs by hyper-forward translocation mechanism. For example, in EQ-33, 16- and 17-nt abortive transcripts arose from RNAP hyperforward translocation (the orange bar is higher than the blue, i.e. GreB-resistant) in comparison to 18- and 19-nt abortive transcripts that resulted from RNAP backtracking (the blue bar is higher than the orange, i.e. GreB-rescued). On EQ-36, all 16-19 nt long abortive transcripts (VLATs) are produced by RNAP hyper forward translocation as there was enough 'space' for the RNAP to jump forward as escape occurs.

Based on the analysis described in previous section, the compiled data of the lengths of abortive transcripts produced from RNAP backtracking versus stalling versus hyper forward translocation are summarized in Table 4. As well, Table 4 contains the estimate of the distances between the RNAP active center and the *EcoRI* binding site when various types of abortive transcripts are made.

Figure 13. The quantification of the results from roadblock transcription experiments. The levels of abortive transcripts released at the site of blockage on EQ-series DNA were quantified using ImageQuant 5.2 software. The height of the transcript levels in the presence of roadblock (blue bar) and those in the presence of roadblock and GreB (orange bar) were drawn in proportion to reflect the intensity of each abortive transcript band. The grey vertical line denotes the promoter escape position at +19. From the active center to the downstream edge, the RNAP covers ~14-15 bp and from the binding site to the upstream edge, the roadblock protein covers ~2 bp. The boundaries are shown in turquoise (RNAP) and yellow (EcoRI). From EQ-32 to EQ-43, the levels of GreB-resistant VLATs (transcript levels in the presence of the roadblock and GreB) increased as the RNAP was allowed more space to jump forward and release the VLATs by hyper-forward translocation mechanism. This figure was constructed by T. K. Vallery '10.



The summary data in Table 4 allow us to glean several very interesting conclusions: 1. When RNAP cannot reach the +16 to +19 positions, no VLATs are formed (see EQ-23, -28, and -32). Thus, VLATs are indeed products of RNAP forward movement. 2. No stalled RNAs, 16-20 nt in length, could be detected, respectively, for EQ-33, -34, -35, -36, and -37 templates, identifying the RNAP complexes at these positions as being on the cusp of escape, when the initial transcribing complex is of such high instability that the RNAP is forced to release its nascent RNA either by backtracking or by hyper forward translocation. The absence of stalled transcripts, therefore, allowed us to identify the exact positions of escape. On the DG203 promoter, escape occurs at multiple positions (when RNAP has transcribed to the +16 to +20 positions), coinciding with the formation of the GreB-resistant VLATs. This observation supports our claim that VLATs are produced during the promoter escape transition. 3. Looking at the inter-distance between the RNAP active center and *EcoRI* binding site, it is clear that excessive scrunching (which reduces the inter-distance to 15-16 bp) causes RNAP backtracking. 4. VLATs resulting from hyper forward translocation are made only when the inter-distance is 17 bp or longer.

Taken together, the summary data confirmed that VLATs are the products of RNAP forward movement during the promoter escape transition. Roadblock experiments with and without GreB revealed a spatial requirement of at least 2 bp, but optimally 4-5 bp, for RNAP to jump forward and produce GreB-resistant VLATs.

Table 4. Summary of roadblock transcription analysis. The lengths of abortive transcripts from RNAP backtracking, stalling, and hyper forward translocation, and the distance in bp between the RNAP active center and *EcoRI* binding site inducing each type of release, are compared. The VLATs were produced on the verge of promoter escape; thus the transition had made the distance between the active center and the downstream edge of the RNAP ~15 bp, i.e., the RNAP was stopped by the roadblock protein when the distance was 15 bp during the transition points (from EQ-33 to EQ-38). The GreB-resistant VLATs were produced only when the distance was at least 16 or 17 bp (EQ-33). The levels of VLATs reached back to GreB alone levels when the distance became ≥ 19 bp (EQ-37 and -38).

DNA	RNA released due to:				Distance between the RNAP active center and <i>EcoRI</i> site inducing:			
	Backtracking	Stalling	Hyper forward translocation		Backtracking	Stalling	Hyper forward translocation	
EQ-23	8, 7 nt	6 nt	-		15, 16 bp	17 bp	-	
EQ-28	14, 12, 11 nt	13 nt	-		14, 16, 17 bp	15 bp	-	
EQ-32	17, 15 nt	16 nt	-		15, 17 bp	16 bp	-	
EQ-33	19, 18 nt	17 nt	16 nt		14, 15 bp	16 bp	17 bp	
EQ-34	Some 19 nt	-	18, 17, 16 nt		15 bp	-	16, 17, 18 bp	
EQ-35	20 nt	-	19, 18, 17 nt		15 bp	-	16, 17, 18 bp	
EQ-36	21, 20 nt	-	19, 18, 17 nt		15, 16 bp	-	17, 18, 19 bp	
EQ-37	22, 21 nt	-	20-17 nt		16 bp	-	17-20 bp	
EQ-38	23, 22 nt	-	21-17 nt		15, 16 bp	-	17-21 bp	
EQ-41	27 nt	26, 25 nt	24 nt		14 bp	15, 16 bp	17 bp	
EQ-43	28 nt	27 nt	26 nt		15 bp	16 bp	21 bp	

Permanganate Footprinting Experiments

Rationale

Permanganate is a chemical reagent that can differentially oxidize thymine residues in single-stranded form and spare those in double-stranded conformation (Hayatsu and Ukita, 1967; Hayatsu and Iida, 1969). Therefore, using it as a footprinting agent would allow the capture of the transcription bubble expansion as more DNA becomes melted apart in initial transcribing complexes (ITCs). We wanted to investigate how much DNA was melted apart and scrunched in by the RNAP on DG203 and SP_{fullcon} promoters to generate the stress that could give rise to RNAP hyper forward translocation and compare the sizes of transcription bubbles among the four promoters (N25, N25_{anti}, DG203 and SP_{fullcon}; abbreviated N, A, D and S, respectively). For easy reference, their non-template and template sequences from -12 to +50 are included in Figure 21, with the thymine residues highlighted in darker color.

Optimization

The permanganate footprinting experiments were performed using N, A, D and S promoters to compare the sizes of the transcribing complex bubbles among the four to investigate whether the initial transcribing complexes (ITCs) of VLAT-producing templates (D and S) were significantly larger as postulated. Before the actual footprinting experiments were conducted, *in vitro* transcription

experiments were carried out to test the effects of transcription buffer (Figure 14) and KCl concentration (Figure 15) on the abundance of abortive transcripts. The comparison among the old transcription buffer (pH 8.0), new transcription buffer (pH 7.2) and cacodylate buffer (pH 7.0) showed no difference (Figure 14). The KCl concentration, however, affected the amount of abortive transcripts, with the highest amount being produced at the lowest KCl concentration tested (50 mM; see Figure 15).

The earlier footprinting experiments were carried out with radiolabeled DNA prepared by primer extension. However, this DNA-preparatory method did not yield labeled DNA of high enough quality. Although the open complex and transcribing complex bubbles were successfully captured with permanganate footprinting technique, the background level of DNA breakage made it impossible for the results to be critically interpreted (Figure 16). We then switched to use the plasmid DNA labeling procedure (see *Plasmid labeling: Generating single end-labeled fragment from plasmid DNAs* in Methods) and the quality of the radiolabeled DNA significantly improved. However, we kept on losing the ^{32}P -label rapidly in the permanganate reaction from one set of end-labeled DNA, and realized a very important lesson – that if the 5'-overhang sequence of a restriction enzyme cut site bearing the ^{32}P label contained a single-stranded thymine residue on it, then the ^{32}P -tag was subject to rapid removal by permanganate modification and subsequent cleavage. This factor came into play especially in the labeled non-template strand of N25 and N25_{anti} and the labeled template strand of DG203

Figure 14. Gel image of *in vitro* transcription experiments on the footprinting promoter templates N25, N25_{anti}, DG203 and SP_{fullcon}: comparison of buffer conditions. The reaction conditions were 30 nM DNA, RNAP:DNA (1:1) and 200 mM KCl. The transcription reaction for each set (N, A, D and S) was carried out with different transcription buffers and from left to right, 1x old transcription buffer (pH 8.0), 1x new transcription buffer (pH 7.2) and 1x cacodylate buffer (pH 7.0). The experiments were carried out to test different transcription buffer conditions, compared their effects on VLAT production. There was no difference in the level of abortive transcripts on any of the promoter.

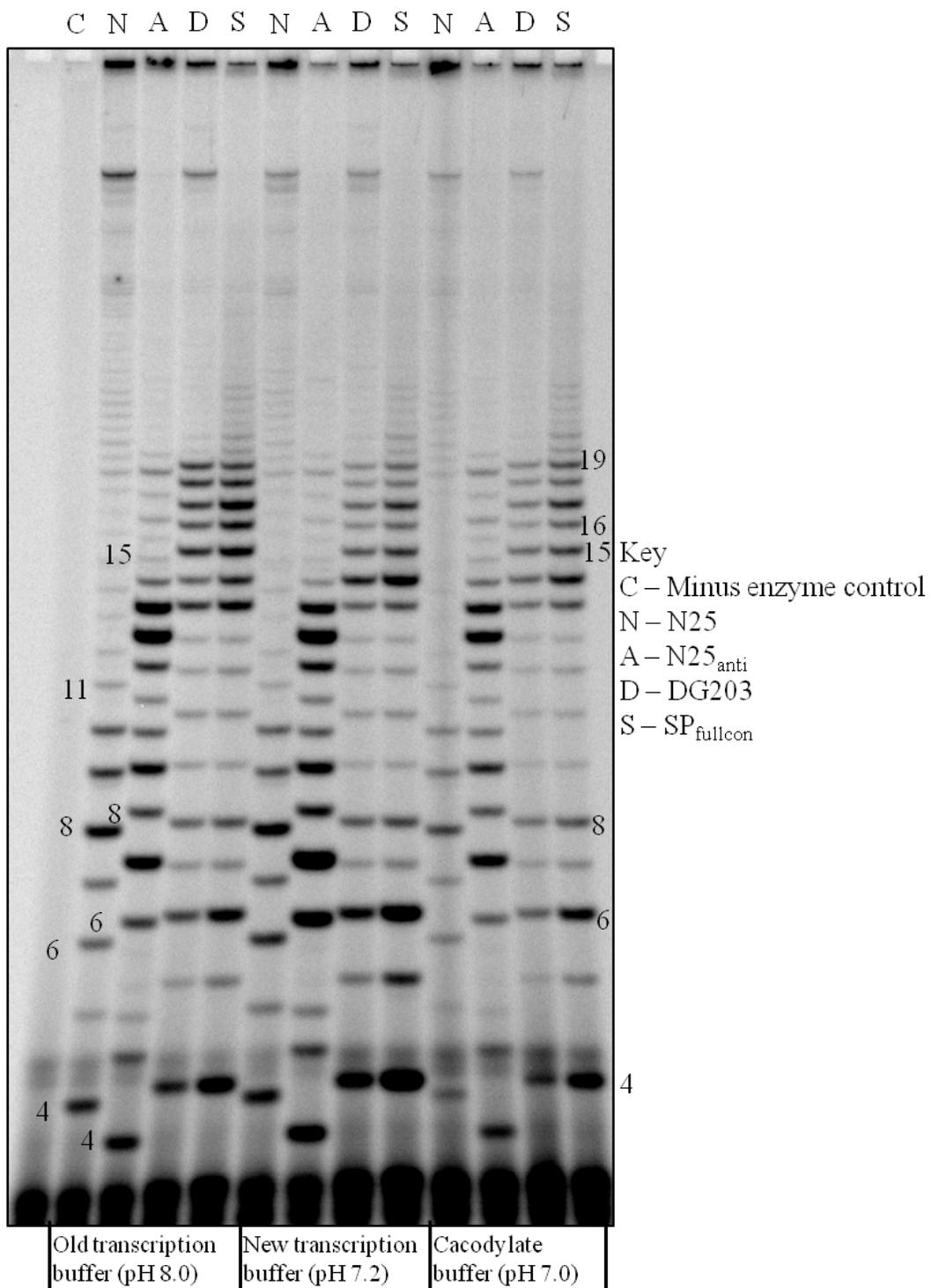


Figure 15. Gel image of *in vitro* transcription experiments on the footprinting promoter templates N25, N25_{anti}, DG203 and SP_{fullcon}: KCl titration. The reaction conditions were 30 nM DNA, RNAP:DNA (1:1) and 1x cacodylate buffer (pH 7.0). The transcription reaction for each set (N, A, D and S) was carried out with different salt concentrations (200 mM, 100 mM and 50 mM KCl). The experiments were carried out to compare the effects of KCl concentrations on VLAT production. The RNAP aborted more transcripts on all promoters in experiments with 50 mM KCl. The full-length production on N25 and DG203 promoters was also decreased as lower KCl concentration was used.

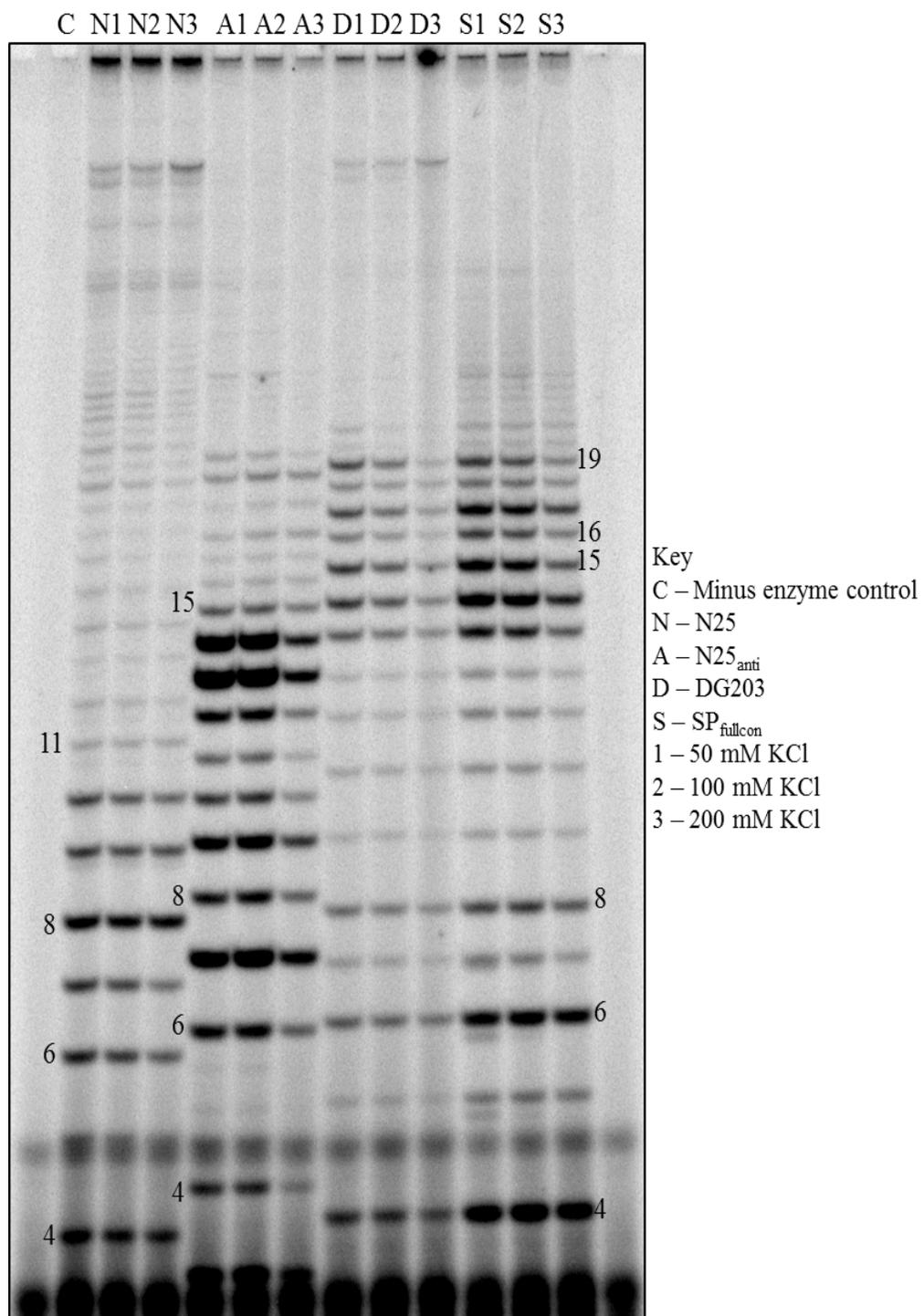
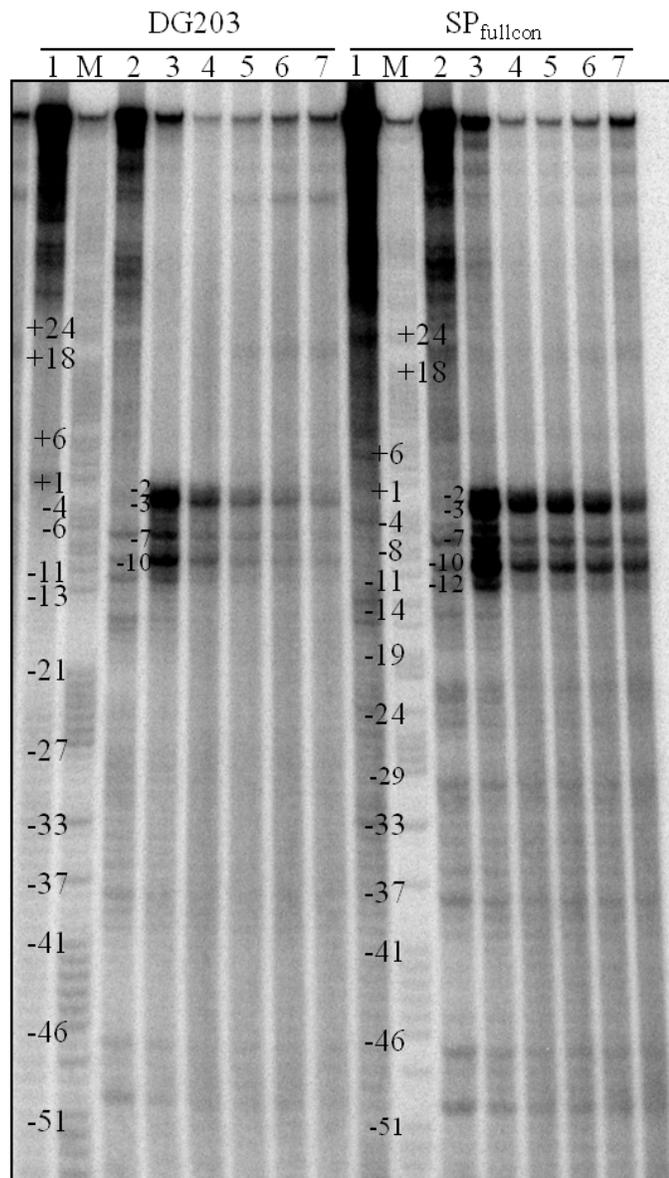


Figure 16. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled DG203 and SP_{fullcon} promoter fragments prepared by primer extension. The DNA was labeled on the non-template strand, so the transcription bubble expanded in the upward (↑) direction. The reaction conditions were 30 nM labeled DNA (286 cpm/fmol), RNAP:DNA (3:1), 1x old transcription buffer (pH 8.0) and 50 mM KCl. The reactions proceeded at 37 °C. Into the DNA/transcription buffer/KCl solution was added the RNAP and the open complex was allowed to form for 10 min. Then heparin and NTP were added to the final concentrations of 10 µg/mL and 100 µM, respectively (minus NTP for open complexes). For transcribing complexes, KMnO₄ was added to 9 mM final concentration after the specified NTP reaction time. Although the open and transcribing complexes were captured, the DNA only and DNA+KMnO₄ only controls revealed DNA breakage and possibly transient breathing of DNA at the AT-rich sequences. The purity of the radiolabeled DNA is extremely important for low background in footprinting reactions. Thus, plasmid labeling method was pursued to prepare the DNA templates for footprinting reactions.



RNAP	-	-	+	+	+	+	+	-	-	+	+	+	+	+
NTP	-	-	-	0'	1'	3'	5'	-	-	-	0'	1'	3'	5'
KMnO ₄	-	+	+	+	+	+	+	-	+	+	+	+	+	+

Key

M – Maxam-Gilbert (G+A) Sequencing Ladder

1 – DNA only control

2 – DNA + KMnO₄ only control

3 – Open complex

4 to 7 – Transcribing complexes with 0, 1, 3 and 5 min of NTP reaction time

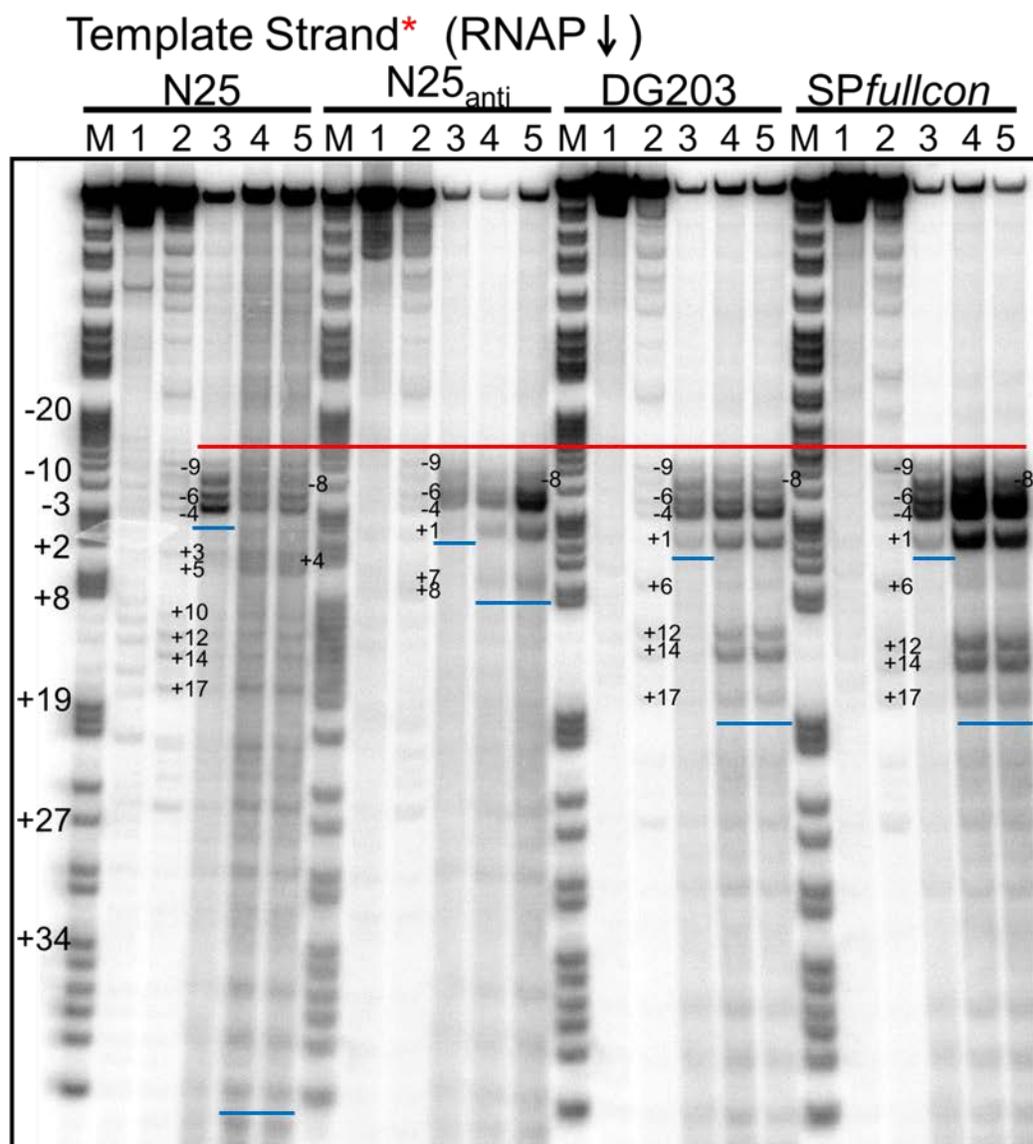
and SP_{fullcon} where the 5' overhangs contained a thymine. Fortunately, this drawback could be overcome by careful adjustment of KMnO₄ concentration and reaction duration to achieve single-hit reaction conditions where, on each labeled DNA molecule, only one of the thymine bases would be modified. After different reaction conditions were fine-tuned (results not shown), the optimal transcription conditions for permanganate footprinting were determined to be RNAP:DNA (10:1), 1x old transcription buffer (50 mM Tris-HCl, pH 8, 10 mM MgCl₂, 10 mM β-mercaptoethanol, 10 μg/mL acetylated BSA), 50 mM KCl, 10 μg/mL heparin, 100 μM NTP, and 4.5 mM KMnO₄. Under this condition, the open complexes as well as the transcribing complexes of N, A, D and S promoters were mapped, and the results from the template strand and non-template strand are shown in Figures 17 and 18, respectively.

Results on Template Strand

The sizes of the open complex bubbles were the same, from ~ -9 to ~ +1, for all four promoters. It should be noted that the permanganate footprinting can only provide the minimum estimate of the bubble size, since the residue at the edge of the bubble may not be a thymine. Nonetheless, we could see the expansion of the transcription bubbles upon the addition of NTP. In the case of the N25 promoter, almost all of the thymine residues on the template strand were modified by KMnO₄, due to the facility at escape and subsequent synthesis of the full-length transcripts by the productive RNAP (Figure 15). Thus, for the N25

Figure 17. Gel image of permanganate footprinting experiments on 5'-template strand end-labeled promoter fragments derived from plasmid DNA.

The DNA was labeled on the template strand, so the transcription bubble expanded in the downward (\downarrow) direction. Each reaction contained 2000 cpm/fmol of DNA. After various reaction conditions such as RNAP:DNA ratios, KMnO_4 concentrations, reaction durations for +NTP transcription and KMnO_4 treatment, were tested for single-hit kinetics (results not shown), the following reaction conditions were found to generate optimal permanganate footprints: RNAP:DNA (10:1), old transcription buffer (pH 8.0), 50 mM KCl, 10 $\mu\text{g/mL}$ heparin, 100 μM NTP and 4.5 mM KMnO_4 for template-strand-labeled DNA. The incubations proceeded at 37 °C. Into the DNA/transcription buffer/KCl solution was added the RNAP and the open complex was allowed to form for 10 min. Then an aliquot of the reaction mixture was added to heparin+ KMnO_4 for open complex footprint and heparin+NTP added to the rest of the open complex mixture. After one minute reaction time for each, KMnO_4 Stop Solution was added to the open-complex footprinting reaction and KMnO_4 added to the transcribing-complex mixture. From the latter, aliquots were withdrawn at 30 sec and 2 min, and added to KMnO_4 Stop Solution. The upstream edges of the DNA bubbles were the same for all four promoter templates as shown by the red line. The downstream edges (the blue line) for the four promoters were different. Although the sizes of the open complex bubbles were similar, the RNAP formed the largest initial transcribing complex bubbles on VLAT-producing DG203 and $\text{SP}_{\text{fullcon}}$.



Key

M – Maxam-Gilbert (G+A) Sequencing Ladder

1 – DNA only control

2 – DNA + KMnO_4 control

3 – Open Complex (1 min)

4 – Transcribing Complex (30 sec)

5 – Transcribing Complex (2 min)

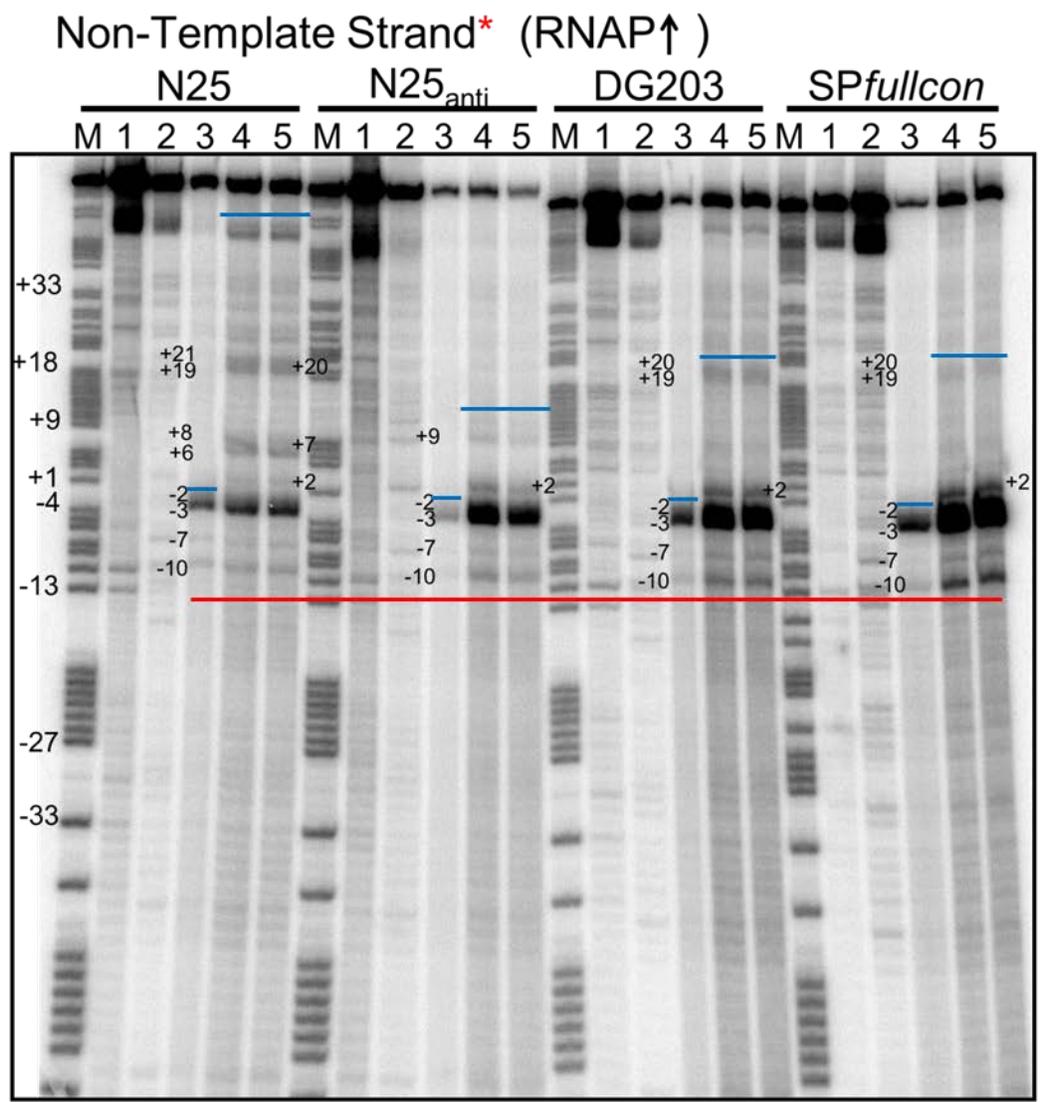
In parentheses are the reaction durations of KMnO_4 .

template, we were visualizing all of the elongation complexes as well (Figure 17). By contrast, on N25_{anti} promoter which escapes very little and performs abortive initiation up to +15 (Figure 15), the transcribing complex bubble extends from -9 to +8 only. (The next thymine residue on N25_{anti} template strand is at +22). On both DG203 and SP_{fullcon} promoters, the transcribing complex bubbles ranged from -9 to +17 (Figure 17). Compared to N25 promoter, few full-length transcripts were generated from DG203 and none at all from SP_{fullcon} (Figure 15). Thus, the transcription bubbles captured were of initial transcribing complex origin on SP_{fullcon}.

Results on Non-template Strand

Without further adjustment to the reaction conditions, footprinting experiments were conducted for promoters labeled on the non-template strands. Unfortunately, careful inspection of the results revealed two features that suggested that the reaction conditions might have exceeded single-hit requirement: 1) the intensity of the DNA bands on the open-complex lanes were not as strong, and 2) the thymine residues (+12, +14 and +15) on N25_{anti} that were expected to be captured in the initial transcribing complex bubbles were not detected (Figure 18). Therefore, to optimize the reaction conditions, KMnO₄ (Figure 19) and NTP (Figure 20) titration reactions were conducted on N25_{anti} promoter. The results indicated that much clearer footprints were obtained from 1 mM KMnO₄, confirming that at 4.5 mM KMnO₄, we have exceeded the single-hit conditions.

Figure 18. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled promoter fragments derived from plasmid DNA. The DNA was labeled on the non-template strand, so the transcription bubble expanded in upward (\uparrow) direction. Each reaction contained 667 cpm/fmol of DNA. The reactions were carried out under the same conditions as for template-strand-labeled promoter templates as described in the legend to Figure 17. The upstream edges of the DNA bubbles were the same for all four promoter templates as shown by the red line. The downstream edges (the blue line), however, were different: although the sizes of the open complex bubbles were similar, the RNAP formed the largest initial transcribing complex bubbles on VLAT-producing DG203 and SP_{fullcon}. Due to the observed weaker open-complex footprints and the smaller-than-expected size of the N25_{anti} initial transcribing complex bubble, we suspected that the single-hit reaction conditions were not met for non-template-strand-labeled promoters. Therefore, KMnO₄ and NTP titrations were carried out to investigate this possibility.

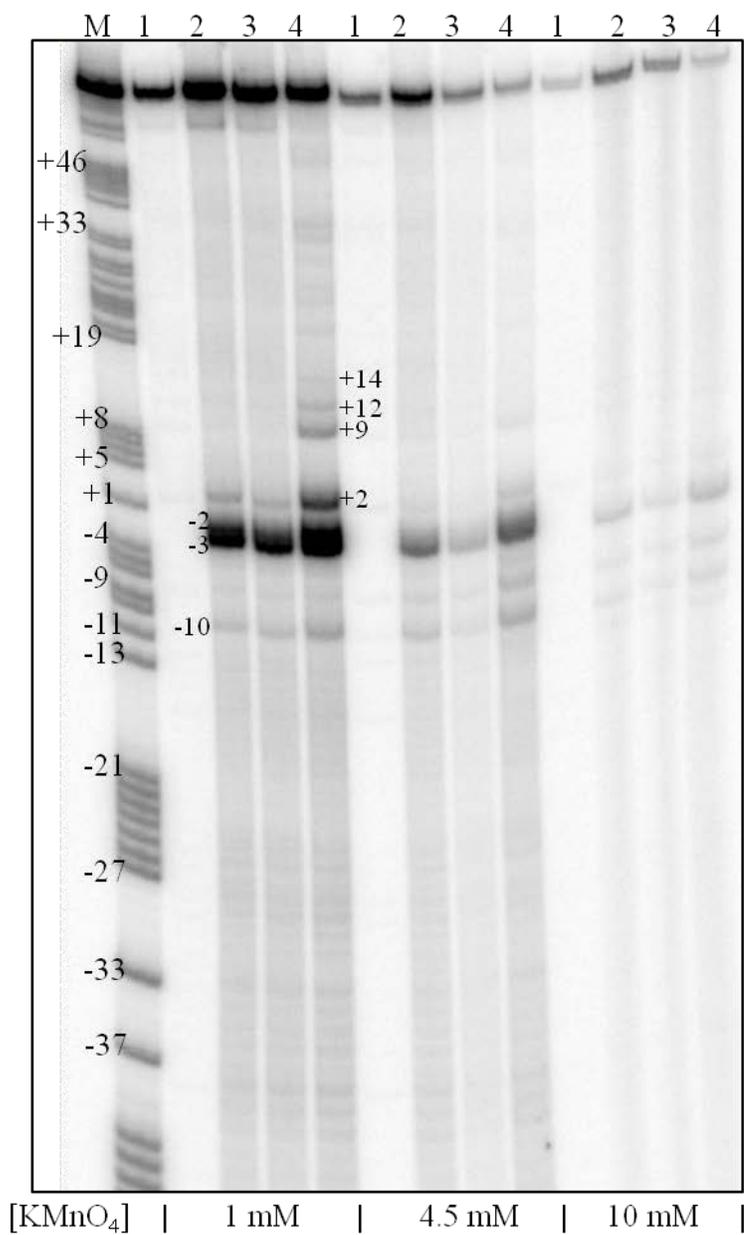


Key

- M – Maxam-Gilbert (G+A) Sequencing Ladder
- 1 – DNA only control
- 2 – DNA + KMnO₄ control
- 3 – Open Complex (1 min)
- 4 – Transcribing Complex (30 sec)
- 5 – Transcribing Complex (2 min)

In parentheses are the reaction durations of KMnO₄.

Figure 19. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled N25_{anti} promoter fragment prepared from plasmid DNA: KMnO₄ titration. The DNA was labeled on the non-template strand, so the transcription bubble expanded in upward (↑) direction. Each reaction contained 1100 cpm/fmol of DNA. This set of experiments was carried out with different concentrations of KMnO₄ (1, 4.5 and 10 mM) to find conditions consistent with single-hit kinetics. The reaction conditions were RNAP:DNA (10:1), 10 μg/mL heparin and 0.1 mM NTP. The incubation temperature was 37 °C and durations were 10 min for RNAP, 30 sec after heparin+NTP addition, 30 sec and 1 min for KMnO₄ reaction with the open complex and 30 sec for KMnO₄ reaction with the transcribing complex. Only the lowest concentration of KMnO₄ (1 mM) showed the single-hit condition.



Key

M – Maxam-Gilbert (G+A) Sequencing Ladder

1 – DNA+KMnO₄ control

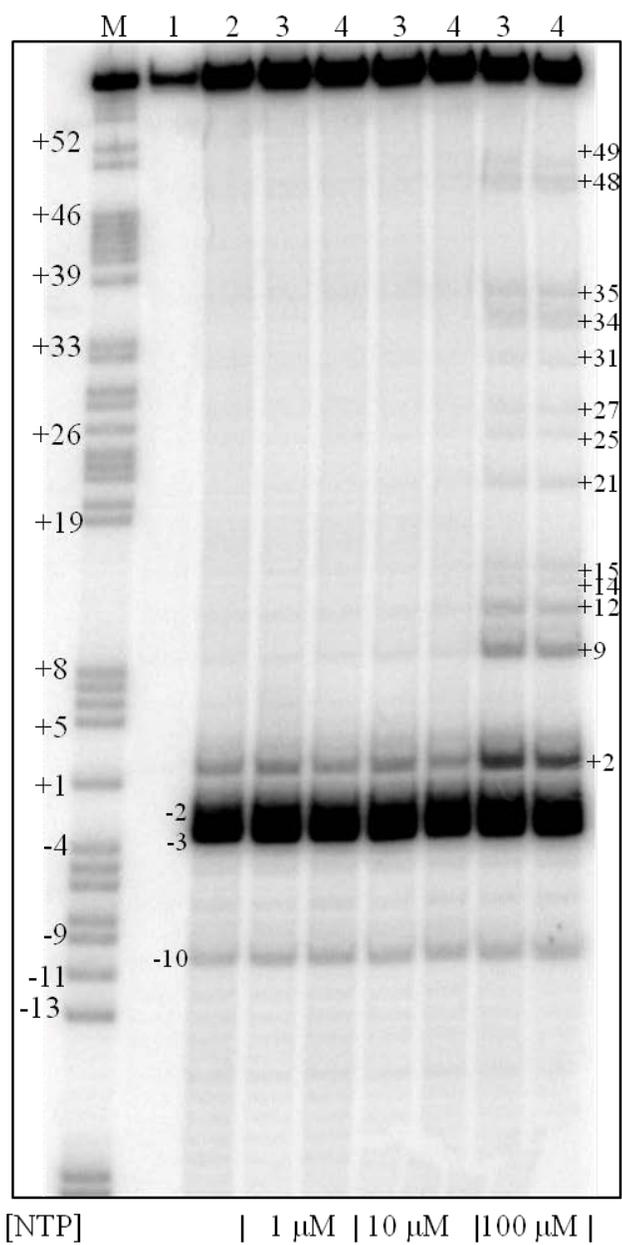
2 – Open complex (30 sec)

3 – Open complex (1 min)

4 – Transcribing complex (30 sec)

In parentheses are the reaction durations of KMnO₄.

Figure 20. Gel image of permanganate footprinting experiments on 5'-nontemplate strand end-labeled N25_{anti} promoter fragment prepared by plasmid labeling: NTP titration. The DNA was labeled on the non-template strand, so the transcription bubble expanded in the upward (↑) direction. Each reaction contained 1100 cpm/fmol of DNA. This set of experiments was carried out with different concentrations of NTP (1, 10 and 100 μM). The reaction conditions were RNAP:DNA (10:1), 10 μg/mL heparin and 1 mM KMnO₄. The incubation temperature was 37 °C and durations were 10 min for RNAP open complex formation, 30 sec for heparin+NTP addition, 30 sec for KMnO₄ reaction with open complex, and 30 sec and 2 min for KMnO₄ reaction with the transcribing complex. Only the highest concentration of NTP (0.1 mM) allowed the RNAP to synthesize RNA, undergo promoter escape and transition into the elongation phase.



Key

M – Maxam-Gilbert (G+A) Sequencing Ladder

1 – DNA+KMnO₄ control

2 – Open complex (30 sec)

3 – Transcribing complex (30 sec)

4 – Transcribing complex (2 min)

In parentheses are the reaction durations of KMnO₄.

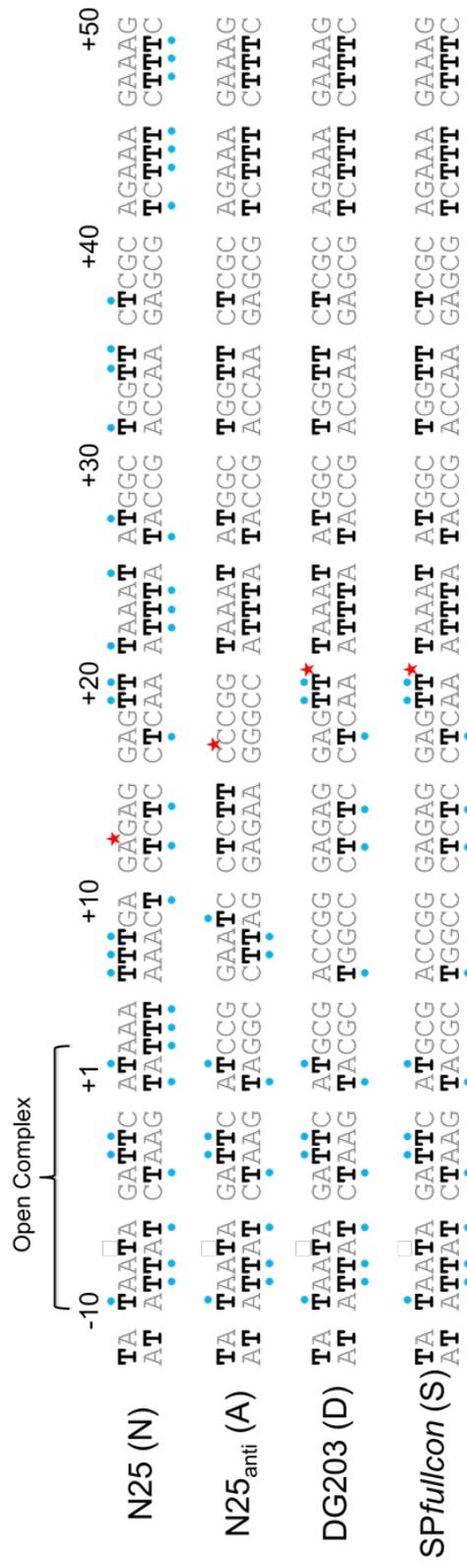
When the non-template strand was subsequently footprinted with 1 mM KMnO₄, and with higher concentration of NTPs which enabled more transcription-translocation-scrunching, the aforementioned T's were all captured (Figures 19 and 20). It is interesting that the template and non-template strand actually displayed different [KMnO₄] requirements for single-hit kinetics; the differences correspond with the more accessible nature of the non-template strand and the more “buried” nature of the template strand in the open/transcribing complexes (Murakami et al., 2002b). However, under any permanganate reaction conditions, the modification of the thymine residue at -7 was not observed on any of the promoters. This puzzle was solved when we realized that the non-template strand -7 T was involved in σ_2 /-10 element interaction and the -7 T was flipped out and bound into a specific protein pocket, presumably rendering it protected from KMnO₄ modification (Feklistov and Darst, 2011).

Summary

The data from the permanganate footprinting experiments were summarized in Figure 21 – the footprinted thymine residues are marked with blue dots. The results obtained from permanganate footprinting experiments were remarkably consistent with the data gathered from the *in vitro* transcription reactions (see Figure 15). It can be concluded that abortive initiation involved DNA scrunching on N25_{anti} and VLAT-producing promoters (D and S), the RNAP scrunched in excess amount of DNA during the initiation phase due to its

delayed escape on DG203 and SP_{fullcon} promoters, and the stress generated during this scrunching might be responsible for propelling the RNAP forward to produce GreB-resistant VLATs.

Figure 21. Summary of permanganate footprinting experiments. The non-template and template strand DNA sequences from -12 to +50 are shown for footprinting promoters. The modifiable thymine residues (T's) on both strands are in darker color. The red star denotes the position of the promoter escape (see Figure 15). The sizes of the open complexes were the same whereas the T's modified in the presence of NTP during permanganate reactions are marked with blue dots. The sizes of transcribing complex bubbles captured in permanganate footprinting experiments accurately reflected the proposed mechanism of RNAP hyper-forward translocation and also the biochemical data in Figure 15.



★ Positions of promoter escape

• The thymine residues (T) modified in permanganate reaction

DISCUSSION

In any organism, appropriate spatial and temporal regulation of gene expression is essential for development as well as for responses to environmental stimuli. Even though this regulation occurs at multiple levels, recent studies have found that some of the rapid and major fine-tunings happen at the transition from transcription initiation to elongation in species ranging from *E. coli* to humans (Wade and Struhl, 2008). In spite of being mechanistically different, prokaryotic and eukaryotic transcriptional initiation have analogous steps such as promoter recognition, open complex formation, abortive initiation and promoter escape. Analysis using chromatin immunoprecipitation (ChIP) on a genome-wide scale indicated that in *Drosophila* and mammalian systems, RNA polymerase II (Pol II) is poised at promoter-proximal regions of a large number of genes. This Pol II stalling is also shown to allow rapid and precise adjustment of transcription level to the environmental conditions; the Pol II is released into the elongation phase only when ‘on-signals’ for the genes are present (Nechaev and Adelman, 2008). In parallel, similar studies in *E. coli* revealed that the expressions of certain genes are tightly controlled by keeping their promoter escape steps rate-limiting (Reppas et al., 2006). Contrary to the previous paradigm that transcription initiation is regulated at the step of RNAP recruitment to the promoter, these findings illuminate the important biological roles of abortive initiation and promoter escape.

The phenomenon of abortive initiation or the release of short nascent transcripts by RNAP during transcription initiation was first observed about three decades ago. Since then, a plethora of studies using many different lines of investigation have led to the delineation of a mechanism. In this mechanism, abortive transcripts that range 2-15 nt in length are the products of RNA polymerase bound at a strong promoter, forming complexes that are highly capable of *de novo* initiation but poorly able to escape. Such complexes are caught performing repetitive rounds of abortive initiation prior to escape, producing abortive transcripts via the DNA scrunching–RNAP backtracking mechanism (Cheetham and Steitz, 1999; Kapanidis et al., 2006; Revyakin et al., 2006). Abortive transcripts produced by this mechanism are susceptible to GreB-mediated rescue through RNAP cleavage and re-elongation (Hsu et al., 1995; Opalka et al., 2003)..

There exists potentially another mechanism of abortive initiation by *E. coli* RNA polymerase distinct from the above well-established one; it was first described by Chander et al. (2007). This novel mechanism called RNAP hyperforward translocation was proposed to explain the occurrence of very long abortive transcripts (VLATs) that are 16-19 nt in length. Not only are these RNAs longer than the usual abortive transcripts (2-15 nt), they are also resistant to cleavage-reelongation rescue by GreB. The persistence of VLATs in the presence of GreB led to the hypothesis of an alternate mechanism of abortive initiation (Chander et al., 2007).

The hyperforward translocation mechanism postulates that: 1. The VLATs are produced during the initiation-elongation transition occurring at the +16 to +19 stage; 2. To escape at such a late stage, the RNAP is required to scrunch in excessive amount of DNA (up to ~19 bp) leading to a highly stressed intermediate; 3. Hence when the RNAP-promoter contacts are disrupted, the large amount of stress energy released forces the upstream DNA out of the enzyme and propels the RNAP to move forward by more than 1 nt (i.e. hyper translocation); 4. RNAP hyperforward translocation results in a complex where the RNAP active center has slid past the nascent RNA 3'-OH end, generating an arrested complex; and 5. As a result of RNAP hyperforward translocation, the 3'-OH end of the nascent RNA would be placed further upstream of the active center into the RNA exit channel (Chander et al., 2007).

Work in the Hsu lab in recent years has been focused on gathering evidence to provide proof for this novel abortive initiation mechanism. Four different approaches are undertaken to achieve this goal: 1. Roadblock transcription experiments to monitor RNAP movement through the promoter escape transition; 2. Permanganate footprinting experiments to probe the expansion of the transcription bubble due to scrunching; 3. Exonuclease III footprinting to monitor the changes in upstream and downstream boundaries of the RNAP during early transcription; and 4. Photocrosslinking experiments to locate the 3'-OH end of the VLAT RNA, presumably in the RNA exit channel.

This study describes the evidence I have gathered through the *Roadblock Transcription* approach and the *Permanganate Footprinting* approach, both of which support the hyperforward translocation mechanism. In the Roadblock Transcription experiments, we used the E₁₁₁Q EcoRI as the roadblock protein. This protein was shown previously to stably block the RNAP movement during the elongation phase when the RNAP active center comes within 14 bp of the EcoRI binding site (Pavco and Steege, 1990; 1991). On our EQ-series DNA templates, and under our reaction conditions probing the movement of RNAP during the initiation phase, blocking RNAP forward movement led to the formation of three and sometimes four different RNAs in a cluster. Although stoppage occurs at multiple positions, the lack of full-length transcripts in the presence of EcoRI ascertained that the RNAP forward movement was stalled. Characterizing the various RNAs by GreB susceptibility or resistance revealed that the RNAP could abort nascent transcripts by three means: backtracking, stalling and hyper-forward translocation. The compiled data of roadblock experiments on all EQ-series templates, shown in Figure 13 and Table 4, confirmed that VLATs are products of RNAP forward movement, formed during the promoter escape transition. In this study, we were able to identify the exact stretch of template where escape occurs—at +16 to +20 positions. The initial transcribing complexes at these positions are on the cusp of escape and cannot give rise to stably stalled RNAs, but either backtracked RNA or hyper forward tracked RNAs. In addition, analyzing the distance between the RNAP active

center and the *EcoRI* binding site showed that excessive scrunching induces backtracking, whereas hyperforward translocation can occur only when there is a spatial allowance of 4-5 bp ahead of the escape transition point. Overall, our roadblock transcription experiments confirm the dynamic nature of the RNAP in initial transcribing complexes (ITCs), in contrast to the uniform ternary elongation complex structures (Pavco and Steege, 1990). The dynamic nature of the ITCs is the direct consequence of RNAP translocation during the initiation stage by DNA scrunching (Revyakin et al., 2006).

The occurrence of DNA scrunching in the promoters we analyzed is further confirmed by footprinting experiments with Exonuclease III and potassium permanganate. The Exonuclease III footprinting showed that on $SP_{fullcon}$ promoter, the RNAP upstream boundary in open complex did not shift upon the addition of NTP (Lee and Hsu, unpublished results). However, $KMnO_4$ footprinting showed that the transcription bubble expanded downstream to +20. Since RNAP cannot move off the promoter region, the expanded bubble region must be derived from scrunching which compacts a large amount of DNA within the enzyme generating an enormous amount of stress. It is this stress energy that ultimately leads to escape, causing a fraction of the RNAP to hyper translocate. . In permanganate footprinting, we detected a differential sensitivity to permanganate modification by the template- versus non-template strand of the promoter DNA, consistent with the accessibility of the strands. Thus, the non-template strand, being closer to the surface of the transcribing complex is more

sensitive and readily modified at lower KMnO_4 concentration. The template strand is more “buried” than the non-template strand and is only modified at higher KMnO_4 concentration. We also reconciled the absence of thymine residue modification at the -7 position on the non-template strand (Figures 18 and 21). Recent structural studies showed the involvement of σ_2 domain bound to single-stranded DNA bearing the -10 element sequence (TATAAT); flipping A (-11) and T (-7) into protein pockets is critical during the DNA melting process to form the initial open complex. Therefore, in both open and transcribing complexes, T (-7) would be protected from being modified by KMnO_4 (Feklistov and Darst, 2011). Thus, although permanganate footprinting technique is limiting in that it could only show the minimal size of the transcription complex bubbles (the last nucleotide of the bubble might not necessarily be thymine), the results indicated that the RNAP scrunches in excess amount of DNA (at least up to +20) prior to escape on VLAT-producing promoters, $\text{SP}_{\text{fullcon}}$ in particular. This, in turn, supported the mechanism that the stress generated from excessive scrunching led to unproductive disruption of the upstream promoter contacts.

In conclusion, this study provided several lines of indirect evidence of the hyperforward translocation mechanism by supporting that VLATs were produced from RNAP forward movement during escape transition into elongation phase and that the RNAP scrunches in a larger amount of DNA on VLAT-producing promoters than on N25 and N25_{anti} that did not produce VLATs. The direct evidence is being gathered by *Photocrosslinking the VLATs to the RNAP* –

locating the 3'-end of the VLATs upstream of the active center in the exit channel would be a definitive proof that the RNAP forward translocated by more than one nucleotide to release the VLATs (Wieland, 2012).

REFERENCES

- Barne, K. A., Bown, J. A., Busby, S. J. and Minchin, S. D. (1997). Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended -10' motif at promoters. *EMBO. J.* **16**, 4034–4040.
- Borukhov, S. and Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends Microbiol.* **16**, 126–134.
- Brunner, M. and Bujard, H. (1987). Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO. J.* **6**, 3139–3144.
- Burgess, R. R. and Jendrisak, J. J. (1975). A procedure for the rapid, large-scale purification of *Escherichia coli* DNA-dependent RNA polymerase involving Polymyxin P precipitation and DNA-cellulose chromatography. *Biochemistry* **14**, 4634–4638.
- Carpousis, A. J. and Gralla, J. D. (1980). Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation *in vitro* at the *lacUV5* promoter. *Biochemistry* **19**, 3245–3253.
- Carpousis, A. J. and Gralla, J. D. (1985). Interaction of RNA polymerase with *lacUV5* promoter DNA during mRNA initiation and elongation. Footprinting, methylation, and rifampicin-sensitivity changes accompanying transcription initiation. *J. Mol. Biol.* **183**, 165–177.
- Carpousis, A. J., Stefano, J. E. and Gralla, J. D. (1982). 5' nucleotide heterogeneity and altered initiation of transcription at mutant *lac* promoters. *J. Mol. Biol.* **157**, 619–633.
- Chamberlin, M. J., Nierman, W. C., Wiggs, J. and Neff, N. (1979). A quantitative assay for bacterial RNA polymerases. *J. Biol. Chem.* **254**, 10061–10069.
- Chander, M., Austin, K. M., Aye-Han, N-N., Sircar, P. and Hsu, L. M. (2007). An alternate mechanism of abortive release marked by the formation of very long abortive transcripts. *Biochemistry* **46**, 12687–12699.
- Chander, M. and Hsu, L. M. (manuscript in submission). An AT-rich spacer enhances intrinsic and α CTD-dependent release of very long abortive transcripts (VLATs) during promoter escape.
- Cheetham, G. M. and Steitz, T. A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* **286**, 2305–2309.

- Choy, H. E. and Adhya, S. (1993). RNA polymerase idling and clearance in gal promoters: use of supercoiled minicircle DNA template made in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 472–476.
- Engler, C., Gruetzner, R., Kandzia, R. and Marillonnet, S. (2009). Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE* **4**, e5553.
- Estrem, S. T., Gaal, T., Ross, W. and Gourse, R. L. (1998). Identification of an UP element consensus sequence for bacterial promoters. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9761–9766.
- Estrem, S. T., Ross, W., Gaal, T., Chen, Z. W. S., Niu, W., Ebright, R. H. and Gourse, R. L. (1999). Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit. *Genes Dev.* **13**, 2134–2147.
- Feklistov, A. and Darst, S. A. (2011). Structural basis for promoter -10 element recognition by the bacterial RNA polymerase σ subunit. *Cell* **147**, 1257–1269.
- Feng, G. H., Lee, D. N., Wang, D., Chan, C. L. and Landick, R. (1994). GreA-induced transcript cleavage in transcription complexes containing *Escherichia coli* RNA polymerase is controlled by multiple factors, including nascent transcript location and structure. *J. Biol. Chem.* **269**, 22282–22294.
- Gaal, T., Ross, W., Estrem, S. T., Nguyen, L. H., Burgess, R. R. and Gourse, R. L. (2001). Promoter recognition and discrimination by $E\sigma^S$ RNA polymerase. *Mol. Microbiol.* **42**, 939–954.
- Hawley, D. K. and McClure, W. R. (1983). Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res.* **11**, 2237–2255.
- Hayatsu, H. and Ukita, T. (1967). The selective degradation of pyrimidines in nucleic acids by permanganate oxidation. *Biochem. Biophys. Res. Commun.* **29**, 556–561.
- Hayatsu, H. and Iida, S. (1969). Studies on the chemical modifications of nucleic acids. The permanganate oxidation of thymine. *Tetrahedron Lett.* 1031–1034.
- Hsu, L. M. (2002a). Open season on RNA polymerase. *Nat. Struct. Biol.* **9**, 502–504.
- Hsu, L. M. (2002b). Promoter clearance and escape in prokaryotes. *Biochim. Biophys. Acta* **1577**, 191–207.
- Hsu, L. M. (2009). Monitoring abortive initiation. *Methods* **47**, 25–36.
- Hsu, L. M., Vo, N. V. and Chamberlin, M. J. (1995). *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene

expression *in vivo* and *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11588–11592.

Hsu, L. M., Vo, N. V., Kane, C. M. and Chamberlin, M. J. (2003). *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 1. RNA chain initiation, abortive initiation, and promoter escape at three bacteriophage promoters. *Biochemistry* **42**, 3777–3786.

Hsu, L. M., Cobb, I. M., Ozmore, J. R., Khoo, M., Nahm, G., Xia, L., Bao, Y. and Ahn, C. (2006). Initial transcribed sequence mutations specifically affect promoter escape properties. *Biochemistry* **45**, 8841–8854.

Jiang, Y. (2009) Testing the occurrence of forward hyper-translocation during the promoter escape transition. [Senior Thesis]. Mount Holyoke College, South Hadley, MA-01075.

Kammerer, W., Deuschle, U., Gentz, R. and Bujard, H. (1986). Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *EMBO. J.* **5**, 2995–3000.

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

King, K., Benkovic, S. J. and Modrich, P. (1989). Glu-111 is required for activation of the DNA cleavage center of *EcoRI* endonuclease. *J. Biol. Chem.* **264**, 11807–11815.

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

Krummel, B. and Chamberlin, M. J. (1989). RNA chain initiation by *Escherichia coli* RNA polymerase. Structural transitions of the enzyme in early ternary complexes. *Biochemistry* **28**, 7829–7842.

Maxam, A. M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* **65**, 499–560.

McClarín, J. A., Frederick, C. A., Wang, B. C., Greene, P., Boyer, H. W., Grable, J. and Rosenberg, J. M. (1986). Structure of the DNA-*EcoRI* endonuclease recognition complex at 3 Å resolution. *Science* **234**, 1526–1541.

McClure, W. R. (1980). Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5634–5638.

McClure, W. R. (1985). Mechanism and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* **54**, 171–204.

- Mooney, R. A., Darst, S. A. and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Mol. Cell* **20**, 335–345.
- Mulligan, M. E., Hawley, D. K., Entriken, R. and McClure, W. R. (1984). *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. *Nucleic Acids Res.* **12**, 789–800.
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. (2002a). Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285–1290.
- Murakami, K. S., Masuda, S. and Darst, S. A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296**, 1280–1284.
- Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wriggers, W. and Darst, S. A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* **114**, 335–345.
- Pavco, P. A. and Steege, D. A. (1990). Elongation by *Escherichia coli* RNA polymerase is blocked *in vitro* by a site-specific DNA binding protein. *J. Biol. Chem.* **265**, 9960–9969.
- Pavco, P. A. and Steege, D. A. (1991). Characterization of elongating T7 and SP6 RNA polymerases and their response to a roadblock generated by a site-specific DNA binding protein. *Nucleic Acids Res.* **19**, 4639–4646.
- Revyakin, A., Liu, C., Ebright, R. H. and Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314**, 1139–1143.
- Roberts, J. W. (2006). Biochemistry. RNA polymerase, a scrunching machine. *Science* **314**, 1097–1098.
- Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K. and Gourse, R. L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**, 1407–1413.
- Singh, S. S., Typas, A., Hengge, R. and Grainger, D. C. (2011). *Escherichia coli* σ^{70} senses sequence and conformation of the promoter spacer region. *Nucl. Acids Res.* **39**, 5109–5118.
- Steitz, T. A. and Steitz, J. A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 6498–6502.
- Straney, D. C. and Crothers, D. M. (1987). A stressed intermediate in the formation of stably initiated RNA chains at the *Escherichia coli lacUV5* promoter. *J. Mol. Biol.* **193**, 267–278.

- Vallery, T. K. (2010) Investigating the mechanism of very long abortive transcript (VLAT) formation via hyper forward translocation by *E. coli* RNA polymerase (RNAP). [Senior Thesis]. Mount Holyoke College, South Hadley, MA-01075.
- Vassilyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M. N., Borukhov, S. and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712–719.
- Vo, N. V., Hsu, L. M., Kane, C. M., and Chamberlin, M. J. (2003). *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* **42**, 3798–3811.
- Wieland, R. S. (2012) Crosslinking very long abortive transcripts (VLATs) to *E. coli* RNA polymerase: an approach to elucidate forward hyper-translocation. [Senior Thesis]. Mount Holyoke College, South Hadley, MA-01075.
- Yuzenkova, Y., Tadigotla, V. R., Severinov, K. and Zenkin, N. (2011). A new basal promoter element recognized by RNA polymerase core enzyme. *EMBO J.* **30**, 3766–3775.