TESTING THE OCCURRENCE OF FORWARD HYPER-TRANSLOCATION DURING THE PROMOTER ESCAPE TRANSITION

By

Yunnan Jiang

This paper was prepared under the direction of

Dr. Lilian M. Hsu

Professor in Biochemistry

For 8 credits

Research support for this project was provided by a NSF-RUI grant (No. MCB 0418316) awarded to L.M. Hsu

ACKNOWLEDGEMENTS

Firstly, I would like to thank Professor Lilian Hsu. Without your help and wisdom, this project would not have been possible. I owe an immense debt to you for helping me, motivating me and challenging me. Your most invaluable support and kind encouragement helped me through difficult times. Your phenomenal teaching inspired my interest in science and medicine. In everything you taught me in class and in lab, I learned in how to become a better scientist. I cannot thank you enough for all the help you have most generously given me in every possible way and the immeasurable input in preparing me for the journey ahead.

I would like to thank all members of the Hsu lab, especially Nwe-Nwe, Elina, Jessica, Hilary, Sara and Ahri for your help and for making long late nights in the lab exciting and fun.

I would like to thank Professor Wei Chen for your invaluable support all through my time in Mt Holyoke. I am deeply grateful to you for introducing me to research and for helping me discover my interest in science, every step along the way. In the past four years, you have given me incalculable amount of guidance and advice to become a better student and person, for which I am truly thankful.

I would like to thank Professor Craig Woodard for all your support that helped me achieve my goals and for your patience and effort to answer my countless questions since I was a student in your class. I would like to thank Professor Jeffrey Knight for introducing me to the studies of molecular biology, in which I found my passion, and for all the thoughtful advice you have given me all through my time at Mt Holyoke.

Thank you to Professor Margaret Robinson for your support and encouragement in helping me make crucial decisions especially in the past year, and for helping me realize my journey that begins after Mt Holyoke. Thank you to Professor Sheila Brown for helping me realize my dream and the tremendous amount I learned from you.

I would like to thank Mt. Holyoke College Biology department and Dianne Baranowski in the Biochemistry department.

I would like to thank my friends and family members for sharing my joy, my dream, my difficulties and my silliness and most importantly for always believing in me.

Lastly I want to thank my parents for giving me the maximum freedom to realize my dreams and for always being there to support me. I could not have been more fortunate as your daughter. Mom, for risking your life to give birth to me and for

accepting all my flaws. Dad, for showing me all my flaws and teaching me to constantly challenge myself. Thank you for loving me the way you have.

最后，感谢我的父母。谢谢你们给我最广阔的发展空间，帮助我实现我的梦想。能做你们的女儿是我最大的幸福。妈妈，谢谢你冒生命危险给我一次人生的机会，和一直以来对我最大的包容。爸爸，谢谢你教我认识自我，不断进取。谢谢你们对我无保留的爱。

TABLE OF CONTENTS

LIST OF FIGURES

List of Tables

ABSTRACT

The macromolecular process of RNA transcription is composed of three phases, initiation, elongation and termination. At many promoters, RNA polymerase (RNAP) repetitively aborts short RNAs 2 to 15 nucleotides (nts) in length before successfully making the transition to the elongation phase. Mutations in the initial transcribed sequence (ITS) region, from +3 to +10, of $E\sigma^{70}$-dependent *T5 N25* promoter lengthen the abortive initiation program, and lead to the formation of the very long abortive transcripts (VLATs) of 16 to 21 nts. Recent published work done in the lab suggests that the VLATs formation may result from forward hyper-translocation movement of RNAP during the escape transition. This project is aimed to test the physical occurrence of forward hyper-transcription by binding a non-cleaving $E_{111}Q$ *Eco*RI protein roadblock to stall the forward translocating RNAP and thus prevent the formation of VLATs.

In this experiment, the front edge of RNAP is 14 bp downstream of the active site, and would bump into a bound *Eco*RI dimer whose binding site is centered 5 bp downstream. Thus, the predicted position of the VLAT19 roadblock would be at (19+14+5 =) +38. I constructed six templates with an *Eco*RI site at +25, +30, +34, +45, +55, and +60 by PCR. (All templates run off at +60, so the +60 construct is a control without an *Eco*RI site.) Transcription with or without a 10:1 molar excess of *Eco*RI (to DNA) showed the evidence of roadblock activity, in that the full-length transcript is greatly diminished in the presence of *Eco*RI.

We also obtained elevated abortive RNA synthesis at 6-8 nt, 11-13 nt, and 15-17 nt, predicted for a roadblock placed at the +25, +30, and +34 templates, respectively. However, we observed VLATs formation on all of the templates, indicating that the $E_{111}Q$ *Eco*RI roadblock activity was not complete. This result raised the possibility that the early *Eco*R I roadblocks cannot completely stall the "powerful" forward-tracking RNAP. To validate this conclusion, I tested the stability of the bound Q_{111}*Eco*RI protein before and during transcription using electrophoretic mobility shift assay (EMSA). EMSA experiments showed restriction cutting of $E_{111}Q$ *Eco*RI in high Mg^{+2} and pH environment, which can account for the incomplete stalling. We found conditions that minimize the restriction cutting activity of Q_{111}*Eco*RI, and will be re-testing RNAP forward hyper-translocation under conditions that ensure the stable presence of the non-cleaving $E_{111}Q$ *Eco*RI roadblock.

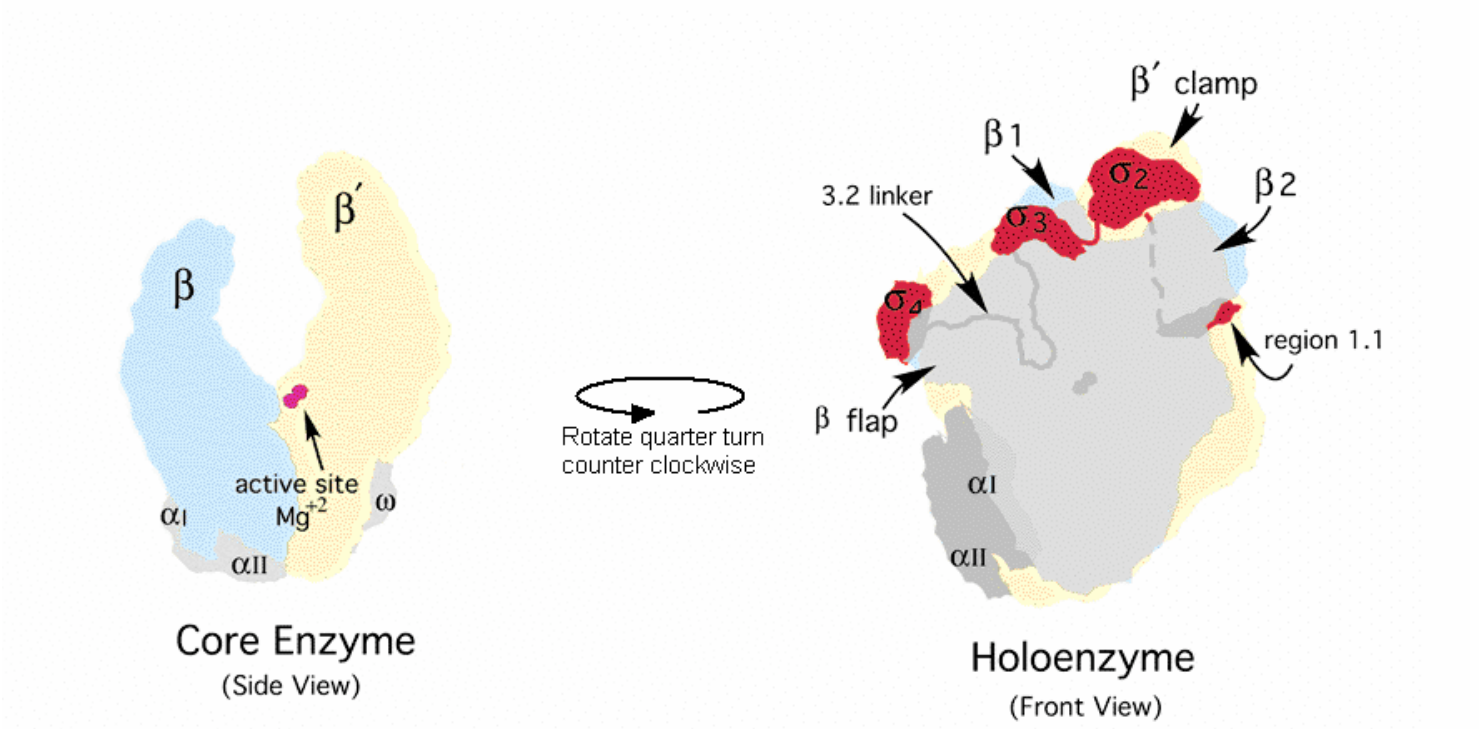
INTRODUCTION

Transcription is the first step of gene expression. DNA is transcribed into RNA by RNA polymerase (RNAP) according to the complimentary base pairing rule. A cycle of transcription involves three phases: initiation, elongation and termination. During initiation, RNAP binds to the promoter region of a gene and starts *de novo* synthesis of RNA oligonucleotides. During elongation, RNAP continues to add nucleotides to the growing transcript in the 5' to 3' direction. Finally, RNAP stops transcription upon encountering a termination signal, releases the full-length RNA transcript, and disassembles from the template, thus concluding a round of transcription.

Of the three phases, transcription initiation is the most highly regulated. Its regulation is crucial for the proper timing and extent of gene expression. In this study, we will focus on transcription initiation with *E. coli* bacterial RNAP.

The *E. coli* bacterial RNA polymerase (RNAP) is a multisubunit enzyme that exists in two forms, the core enzyme and holoenzyme as shown in Figure 1. The core enzyme consists of four different subunits, α, β, β', and ω, in the molar ratio of 2:1:1:1. The holoenzyme is comprised of the core enzyme ($\alpha_2\beta\beta'\omega$) and a σ factor, which is responsible for promoter recognition. While the core enzyme can start transcription on any stretch of DNA sequence, the holoenzyme initiates transcription only at promoters. There exist a variety of σ factors in *E. coli* that differ in their molecular weight and the specific class of gene promoters they

Figure 1. *RNA polymerase is found in the core and holoenzyme form.* Cartoon depiction of *E. coli* RNA polymerase core enzyme (left) and holoenzyme (right). It is apparent in this diagram that the core structure resembles a crab claw formed by β and β' subunits around a central channel, where the active site is located in the middle of the channel on the floor. Subunits α_1, α_2 and ω sit on the underside of the enzyme. In the holoenzyme σ^{70} rests along the top of the complex almost as if holding the "claw" together.


β´
β
active site
Mg+2
ω
αI
αII
Core Enzyme
(Side View)
Rotate quarter turn
counter clockwise
β´ clamp
β1
σ2
β2
3.2 linker
σ3
σ4
region 1.1
β flap
αI
αII
Holoenzyme
(Front View)

recognize. In this study, we utilize holoenzyme containing sigma-70 ($E\sigma^{70}$, M.W. of σ^{70} is 70 KDa), the major σ subunit in *E. coli*. $E\sigma^{70}$ is responsible for transcribing the housekeeping genes, which comprise approximately 70% of *E. coli* transcription during normal growth (Helmann and Chamberlin, 1988).

More than 300 $E\sigma^{70}$ promoters have been identified (Lisser and Margalit, 1993). An $E\sigma^{70}$ promoter is usually 80 bp long, spanning nucleotides -60 to +20 in reference to the +1 position which corresponds to the start site of transcription. All transcribed sequences are labeled with positive numbers which increase in the 5' to 3' direction and include the initial transcribed sequences (ITS) spanning nucleotides +1 to +20. Promoter sequences upstream of the transcription start site are untranscribed and given negative numbers that increase with distance from the start site. The upstream region spanning -60 to -1 is designated the promoter recognition region (PRR). Within PRR, there exist two hexameric sequences around the -35 and -10 positions called the -35 and -10 elements (also denoted as the -35 and -10 boxes). Sequence compilation and analysis of hundreds of *E. coli* promoters revealed sequence conservation at these two hexameric regions, yielding a consensus of TTGACA for the -35 element and TATAAT for the -10 element. The two boxes are known to interact with σ^{70} through major groove contacts and are separated by a spacer region, optimally at 17 bp to accommodate the twist in DNA double helix. Another RNAP and DNA interaction region exists

Figure 2. The non-template strand nucleotide sequence of *T5 N25* and *T5 N25*$_{anti}$ sequence. The bold regions represent the -10 and -35 elements. *N25* and *N25*$_{anti}$ share identical upstream sequences from -60 to -1, but differ from +3 to +20 region with A ↔ C and T ↔ G alteration.

***T5 N25* promoter**

-60 -35
5’ TCGAGGGAAA TCATAAAAAA TTTAT**TTGCT**
-10
TTCAGGAAAA TTTTTCTG**TA TAAT**AGATTC
+1
ATAAATTTGA GAGAGGAGTT TAAATATGGC
+67
TGGTTCTCGC AGAAAGCTTC TGCAGGTACC 3’

ITS (+1 ~ +20)
T5N25: 5’-^{1}ATAAATTTGA GAGAGGAGTT^{20}-3’
T5N25$_{anti}$: 5’-^{1}ATCCGGAATC CTCTTCCCGG^{20} -3’

downstream of the -10 region, termed the discriminator region (DIS) that is 6 to 8 bp in length. Some promoters also contain an A/T-rich sequence, known as the UP element upstream of the -35 region. An example of promoter sequences is shown in Figure 2.

Transcription initiation occurs in two stages. In the first stage, RNA polymerase binds to the promoter to form a closed complex (RP_c) that is catalytically inactive. Subsequently, a ~13-bp stretch of DNA from -11 to +2/+3 is melted to form an open complex (RP_o) that is catalytically active for phosphodiester bond formation. The open complex is poised for *de novo* initiation.

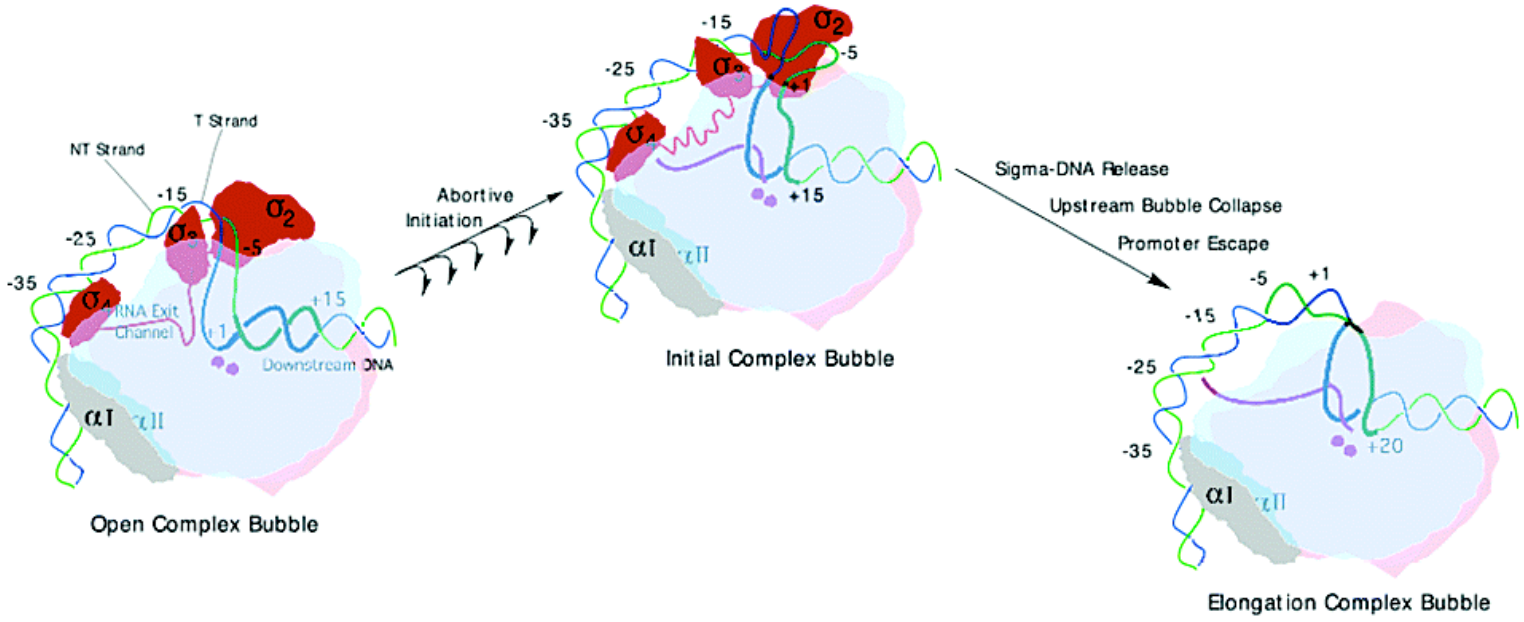
In the second stage, upon the addition of free nucleoside triphosphates (NTP), transcription complex starts template directed *de novo* synthesis of RNA transcripts of 2- to 15-nt long. At many promoters, the RNA polymerase repetitively releases short RNAs 2~15 nt long called abortive transcripts. This process is called abortive initiation, which results from the repetitive cycling of RNA polymerase to the open complex conformation after releasing the short RNA sequences. Abortive initiation ceases when RNAP successfully negotiates promoter escape which is characterized by the collapse of the open complex transcription bubble following the release of RNAP-promoter DNA contacts, RNAP translocation away from the promoter region, and proper location of the nascent transcript in the RNA exit channel. The escape from abortive cycling by RNAP to productive synthesis is called promoter escape, now known as the

transition step from transcription initiation to elongation.

Scrunching model

Structural analysis showed that transcription initiation utilizes a scrunching model of translocation (Kapanidis *et al*., 2006; Revyakin *et al*., 2006) as shown in Figure 3. RNAP binds to the -35 and -10 element of the promoter and stays bound during the entire transcription initiation process. A 14 nt stretch of DNA is melted from -11 to +2/+3 to create a transcription bubble. Upon the addition of NTPs, RNAP continues to unwind and pulls in the downstream DNA while remaining bound to the PRR region. This results in enlargement of the transcription bubble. As more and more DNA is scrunched in, there is increasing amount of stress accumulated in the transcription complex. This stress eventually drives the rewinding of upstream promoter region and the energy released dislodges the PRR region from RNAP. The concurrent events of the collapse of the open complex bubble and promoter release marks promoter escape. When RNAP cannot successfully negotiate promoter escape, abortive transcripts are formed.

Figure 3. Schematic diagram illustrating the scrunching mechanism during transcription initiation. Image borrowed from Hsu et al., 2006. During transcription initiation, a 13 base pair DNA region is melted to create a transcription bubble which is expanded during initiation followed by upstream bubble collapse during initiation-elongation. This diagram illustrates the expansion of the transcription initiation and its collapse to form the elongation complex. The subunits of RNAP are depicted in color as follows: gray, α; blue, β; pink, β'; and red, σ. Mg^{2+} catalytic center is marked purple. Nucleic acids are depicted in color as follows: NT strand, green; T strand, blue; and RNA, purple. The open complex bubble is denoted in thick lines. The ITS region is highlighted in thickest lines.


NT Strand
T Strand
-15
-25
-35
-5
+15
+1
RNA Exit Channel
Downstream DNA
Open Complex Bubble
Abortive Initiation
Initial Complex Bubble
Sigma-DNA Release
Upstream Bubble Collapse
Promoter Escape
+20
Elongation Complex Bubble

Kinetic Analysis

The multi-step transcription initiation reaction can be examined by kinetics analysis as shown in Figure 4. In the first stage of transcription, the binding of RNA polymerase with promoter to form an RP_c is controlled by an equilibrium constant, K_B. The isomerization of RP_c to RP_o is described with a forward reaction rate constant, k_2, and a reverse rate constant, k_{-2}. On many promoters, as with the case of *T5 N25* promoter, the formation of RP_o is stable enough that its formation is irreversible; thus, k_{-2} is negligible. In assessing the activity of a promoter, the first stage of promoter binding and open complex formation can be described by $K_B \times k_2$, an in vitro indicator of promoter strength and chain initiation frequency. It is known that large $K_B \times k_2$ is associated with promoters containing consensus-like -35 and -10 elements (McClure, 1985). The more consensus-like the core elements, the stronger binding and faster formation of the open complex.

The second stage of transcription initiation involves abortive initiation and promoter escape and is shown as a multi-step process between RP_o and the stable ternary elongation complex (TEC). Each step involves the incorporation of a new phosphodiester bond, followed by (scrunching) translocation to realign the polymerase active center with the 3'-OH end of RNA so that the next NTP can be brought into the active site for incorporation (Vassylyev *et al.*, 2007). During this nucleotide incorporation cycle, RNAP can abort transcription by releasing the nascent transcript. Differential stress in the initial transcribing complexes (ITC)

Figure 4. *Kinetic diagram describing transcription initiation*. In this figure R represents RNA polymerase, P represents promoter. RP_C represents the closed complex, RP_O represents the open complex. TEC represents the transcription elongation complex. In short, binding occurs of R and P to form $RP_{C,}$ isomerization takes place to form RP_O, and upon addition of NTP, the process can proceed through multiple steps of abortive initiation towards the TEC complex. The process is further explained in the text.

R + P
k1
k-1
RPc
k2
k-2
RPo
NTP
2
3
4
5
6
kE
?
TEC
σ

leads to stochastic release of the nascent RNAs until promoter escape occurs. Therefore, the longest abortive transcript represents the position of promoter escape (Hsu *et al*., 2003). The overall rate constant governing the transition from RP_o to TEC is k_E. In this study, we will focus our attention on the second stage of transcription initiation – abortive initiation and promoter escape.

Abortive initiation and promoter escape.

The production of abortive transcripts was first observed in transcription reactions containing only the first two complementary NTP substrates. Missing the full sets of NTPs prevented the elongation of RNA transcripts, but also resulted in unexpected release of 2-nt nascent RNA (Johnston and McClure, 1976). To validate the existence of abortive transcripts, Gralla and coworker utilized high-resolution gel electrophoresis and captured abortive products of 2 ~ 6 nt long in the presence of all four NTPs (Carpousis and Gralla, 1980). The study also showed that only a fraction of *de novo* initiated chains can reach full length. This observation led to the proposal of a cycling model in which RNAP cycles to produce short RNA transcripts during the early steps of transcription initiation and that the production of a long RNA transcript is essentially an escape from the cycling process. This study gave way to a series of analysis on abortive initiation.

The observed abortive cycling indicates the existence of a rate-limiting step at the conversion from nascent RNA production to full-length RNA synthesis, now known as the promoter escape stage. Carpousis and coworker

measured the rate constants of *lac*UV5 and T7 promoters by monitoring the frequency of formation of the first phosphodiester bond. They found that on *lac*UV5 promoter, the slowest step is after the formation of the open complex, therefore validating a rate limiting stage at promoter escape (Gralla and Carpousis, 1980).

Among σ^{70} promoters, *T5 N25* promoter is found to be highly abortive. *T5 N25* readily forms a stable open complex with a large $K_B \times k_2$, but has a small k_E, therefore it is rate limited at promoter escape. It is observed that *T5 N25* produces 97% abortive transcripts and only 3% productive synthesis (Hsu *et al*., 2003). In this study, we utilize *T5 N25* to understand the mechanism of abortive initiation and promoter escape due to its highly abortive nature.

PRR regulates total chain initiation.

There are several factors that affect abortive initiation and promoter escape. Promoter escape depends on how well RNAP is anchored at the PRR region (-60~-1). During promoter recognition, σ2, σ3 and σ4 domains bind to the -35 to -10 region connected by a 17 base pair (bp) spacer – σ2 to -10, σ3 to -14/15, and σ4 to -35 (Murakami *et al*., 2002). In addition, σ1.2 has been reported to bind to the -5G or -7G base in the discriminator region connecting the -10 box to +1 (Haugen *et al*., 2006; 2008). On some promoters, the UP element interacts with the α-CTD of RNAP (Ross *et al*., 1998). Footprint analysis showed that

promoter escape is accompanied by releasing all of these contacts (Krummel and Chamberlin, 1992).

Strong PRR-polymerase binding presents a barrier for promoter escape by preventing RNAP from releasing the PRR contacts. Analysis with *lac*UV5 and *T5 N25* promoter variants that differ in -35 and -10 elements showed that the closer the promoter sequence is to the consensus, the poorer the escape (Vo *et al.*, 2003). The consensus sequence is associated with stronger binding, thus poor promoter release. A polymerase that cannot undergo promoter escape releases abortive transcripts, thereby increasing the abortive to productive RNA ratio (APR) on a given promoter.

In the case of *T5 N25* promoter, its highly abortive nature is due to the near-consensus -35 element (TTGCTT) and the consensus -10 element (TATAAT). This explains the large K_B x k_2 value and small k_E, making *T5 N25* rate limited at promoter escape. Further sequence variation showed that the -10 and -35 elements influence the abortive probability at positions 6~10 and 11~15, respectively (Vo *et al.,* 2003). The discriminator region and spacer region influence the abortive probability at positions 2~5 and 11~15, respectively.

ITS regulates the relative extent of abortive to productive synthesis

The effect of ITS on promoter escape was first observed by Bujard and cowokers on *T5 N25* promoter (Kammerer *et al.*, 1986). They constructed an "anti" version of the ITS (involving A↔C and G↔T changes) which rendered

the *N25* promoter 10 fold weaker in mRNA synthesis without affecting its K_B x k_2. This result showed that ITS affects transcription initiation after the open complex formation step.

In this study, we call this "anti" version the $N25_{anti}$ promoter. Further analysis showed that $N25_{anti}$ and *N25* perform comparable total transcription initiation but very different abortive to productive ratio (Hsu *et al.,* 2003). The reduced amount of full-length RNA synthesis on $N25_{anti}$ results in increased amount of abortive synthesis. On an escape rate-limited promoter, ITS affects the extent of productive synthesis vs. abortive initiation.

It was also found that the "anti" conversion extended the abortive ladder from 10 nt at *N25* promoter to 15 nt at $N25_{anti}$ promoter (Hsu *et al.*, 2003). Since the longest abortive transcript represents the position of promoter escape, changing the ITS sequence to that of *N25anti* delayed promoter escape from +10 to +15. The authors postulated that this delay may result from higher CG content in the ITS sequence of $N25_{\text{anti}}$. The additional hydrogen bond between CG increases the energy required to separate the base pairs and keep it open after escape during elongation. Thus the high CG content in $N25_{anti}$ ITS sequence makes it harder to undergo promoter escape.

To further understand the effect of ITS on promoter escape, Hsu *et al.* (2006) constructed ~40 randomized ITS variants of *N25* promoters which rendered a ~25 fold change in productive synthesis from the naturally occurring *N25* promoter. An inverse relationship is found between total transcription

initiation and productive to abortive ratio among the ITS variants, suggesting that failure at promoter escape results in repetitive abortive initiation. The study also showed that the higher the purine content in the nontemplate strand (NT), the lower the APR. All of the ITS showed specific abortive profiles suggesting that ITS has direct impact on the length of abortive initiation and amount of abortive synthesis at each position.

On the whole, sequence variation indicates that PRR determines the rate-limiting step, whether it is at promoter binding or promoter escape. ITS fine-tunes the transcription initiation-elongation transition by altering the extent of productive synthesis vs. abortive transcription.

Mechanisms of abortive initiation

There are several known mechanisms of abortive initiation. Short abortive transcripts (<5 nt) are released due to the highly unstable nature of the early transcription complex. The scrunching process involves significant accumulation of stress making the initial transcription complex highly unstable and dynamic. As each additional base pair is scrunched into the transcription bubble, the additional stress is stabilized by the extension of the RNA-DNA hybrid (heteroduplex). RNA-DNA hybrid of $\geq$ 5 bp will stabilize a backtracked complex from releasing abortive transcripts. Short abortive transcripts less than 5 nt are quickly released. (Hsu *et al*., 2006).

Nascent transcripts longer than the heteroduplexed length (of 8 bp) bump into the σ3.2 linker that occupies the RNA exit channel (Kulbachinskiy and Mustaev, 2006). At this length, the 5' end of the nascent RNA starts to peel off from the template strand and is directed into the RNA exit channel. The σ3.2 linker occupies the RNA exit channel forming a physical barrier for promoter escape as shown in Figure 1. If the nascent RNAs fail to dislodge the σ3.2 linker, they are released as abortive transcripts. However σ3.2 cannot be the sole cause of abortive initiation, as abortive transcripts longer than 8 nt are frequently observed among $E\sigma^{70}$ promoters.

GreB studies showed that abortive transcripts ≤15 nt are formed by backtracking. GreB is a transcription factor that can bind to the secondary channel of the RNAP and stimulate the intrinsic cleavage activity of the active site. In the presence of GreB, the level of abortive transcripts up to15 nt long is greatly diminished; this is accompanied by increased amount of full-length transcript (Hsu *et al.*, 2006). GreB mediated cleavage and rescue proves that before their release, abortive transcripts are bound up in a backtracked complex. Backtracking results in a 3' tail extruding into the secondary channel with the 3'-OH of the RNA transcript positioned away from the active site unable to incorporate a new NTP. The 3' tail is cleaved in the presence of GreB and the 5' end is rescued by being extended to the full length. However GreB rescue requires stable bonding interactions between RNA and DNA. Therefore short abortive transcripts (<5 nt)

with less than 5 bp of RNA-DNA heteroduplex in the backtracked RNA are quickly released before GreB-mediated rescue can occur.

Very Long Abortive Transcripts (VLATs)

A subset of random-ITS variants that differ from *N25* only from +3 to +10 were found to produce abortive transcripts up to 19- or 20-nt long (Chander *et al.*, 2007). Abortive transcripts 16~19 nt long are very unusual for two reasons. First, the nascent RNA should have maximized its binding stability with the enzyme by 16 nt and would be emerging from the exit channel to the enzyme exterior (Komissarova and Kashlev, 1998). Second, abortive transcripts 16 to 19 nt long were not rescued by GreB presence indicating they were not formed by backtracking (Chander *et al.*, 2007). These results suggest that abortive transcripts 16~19 nt are formed by a mechanism entirely different from abortive transcripts studied before (≤16 nt). From here on, we shall refer to abortive transcripts 16~19 nt as very long abortive transcripts (VLATs). Most of the VLAT studies were conducted on the *DG203* promoter, which is a specific ITS variant of *N25* at the +3 to +10 region.

VLATs were confirmed to be abortive transcripts because, unlike pause transcripts, they are not susceptible to high [NTP] chase to form the full-length RNA. Transcription with RNAP immobilized on agarose beads confirmed that they were released RNAs (Chander *et al.*, 2007).

Studies showed that the production of VLATs depends on strong contacts between RNAP and the promoter recognition region, especially at -35/spacer/-10 elements (Chander *et al*., 2007). An AT rich spacer region is required but not sufficient for the production of VLATs. A UP element-like AT rich sequence was found in the spacer region of *N25* that attributed to high level of VLAT formation, due to possible contacts between the spacer region UP element and the α-CTD during promoter escape (Chander and Hsu, submitted manuscript). Transcription with RNAP mutant with α-CTD deletion reduced VLAT synthesis by 50%.

VLATs are formed by a mechanism distinct from backtracking. We speculate that VLATs may be formed by RNAP translocating forward by more than 1 bp. Such a movement would place the 3'-OH group upstream of the active site (i.e., the *i* site), unable to incorporate the incoming NTP positioned at the *i+1* site. In the scrunching model, as more and more DNA is unwound and scrunched into the transcription bubble, there is accumulated stress. This stress demands relief, which can come from rewinding the enlarged transcription bubble. Rewinding, however, can take place at the upstream or downstream boundary of the bubble. When the upstream portion of the transcription bubble rewinds, RNAP undergoes promoter escape and translocates forward. When the downstream portion of the transcription bubble rewinds, RNAP backtracks and releases abortive transcripts. However if there is too much stress scrunched in the RNAP (such as when RNAP has to transcribe to the 17^{th}-19^{th} positions to achieve escape), rewinding of the upstream portion of the transcription bubble may propel

a fraction of RNAP to hyper forward jump by more than one nucleotide resulting in 3'-OH being several nucleotides always from the active site. As a result RNA is peeled off from the shortened RNA-DNA hybrid and released as VLATs. Previous studies indicated that a less than 5 bp heteroduplex is required for abortive release during RNAP backtracking (Chander *et al.*, 2007). Thus, RNAP may need to hyper forward-translocate ≥3 nt to release VLATs. In this study we propose to test the physical occurrence of RNAP hyper forward-translocation during VLAT formation utilizing a non-cleaving $E_{111}Q$ *Eco*R I roadblock.

Forward hyper-translocation with $E_{111}Q$ *Eco*R I roadblock

To directly test the physical occurrence of RNAP hyper forward-translocation, we are utilizing a non-cleaving $E_{111}Q$ *Eco*R I as roadblock. A roadblock is a protein bound very stably in the downstream region of the template DNA to stall the forward movement of RNAP, thus preventing the formation of VLATs. *Eco*R I restriction enzyme recognizes a GAATTC binding site and cleaves between G and the first A on both strands of DNA. The $E_{111}Q$ *Eco*R I mutant contains a single amino acid substitution at position 111 from Glu(E) to Gln(Q) that renders the enzyme defective in cleavage activity; however, the E111Q mutant protein retains increased specific binding affinity for the wild-type *Eco*R I recognition sequence. To install the roadblock, we simply need to introduce an *Eco*R I binding sequence to the promoter template. The cleavage activity is reduced by factors of 60,000 to 20,000 relative to wild type, while the

binding affinity is increased by ~1000 fold (Wright *et al.*, 1989). Because of its small protein size and high specific binding affinity, $E_{111}Q$ *Eco*R I makes an ideal roadblock for this study.

Steege and coworkers found that the front edge of $E_{111}Q$ *Eco*R I is able to stall transcription during the elongation phase at 14 bp downstream of the RNAP active site (Pavco and Steege, 1990). The *Eco*R I dimer occupies a 10 bp stretch of DNA when bound to the specific binding site. In this study, we reference the position of the *Eco*R I roadblock by its +3 base on the recognition sequence which is located 5 bp downstream of its front edge. In this experiment, the predicted position of $E_{111}Q$ *Eco*R I to stall $VLAT_{19}$ will be (19 + 14 + 5 =) +38. We constructed six templates with *Eco*R I sites at +25, +30, +34, +45, +55 by PCR. The results of my investigation are reported in the following sections.

MATERIALS AND METHODS

MATERIALS

Enzymes

RNA polymerase holoenzyme, fraction PC-45 (5.6 μM), was purified from *E. coli* strain RL721 by M. Chander and L. Hsu (March, 2007). It was shown to contain 95% active molecules (Chamberlin *et al*., 1983). RNA polymerase holoenzyme is stored at -20 ^{o}C in 50% glycerol, 40 mM KPO_4 buffer, pH 8.0, 100 μM Na_2EDTA, 1 mM DTT, 15 mM $MgCl_2$. $E_{111}Q$ *Eco*R I was generously provided by Dr. Artsimovitch (Ohio State University). T4 polynucleotide kinase (T4 PNK) and *Taq* DNA polymerase were purchased from New England Biolabs.

Plasmids

Purified plasmid DNA, pKK-*DG203*, prepared by L. M. Hsu was used in polymerase chain reaction (PCR) to generate *DG203* and its variant promoter fragments . This plasmid was constructed by cloning the *N25-DG203* promoter in front of the promoterless chloramphenicol acetyl transferase (CAT) gene in the vector plasmid pKK232-8 (Hsu *et al*., 2006).

PCR Primers

The primers used in PCR amplification were purchased from Integrated DNA Technologies, Inc. and stored as 10 μM solutions in TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM Na_2EDTA) at -20 °C. *Eco*R I binding sites are introduced to *DG203*-RI primers at various positions shown in the underlined sequences. *N25*-u (XME) primer was used to eliminate an *Eco*R I site in the PRR region. The sequences of the primers were as follows:

***N25*-u (XE):** *DG203 upstream primer extending from -85 to -60 (25 nt)*;

5'-CCCTCGAGGA ATTCCCGGGG ATCCG-3'

***N25*-u (XME):** *DG203-RI variants upstream primer extending from -85 to -57 (28 nt);* 5'-CCCTCGAGCG ATTCCCGGGG ATCCGTCG-3'

***DG203-RI25*d:** *DG203-RI25 downstream primer extending from +60 (5') to +4 (3') (57 nt); 3'-* GCTGGCC CTCTCCTCAA ATCTTAAGCG ACCAAGAGCG TCTTTCGAAG ACGTCTTGCC-5' (underlined sequences indicate *Eco*R I binding site)

***DG203-RI30*d:** *DG203-RI30 downstream primer extending from +60 (5') to +9 (3') (52 nt);* 3'-CC CTCTCCTCAA ATTTATACG ACCAAGACTT AAGTTCGAAG ACGTCTTGCC-5'

***DG203-RI34*d:** *DG203-RI34 downstream primer extending from +60 (5') to +11 (3') (50 nt);* 3'-CCTCTCCTCAA ATTTATACG ACTTAAGGCG TCTTTCGAAG ACGTCTTGCC-5'

***DG203-RI45*d:** *DG203-RI45 downstream primer extending from +60 (5') to +22 (3') (39 nt);* 3'-TTATACCG ACCAAGAGCG TCCTTAAGAG ACGTCTTGCC-5'

***DG203-RI55*d:** *DG203-RI55 downstream primer extending from +60 (5') to +32 (3') (29 nt);* 3'- CCAAGAGCG TCTTTCGAAG ACCTTAAGCC-5'

***DG203-RI60*d:** *DG203-RI60 downstream primer extending from +60 (5') to +38 (3') (29 nt);* 3'-GCG TCTTTCGAAG ACGTCTTGCC-5'

METHODS

Preparation of *DG203* variants by PCR amplification

PCR amplification of templates: DG203 and *DG203-RI* variants were amplified with *Taq* DNA polymerase by PCR. The reaction was carried out in 200 μL containing 2 μL pKK-*DG203* plasmid DNA template, 1x Taq DNA polymerase buffer (50 mM KCl, 10 mM Tris-HCl, pH 9, 0.1% Triton X-100), 0.25 μM *N25*-u(XE) primer, 0.25 μM *N25*-d(PH) primer, 250 μM dNTPs, 1.5 mM $MgCl_2$, and ddH_2O. After adding 1 μL Taq DNA polymerase, the reaction mixture was placed in a thermocycler (Perkin Elmer Model 480) for 30 cycles at 1 min each at 94 °C, 53 °C, and 72 °C per cycle.

Ethanol Precipitation: After PCR amplification, each 200 μL reaction mixture was added 20 μL 3 M NaAc, 600 μL EtOH and left at -20 °C for overnight. Then the sample was spun for 15 min at 4 °C to collect the pellet. The

pellets were dissolved in 45 μL STE buffer (0.1 M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM Na_2EDTA) and used directly for gel extraction purification.

Gel extraction: For sample loading, 5 μL 40% glycerol loading dye (containing 0.1 M Na_2EDTA and 0.5% bromophenol blue [BPB]) was added to each sample for gel electrophoresis fractionation (final glycerol concentration was 4%). Electrophoresis was carried out with polyacrylamide native gel (8% acrylamide:bisacrylamide [19:1]) in 1X TBE (89 mM Tris Base, 89 mM Boric Acid, 2.5 mM Na_2EDTA, pH 8.3) buffer. Gel was poured with 1.5 mm spacers and the entire volume of a PCR product from previous steps (~50 μL) was loaded into a single one-inch wide well and electrophoresed at 200 V for ~ 2.5 hours, or until the BPB had migrated 10 cm. Under UV illumination, the150-bp band was cut out, placed in a 15-mL Falcon centrifuge tube, crushed and soaked (with shaking at 37 ^{o}C) in STE buffer for 24 h.

Isopropanol precipitation: The gel-extraction sample (above) was spun down at maximum speed (3075 x g) for 10 min. The supernatant is precipitated with equal volume of isopropanol at -20 °C overnight or -80 °C for at least 2 hours. The sample is then spun down for 15 min at 4 °C. The pellet is dissolved in 100 μL TE buffer followed by phenol-chloroform-isoamyl alcohol extraction.

Phenol-chloroform-isoamyl alcohol (25:24:1) extraction: To the sample above, equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) mixture was added and mixed vigorously on vortex for 1 min, then followed by a

microfuge spin for 2 min. The aqueous phase was collected containing the *DG203* variants.

Chloroform: isoamylalcohol (24:1) extraction: To the aqueous sample recovered above, an equal volume of chloroform: isoamylalcohol (24:1) was added, mixed vigorously by vortexing for 10-20 sec, followed by spinning for 2 min. The aqueous layer was mixed with 1/10 volume of 3 M NaAc and 3 volumes of ethanol for precipitation at -20 ^{o}C overnight.

The sample was spun at 4 °C for 15 min to collect the DNA pellet. The pellet was rinsed with 500 μL ice-cold 70% EtOH followed by a spin at 4 °C for 5 min. The pellet was dried in a Speed Vac rotary vacuum desiccator for 15 min and re-dissolved in 100 μL TE. A 3-μL aliquot of the sample was mixed with 7 μL stop dye solution and was analyzed by 2% agarose gel electrophoresis in 1x TAE buffer (40 mM Tris-acetate, 1 mM Na_2EDTA, pH 8.0) containing 0.5 μg/mL of ethidium bromide (EtdBr) at 100 V. The concentration of the DNA was determined with Nanodrop.

Radioactive Labeling of *DG203-RI25*

Labeling of PCR primers: T4 PNK was used to label the PCR primer *N25*-u (XME) at the 5' end with [γ-^{32}P] ATP. Labeling reactions were carried out in a 40 μL volume containing 2.5 μM of the primer, 1 X T4 PNK buffer (70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 2.5 μM [γ-^{32}P] ATP (NEG-035C, diluted with unlabeled ATP to give a specific activity of ~3800 Ci/mmol) and 1

μL of T4 PNK enzyme (10 units/μL). The reaction was incubated at 37 °C for 30 minutes and used directly in a PCR reaction.

PCR amplification with labeled primers: Single end-labeled *DG203-RI25* promoter templates were generated by PCR amplification in 200 μL volume containing 20 ng of pKK-*DG203* plasmid DNA template, 0.25 μM ^{32}P-labeled *N25*-u(XME) upstream primer, 0.25 μM *DG203-RI25*d downstream primer, 0.25 mM dNTP, 1X *Taq* DNA polymerase reaction buffer, 1.5 mM $MgCl_2$ and 2.5 units of *Taq* DNA polymerase. The thermal amplification was carried out in 30 cycles of 1 minute incubation at 94 °C for template DNA denaturation, 1 minute incubation at 53 °C primer for primer-template annealing, and 1 minute incubation at 72 °C for primer extension. The labeled promoter templates were purified by gel extraction.

Purification of PCR products by gel extraction: The reaction volume was added with 1 M NaCl to a final concentration of 0.1 M and 40% glycerol loading dye (containing 0.1 M EDTA and 0.5% BPB) to a final glycerol concentration of 4%. Electrophoresis and gel extraction was carried out as described above (see under *Preparation of DG203 Variants by PCR Amplification*).

In Vitro Transcription

DG203 and *DG203-RI* templates were used to perform *in vitro* transcription. The standard reaction was carried out in 10 μL containing 30 nM DNA, 1x transcription buffer III(10) (50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 10

mM β-mercaptoethanol, 10 μg/mL acetylated bovine serum albumin[BSA]), 40~200 mM KCl, 100 μM NTP, [γ-^{32}P] ATP (NEG002Z, added to ~5 cpm/fmol), diethyl pyrocarbonate (DEPC)-treated H_2O and 1 μL *E. coli* RNA polymerase fraction PC-45 (freshly diluted with RNAP diluent [10 mM Tris-HCl, pH 8.0, 10 mM β-mercaptoethanol, 10 mM KCl, 5% (v/v) glycerol, 0.1 mM Na_2EDTA, 0.4 mg/mL acetylated BSA, 0.1% (v/v) Triton X-100] to 500 nM). The reaction mixture was incubated at 37 °C for 10 min, terminated with the addition of 100 μL stop solution (1 mg/mL nuclease-free glycogen, 10 mM Na_2EDTA, 0.3 M sodium acetate) and precipitated overnight with 330 mL EtOH. The sample was spun at 4 °C for 15 min to collect the pellet, followed by drying with in a SpeedVac desiccator for 15 min. The pellet was resuspended in 10 μL of FLB (formamide loading buffer: 80% (v/v) deionized formamide, 1X TBE, 10 mM Na_2EDTA, 0.08% xylene cyanol (XC), 0.08% amaranth) for analysis by denaturing polyacrylamide gel electrophoresis.

$E_{111}Q$ EcoR I roadblock studies: For *in vitro* transcription reaction in the presence of roadblocks, $E_{111}Q$ *Eco*R I protein was added into the transcription reaction mixture, prior to the addition of RNAP, at various concentrations: 30 nM ([DNA] : [EcoRI]=1:1), 100 nM (1:3.3), 300 nM (1:10), 1 μM (1:33), and allowed to bind by incubating at 37 °C for 10 min.

Gel Shift Assay

Gel shift assay of 5' end-labeled *DG203-RI25* template was carried out to analyze the binding characteristics of $E_{111}Q$ *Eco*R I. The total volume for the binding reaction was 10 μL containing 4.8 μM ^{32}P-*DG203-RI25* DNA (473.6 cpm/fmole), 1x transcription buffer III(10), 40 mM ~ 200 mM KCl, DEPC-H_2O, and $E_{111}Q$ *Eco*R I diluted to the necessary concentration. The reaction mixture was incubated at 37 °C for 15 minutes. The reaction was terminated with 1 μL glycerol stop dye (40% glycerol, 0.1 M Na_2EDTA, 0.2% each of XC and BPB), and immediately analyzed by native PAGE.

Gel Shift Assay was also utilized to analyze the binding characteristics of *DG203-RI25* with RNAP. The reaction mixture of 10 μL containing 4.8 μM ^{32}P-*DG203-RI25* DNA (473.6 cpm/fmole), 1 x transcription buffer III(10) and RNAP diluted to the appropriate concentration was incubated at 37 °C for 15 minutes. The reaction was terminated with 1 μL glycerol stop dye, and immediately analyzed by native PAGE.

To minimize $E_{111}Q$ *Eco*R I cleavage activity, the transcription buffer was varied in pH (from pH 7.2 to pH 8) and $MgCl_2$ concentration (from 5 mM to 10 mM).

Polyacrylamide gel electrophoresis

For transcription reactions: A 4-μL aliquot of [γ-^{32}P]-ATP labeled RNAs (recovered and dissolved in 10 μL FLB) was fractionated in denaturing

polyacrylamide gels (23% acrylamide:bisacrylamide [10:1] containing 7 M urea in 1 x TBE [89 mM Tris Base, 89 mM Boric Acid, 2.5 mM Na_2EDTA, pH 8.3). A 40-mL volume of the gel mixture was polymerized with the addition of 200 μL of 10% ammonium persulfate (APS) and 20 μL of tetraethylmethylene diamine (TEMED). Electrophoresis was accomplished with a steady salt gradient (top buffer: 1x TBE; bottom buffer: 0.3 M NaAc in 1x TBE) at 35 watts until the amaranth dye, which migrates with 2-nt RNAs, has reached approximately 1 cm from the bottom edge of the gel. The gel was exposed to a phosphor screen and visualized on a GE Storm 820 Phosphorimager.

For gel shift: The gel shift assay was analyzed with native PAGE gel (4%, 37.5:1). The gel mixture was prepared in 40 mL total volume containing 4 mL of 40% (37.5:1) acrylamide-bisacrylamide and 1x TBE buffer. The gels were polymerized with the addition of 200 μL of 10% APS and 20 μL of TEMED. The gel was run in 1x TBE buffer at 200 V until the XC dye has migrated half way into the gel. The gels were transferred to 3MM blotter paper and dried at 80 °C under vacuum for 45 min. The dried gels were exposed and visualized using a phophorimager.

Agarose gel electrophoresis

The purified promoter template prepared by PCR was analyzed with 2% agarose gel in TAE buffer + 0.5 μg/mL EtdBr to confirm the presence of correctly-sized PCR products. A 100-bp DNA ladder in glycerol loading dye was

used to index the length of the promoter. Each DNA sample was loaded in the gel with glycerol loading dye. The gel was run at 100 V for approximately 25 min and visualized on a ultra-violet transilluminator.

RESULTS

Abortive transcripts ≤15 nt long are formed by backtracked initial transcribing complexes. A subset of random-ITS variants of *N25* from +3 to +10 positions produced very long abortive transcripts (VLATs) of 16 ~ 19 nt. VLATs are highly unusual due to their undiminished level in the presence of GreB, which rescues abortive RNAs ≤15 nt that arose from RNAP backtracking. The GreB-resistant nature of VLATs suggests they are formed by a mechanism distinct from backtracking. We propose that VLATs are formed by RNAP hyper forward translocation during promoter escape, a phenomenon that has not been observed previously. In this study we propose to test the physical occurrence of RNAP hyper forward-translocation during VLAT formation utilizing a non-cleaving $E_{111}Q$ *Eco*R I roadblock. *DG203* promoter, the specific *N25*-random ITS variant that produces the most robust level of VLATs, is chosen for detailed analysis.

DG203-RI templates

*Eco*R I binding site (GAATTC) was introduced into *DG203* templates through PCR primers. In this study, we reference the position of the *Eco*R I roadblock by its +3 base on the recognition sequence which, when bound by an *Eco*R I dimer, is 5-bp downstream of its front edge, Thus bound, the *Eco*R I dimer blocks transcription by contacting the surface of RNAP located 14 base pairs downstream of the active site (Pavco and Steege, 1990). Therefore the predicted

Figure 5. The non-template strand sequence of *DG203* and *DG203*-RI templates as amplified in the PCR reaction. *DG203* and *DG203*-RI templates share the exact PRR region. Sequences after the initial transcribed sequences of *DG203*-RI templates were shown. Underlined regions represent *Eco*R I binding sites. *RI60* is constructed without an *Eco*R I binding site as control.

(A) *DG203* promoter synthesized by PCR (142-bp) 5' → 3'

-85

5' CCCTC GAGGAATTCC CGGGGATCCG TCGAGGGAAA

-35

TCATAAAAAA TTTAT**TTGCT** **T**TCAGGAAAA

-10 +1

TTTTTCTG**TA** **TAAT**AGATTC ATGCGACCGG

GAGAGGAGTT TAAATATGGC TGGTTCTCGC

+57

AGAAAGCTTC TGCAGAA 3'

(B) *DG203*-RI variants synthesized by PCR (145-bp) 5' → 3'

RI-25 $^{+1}$ATGCGACCGG GAGAGGAGTT TAGAATTCGC TGGTTCTCGC
AGAAAGCTTC TGCAGAACGG^{+60}

RI-30 $^{+1}$ATGCGACCGG GAGAGGAGTT TAAATATGAA TTCTTCTCGC
AGAAAGCTTC TGCAGAACGG^{+60}

RI-34 $^{+1}$ATGCGACCGG GAGAGGAGTT TAAATATGGC TGAATTCCGC
AGAAAGCTTC TGCAGAACGG^{+60}

RI-45 $^{+1}$ATGCGACCGG GAGAGGAGTT TAAATATGGC TGGTTCTCGC
AGGAATTCTC TGCAGAACGG^{+60}

RI-55 $^{+1}$ATGCGACCGG GAGAGGAGTT TAAATATGGC TGGTTCTCGC
AGAAAGCTTC TGGAATTCGG^{+60}

RI-60 $^{+1}$ATGCGACCGG GAGAGGAGTT TAAATATGGC TGGTTCTCGC
AGAAAGCTTC TGCAGAACGG^{+60}

position of the *Eco*R I site for binding $E_{111}Q$ *Eco*R I to stall the synthesis of $VLAT_{19}$ will be (19 + 14 + 5 =) +38. Based on these structural parameters, we constructed *DG203-RI* templates with *Eco*R I binding sites at +34 (*DG203-RI34*), +45 (*DG203-RI45*), and +55 (*DG203-RI55*) by PCR as shown in Figure 5. A preexisting *Eco*R I binding site at the upstream PRR sequence of *DG203* was eliminated through the use of a new upstream PCR primer. *DG203-RI60* is constructed without an *Eco*R I binding site and serves as a control template. Transcription initiated on *DG203-RI* promoters but blocked by $E_{111}Q$ *Eco*R I protein bound to the *Eco*R I sites on *DG203-RI* promoters should yield stalled transcripts that reflect the position of the bound protein. The predicted lengths of the stalled transcripts are listed in Table 1.

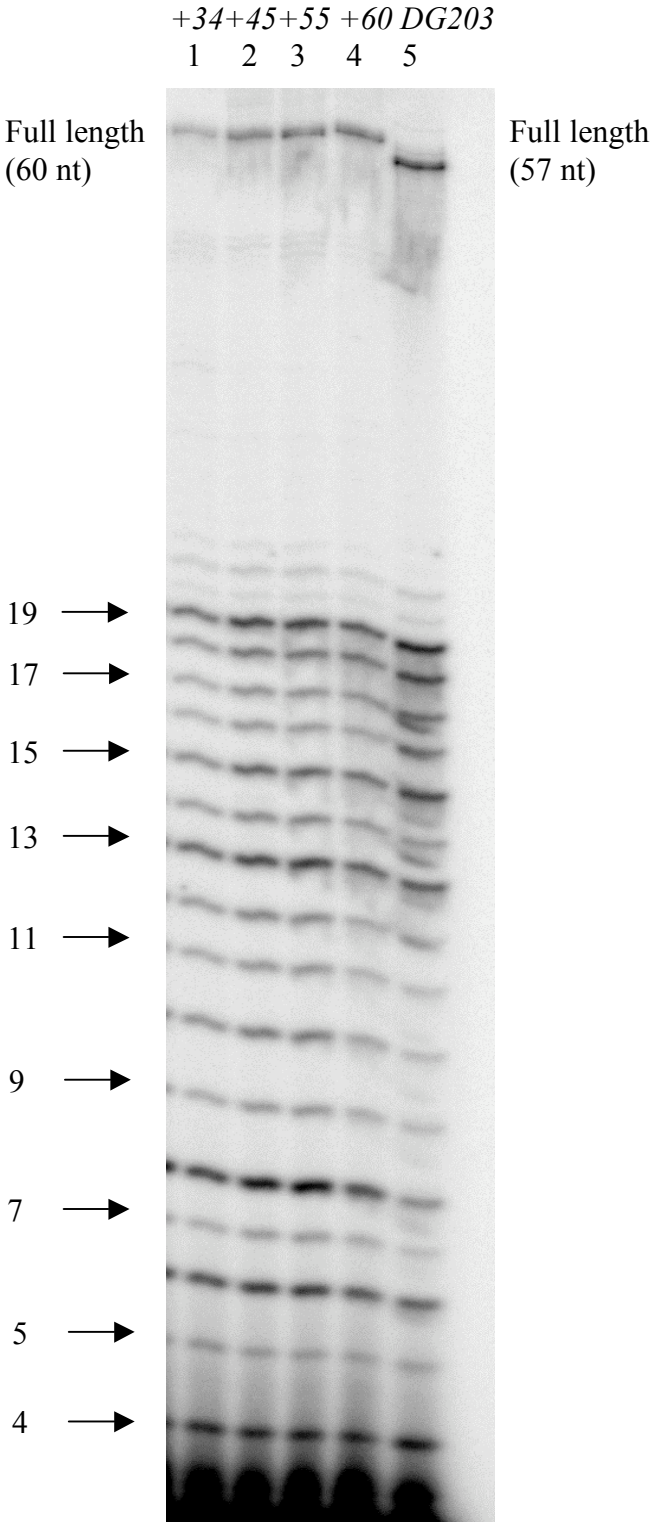
In Vitro transcription of *DG203-RI34, -RI45, -RI55* promoters

In vitro transcription was performed with *DG203-RI* promoters to confirm that they can transcribe normally. Transcription reactions were analyzed on a 23% (10:1)/7M urea polyacrylamide gel as shown in Figure 6. The expected run-off transcript for *DG203* is 57 nt, while the expected run-off transcript for *DG203-RI* promoters are 60 nt because of three additional nucleotides at the 3' end of the non-template strand (NT). Each *DG203-RI* promoter produced abortive ladder with VLATs up to 19 nt in length and full length transcripts. There appears to be no difference in the abortive profile and full-length production between *DG203* and *DG203-RI* variants.

Table 1. The predicted position of stoppage site on *DG203-RI* templates and the observed results from *in vitro* transcription in the presence of $E_{111}Q$ *Eco*R I.

Promoter	Predicted stoppage site	Result
RI-25	+ 7	Partial blockage at +6~+8
RI-30	+12	Partial blockage at +11~+13
RI-34	+16	Partial blockage at +15~17
RI-45	+27	Stable stoppage at +26, +27

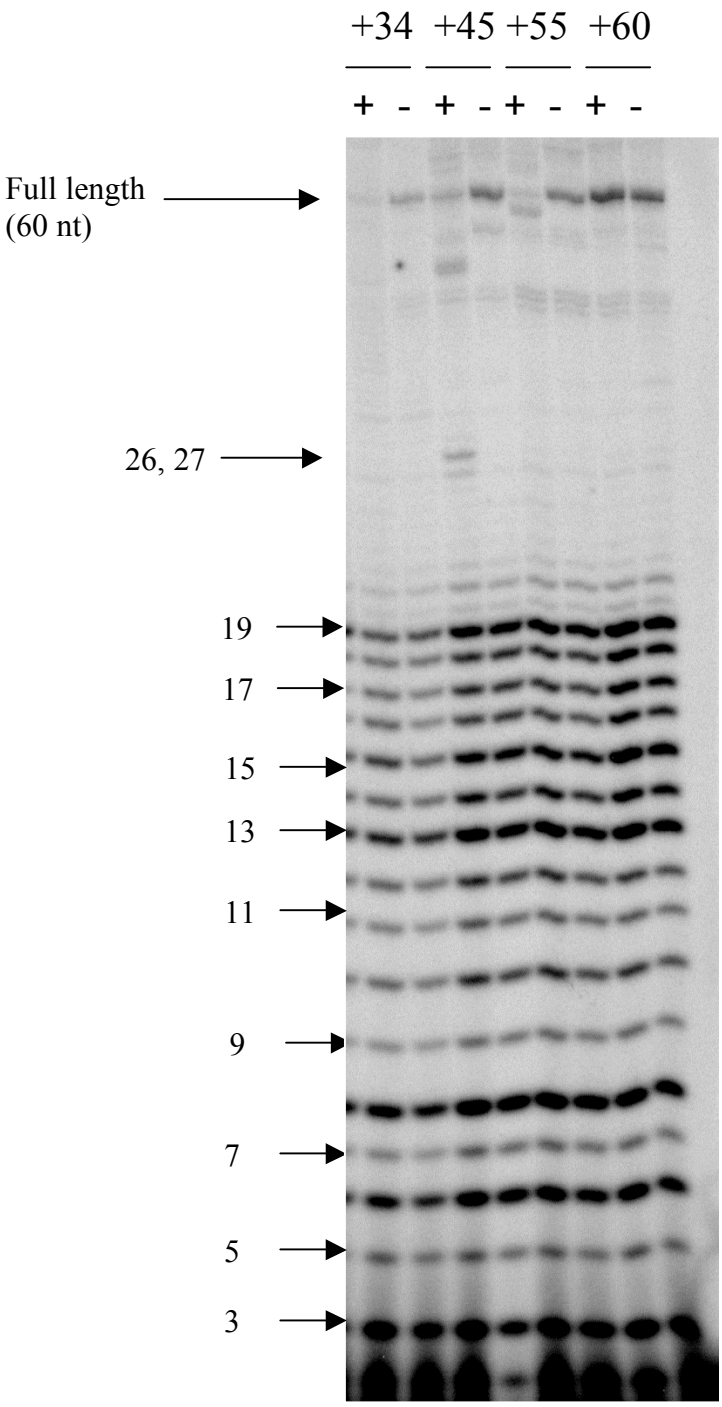
Figure 6. *In vitro* transcription of *DG203-RI34, DG203-RI45, DG203-RI55, DG203-RI60* templates in standard transcription condition. Lane 1 contains *RI34*, lane 2 contains *RI45,* lane 3 contains *RI55,* lane 4 contains *RI60,* lane 5 contains *DG203*.


+34 +45 +55 +60 DG203
1 2 3 4 5
Full length
(60 nt)
Full length
(57 nt)
19
17
15
13
11
9
7
5
4

In vitro transcription on *DG203-RI34*, *-RI45, -RI55* templates associated with $E_{111}Q$ *Eco*R I protein

Having established that *DG203-RI* promoters transcribe the same as *DG203* promoter, they were utilized to analyze the physical occurrence of hyper forward translocation of RNAP by *in vitro* transcription in the presence of $E_{111}Q$ *Eco*R I enzyme. This is done by initiating transcription with RNAP on *DG203-RI* templates pre-bound in the presence of 3 fold molar excess of $E_{111}Q$ *Eco*R I. *DG203-RI60* template is constructed as a control template without any *Eco*R I binding site. The level of its run-off transcripts was not affected by $E_{111}Q$ *Eco*R I presence, confirming that normal transcription was not affect by the $E_{111}Q$ *Eco*R I as shown in Figure 7. With *DG203-RI34*, *DG203-RI45*, *DG203-RI55*, the presence of $E_{111}Q$ *Eco*R I diminished the level of full-length products indicating $E_{111}Q$ *Eco*R I bound and stalled RNAP transcription. However the position of stalled bands yielded many puzzling results. With *DG203-RI45* promoter (predicted stoppage site +27), the presence of $E_{111}Q$ *Eco*R I blocked transcription at approximately +26 to +27 position resulting in stalled bands of 26 and 27 nt long. However, an additional band between +27 and the run-off transcript was also observed. Further more, full-length production on *DG203-RI45* promoter was reduced but not eliminated, indicating incomplete blockage. With *DG203-RI55* promoter (predicted stoppage site +37), a stalled band of 55 nt long is observed in the presence of $E_{111}Q$ *Eco*R I. With *DG203-RI34* (predicted stoppage site +16),

Figure 7. *In vitro transcription* study with $E_{111}Q$ *Eco*R I roadblock with *DG203-RI34, DG203-RI34, DG203-RI45, DG203-RI55, DG203-RI60*. *In vitro* transcription was performed with (+) and without (-) $E_{111}Q$ *Eco*R I roadblock with each template. Partial blockage was observed at +15 ~ +17 on *RI34* templates, at +26 and +27 on *RI45* templates.


+34 +45 +55 +60
+ - + - + - + -
Full length
(60 nt)
26, 27
19
17
15
13
11
9
7
5
3

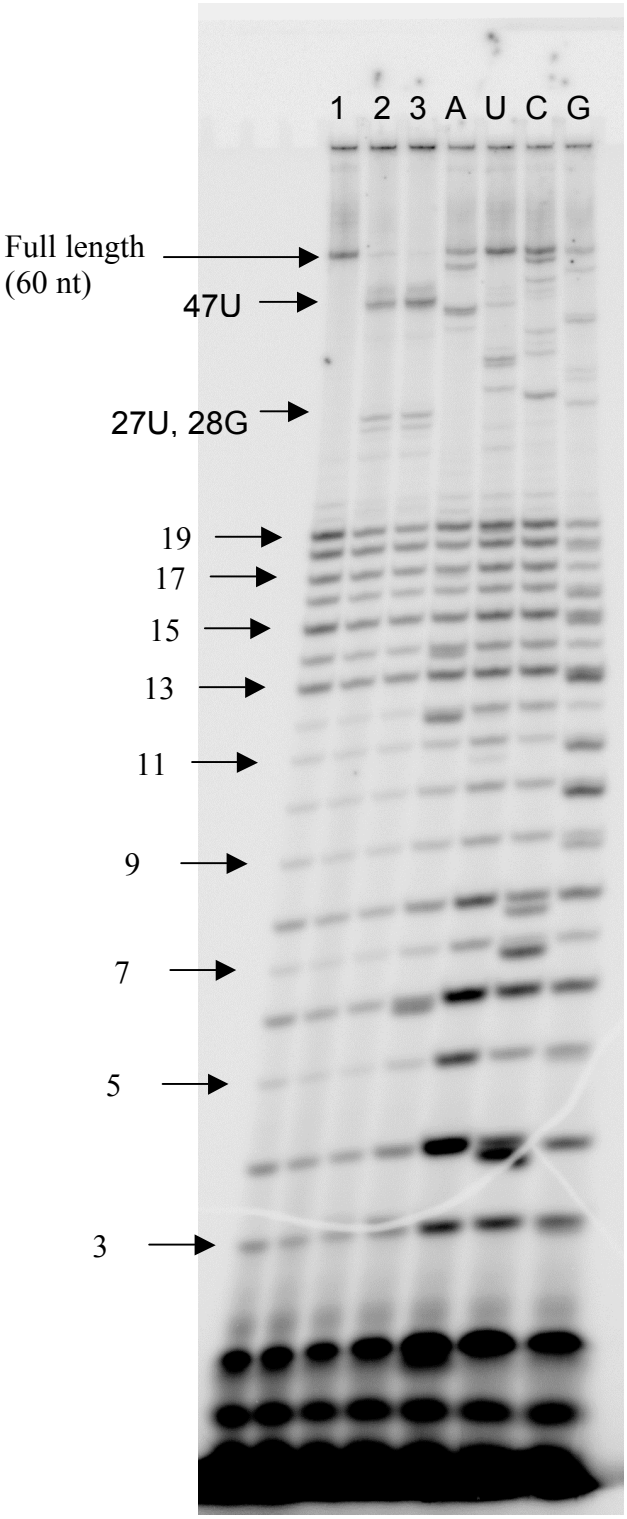
VLATs longer than 16 nt was formed. Increased amount of abortive transcripts from 15 to 17 nt long was observed, again indicating incomplete blockage.

Previous studies showed that $E_{111}Q$ *Eco*R I binding activity is affected by KCl concentration. To ensure reaction was carried out in the optimal KCl concentration, a range of KCl concentration (40 mM ~ 200 mM) was tested with *DG203-RI34*, *-RI45*, and *-RI55* promoters. No significant improvement was observed (figure not shown).

3'-deoxynucleoside triphosphate (3'-dNTP) sequencing on *DG203-RI45* template

To determine the exact sequence position of the new bands that appeared with *RI45* templates in the presence of $E_{111}Q$ *Eco*R I, 3'-dNTP sequencing was performed. Transcription in the presents of [γ-^{32}P] ATP, sufficient amount of NTP and one of the four 3'-dNTPs proceed normally until occasionally a 3'-dNTP is incorporated. Incorporation of dNTP at the 3' end of the RNA transcript terminates transcription due to the lack of an OH group at the 3' position for the formation of phosphodiester bond. The terminated transcript is released and appears slight below transcripts with regular NTPs at the 3' end of the same length by gel electrophoresis, due to the lack of an -OH group. In this way, the position of a band can be read out by transcribing alongside four sets of transcription each with the presence of one 3'-dNTP. Sequencing with 3'-dNTP showed that the additional bands were at +27U, +28G and +47U position as shown in Figure 8.

Figure 8. 3'-deoxynucleoside triphosphate (3'-dNTP) sequencing of *DG203-RI45* transcripts. Lane 1 contains transcription reaction without $E_{111}Q$ *Eco*R I or 3'-dNTP; lane 2, 3 contains reaction with $E_{111}Q$ *Eco*R I without 3'-dNTP. Lane A, U, C, G contains 3'-dNTP sequencing reactions without the presence of $E_{111}Q$ *Eco*R I: lane A contains reaction with 3'-dATP; lane U contains reaction with 3'-dUTP; lane C contains reaction with 3'-dCTP; lane G contains reaction with 3'-dGTP. The additional bands were at positions +27U, +28G, +47U.


1 2 3 A U C G
Full length
(60 nt)
47U
27U, 28G
19
17
15
13
11
9
7
5
3

Titration of $E_{111}Q$ *Eco*R I concentration for maximum blockage

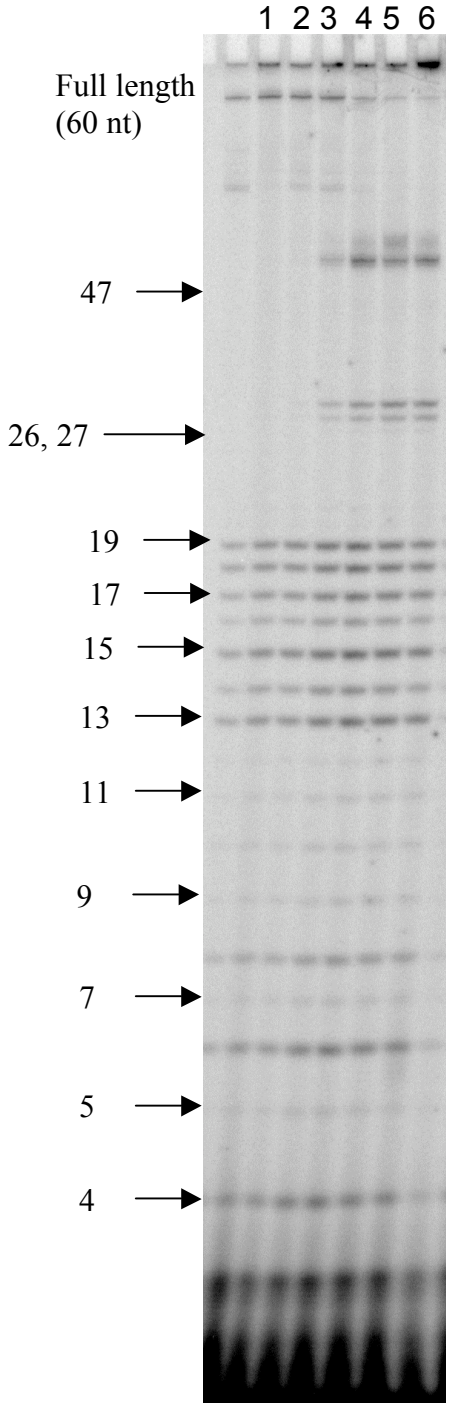
To test whether incomplete blockage was due to incomplete occupation of *Eco*R I site, a series of $E_{111}Q$ *Eco*R I concentration was used to titrate for maximum blockage. Steady blockage was observed at 30-fold molar excess of $E_{111}Q$ *Eco*R I on *DG203-RI45* promoters characterized by the absence of full-length transcripts and plateau level of stalled transcripts as shown in Figure 9a.

Partial blockage was observed with *DG203-RI34* promoters even after the amount of blockage transcripts has plateaued as shown in Figure 9b. With increased amount of *Eco*R I concentration, the amount of abortive transcripts of 15~17 nt long increased and plateaued with 10-fold molar excess of $E_{111}Q$ *Eco*R I as shown in Figure 10. However, VLATs and full-length products were still formed on the *DG203-RI34* promoter. Starting at 10-fold molar excess of $E_{111}Q$ *Eco*R I, an additional band of 24 nt appeared. Its level increased as $E_{111}Q$ *Eco*R I concentration increased. At 60-fold molar excess of $E_{111}Q$ *Eco*R I, an additional band was observed below the full-length product which may have resulted from nonspecific binding of *Eco*R I. Overall blockage with *DG203-RI34* promoter was incomplete.

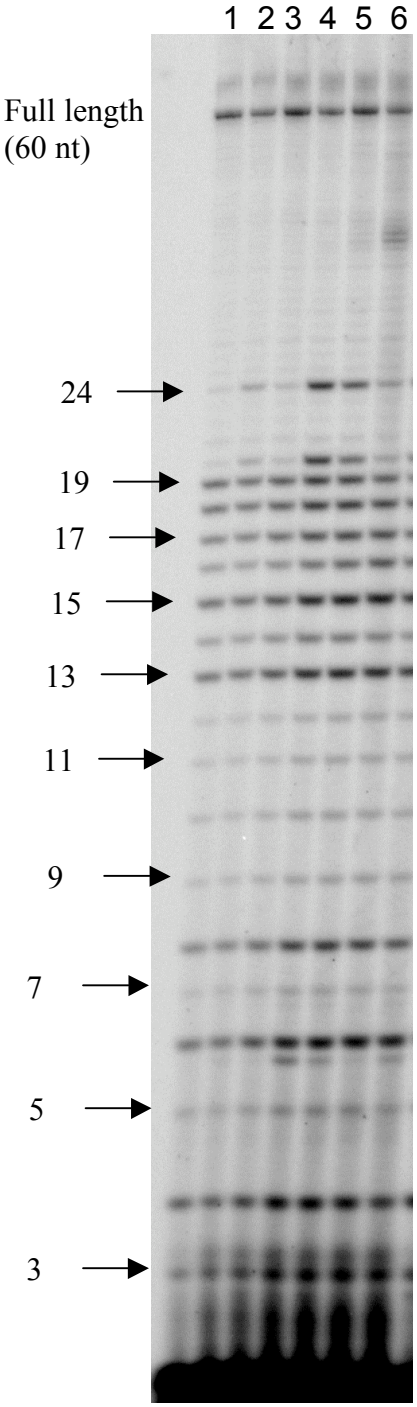
The different ability of $E_{111}Q$ *Eco*R I to stall transcription on *DG203-RI34* and *DG203-RI45* promoters raised the question as what would happen if $E_{111}Q$ *Eco*R I was placed further upstream of +34.

Figure 9. *In vitro* transcription of *DG203-RI45* and *DG203-RI34* promoter templates under varying $E_{111}Q$ *Eco*R I concentration. Transcription was performed without $E_{111}Q$ *Eco*R I (lane 1), with $E_{111}Q$ *Eco*R I to DNA molar ratio of 1:1 (lane 2), 3:1 (lane 3), 10:1 (lane 4), 30:1 (lane 5), and 60:1 (lane 6). a. *In vitro* transcription with *DG203-RI45* template. b. *In vitro* transcription with *DG203-RI34* template.

a. *DG203-RI45*



b. *DG203-RI34*



In vitro transcription on *DG203-RI25* and *DG203-RI30* templates associated with $E_{111}Q$ *Eco*R I protein

Two more promoters were constructed to contain *Eco*R I binding sites at +25 (*DG203-RI25*, predicted blockage site at +7) and +30 (*DG203-RI30*, predicted blockage site at +12). *In vitro* transcription confirmed that *DG203-RI25* and *DG203-RI30* promoters showed the same transcription profile as *DG203* promoter as shown in Figure 10. Then *in vitro* transcription was performed with *DG203-RI* promoters in the presence of 10-fold molar excess of $E_{111}Q$ *Eco*R I as shown in Figure 11.

VLATs were still formed on both *RI25* and *RI30* templates. In the presence of $E_{111}Q$ *Eco*R I, *RI25* showed increased amount of transcripts 6 to 8 nt long with diminished amount of abortive products after +8 and absence of full-length product as shown in Figure 12. In the presence of $E_{111}Q$ *Eco*R I, *RI30* showed increased amount of transcripts 11 to 13 nt long, diminished level of abortive transcripts after +13 and full length products as shown in Figure 11.

To further quantify the data, abortive yield (AY) and productive yield (PY) were calculated and graphed to represent the relative amount of transcripts at each position as shown in Figure 12. As shown in Figure 12a, the presence of $E_{111}Q$ *Eco*R I with *RI25* clearly increased AY of 6 to 8 nt long accompanied with diminished transcription of longer transcripts, indicating transcription was blocked at position +6 to +8. As shown in Figure 12b, the presence of $E_{111}Q$ *Eco*R

Figure 10. *In vitro* transcription in standard condition with *DG203-RI25, DG203-RI30, DG203-RI34, DG203-RI45, DG203-RI55, DG203-RI60, DG203* promoter templates.

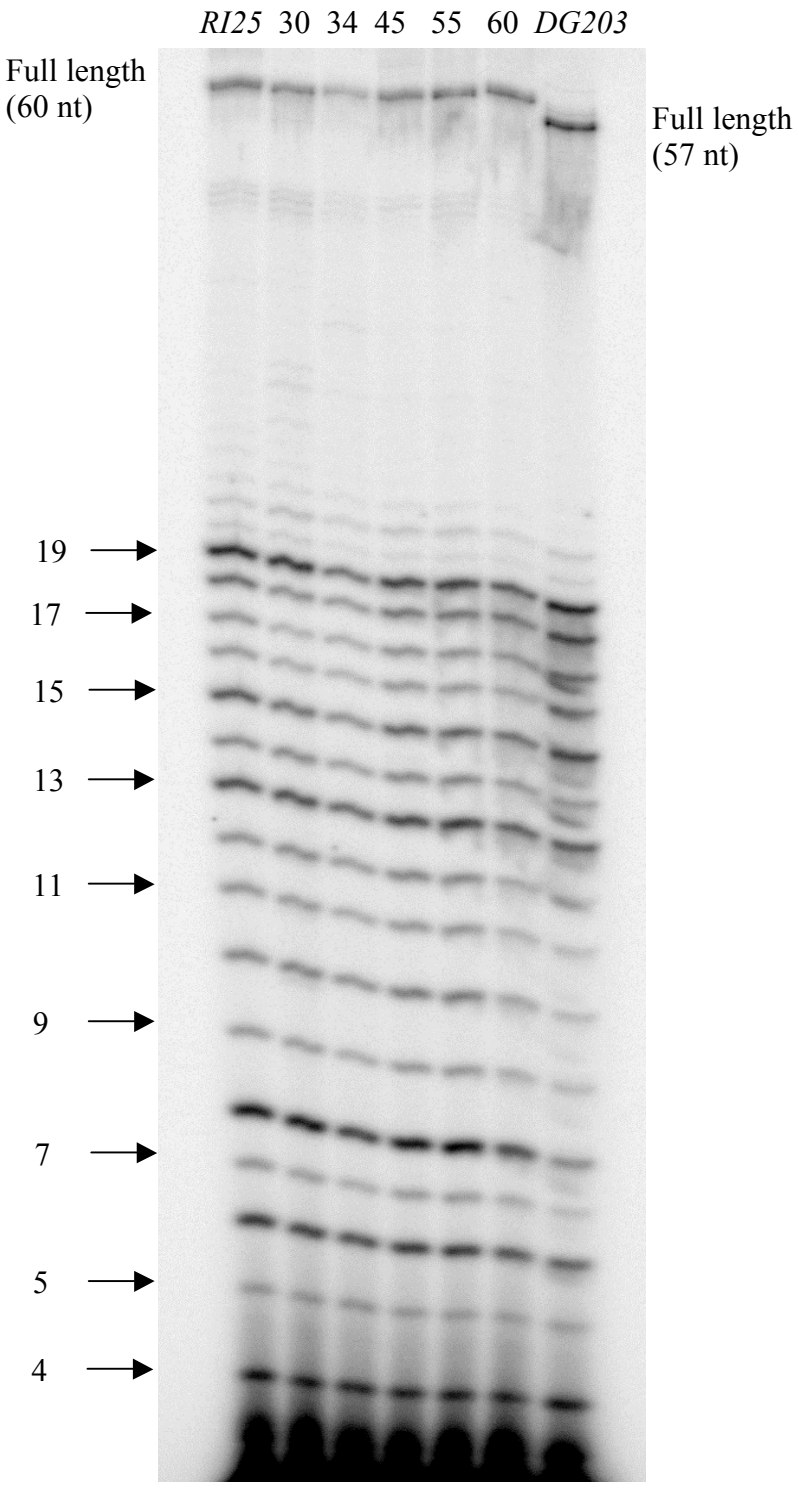

RI25 30 34 45 55 60 DG203
Full length
(60 nt)
Full length
(57 nt)
19
17
15
13
11
9
7
5
4

Figure 11. *In vitro* transcription with *DG203-RI25, DG203-RI30, DG203-RI34, DG203-RI45, DG203-RI55, DG203-RI60* templates associated with $E_{111}Q$ *Eco*R I roadblock. *In vitro* transcription was performed with (+) and without (-) $E_{111}Q$ *Eco*R I roadblock. Partial blockage was observed at +5 ~ +7 on *RI25*, +11~+13 on *RI30*, +15 ~ +17 on *RI34* templates, +26 and +27 on *RI45* templates.

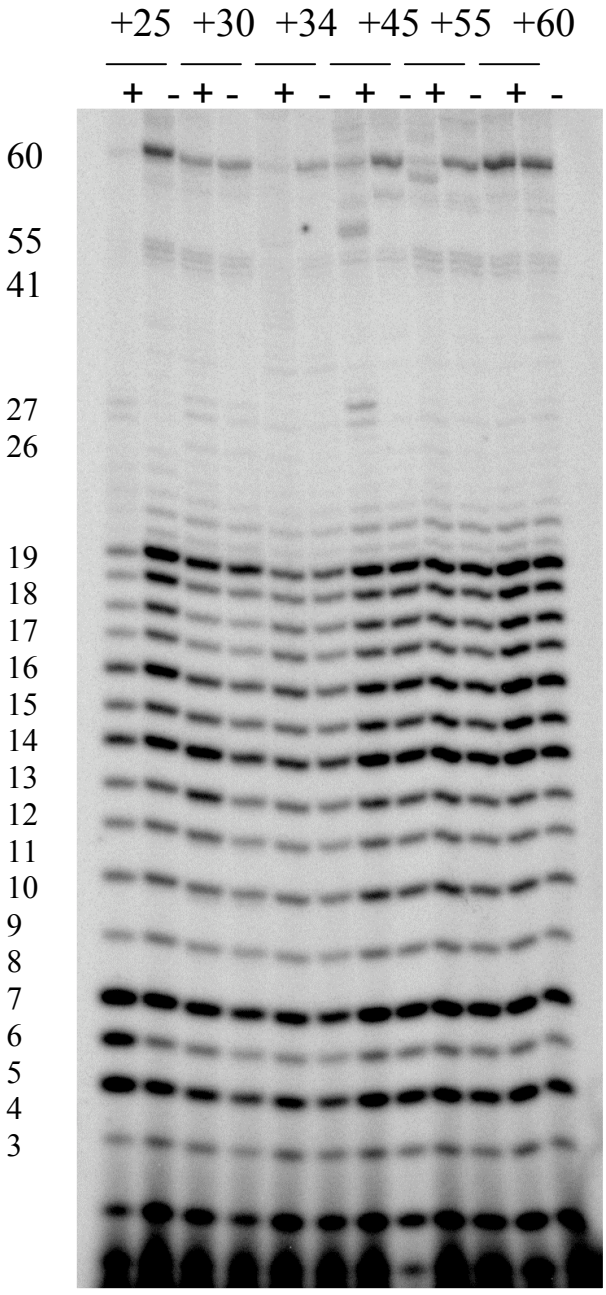
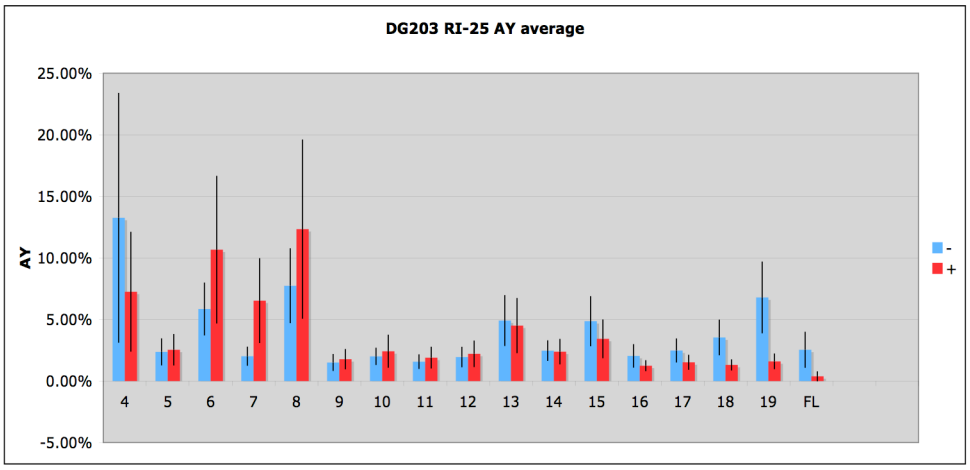
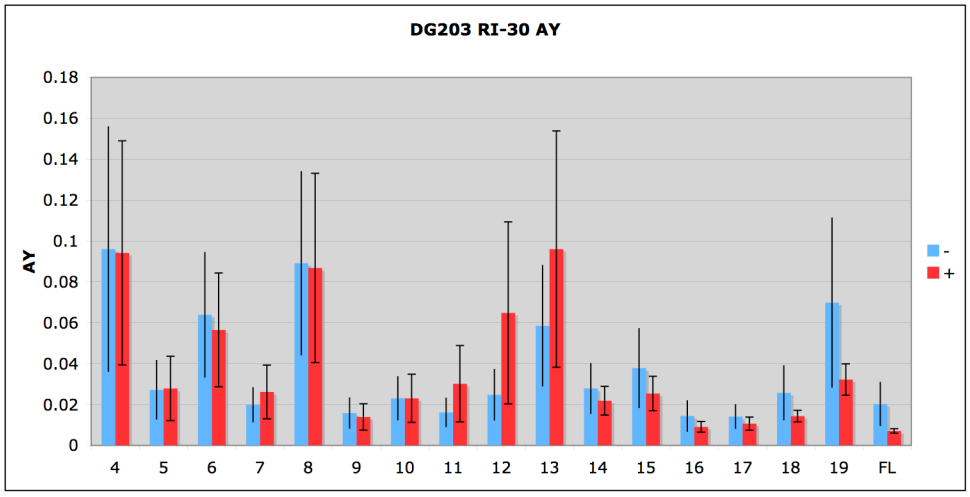

+25 +30 +34 +45 +55 +60
+ - + - + - + - + - + -
60
55
41
27
26
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3

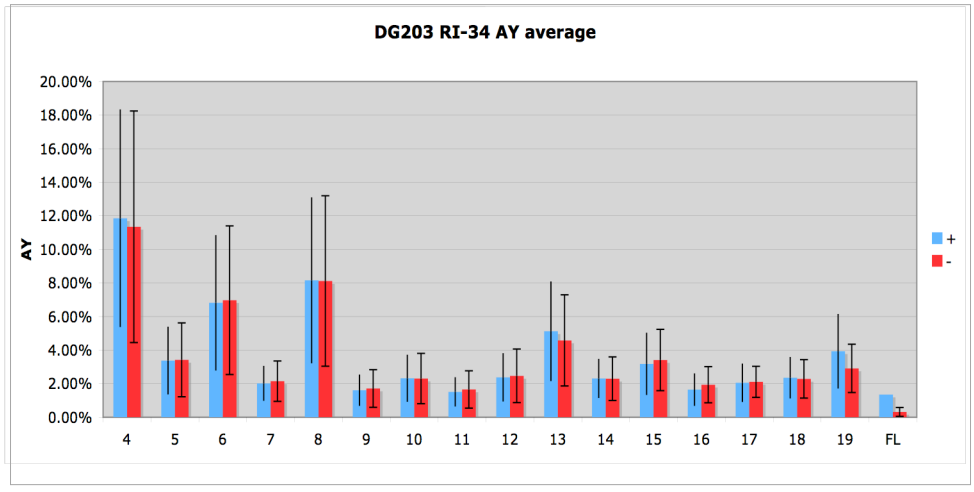
Figure 12. Abortive yield profiles of *DG203-RI25, DG203-RI30,* and *DG203-RI34* promoter templates associated with $E_{111}Q$ *Eco*R I. Abortive profile with the presence of $E_{111}Q$ *Eco*R I is denoted red, abortive profile without $E_{111}Q$ *Eco*R I is denoted blue. Abortive yield was calculated for the abortive RNAs and graphed. The larger the abortive yield, the more abortive transcripts were produced.


DG203 RI-25 AY average
25.00%
20.00%
15.00%
10.00%
5.00%
0.00%
-5.00%
AY
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 FL
-
+

a


DG203 RI-30 AY
0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
AY
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 FL
-
+

b


DG203 RI-34 AY average
20.00%
18.00%
16.00%
14.00%
12.00%
10.00%
8.00%
6.00%
4.00%
2.00%
0.00%
AY
4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 FL
+
-

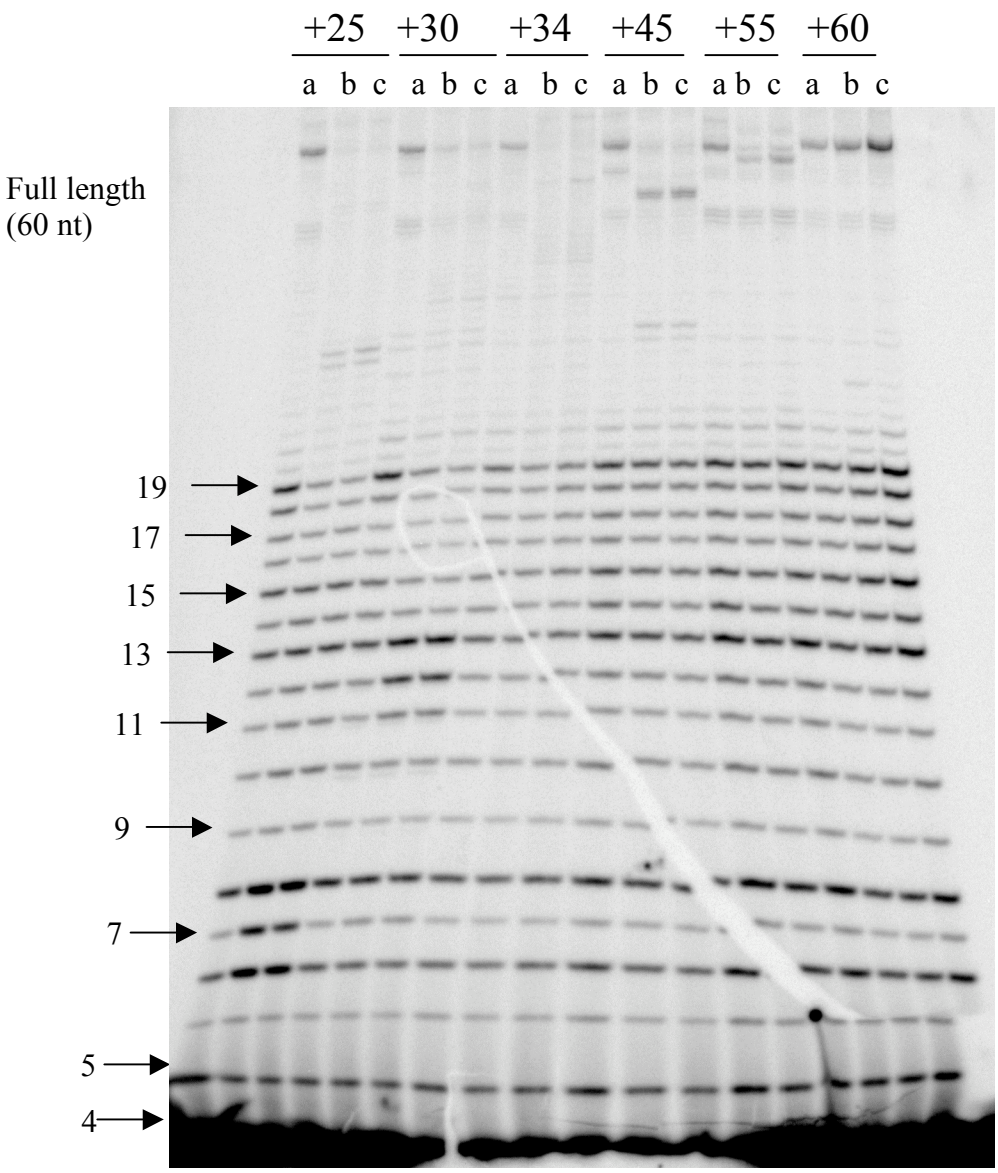
c

I with *RI30* increased AY of 11 to 13 nt long with diminished transcription afterwards, indicating blockage at +11 to +13. As shown in Figure 12c, the AY of VLAT19 and PY of *RI34* diminished in the presence of $E_{111}Q$ *Eco*R I. Increase of transcription with *RI34* promoter in the presence of *Eco*R I at 16 was not prominent. To amplify this difference, *in vitro* transcription in the presence of [α-^{32}P]-CTP with promoters associated with $E_{111}Q$ *Eco*R I was performed. Because [α-^{32}P]-CTP labels every C in the RNA transcript unlike [γ-^{32}P]-ATP that only labels at the 5' end, RNA transcripts are labeled to a higher specific radioactivity. No significant difference was observed.

The formation of VLATs on both *RI25* and *RI30* promoters indicates partial blockage. To test if partial blockage was due to incomplete occupation of *Eco*R I binding sites, the molar excess of $E_{111}Q$ *Eco*R I was increased to 30 fold. As shown in Figure 13, increasing $E_{111}Q$ *Eco*R I to 30-fold molar excess did not significantly increase blockage activity, indicating complete occupation of *Eco*R I binding sites during transcription.

Overall, partial blockage was observed with *RI25*, *RI30* and *RI34* templates in the presence of $E_{111}Q$ *Eco*R I. Complete blockage was observed with *RI45* templates. In the case of *RI45*, the stalled transcripts of 26 to 27 nt long were produced when RNAP has undergone promoter escape to perform stable elongation. In the case of *RI25, RI30* and *RI34* templates, partial blockage was observed during transcription initiation. This result suggests that $E_{111}Q$ *Eco*R I can only block transcription completely during elongation, but unable to block the

Figure 13. *In vitro* transcription of *DG203-RI25, DG203-RI30, DG203-RI34, DG203-RI45, DG203-RI55, DG203-RI60* in the presence of $E_{111}Q$ *Eco*R I roadblock. a. Transcription without $E_{111}Q$ *Eco*R I; b. Transcription with $E_{111}Q$ *Eco*R I to DNA molar ratio of 10; c. Transcription with $E_{111}Q$ *Eco*R I to DNA molar ratio of 30.


+25 +30 +34 +45 +55 +60
a b c a b c a b c a b c a b c a b c
Full length
(60 nt)
19
17
15
13
11
9
7
5
4

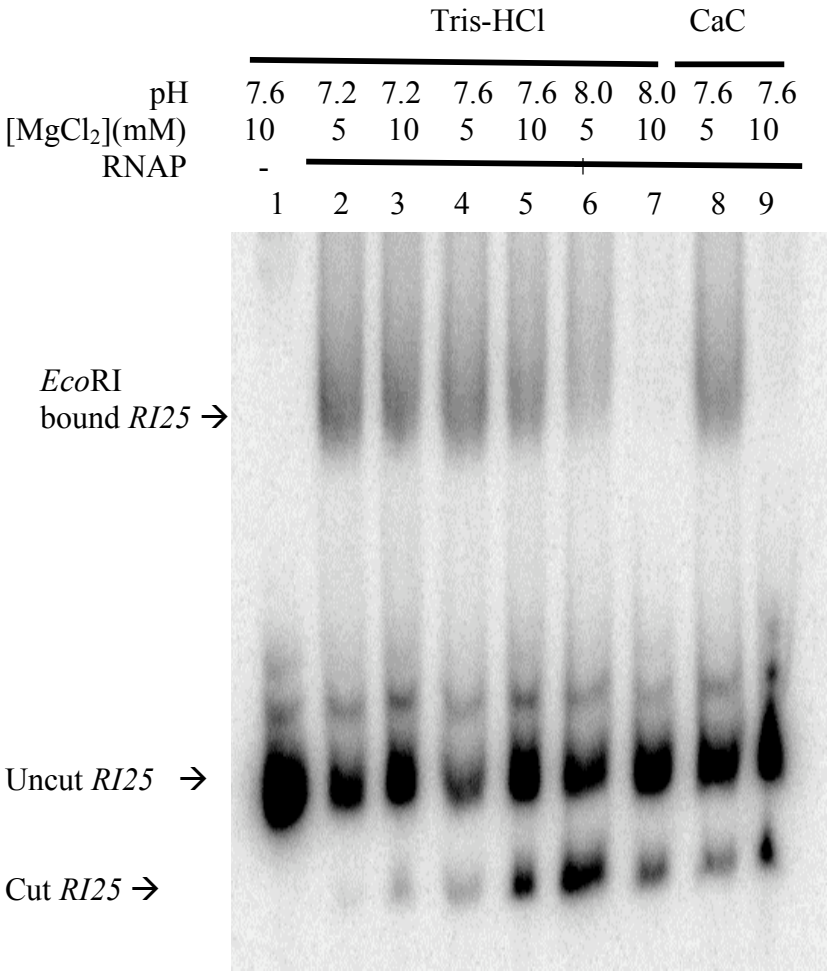
"powerful" transcription initiation complex. Likely, RNAP bumps off $E_{111}Q$ *Eco*R I and continues on to synthesize VLATs and full-length products. Further *RI25* and *RI30* showed marked increase of stalled transcripts in comparison to *RI34*. With *RI34* promoter, stalled transcripts were formed at +15 to +17 position. At these positions, RNAP is at the transition stage of promoter escape. The better blockage performed on *RI25* and *RI30* promoters in comparison to that of *RI34* indicates that RNAP forward movement is even more powerful during the formation of VLATs than during the earlier initial transcription steps.

Gel shift assay to measure binding

The above results of *in vitro* transcription of *DG203-RI* promoters suggest that the presence of $E_{111}Q$ *Eco*R I may not be able to stall the "powerful" forward movement of RNAP during transcription initiation. Essentially RNAP translocates with much more power during transcription initiation than during elongation due to elevated stress. However for this conclusion to be true, *Eco*R I binding sites need to be occupied the entire time during transcription. If $E_{111}Q$ *Eco*R I does not stay bound to the promoter 100% during transcription, only partial blockage will be observed.

To test the binding characteristics of $E_{111}Q$ *Eco*R I on *DG203-RI* promoter, gel shift assay was used. *DG203-RI25* was chosen for detailed analysis due to its clear partial blockage characteristics. To perform the gel shift assay, *DG203-RI25* promoter was first radioactively labeled on the 5' end of the

Figure 14. Gel shift assay of $E_{111}Q$ *Eco*R I binding to *DG203-RI25* in different buffer conditions to minimize its cleavage activity. Buffer conditions varied in pH 7.2~8.0, $MgCl_2$ concentration 5 mM ~ 10 mM, and buffer type (Tris-HCl or CaC buffer). Standard transcription is carried out in pH 8.0, 10 mM $MgCl_2$ concentration shown in lane 7. Cleavage activity was optimized in Tris-HCl, pH 7.2 buffer containing 5 mM $MgCl_2$.


Tris-HCl
CaC
pH 7.6 7.2 7.2 7.6 7.6 8.0 8.0 7.6 7.6
[MgCl2](mM) 10 5 10 5 10 5 10 5 10
RNAP -
1 2 3 4 5 6 7 8 9
EcoRI bound RI25 →
Uncut RI25 →
Cut RI25 →

nontemplate strand with [γ-^{32}P]-ATP using T4 polynucleotide kinase (PNK). Unincorporated nucleotides and PNK protein were removed from the ^{32}P-labeled promoter using polyacrylamide gel extraction. After cleaning, the radioactivity associated with the template was 50,000 cpm/μL. For gel shift assays, 100,000 cpm of labeled promoter was used for each reaction. The labeled promoter template was incubated with $E_{111}Q$ *Eco*R I for 15 minutes to enable binding, immediately followed by gel electrophoresis on native 4% polyacrylamide (37.5:1) gels. The absence of urea in the gel mixture ensures that there is no denaturing effect on the promoter-protein complex formed.

Unexpectedly, residual $E_{111}Q$ *Eco*R I cleavage activity was observed as shown in Figure 15. After $E_{111}Q$ *Eco*R I cleaves at the binding site, it would fall off the promoter, thus exerting only partial blockage. Transcription on truncated promoters would result in shorter run-off transcripts; this explains the additional band we observed with *DG203-RI45* at about +41.

To eliminate the residual cleavage activity of $E_{111}Q$ *Eco*R I, reaction condition of different pH, $MgCl_2$ concentration, Tris-HCl or sodium cacodylate (CaC) buffer were tested. As shown in Figure 15, decreasing $MgCl_2$ concentration decreased $E_{111}Q$ *Eco*R I cleavage activity significantly. Decreasing pH from 8 to 7.2 decreased $E_{111}Q$ *Eco*R I cleavage activity. Less cleavage activity was observed in Tris-HCl buffer in comparison to CaC buffer. Cleavage activity was minimized at pH 7.2, 5 mM $MgCl_2$ concentration in Tris-HCl buffer.

DISCUSSION

The goal of the study was to test the physical occurrence of RNAP hyper forward-translocation during VLATs (16~19 nt), a mechanism distinct from backtracking associated with abortive transcripts ≦15 nt. The study was done with a *DG203* promoter utilizing a non-cleaving $E_{111}Q$ *Eco*R I roadblock. Originally we expected $E_{111}Q$ *Eco*R I roadblock to block the forward movement of RNAP, thus preventing the formation of VLATs. We constructed *DG203-RI* promoters with *Eco*R I binding sites at +25 (*RI25*), +30 (*RI30*), +34 (*RI34*), +45 (*RI45*), +55 (*RI55*).

To our surprise, VLATs formed on all five templates in the presence of $E_{111}Q$ *Eco*R I protein, even on promoters with a predicted stoppage site before +16. Based on the predicted stoppage site, $E_{111}Q$ *Eco*R I should block transcription on *RI25*, *RI30*, *RI34*, *RI45*, *RI55* at +7, +12, +16, +22, +27, respectively. The *DG203* template undergoes promoter escape at +19 position. Therefore RNAP would bump into $E_{111}Q$ *Eco*R I roadblock during transcription initiation on *RI25*, *RI30*, and *RI34* promoters, or during stable elongation phase on *RI45* and *RI55* promoters. $E_{111}Q$ *Eco*R I was able to stall transcription during the elongation phase as with *RI45* promoter, but only show partial blockage during transcription initiation with *RI25*, *RI30*, and *RI34* promoters. This unexpected result changed the course of our study and raised the possibility that $E_{111}Q$ *Eco*R I may be utilized to probe the force associated with RNAP forward movement

during different stages of transcription. It is possible that $E_{111}Q$ *Eco*R I is unable to stall the "powerful" forward translocating RNAP during abortive initiation and promoter escape, a direct piece of evidence that there is increased amount of stress in the initial transcription complex. Interestingly, in comparison to *RI25* and *RI30* promoters, even less stoppage was observed with *RI34* promoter, of which the stoppage sites (+15 ~ +17) correspond to where VLAT formation occurs. This result indicates that there is even more stress and movement during the formation of VLATs on the brink of promoter escape transition than the early transcription initiation steps. If validated, the *in vitro* transcription results in the presence of $E_{111}Q$ *Eco*R I would be a direct piece of evidence that there is accumulation of force as RNAP forward translocates during transcription initiation that reaches a maximum at the promoter escape transition point, and undergoes a sudden fall as the initial transcription complex collapses into a stable elongation phase. This will indirectly support the proposed forward translocation mechanism for the formation of VLATs, as increased stress is the cause of RNAP hyper forward translocation.

To validate this conclusion, we needed to confirm that $E_{111}Q$ *Eco*R I was indeed bumped off from the promoter by the "powerful" RNAP during transcription initiation. It is possible that $E_{111}Q$ *Eco*R I may not stay bound to the binding site 100% allowing RNAP to transcribe past the blockage site. $E_{111}Q$ *Eco*R I binding characteristics were analyzed with gel shift assay. Surprisingly $E_{111}Q$ *Eco*R I showed residual cleavage activity in the normal transcription

reaction condition. After cleaving at the binding site, $E_{111}Q$ *Eco*R I falls off the template allowing RNAP to transcribe pass the stoppage site producing run-off transcripts on the truncated promoter. This was seen with *RI34* and *RI45* promoters under high *Eco*R I molar excess. To minimize the cleavage activity, a range of transcription buffer condition were tested and the optimized condition was found to be pH 7.2 and 5 mM $MgCl_2$ concentration in Tris-HCl buffer. Previous studies indicate that this buffer condition allows abortive initiation and promoter escape (Hsu *et al*., 2003). This result correlates with conditions used in previously published studies (Wright *et al.*, 1989). However in their studies, the observed cleavage activity with pH 8.5 was much less in comparison to ours. The rate constants for cleavage activity on the first and second strand of DNA were $4.3 \pm 0.9 \times 10^{-2}$ and $3.4 \pm 0.7 \times 10^{-2}$ min^{-1} for the cutting activity on each strand of DNA. This is a 10,000-fold decrease from wild type cleavage activity, which would not be visible on gel shift in a 15-minute reaction. We attribute this marked difference to the significant difference in the length of DNA, as linearized pBR322 DNA (4363 bp) was used in their study, while 140-bp templates were used in our study.

Having established conditions that are compatible with transcription and allow binding of $E_{111}Q$ *Eco*R I to the DNA templates without cleavage activity, the next step of the study will determine the effect of these bound proteins on VLATs formation. *In vitro* transcription in the presence of $E_{111}Q$ *Eco*R I will be performed in the newly optimized condition. If indeed *Eco*R I roadblocks do not

stall the advancing RNAP, it would be interesting to "titrate" the distance over which RNAP becomes a "stallable" machine. Multiple roadblocks can be placed to assess the force associated with an "escaping" RNAP.

In summary, RNAP behaved clearly differently during transcription initiation, promoter escape and stable elongation. It forward translocates with much more force during promoter escape than stable elongation. The increased amount of force propels RNAP hyper forward-translocation more than one nucleotide at a time. The study further clarifies how an escaping RNAP molecule behaves.

REFERENCES

Carpousis, A. J. and Gralla, J. D. (1980) Cycling of ribonucleic acid polymerase to produce oligonucleotides during initiation in vitro at the lac UV5 promoter. *Biochemistry* **19,** 3245-53.

Chander, M., Austin, K. M., Aye-Han, 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,** 12687-99.

Gralla, J. D., Carpousis, A. J. and Stefano, J. E. (1980) Productive and abortive initiation of transcription in vitro at the lac UV5 promoter. *Biochemistry* **19,** 5864-9.

Haugen, S. P., Berkmen, M. B., oss, W., Gaal, T., Ward, C. and Gourse, R. L. (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125,** 1069-82.

Haugen, S. P., oss, W., Manrique, M. and Gourse, R. L. (2008) Fine structure of the promoter-sigma region 1.2 interaction. *Proc. Natl. Acad. Sci. USA* **105,** 3292-7.

Helmann, J. D. and Chamberlin, M. J. (1988) Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **57,** 839-72.

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,** 8841-54.

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,** 3777-86.

Johnston, D.E. and W.R. McClure In: R. Losick and M. Chamberlin, Editors, *RNA Polymerase*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1976), pp. 413–428.

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,** 1144-7.

Komissarova, N. and Kashlev, M. (1998) Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **95,** 14699-704.

Krummel, B. and Chamberlin, M. J. (1992) Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J. Mole. Biol.* **225,** 239-50.

Kulbachinskiy, A. and Mustaev, A. (2006) Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J. Biol. Chem.* **281,** 18273-6.

Lisser, S. and Margalit, H. (1993) Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Research* **21,** 1507-16.

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

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296,** 1285-90.

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,** 9960-9.

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,** 1139-43.

Ross, W., Aiyar, S. E., Salomon, J. and Gourse, R. L. (1998) *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *J. Bacteriol.* **180,** 5375-83.

Vassylyev, D. G., Vassylyeva, M. N., Zhang, J., Palangat, M., Artsimovitch, I. and Landick, R. (2007) Structural basis for substrate loading in bacterial RNA polymerase. *Nature* **448,** 163-8.

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,** 3798-811.

Wright, D. J., King, K. and Modrich, P. (1989) The negative charge of Glu-111 is required to activate the cleavage center of *EcoR* I endonuclease. *J. Biol. Chem.* **264,** 11816-21.