

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

Doris Priscilla Tabassum  
Signature

05/12/2010  
Date

PROMOTER ESCAPE BY *E. COLI* RNA  
POLYMERASE: ITS DEPENDENCE ON  
UPSTREAM REWINDING POTENTIAL OF THE  
-10/DISCRIMINATOR (DIS) REGION

BY:  
DORIS PRISCILLA TABASSUM

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

Department of Biochemistry  
South Hadley, MA 01075

2010

This thesis was prepared under the direction of:

Dr. Lilian M. Hsu  
Professor of Biochemisry  
for eight credits.

*For Nanu and Dadu, my two grandmothers, who taught me how to be resilient and unswerving no matter what the setbacks are. I shall make you proud.*

## ACKNOWLEDGMENTS

To Professor Hsu, for trusting me more than I trusted myself. From day one, you have always been there to answer my questions and guide me to the right path. You have inspired me to push beyond my limits and keep working hard. It has been an honor to have you as my adviser and mentor.

To my family and friends for believing in my abilities and allowing me to explore new and unknown territories. Your continuous encouragement has given me the strength to make it through each day.

To my lab-mates, Tenaya, Casey and Ahri, for making my journey all the more memorable. Transcription would not have as much fun without you guys.

To the Mount Holyoke College Faculty and Staff for being a wonderful resource and giving me more than I could have asked for. You have helped me grow and mature into the scientist I am today.

And finally, I would like to give a special thanks to HHMI (grant#52006307) and NSF (grant#RUI-0841452) for funding my research.

## TABLE OF CONTENTS

	PAGE
<b>ABSTRACT</b>	1
<b>INTRODUCTION</b>	3
<b>MATERIALS AND METHODS</b>	
PROMOTER TEMPLATE CONSTRUCTION	17
PURIFICATION OF THE PROMOTER CONSTRUCTS	28
TRANSCRIPTION ASSAYS	29
VARYING THE NUCLEOTIDE CONCENTRATION FOR THE GC-SERIES	30
TESTING FOR PAUSED TRANSCRIPTS IN DT-12 VIA NTP CHASE AND HEPARIN	31
TESTING THE SENSITIVITY OF VLATs TO GREB	31
DENATURING PAGE	32
QUANTITATION OF PRODUCTS	33
<b>RESULTS</b>	
<b>THE DT-SERIES</b>	34
RATIONALE	34
EFFECT OF THE UPSTREAM REWINDING POTENTIAL ON THE ABORTIVE PROFILE OF N25 USING TEMPLATE MISMATCHES	41
EFFECT OF THE UPSTREAM REWINDING POTENTIAL ON THE ABORTIVE PROFILE OF N25 USING NON-TEMPLATE MISMATCHES	66
SUMMARY OF FINDINGS FROM THE DT-SERIES	71
<b>THE GC-SERIES</b>	74
RATIONALE	74
TRANSCRIPTION OF THE GC-PROMOTERS WITH [ $\gamma$ - <sup>32</sup> P]-ATP LABEL	78
TRANSCRIPTION OF THE GC-PROMOTERS WITH [ $\alpha$ - <sup>32</sup> P]-UTP LABEL	97
SUMMARY OF FINDINGS FROM THE GC-SERIES	111
<b>DISCUSSION</b>	112
<b>REFERENCES</b>	123

## FIGURES AND TABLES

### FIGURES

	PAGE
1. Transition of the RNAP-promoter complex from the initiation to the elongation phase of transcription	9
2. Rewinding of the expanded transcription bubble in an initial transcribing complex (ITC)	14
3. Sequences of DT-3 through 12, <i>N25</i> mutants with sequence changes on the template strand	38
4. Sequences of DT-14 through 16, <i>N25</i> mutants with sequence changes on the non-template strand	40
5. Gel image of transcription reactions carried on the DT-series with 1-3 bp mismatches	43
6. Comparison of the abortive profiles of DT-2 (wt- <i>N25</i> ) with the constructs having small mismatches-DT-3 (1-bp mm at -6), DT-4 (2-bp mm from -6 to -5) and DT-5 (3-bp mm from -6 to -4)	45
7. Gel image of transcription reactions carried on the DT-series with 3 bp mismatches	47
8. Comparison of the abortive profiles of DT-2 (wt- <i>N25</i> ) with the constructs having 3 bp mismatches: DT-5 (mm -6 to -4), DT-7 (mm -11 to -9), DT-8 (mm -8 to -6) and DT-9 (mm -3 to -1)	49
9. The “anti” algorithm (A to C and G to T substitutions) changes could give rise to staggered hybridization due to Watson-Crick basepairing by skipping a base	52
10. Gel image of transcription reactions carried out with DT-series containing larger than 5-bp mismatches	54
11. Abortive profile of the DT-series with the larger mismatches	56
12. Gel image confirming that the products of length 19 and 21 nts produced by DT-12 are not due to paused transcripts	59
13. Gel image with [ $\gamma$ - <sup>32</sup> P]-ATP label confirming that the VLATs produced by DT-12 are resistant to GreB rescue	63
14. Gel image with [ $\alpha$ - <sup>32</sup> P]-UTP label confirming that the VLATs produced by DT-12 are resistant to GreB rescue	65

15.	Gel image of transcription reactions of the DT-series templates with mismatches on the DIS-region, non-template strand	68
16.	Abortive profiles of the DT-series with mismatches on the DIS-region, non-template strand	70
17.	Sequences of GC-promoters, <i>N25</i> mutants with GC-rich DIS regions compared with wt- <i>N25</i> and the <i>rrnB</i> P1 promoter.	76
18.	Gel image comparing the GC promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt- <i>N25</i> )	80
19.	Comparison of the abortive profiles of GC promoters with DT-2 (wt- <i>N25</i> )	82
20.	Gel image comparing transcription of GC-6, -7, and DT-2 (wt- <i>N25</i> ) under variable NTP concentrations (100, 250 and 500 $\mu$ M NTP) and [ $\gamma$ - <sup>32</sup> P]-ATP label	85
21.	Comparison of abortive profiles of GC-6 and -7 transcription with DT-2 (wt- <i>N25</i> ) under variable NTP concentrations (100, 250 and 500 $\mu$ M NTP) and [ $\gamma$ - <sup>32</sup> P]-ATP label	87
22.	Ratio of total transcription initiation events at higher [NTP] relative to that with 100 $\mu$ M NTP	89
23.	Gel image comparing GC-6 and -7 transcriptions with DT-2 (wt- <i>N25</i> ) under varying KCl concentrations (20, 50, 100 and 200 mM)	92
24.	Comparison of the abortive profiles of GC-6 and -7 with DT-2 (wt- <i>N25</i> ) under varying KCl concentrations (20, 50, 100 and 200 mM)	94
25.	Comparison of the abortive profiles of DT-2, GC-6, and -7 at different KCl concentrations (20, 50, 100 and 200 mM)	96
26.	Gel image comparing the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt- <i>N25</i> )	99
27.	Comparison of the abortive profiles of the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt- <i>N25</i> )	101
28.	Gel image comparing the GC-promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt- <i>N25</i> )	105
29.	Comparison of the abortive profiles of the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt- <i>N25</i> )	107
30.	Comparison of the abortive profiles of DT-2 (wt- <i>N25</i> ) and the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) under different [NTP] and [KCl] combinations	109



## **TABLES**

	<b>PAGE</b>
1. Sequence of the oligonucleotides used in the construction of the DT- and GC-promoters	20
2. Sequence of the DT-promoters	23
3. Sequence of the GC-promoters	27

## ABSTRACT

Promoter escape governs the transition of RNA polymerase (RNAP) from the initiation phase to elongation. Being rate-limited at this step, the phage T5 *N25* promoter undergoes a high degree of abortive cycling prior to escape. Escape occurs at different positions depending on the initial transcribed sequence (Hsu *et al.*, 2006). As transcription proceeds, DNA scrunching (Revyakin *et al.*, 2006) leads to transcription bubble expansion to create a strained initial transcribing complex which, in turn, can result in successful escape or abortive release depending on which end of the expanded bubble rewinds. We hypothesized that changes in the upstream (-10/ DIS region) DNA rewinding potential will affect the escape position of *E. coli* RNAP on *N25* promoters.

To decrease the upstream rewinding potential, we introduced mismatched stretches within the -10/DIS region of the promoter and predicted that these mutants (DT-series promoters) will have a delayed position of escape and show less full-length synthesis. On another set of promoters, we increased the upstream rewinding potential by using GC-rich DIS regions that are 6-8 bp in length. We expected these latter mutants (GC-series promoters) to exhibit earlier or even increased rates of escape. We analyzed the constructs

by *in vitro* transcription with [ $\gamma$ - $^{32}$ P]-ATP label followed by denaturing PAGE and phosphorimaging.

Our results with the DT-series showed that escape profiles are affected only when upstream rewinding is severely compromised with mismatched regions that are 5 bp or larger. On the contrary, the data from the GC-series did not follow our predicted outcome but instead indicated rate limitation at the open complex formation step similar to that seen in the *rrnB* P1 promoter (Barker and Gourse, 2001) further alluded from sensitivity to [NTP] and [KCl] in the reaction mixture. We can thus conclude that our hypothesis is too simplistic in predicting the role of the upstream bubble collapse in the promoter escape of *E. coli* RNAP. It is more likely that the energy for relinquishing promoter contacts is obtained from DNA scrunching (Revyakin *et al.*, 2006) rather than upstream rewinding. The data also indicate the possibility that promoter escape by *E. coli* RNAP can occur with or without the prior collapse of the open complex bubble.

## INTRODUCTION

Transcription is an indispensable process in gene expression whereby the information stored in DNA is decoded in the form of an RNA transcript. Hence, it is not at all surprising to see that this process is always maintained under strict regulation and usually depends on a plethora of associated factors, many of which are yet to be deciphered. This research aims to provide some insight into one of the aspects of the transcriptional regulation in the prokaryotic system using the *Escherichia coli* RNA polymerase (RNAP) that carries out template-dependent *de novo* synthesis of RNA.

Transcription can be divided into three main stages: initiation, elongation, and termination. Each of these stages is usually orchestrated by a series of highly coordinated steps, many of which often determine the success of completion of the entire process. During the initiation phase of transcription, the RNAP recognizes specific sequences in the promoter region of the DNA template, especially at the -35 and the -10 hexanucleotides, and binds to them through the  $\sigma$ -subunit. The closer the sequences of these regions are to the consensus (Hawley and McClure, 1983), the higher is the binding affinity (Roberts and Roberts, 1998). The resultant complex is highly unstable and is referred to as the closed complex (RP<sub>C</sub>) (Chamberlin, 1974). The subsequent step involves melting of ~13 bp of DNA, spanning the start site of transcription, which is brought within the enzyme to the active site of

the RNAP forming the open complex (RP<sub>O</sub>). At this point, the RNAP can begin *de novo* synthesis of the RNA molecule from the +1 site, converting this entire complex into an initial transcribing complex (ITC; Krummel and Chamberlin, 1989). With *de novo* initiation, the RNAP can perform repetitive incorporation of nucleoside monophosphates (NMPs) to grow the transcript, proceed into the elongation phase, and eventually encounter a termination signal which halts transcription leading to the release of the full-length transcript and the enzyme from the DNA. The released RNAP can then bind to another promoter and initiate a new cycle of transcription (Das, 1993).

The transition of a transcribing complex from the initiation to the elongation phase, however, is quite a bit more complicated than alluded to above. In fact, this process involves a highly coordinated sequence of events, many of which still remain poorly understood. Previous studies with single-molecule DNA nanomanipulation (Revyakin *et al.*, 2004) and fluorescence resonance energy transfer (FRET) (Kapanidis *et al.*, 2006) have shown that, during the earliest stages of transcription initiation, and right through transition into elongation, a process called DNA scrunching takes place within the RNA polymerase (Revyakin *et al.*, 2006; Roberts, 2006). DNA footprinting assays have further revealed that although the upstream protection boundary of the DNA by the RNAP in both the RP<sub>O</sub> and the ITC are identical, the downstream boundary advances in the ITC (Carpousis and Gralla, 1985; Krummel and Chamberlin, 1992). This indicates the RNAP does not release its promoter contacts during the transformation from the RP<sub>O</sub> to the ITC but

rather pulls more of the template DNA within itself in order to continue transcription. Potassium permanganate sensitivity assays have confirmed that the process of DNA scrunching leads to a simultaneous expansion of the melted DNA bubble in order to expose the template sequences (Revyakin *et al.*, 2006). This need for the continual maintenance of the DNA bubble in this unwound conformation by the RNAP in addition to the maintenance of the enzyme-promoter contacts, inevitably gives rise to a highly stressed intermediate.

The accumulating stress within the RNAP arising due to DNA scrunching renders the enzyme-promoter complex highly unstable energetically which must be soon alleviated. This is usually accomplished by the rewinding of the melted DNA bubble at either boundary. At this point, if the downstream end of the bubble (including the initial transcribed sequence or ITS) is able to rewind first, it causes a backtracking movement of the RNAP during which the RNA 3' end is displaced from the RNA-DNA hybrid, extruded past the active site and directed into the secondary channel. This forms an arrested complex where the 3'-OH end of the nascent RNA is disengaged from the active site (Komissarova and Kashlev, 1997; Hsu, 2002a; Chander *et al.*, 2007). In the arrested ITCs, the backtracked nascent RNAs are held by insufficient RNA-DNA hybrid length, and thus, are spontaneously released as short transcripts (usually 2-15 nts long) thereby resulting in unsuccessful transcription. This phenomenon is referred to as abortive initiation (Munson and Reznikoff, 1981). Additionally, experiments in the

presence of heparin, which sequesters free RNAP and prevents it from being recycled, have confirmed that during this repetitive abortive release, the RNAP does not let go of the promoter region (Carpousis and Gralla, 1980). Alternative to abortive release of the backtracked RNA, an intrinsic endonucleolytic activity of RNAP can cleave the 3'-extruded RNA at the active site to regenerate a new 3'-OH group, allowing the 5'-piece of the arrested RNA to undergo re-elongation (Orlova *et al.*, 1995). This intrinsic cleavage activity of RNAP is greatly stimulated by transcript cleavage factors, such as GreA and GreB, which bind in the RNAP secondary channel to help re-establish the process of transcription (Opalka *et al.*, 2003). The presence of GreA or GreB in the transcription reaction essentially rescues the transcript synthesis and reduces the likelihood of abortive release (Hsu *et al.*, 1995; Hsu *et al.*, 2006).

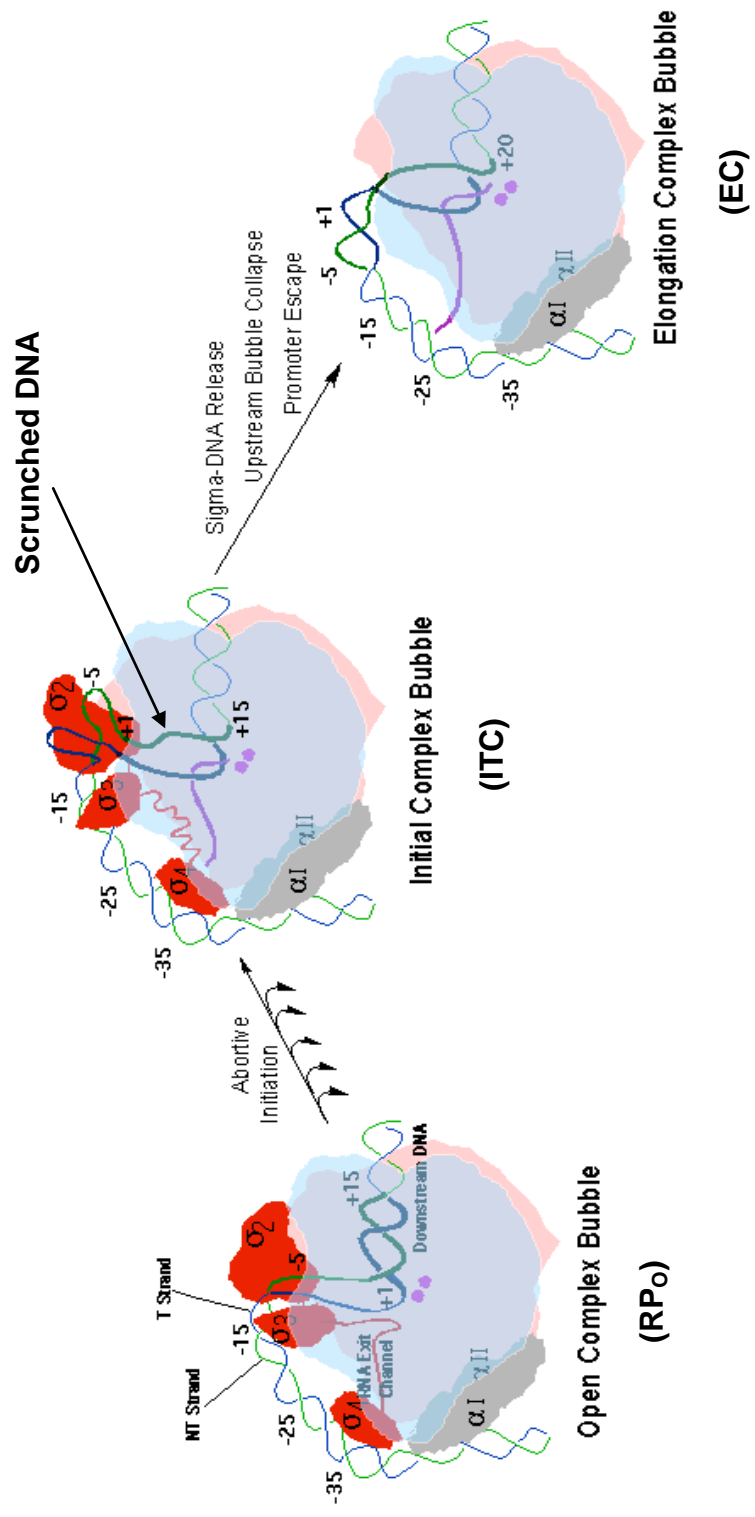
By the same token, in order to relieve the strained ITC, the upstream end (spanning the -10 box and the discriminator or DIS region) of the bubble can also collapse. This, in turn, leads to the productive release of the cumulative stress whereby the RNAP is able to relinquish its hold on the promoter (Chander *et al.*, 2007). The latter sequence of events leads to promoter escape which marks the transition from initiation to elongation and is often, but not always, accompanied by a concomitant release of the  $\sigma$ -factor from the RNAP (Hsu, 2002b). Although abortive initiation usually occurs repeatedly before successful promoter escape, once the elongation phase is achieved, the RNAP can continue the synthesis of the full-length transcript

without any hindrance and the final product is subsequently released during termination (Munson and Reznikoff, 1981; Hsu, 2002a). The transition of the RNAP from the initiation to the elongation phase of transcription is illustrated in Figure 1 which highlights some of the described interactions between the RNAP polymerase and the promoter during abortive and productive initiation.

It is interesting to note that, contrary to the promoter recognition process, the closer the sequences of the -35 and -10 boxes are to the consensus, the more difficult is the promoter escape (Hsu, 2002a; Vo *et al.*, 2003). Additionally, various other factors can also govern abortive initiation and promoter escape. These can be classified both as intrinsic and extrinsic elements. The intrinsic elements are the ones that are present within the DNA template, such as the sequences within the promoter recognition region (PRR), the initial transcribed sequences (ITS) as well as the conformation of the DNA, i.e. whether or not it is present in a supercoiled form. The extrinsic elements, on the other hand, include factors such as the presence of activator or repressor proteins, KCl concentration (the higher it is, the more stable is the open complex), NTP concentration (the higher it is, the more frequent is the initiation), and the structure of the RNAP (Hsu, 2002a).



**Figure 1. Transition of the RNAP-promoter complex from the initiation to the elongation phase of transcription.** The figure illustrates the interactions between the RNAP and the promoter in the open complex (RP<sub>O</sub>) leading to DNA scrunching in the initial transcribing complex (ITC). The strain accumulating due to scrunching can be relieved by bubble collapse at either boundary eventually leading to repeated rounds of abortive release or promoter escape to the elongation complex (EC). (Abbreviations used: T: template; NT: non-template)



Nevertheless, it is a combination of all of the intrinsic and extrinsic factors mentioned that determines the abortive RNA profile- the collection of abortive products generated during transcription of a particular promoter and their relative abundance. The products can be separated by gel electrophoresis and visualized as an abortive ladder. The position of escape is marked by a disappearance of the ladder after a particular length whereupon only the full-length transcript is detectable. It is also worthy of mention that, despite the seemingly stochastic nature of abortive initiation, the abortive RNA profile obtained at each promoter is reproducible (Hsu, 2002a).

As can be expected, the process of promoter escape varies greatly from one promoter to the next. In this project, we have focused mainly on the bacteriophage promoter T5 *N25* which utilizes the host RNAP (in this case *E. coli* RNAP) to express its early genes. The -35 and -10 sequences of this promoter are very similar to the consensus (Hawley and McClure, 1983) making it particularly rate-limited at the promoter escape step (Knaus and Bujard, 1990). Regardless of this rate-limitation, productive synthesis occurs readily on *N25* and escape occurs at the +12 position (Hsu *et al.*, 2003).

A paper published by Hsu *et al.* in 2006 has demonstrated that the rewinding potential of *N25*, and hence the abortive profile and promoter escape, is altered if the ITS is randomly mutagenized from + 3 to + 19. The investigators examined ~40 such ITS mutants and noticed that these mutants

underwent a much delayed promoter escape around the +15/+16 position compared to the +12 for the wt-*N25*. This pattern of delayed escape was similar to that of the *N25<sub>anti</sub>* promoter, where the first 20 nts of the ITS of *N25* is changed by incorporating A ↔ C and G ↔ T substitutions. The *N25<sub>anti</sub>* promoter is also ~10-fold weaker in mRNA production compared to the wt-*N25* and escapes only at the +16 position. Also of note here is another promoter *DG203*, which differs from *N25* in the sequence from +3 to +10 and has a relatively GC-rich ITS. *DG203* undergoes promoter escape at an even later position (+20) suggesting the presence of some correlation between the sequence composition at the ITS and the pattern of promoter escape (Hsu *et al.*, 2006; Chander *et al.*, 2007). The ITS from +1 to +20 of *N25*, *N25<sub>anti</sub>* and *DG203* are given as follows:

	+1	+20
<i>N25</i> :	ATAAATTTGA	GAGAGGAGTT
<i>N25<sub>anti</sub></i> :	ATCCGGAATC	CTCTTCCCGG
<i>DG203</i> :	ATGCGACCGG	GAGAGGAGTT

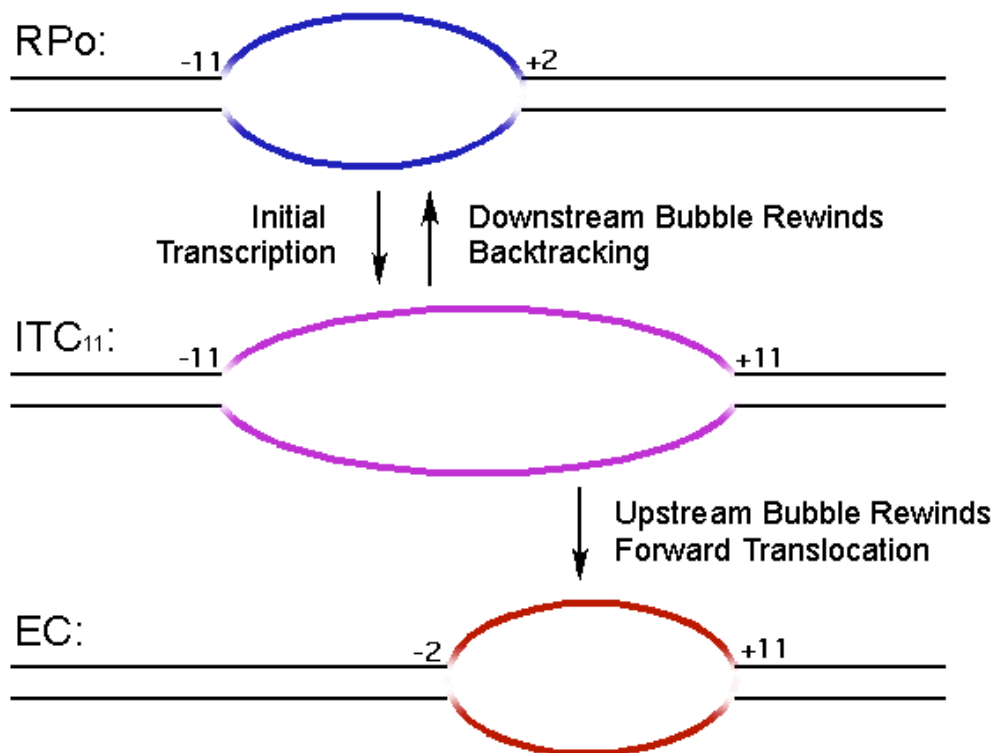
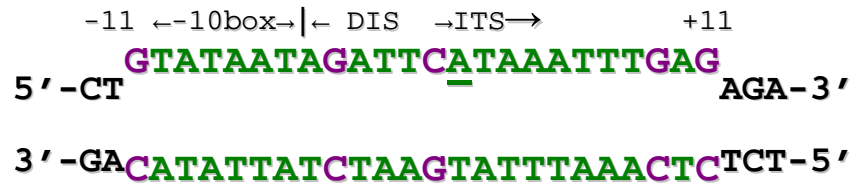
The revelation of an altered pattern of escape in the ITS mutants led to a closer examination of the relative sequences of the DNA bubble in the ITC on the cusp of escape. In case of wt-*N25*, this is ITC<sub>11</sub> whose bubble spans -11 to +11 and the sequence, ATAATAGATTC ATAAATTTGAG, contains iso-energetic upstream and downstream halves. Since the two halves contain equal A/T and G/C basepairs, rewinding of the DNA at either boundary of the ITC<sub>11</sub> bubble is equally likely. It has been speculated that the promoter escape

from *N25* at the +12 position arises primarily from this equipotential ITC bubble rewinding. In contrast, however, in *N25<sub>anti</sub>* and *DG203*, both of which demonstrate delayed promoter escape and lower productive synthesis, the ITS, and hence the downstream end of the ITC, are GC-rich giving rise to a greater likelihood of downstream rewinding. This delayed promoter escape has led to the hypothesis that the relatively greater downstream rewinding potential, compared to that of the upstream end, of the ITC in these two promoters is the reason for more frequent backtracking leading to corresponding decrease in the amount of full-length synthesis.

The cumulative evidence from the ITS mutants eventually led Hsu and colleagues to propose the model for abortive initiation and promoter escape based on the relative rewinding potentials of the upstream and the downstream ends of the ITC bubble. The model is illustrated in Figure 2. They proposed that if the downstream end of the ITC bubble (spanning the ITS) can rewind first, it can lead to backtracking and abortive release. On the other hand, if the upstream end (spanning the -10/DIS region) rewinds first, it can lead to promoter escape and forward translocation into the elongation phase.

**Figure 2. Rewinding of the expanded transcription bubble in an initial transcribing complex (ITC).** The ITC bubble can span from the -11 to the +11 position in *N25* creating a highly stressed intermediate. The sequence illustrates that the ITC<sub>11</sub> bubble is iso-energetic at either boundary. If the downstream end (spanning the ITS) of the bubble can rewind first, it can lead to backtracking and abortive release. But, if the upstream end (consisting of the -10/DIS region) rewinds first, it can lead to promoter escape by forward translocation. The +1 start site is underlined.

## The ITC of T5 N25



Since the effect of *N25* ITS changes on the abortive pattern of *E. coli* RNAP had already been investigated (Hsu *et al.*, 2006), we were interested in looking at the effect of upstream changes on the promoter escape patterns. We hypothesized that changes in the upstream sequences of the ITC bubble (spanning the -11 to -1 positions and encompassing the -10 box/discriminator (DIS) region) may lead to an altered upstream rewinding potential which, in turn, will affect promoter escape patterns of *N25* and related promoters.

Unlike the ITS region changes, it was not practical to randomly change the sequence at the upstream end of the ITC because this region is involved in determining the stability of the open-complex formed between the RNAP and the promoter (Marr and Roberts, 1997) giving us very few options for altering the upstream rewinding potential and testing our hypothesis. Nonetheless, we used two separate approaches to construct two sets of *N25* promoter variants while keeping the ITS unaltered from that of the wt-*N25*. The first approach involved “decreasing” the upstream rewinding potential for which we incorporated mismatch bubbles within the -10/DIS region. Hence we designed the **DT-series** promoters. With the DT-series, we expected the reduced upstream rewinding to promote either a lower rate of escape, delayed position of escape and/or a decrease in the amount of full length transcript synthesis. The second approach involved “increasing” the upstream rewinding potential by enriching the GC-content and length of the DIS region of *N25*. Hence the **GC-series** promoters were synthesized. With the GC-series, we expected the increased upstream rewinding to result in either a destabilized open complex



formation leading to a drop in amount of the shorter abortive fragments or a shorter abortive ladder arising from earlier escape (Barker *et al.*, 2001a). The promoters were constructed by primer extension of long oligonucleotides and analyzed via *in vitro* transcription reactions followed by denaturing PAGE to compare the abortive profiles both qualitatively and quantitatively. The results of the investigation are described in this work.

## MATERIALS AND METHODS

### *PROMOTER TEMPLATE CONSTRUCTION:*

The promoter constructs were prepared by primer extension of two annealed long oligonucleotides. Each of the constructs was prepared from two 100- $\mu$ L reactions. Each reaction was set up in two halves; the first half, the primer mixture, consisted of 10  $\mu$ L each of the upstream and downstream primers (each is 10  $\mu$ M), 5  $\mu$ L of 10X NEB Buffer 2 (500 mM NaCl, 100 mM Tris-HCl, 100 mM MgCl<sub>2</sub>, 10 mM DTT at pH 7.9), and 25  $\mu$ L of doubly distilled water (dd-H<sub>2</sub>O). The primer mixtures were carried through a series of temperature incubation (20' at 70 °C, 10' at 50 °C, 10' at 42 °C and 10' at 37 °C) to encourage annealing before the addition of 50  $\mu$ L of a dNTP solution (250  $\mu$ M in 1X NEB2) and 3  $\mu$ L of Klenow Polymerase I (NEB; 5 units/ $\mu$ L). The mixtures were incubated for 30 minutes at 37 °C followed by reaction termination on ice. The presence of the constructs was confirmed by electrophoresis of a 3- $\mu$ L aliquot of each sample in 7  $\mu$ L of 1X loading dye (10 mM Na<sub>2</sub>EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol, 4% glycerol) in 2% agarose gel for 30 min at 100 V followed by UV detection.

The **DT-series** promoters are 142-bp long (-85 to +57) *N25* promoters containing mismatches in the -10 box/DIS region (TATAAT/AGATTC). The **GC-series** promoters are 147-bp long (-85 to +62) *N25* containing GC-rich substitutions of the DIS sequence. The ITS for both sets of promoters is identical to that of *N25*. The full-length products from the DT-series promoters are 57 nt long whereas those from the GC-series promoters are 62 nt long. The sequences of the primers used and the resulting promoters are given in the following tables.

**Table 1. Sequence of the oligonucleotides used in the construction of the DT- and GC-promoters.** The -35 and -10 hexanucleotides are bold-typed. The +1 start site is underlined. The sequence changes to create the mismatches in the DT-promoters are shown in red. The DIS-regions of the GC-promoters are shown in green.

Primer ID	Boundaries	5'→3'
<i>Upstream Primers</i>		
DT-1	-85 → +13	5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT <b>TTGCTTTCAG</b> GAAAATTTTT CTGTATAATA GATT <u>C</u> ATAAA TTTGAGAG 3'
DT-6	-85 → -11	5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT <b>TTGCTTTCAG</b> GAAAATTTTT CTG 3'
DT-14	-85 → +13	5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT <b>TTGCTTTCAG</b> GAAAATTTTT CTGTATAAT <b>C</b> <b>TCTT</b> CATAAA TTTGAGAG 3'
DT-15	-85 → +13	5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT <b>TTGCTTTCAG</b> GAAAATTTTT CTGTATAATA GACAT <u>A</u> ATAAA TTTGAGAG 3'
DT-16	-85 → +13	5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT <b>TTGCTTTCAG</b> GAAAATTTTT CTGTATAAT <b>C</b> <b>TCCAT</b> <u>A</u> ATAAA TTTGAGAG 3'
<i>Downstream Primers</i>		
DT-2	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT CTATTATACA GAAAATTTTT CC 3'
DT-3	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT <b>CC</b> ATTATACA GAAAATTTTT CC 3'
DT-4	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT <b>TC</b> ATTATACA GAAAATTTTT CC 3'
DT-5	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAC <b>TC</b> ATTATACA GAAAATTTTT CC 3'
DT-7	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT CCATTAT <b>CGC</b> GAAAATTTTT CC 3'
DT-8	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT <b>CCGC</b> TATACA GAAAATTTTT CC 3'
DT-9	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTAT <b>AGGT</b> CCATTATACA GAAAATTTTT CC 3'
DT-10	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT CCATT <b>GCCGC</b> GAAAATTTTT CC 3'

DT-11	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAC CCGGAATACA GAAAAATTTT CC 3'
DT-12	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAG CCGGGGCCGC GAAAAATTTT CC 3'
DT-13	+57 → -26	5' GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAAT CTATTATACA GAAAAATTTT CC 3'
GC-6	+62 → -33	5' CGGGATCCTG CAGAAGCTTT CTGCGAGAAC CAGCCATATT TAAACTCCTC TCTCAAATTT ATGAGCGCAT TATACAGAAA AATTTTCCTG AAAGC 3'
GC-7	+62 → -33	5' CGGGATCCTG CAGAAGCTTT CTGCGAGAAC CAGCCATATT TAAACTCCTC TCTCAAATTT ATGACGCGCA TTATACAGAA AAATTTTCCT GAAAGC 3'
GC-8	+62 → -33	5' CGGGATCCTG CAGAAGCTTT CTGCGAGAAC CAGCCATATT TAAACTCCTC TCTCAAATTT ATGGACGCGC ATTATACAGA AAAATTTTCC TGAAAGC 3'
GC-8-1	+62 → -33	5' CGGGATCCTG CAGAAGCTTT CTGCGAGAAC CAGCCATATT TAAACTCCTC TCTCAAATTT ATAGACGCGC ATTATACAGA AAAATTTTCC TGAAAGC 3'
GC-8-2	+62 → -33	5' CGGGATCCTG CAGAAGCTTT CTGCGAGAAC CAGCCATATT TAAACTCCTC TCTCAAATTT ATAGAGGCGC ATTATACAGA AAAATTTTCC TGAAAGC 3'

**Table 2. Sequence of the DT-promoters.** The -35 and -10 hexanucleotides, and the +1 start site are underlined. The mismatched nucleotides for the DT-promoters are shown in red.

Promoter	Sequence
DT-2 (wt-N25)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATA</b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAT CTAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-3 (mm -6)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATA</b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAC CTAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-4 (mm -6/-5)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATA</b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAC TAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-5 (mm -6/-4)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATA</b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAC TCAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-7 (mm -11/-9)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATA</b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACACGCTAT CTAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>



DT-8 (mm -8/-6)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATAC<b>CGCT</b> CTAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-9 (mm -3/-1)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAT CT<b>GGA</b>TATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-10 (mm -11/-7)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACAC<b>CGCCG</b>T CTAAGTATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-11 (mm -6/-2)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAA <b>GGCCG</b>TATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
DT-12 (mm -11/-2)	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> GATTCATAAA TTTGAGAGAG GAGTTTAAAT GACAC<b>CGCCGG</b> <b>GGCCG</b>TATTT AAACCTCTCTC CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>

<p>DT-14 (mm -11/-9)</p>	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAAT</u></b>C <b>T</b>CTTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAT CTAAGTATTT AAACCTCTCTC CTCAAAATTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
<p>DT-15 (mm -11/-9)</p>	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAATA</u></b> <b>GACAT</b>ATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAT CTAAGTATTT AAACCTCTCTC CTCAAAATTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>
<p>DT-16 (mm -11/-9)</p>	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b><u>TTGCTT</u></b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTAAAAA</p> <p>CTG<b><u>TATAAT</u></b>C <b>TCCAT</b>ATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAT CTAAGTATTT AAACCTCTCTC CTCAAAATTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGCC 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCGG 5'</p>

**Table 3. Sequence of the GC-promoters.** The -35 and -10 hexanucleotides, and the +1 start site are underlined. The DIS-regions are shown in green.

Promoter	Sequence
GC-6	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATG</b> CGCTCATAAA TTTGAGAGAG GAGTTTAAAT GACATATTAG GCGAGTATTT AAACCTCTCT CTCAAATTTA</p> <p>ATGGCTGGCT GGTTCCTCGCA GAAAGCTTCT GCAGGATCCC G 3' TACCGACCGA CCAAGAGCGT CTTTCGAAGA CGTCCTAGGG C 5'</p>
GC-7	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATG</b> CGCGTCATAA ATTTGAGAGA GGAGTTTAAA GACATATTAG GCGCAGTATT TAAACTCTCT CCTCAAATTT</p> <p>TATGGCTGGC TGGTTCTCGC AGAAAGCTTC TGCAGGATCC CG 3' ATACCGACCG ACCAAGAGCG TCTTTCGAAG ACGTCCTAGG GC 5'</p>
GC-8	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATG</b> CGCGTCCATA AATTTGAGAG AGGAGTTTAA GACATATTAG GCGCAGGTAT TTAAACTCTC TCCTCAAATTT</p> <p>ATATGGCTGG CTGGTTCTCG CAGAAAGCTT CTGCAGGATC CCG 3' TATACCGACC GACCAAGAGC GTCTTTCGAA GACGTCCTAG GGC 5'</p>
GC-8-1	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATG</b> CGCGTCTATA AATTTGAGAG AGGAGTTTAA GACATATTAG GCGCAGATAT TTAAACTCTC TCCTCAAATTT</p> <p>ATATGGCTGG CTGGTTCTCG CAGAAAGCTT CTGCAGGATC CCG 3' TATACCGACC GACCAAGAGC GTCTTTCGAA GACGTCCTAG GGC 5'</p>
GC-8-2	<p>5' GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG 3' 3' CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC 5'</p> <p>GGAAATCATA AAAAATTTAT <b>TTGCTT</b>TCAG GAAAATTTTT CCTTTAGTAT TTTTAAATA AACGAAAGTC CTTTTAAAAA</p> <p>CTG<b>TATAATG</b> CGCCTCTATA AATTTGAGAG AGGAGTTTAA GACATATTAG GCGGAGATAT TTAAACTCTC TCCTCAAATTT</p> <p>ATATGGCTGG CTGGTTCTCG CAGAAAGCTT CTGCAGGATC CCG 3' TATACCGACC GACCAAGAGC GTCTTTCGAA GACGTCCTAG GGC 5'</p>

*PURIFICATION OF THE PROMOTER CONSTRUCTS:*

The promoters constructed by primer extension were allowed to precipitate overnight in ethanol in the presence of 0.3 M Sodium Acetate (NaAc). The following day, the samples were centrifuged for 20 minutes at 4 °C. The supernatants were carefully poured off and the pellets from both aliquots of each promoter were re-dissolved in 100 µL TE buffer (10 mM Tris-HCl, pH 8, 1 mM Na<sub>2</sub>EDTA). To this, 100 µL of 10 M Ammonium Acetate (NH<sub>4</sub>Ac) and 600 µL of ethanol were added and the reaction tubes were allowed to stand at room temperature for 30 minutes. The samples were then centrifuged for 30 minutes at room temperature and the supernatants were decanted. The pellets were re-dissolved in 200 µL TE and the constructs were sequentially extracted from the aqueous layers with 200 µL phenyl/chloroform/isoamyl alcohol (25:24:1) followed by 200 µL chloroform/isoamyl alcohol (24:1). The aqueous layers from the latter extraction were then added with 20 µL of 3 M NaAc and the samples were allowed to precipitate at -20 °C with 50% isopropanol (1:1 volume). After at least 3 hours of low temperature incubation, the samples were spun down for 20 minutes at 4 °C. The supernatants were carefully decanted and the pellets were washed with 70% ethanol. The pellets were recovered and dried for 10 minutes in a SpeedVac rotary vacuum desiccator prior to re-suspension in 200 µL TE. 3 µL aliquots of each construct were analyzed by electrophoresis with a 2% agarose gel as before. Each sample was also quantified with the

Nanodrop Spectrophotometer for concentration determination and 300 nM dilutions were prepared for transcription reactions.

*TRANSCRIPTION ASSAYS:*

All *in vitro* transcription reactions were performed in 1.5 mL snap-top Eppendorf® tubes. Each reaction mixture had a total volume of 10  $\mu$ L. A typical reaction consisted of 30 nM promoter DNA, 1X transcription buffer III (50 mM Tris-HCl, pH 8, 10 mM MgCl<sub>2</sub>, 10 mM  $\beta$ -Mercaptoethanol, 10  $\mu$ g/mL acetylated BSA), 200 mM KCl (unless specified otherwise), 100  $\mu$ M NTP containing either [ $\gamma$ -<sup>32</sup>P]-ATP (to label the 5' end) or [ $\alpha$ -<sup>32</sup>P]-UTP (to label at all the uridine residues) at a specific activity of ~10 cpm/fmol. The mixture was brought to a total volume of 9  $\mu$ L using diethyl pyrocarbonate treated water (DEPC-H<sub>2</sub>O). The transcription reactions were initiated with the addition of 1  $\mu$ L of *E. coli* RNA polymerase freshly diluted to 500 nM with enzyme diluent (10 mM Tris-HCl, pH 8, 10 mM  $\beta$ -Mercaptoethanol, 10 mM KCl, 5% v/v glycerol, 0.1 mM Na<sub>2</sub>EDTA, 0.4 mg/mL BSA, 0.1% v/v Triton X-100). A control reaction without the enzyme was also set up with each set of reactions for background correction during quantitation. The tubes were allowed to incubate in a 37 °C water bath for 10 minutes. The reactions were then quenched via addition of 100  $\mu$ L of GES (1 mg/mL glycogen, 10 mM Na<sub>2</sub>EDTA, 0.3 M NaAc). The RNA products were precipitated by incubating the tubes overnight at -20 °C following the addition of 330  $\mu$ L ethanol. The following day, the precipitate was recovered by centrifugation at 13,000 rpm

at 4 °C for ~20 min. The ethanol supernatants were carefully decanted and the pellets were dried under vacuum in a SpeedVac rotary desiccator for 15 minutes. The pellets were then re-dissolved in 10 µL of freshly prepared formamide loading buffer (FLB; 80% deionized formamide, 1X TBE [89 mM Tris base, 89 mM Boric acid, 2.5 mM Na<sub>2</sub>EDTA], 10 mM Na<sub>2</sub>EDTA, 0.08% xylene cyanol, 0.08% amaranth). In order to ensure complete re-suspension, the samples were vortexed vigorously followed by brief centrifugation at least three times. The samples were thus ready to be loaded (in 4-µL aliquots) onto a 23% (10:1) polyacrylamide 7 M urea gel for analysis by electrophoresis.

*VARYING THE NUCLEOTIDE CONCENTRATION FOR THE GC-SERIES:*

The standard protocol for transcription requires 100 µM NTP. However, the yield of RNA products for the GC series under these conditions was very low. In order to increase the signal, NTP was added to the higher concentrations of 250 µM and 500 µM. In an NTP concentration titration experiment like this, it is necessary to maintain the specific activity of the radioactive nucleotide. To achieve this, NTP is added to the reaction from two separate stocks – a 10X stock of the radiolabeled NTP (in our experiments, usually ATP and [ $\gamma$ -<sup>32</sup>P]-ATP) and a similarly 10X stock of a mixture of the other three unlabeled nucleotides (i.e. GTP/CTP/UTP mixture). Because we needed three different 10X stocks of the radioactive NTP, the radioactive NTP stock was first prepared as the 10X stock for the highest NTP concentration tested (i.e. 5 mM at a specific activity of ~5 cpm/fmol) and then diluted to obtain the 10X stocks for the other concentrations (i.e. 2.5 mM and 1 mM).

Prepared this way, the radioactive specificity of the NTP is maintained in all reactions and the phosphorimager signal of a transcription gel can be directly compared. A similar protocol was followed when making the higher concentration NTP mixtures containing [ $\alpha$ - $^{32}$ P]-UTP.

*TESTING FOR PAUSED TRANSCRIPTS IN DT-12 VIA NTP CHASE AND HEPARIN:*

Transcription of DT-12 yielded the higher intensities than normal of abortive transcripts longer than 11 nts. The 19-21 nt products were particularly abundant in this construct. Thus, in order to verify whether these products were indeed very long aabortive transcripts (VLATs) and not paused transcripts, an NTP chase reaction was carried out. To do so, a 10- $\mu$ L aliquot of the reaction mixture was brought to 1 mM NTP mixture after three or ten minutes of transcription, followed by an additional five minutes of incubation at 37 °C before termination with GES. For each time point, an equivalent aliquot was terminated with GES without NTP chase. The NTP chase was carried out with the DT-2 template as a control. For DT-12 another set of reactions was performed in the presence of heparin (10  $\mu$ g/mL; added with the excess NTP) which prevents RNAP recycling by sequestering any free enzyme in the system. All three sets of data were then compared qualitatively.

*TESTING THE SENSITIVITY OF VLATs TO GREB:*

Once it was established that the 19-21 nt products from DT-12 were indeed VLATs and not paused transcripts, their sensitivity to GreB was tested. For such an analysis, the transcription reactions were carried out for DT-2, 10,



11 and 12 in the presence and absence of excess GreB. The GreB protein was isolated previously from IPTG-induced JM109 cells harboring the pGF296 plasmid (Feng *et al.*, 1994) by L. Hsu and was added to the RNA polymerase at 10:1 molar ratio (GreB:RNAP) during the dilution step to make the 10X RNAP (-/+ GreB) stock. The GreB experiments were done with either [ $\gamma$ - $^{32}$ P]-ATP or [ $\alpha$ - $^{32}$ P]-UTP label and the gels were analyzed qualitatively.

#### *DENATURING PAGE:*

Glass plates were carefully cleaned with deionized water and ethanol and then briefly siliconized (~5 min) with RainX prior to assembling into a sequencing gel mold of 0.4 mm thickness. The narrower gels consisting of 18 wells were made with a 40-mL gel mixture consisting of 16.8 g urea dissolved in 4 mL 10X TBE and 23 mL of 40% (10:1) acrylamide-bisacrylamide stock. The mixture was allowed to polymerize for an hour at room temperature following the addition of 400  $\mu$ L of 10% APS (Ammonium Persulfate) and 40  $\mu$ L TEMED (N,N,N',N'-tetramethylethylenediamine). The wider gels containing 27-35 wells were made with 80-mL gel mixture of the same composition. The gel was set up in a high-voltage electrophoresis apparatus and electrophoresis was carried out in a salt gradient buffer using 1X TBE as the top reservoir buffer and 1X TBE with 0.3 M NaAc as the bottom buffer. The salt gradient ensured proper separation of the shorter RNA fragments near the bottom of the gel as well as entry of the longer products into the gel (Hsu, 2009). A pre-run with FLB was carried out for about 10 min to ensure proper assembly of the apparatus. Aliquots of 4  $\mu$ L of each sample

were then loaded onto the wells and electrophoresis was carried out at 35W for 3-5 hr (depending on the size of the gel) until the amaranth dye was about 1 cm from the bottom of the gel. The gel was then prepared for imaging by prying apart the plates, blotting of excess buffer on the gel, and covering the gel with Saran Wrap. It was then exposed to a phosphorimager screen overnight and scanned in a GE Storm Phosphorimager followed by quantitation using the ImageQuant software.

*QUANTITATION OF PRODUCTS:*

Each band on a gel image corresponds to an RNA product of a particular size. Since all the templates were variants of the *N25* promoter with identical ITS known to support escape after the 11<sup>th</sup> position, the abortive ladder was usually characterized from 2-11 nt products. For DT-10, 11 and 12, the ladders were characterized up to 21<sup>st</sup> position. The full-length transcripts were also measured. The bands at each position were enclosed in rectangles and, following background correction using the lane without the enzyme, the intensities were quantitated as ImageQuant Volume (IQV) counts. The IQV values for all the RNAs from a promoter were used to calculate the abortive probability at each of the positions of the abortive ladder (Hsu, 2009). In case of the [ $\alpha$ -<sup>32</sup>P]-UTP labels, each band was “normalized” by dividing the IQV with the number of U’s in the sequence of the particular product. The abortive ladders with [ $\alpha$ -<sup>32</sup>P]-UTP label were characterized from the 3-11 nt products. Each experiment was done three or four times to obtain the average and standard deviation values. The data was illustrated using Microsoft Excel.

## RESULTS

### THE DT-SERIES

#### *RATIONALE*

The promoter escape properties of *E. coli* RNA polymerase depend on a variety of different factors that include the promoter itself, as well as the initial transcribed sequences (ITS) (Hsu, 2002a). The effect of the latter was demonstrated very clearly in the paper by Hsu *et al.* (2006) in which they analyzed the escape patterns of ~40 random initial sequence variants of *N25* and found that all of the mutants exhibited delayed escape at the +15/+16 positions instead of the regular +12 position for the wt-*N25*. This necessitated a closer examination of the mutants which hinted that the change in the abortive ladder may have arisen due to an altered downstream rewinding potential of the initial transcribing complex (ITC) bubble.

The observation of the altered abortive profiles of the ITS mutants led to the proposal of the model described in Figure 2. Consequently, we sought to investigate whether similar results as the ITS mutants would be seen if instead the upstream rewinding potential was changed while keeping the ITS constant. Since the upstream end of the ITC includes the -10 box that is involved in the promoter-RNAP interaction, our options of changing the rewinding potential at this end were quite limited. We thus focused our

attention mostly to the discriminator (DIS) region that spans the -6 to -1 positions of the *N25* promoter and manipulated it in two different ways in order to change the upstream rewinding potential.

The first set of promoters, referred to as the DT series, was constructed with a “reduced” upstream rewinding potential due to mismatch bubbles at various locations of the -10/DIS region. We began our investigation with small mismatched bubbles, 1-3 bp in length and constructed DT-3, 4, and 5. However, the lack of any observable changes in the abortive profiles of these mutants prompted us to not only vary the position of the 3 mismatches (as seen in DT- 7, 8 and 9) but also make them larger (as seen in DT-10, 11 and 12). Figure 3 illustrates the open complex region in these constructs, along with the associated mismatched stretches. (See *Materials and Methods* for the full sequences).

It is important to note that the first few of the constructs, DT-3 through 9) were designed following the “anti” algorithm (i.e., A to C and G to T substitutions) as described in Zhou *et al.* (2007) in order to introduce mismatches. However, very soon, we noticed that the “anti” algorithmic changes may be unsuccessful in preventing rewinding altogether because of the possibility of staggered hybridization between the adjacent nucleotides (this phenomenon is illustrated in Figure 9 later in the section). Therefore, for all of the subsequent DT constructs, i.e. DT-10 onwards, the mismatches were

designed such that the possibility of staggered hybridization within the mismatch bubble was eliminated.

Another important fact to note is that DT-3 through 12 have sequence changes made on the template strand so as not to interfere with the RNAP-promoter interactions that occur with the non-template strand, which could change the escape kinetics altogether (Marr and Roberts, 1997). However, in order to complete our investigation with the DT series, we eventually designed mutants with sequence changes on the non-template strand. Hence, DT-14, 15 and 16 were constructed. With these constructs we were careful to limit the sequence changes to the DIS region so that the interaction between the -10 box and the RNAP will not be affected. Figure 4 illustrates the open complex region in the non-template mismatched constructs (See *Materials and Methods* for the full sequences).

Our hypothesis with the DT-series was that the mismatch bubbles will reduce the rewinding potential of the upstream end of the ITC. Thus in the mismatched mutants, we expected to see either a lower rate of escape, delayed position of escape and/or a decrease in the amount of full length transcript synthesis.

**Figure 3. Sequences of DT-3 through 12, N25 mutants with sequence changes on the template strand.** The DT-2 (wt-N25) sequence is shown for comparison. The open complex region (from -11 to +3) is bold-typed, the +1 start sites are underlined, and the mismatched nucleotides are shown in red.

-10 box ← DIS → +1

DT-2 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (wt-N25) 3'-----GACATATTATCTAAGTATTAA-----5'

*Template Strand Mismatches*

-11 +3

DT-3 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -6) 3'-----GACATATTACCTAAGTATTAA-----5'

DT-4 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -6/-5) 3'-----GACATATTACTTAAGTATTAA-----5'

DT-5 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -6 to -4) 3'-----GACATATTACTCAAGTATTAA-----5'

DT-7 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -11 to -9) 3'-----GACACGCTATCTAAGTATTAA-----5'

DT-8 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -8 to -6) 3'-----GACATATCGCCTAAGTATTAA-----5'

DT-9 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -3 to -1) 3'-----GACATATTATCTGGATATTAA-----5'

DT-10 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -11 to -7) 3'-----GACACGCCGTCTAAGTATTAA-----5'

DT-11 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -6 to -2) 3'-----GACATATTAAAGGCCGTATTAA-----5'

DT-12 5'-----CTGTATAATAGATTCATAAATT-----3'  
 (mm -11 to -2) 3'-----GACACGCCGGGGCCGTATTAA-----5'

**Figure 4. Sequences of DT-14 through 16, N25 mutants with sequence changes on the non-template strand.** The DT-2 (wt-N25) sequence is shown for comparison. The open complex region (from -11 to +3) is bold-typed, the +1 start sites are underlined, and the mismatched nucleotides are shown in red.



-10 box ← DIS → +1

DT-2            5'-----CTGTATAATAGATTCATAAATT-----3'  
 (wt-N25)        3'-----GACATATTATCTAAGTATTTAA-----5'

*Non-Template Strand Mismatches*

                          -11                                                                          +3

DT-14            5'-----CTGTATAATCTCTTCATAAATT-----3'  
 (mm -6 to -4) 3'-----GACATATTATCTAAGTATTTAA-----5'

DT-15            5'-----CTGTATAATAGACATATAAATT-----3'  
 (mm -3 to -1) 3'-----GACATATTATCTAAGTATTTAA-----5'

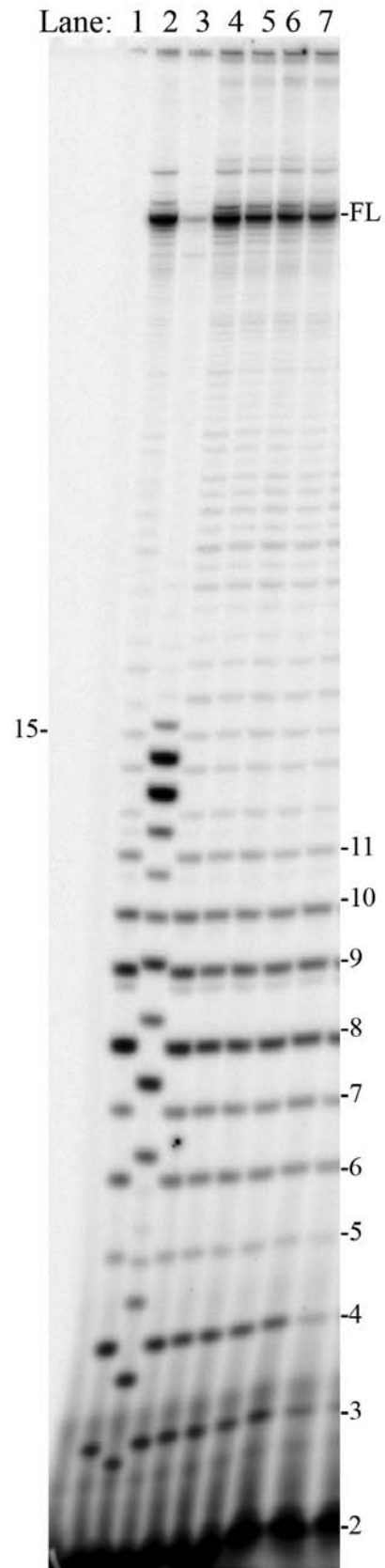
DT-16            5'-----CTGTATAATCTCCATATAAATT-----3'  
 (mm -6 to -1) 3'-----GACATATTATCTAAGTATTTAA-----5'

*EFFECT OF THE UPSTREAM REWINDING POTENTIAL ON THE ABORTIVE PROFILE OF N25 USING TEMPLATE MISMATCHES*

As previously mentioned, we began our analysis of reduced upstream rewinding potential by incorporating small regions of mismatches within the DIS region of the *N25*, keeping the ITS region unaltered. Initially, we compared the full-length synthesis and abortive profiles of DT-3 (1-bp mm at -6), 4 (2-bp mm from -6 to -5) and 5 (3-bp mm from -6 to -4) with DT-2 (wt-*N25*). As a control, we also ran *N25* and *N25<sub>anti</sub>* alongside the DT constructs. Figure 5 demonstrates that the abortive ladders of *N25* and DT-2 were identical, as expected. Additionally, Figures 5 and 6 also show that the small regions of mismatches failed to produce any alteration, neither in the abortive ladder pattern nor the amount of full-length RNA in the *N25* mutants.

This prompted us to experiment with other 3-bp mismatches at different positions within the DIS-region thereby leading to the creation of DT-7 (mm from -11 to -9), DT-8 (mm from -8 to -6), and DT-9 (mm from -3 to -1). However, after transcribing this second set, we found that the 3-bp mismatches, regardless of where they are placed in the -10/DIS region, did not affect the abortive profile significantly. This is illustrated in Figures 7 and 8.

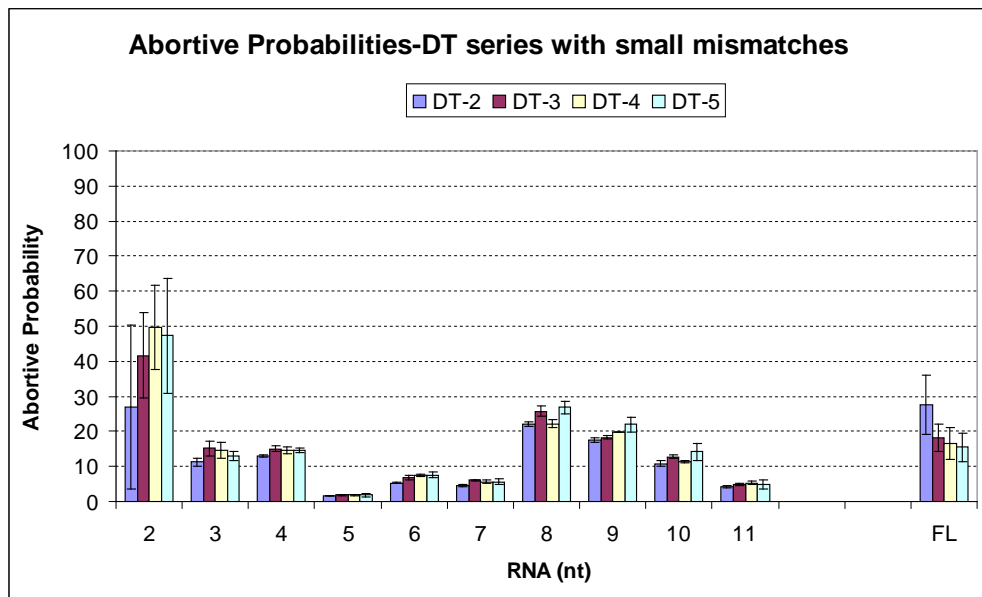
**Figure 5. Gel image of transcription reactions carried on the DT-series with 1-3 bp mismatches.** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P] ATP label and 200 mM KCl for 10 minutes. All the DTs produce a 57-nt long full-length transcript (FL). The FL and abortive ladder of DT-2 (wt-*N25*), DT-3 (1-bp mm at -6), DT-4 (2-bp mm from -6 to -5) and DT-5 (3-bp mm from -6 to -4) can be compared with *N25* and *N25<sub>anti</sub>*. DT-2 has an identical abortive profile as *N25* and escapes at the +12 position whereas on *N25<sub>anti</sub>* escapes at +16 position. No difference in the abortive profiles or FL product of DT-2 and the other mutants (DT-3, 4, and 5) can be seen from the gel image.



KEY

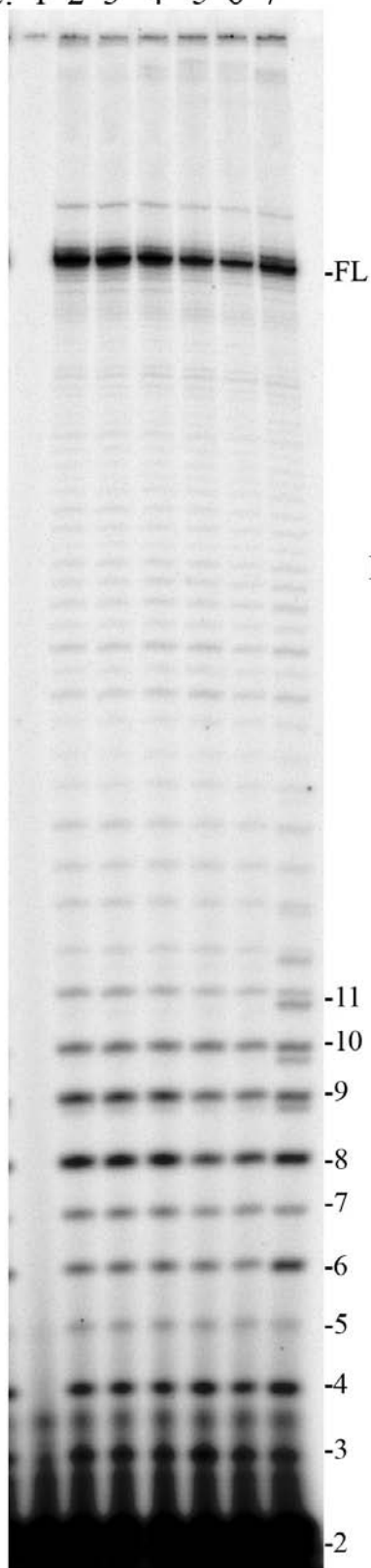
Lane	Sample ID
1	-Enzyme Control
2	N25
3	N25anti
4	DT-2 (wt-N25)
5	DT-3 (MM -6)
6	DT-4 (MM -6/-5)
7	DT-5 (MM -6/-4)

**Figure 6. Comparison of the abortive profiles of DT-2 (wt-N25) with the constructs having small mismatches-DT-3 (1-bp mm at -6), DT-4 (2-bp mm from -6 to -5) and DT-5 (3-bp mm from -6 to -4).** There appears to be no significant difference between the four constructs in terms of the abortive probability or the FL synthesis.



**Figure 7. Gel image of transcription reactions carried on the DT-series with 3 bp mismatches.** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P] ATP label and 200 mM KCl for 10 minutes. All constructs produce a 57-nt long full-length transcript (FL). The FL and abortive ladder of DT-2 (wt-*N25*) can be compared with those of DT-5 (mm -6 to -4), DT-7 (mm -11 to -9), DT-8 (mm -8 to -6) and DT-9 (mm -3 to -1). Two different aliquots of DT-2 were used. No difference in the abortive profiles between the wt-*N25* and the mutants can be observed.

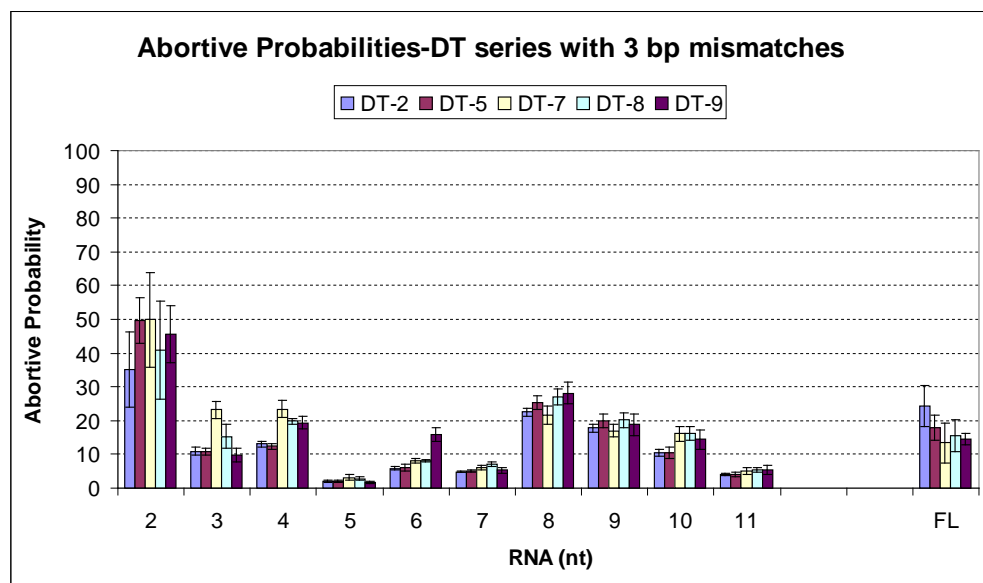
Lane: 1 2 3 4 5 6 7

KEY

Lane	Sample ID
1	-Enzyme Control
2	DT-2 (wt-N25)
3	DT-2 (wt-N25)
4	DT-5 (MM -6/-4)
5	DT-7 (MM -11/-9)
6	DT-8 (MM -8/-6)
7	DT-9 (MM -3/-1)



**Figure 8. Comparison of the abortive profiles of DT-2 (wt-N25) with the constructs having 3 bp mismatches: DT-5 (mm -6 to -4), DT-7 (mm -11 to -9), DT-8 (mm -8 to -6) and DT-9 (mm -3 to -1).** There appears to be no significant difference between the four constructs regardless of the position of the 3-bp mismatch other than a slight difference at the 3-nt and 4-nt bands.



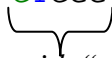
These perplexing results with the small mismatches initiated the construction of yet three more promoters (DT-10 through 12), but this time we used much larger mismatch regions spanning the downstream end of the -10 box as well as the DIS region. Also, while constructing these promoters, we modified the anti-algorithm (Zhou *et al.*, 2007) to avoid any possibility of staggered hybridization in the mismatched region. Figure 9 illustrates how staggered hybridization may occur in one of the mismatch constructs and the steps we took to avoid it.

After transcribing and quantifying this set of DT-promoters, we finally were able to detect changes in the abortive pattern as demonstrated by Figures 10 and 11. First of all, the amount of the full length product decreased along the series indicating that it was progressively harder for the RNAP to undergo promoter escape. Although DT-10 (5-bp mm from -11 to -7) did not show a significant change in the abortive pattern, DT-11 (5-bp mm from -6 to -2) showed a marked increase in the amount of the 8- and 9-nt products. Interestingly, with DT-12, we did not observe any change in the abortive ladder up to the 11<sup>th</sup> position, which is the normal range for *N25* promoter even though it contained the largest mismatch bubble spanning the -11 to the -2 positions. However, we did observe a significant amount of the 19- to 21-nt RNA that are possibly very long abortive transcripts (VLATs; Chander *et al.*, 2007). This led us to speculate if indeed the delayed escape is due to the greatly impaired upstream rewinding potential in this construct.


**Figure 9. The “anti” algorithm (A to C and G to T substitutions) changes could give rise to staggered hybridization due to Watson-Crick basepairing by skipping a base. (A)** The sequence is for DT-5 in which the mismatches are shown in bold and the +1 start site is underlined. As indicated in blue, the A at the -4 position of the non-template (NT) strand could potentially basepair with the T at the -5 position of the template (T) strand. Similarly, as shown in green, the G at the -5 position on the NT-strand could potentially basepair with the C at the -6 position on the T-strand. **(B)** DT-11 sequence using the “anti” algorithm. **(C)** The actual sequence of DT-11 used to avoid potential staggered hybridization.

(A) DT-5 NT 5'-----CTGTATAAT<sup>AG</sup>A<sup>TTC</sup>CA<sup>TAA</sup>AATT-----3'  
 (mm -6 to -4) T 3'-----GACATATTA<sup>CT</sup>C<sup>AA</sup>AGTATTTAA-----5'

(B) DT-11 NT 5'-----CTGTATAAT<sup>AG</sup>A<sup>TT</sup>CA<sup>TAA</sup>AATT-----3'  
 (mm -6 to -2) T 3'-----GACATATTA<sup>CT</sup>C<sup>GG</sup>GTATTTAA-----5'

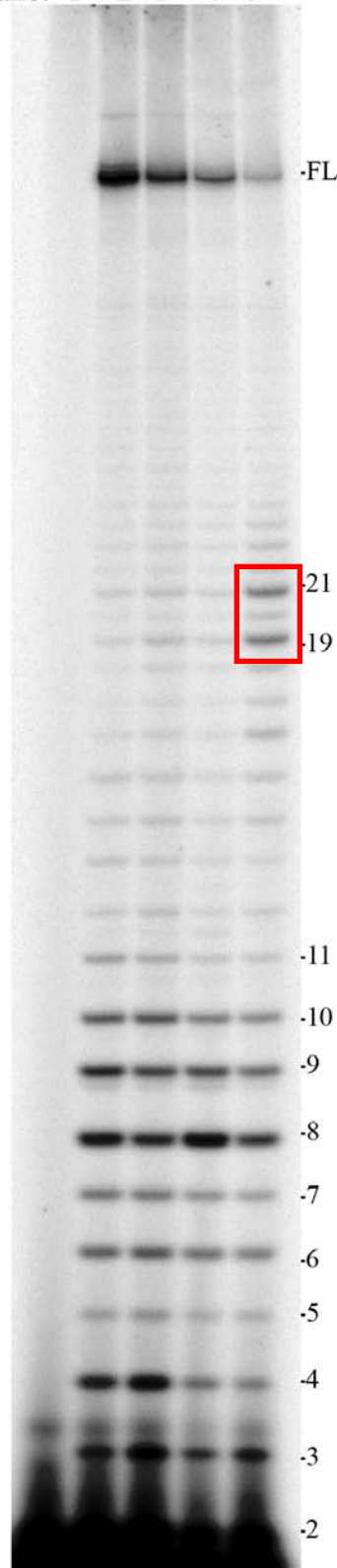
  
*Changes with “anti” algorithm*

(C) DT-11 NT 5'-----CTGTATAAT<sup>AG</sup>A<sup>TT</sup>CA<sup>TAA</sup>AATT-----3'  
 (mm -6 to -2) T 3'-----GACATATTA<sup>AG</sup>G<sup>CC</sup>GTATTTAA-----5'

  
*Changes without “anti” algorithm*

**Figure 10. Gel image of transcription reactions carried out with DT-series containing larger than 5-bp mismatches.** All reactions were performed in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP label and 200 mM KCl for 10 minutes. All the DTs produce a 57-nt long FL transcript. The lanes compare the FL and abortive ladder of DT-2 (wt-*N25*) with DT-10 (5-bp mm from -11 to -7), DT-11 (5-bp mm from -6 to -2) and DT-12 (10-bp mm from -11 to -2). The amount of the full length product appears to decrease along the series. DT-11 also shows a marked increase in the amount of the 8-nt product. Additionally, DT-12 shows a noticeable amount of the products that are 19 to 21 nts in length (enclosed in red) thereby indicating the possible release of very long abortive transcripts.

Lane: 1 2 3 4 5

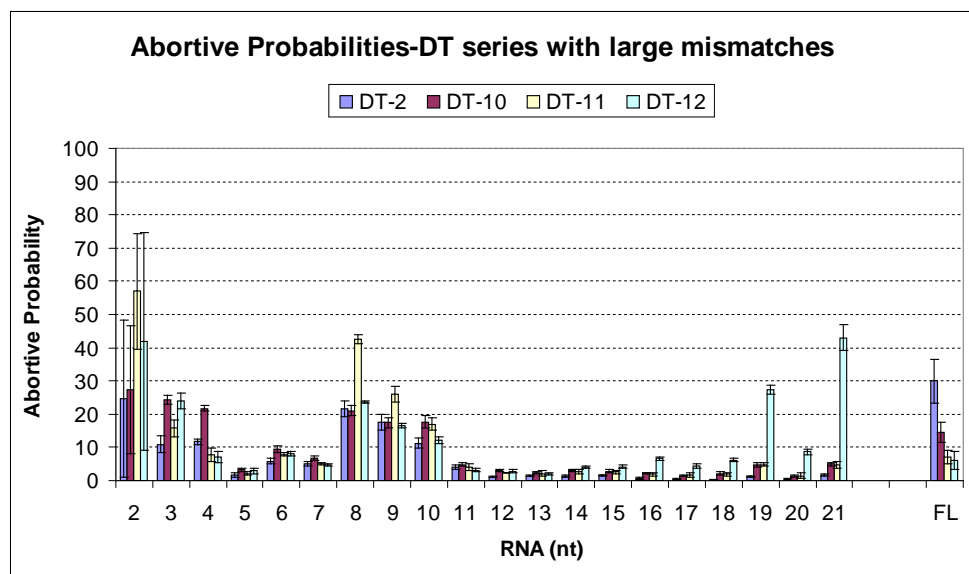
KEY

Lane	Sample ID
1	-Enzyme Control
2	DT-2 (wt-N25)
3	DT-10 (MM -11/-7)
4	DT-11 (MM -6/-2)
5	DT-12 (MM -11/-2)

**Figure 11. Abortive profile of the DT-series with the larger mismatches.**

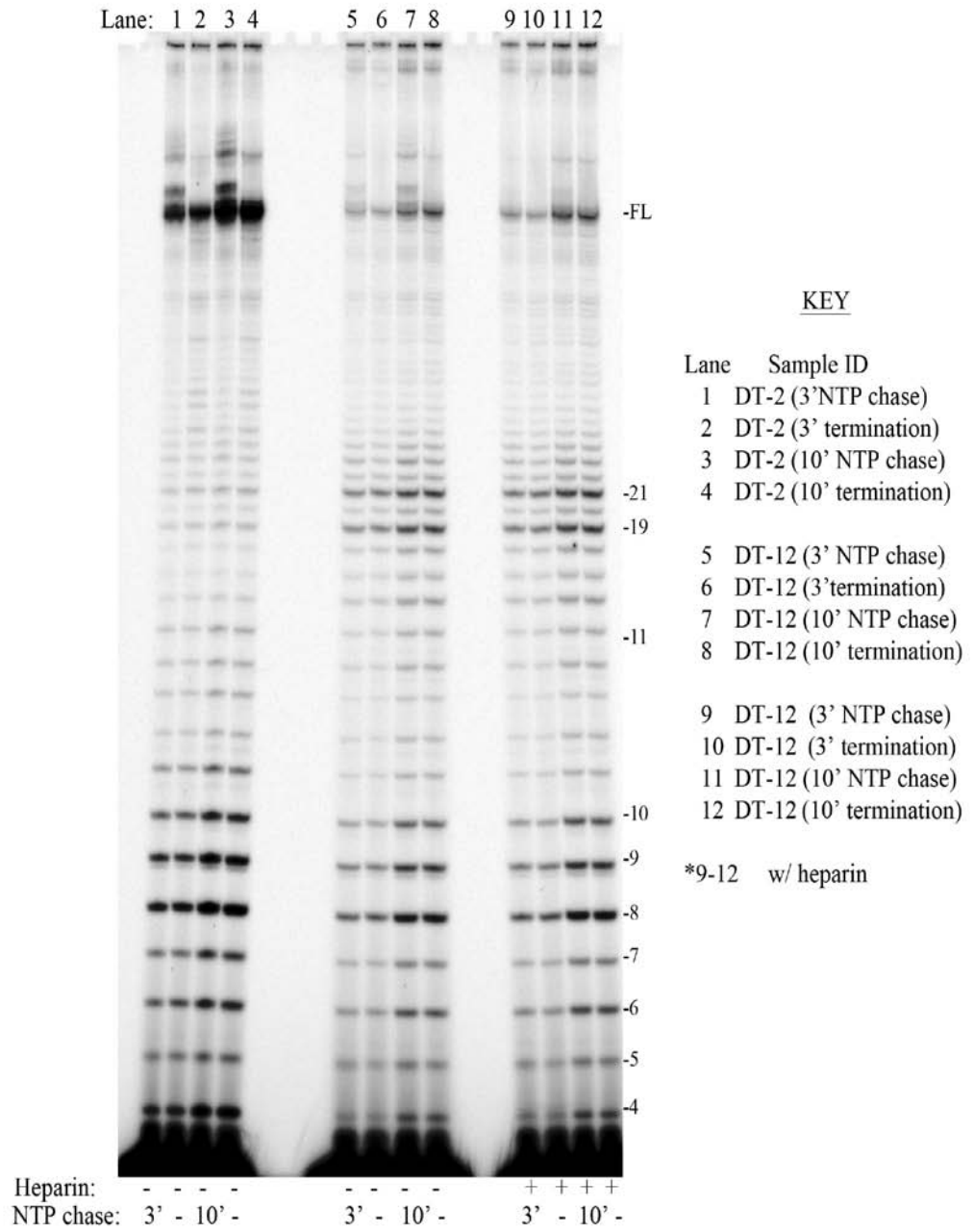
The figure compares the abortive probability along the ladder as well as the extent of full length synthesis between DT-2 (wt-*N25*) with DT-10 (5-bp mm from -11 to -7), DT-11 (5-bp mm from -6 to -2) and DT-12 (10-bp mm from -11 to -2). The amount of the full length product appears to decrease along the series. DT-11 also shows a marked increase in the amount of the 8-nt and the 9-nt products. Unlike all the other constructs, DT-12 shows the presence of longer abortive transcripts along its ladder with a particular abundance of the ones that are 19 and 21 nts in length. This is suggestive of the formation of very long abortive transcripts.





Before we could derive any conclusions about the longer abortive transcripts produced by DT-12, we had to ensure that they were indeed VLATs and not paused transcripts released due to RNAP stalling (Artsimovitch and Landick, 2000; Roberts *et al.*, 1998). Thus we subjected the promoter to an NTP chase experiment. We used the DT-2 promoter as a control and carried out the NTP chase by adding an excess of the NTP mixture to 10  $\mu$ L of the reaction mixture after 3 and 10 minutes respectively, and re-incubated the samples at 37 °C for an additional 5 minutes. We simultaneously terminated a reaction aliquot without NTP chase at the same time points for comparison. If the longer bands were due to transcriptional pausing, we would expect them to be chased to the full length by the excess NTP (Roberts *et al.*, 1998). However, as indicated by Figure 12, we saw no change in the intensity of these bands in the NTP chase reactions. Also, addition of heparin (Chamberlin, 1974), which is a polyanionic competitor that sequesters free RNAP and prevents it from being recycled, did not affect the abortive ladder or the amount of full-length transcript with DT-12. These results indicate that the 19-21 nt RNAs produced in DT-12 are not paused transcripts.

**Figure 12. Gel image confirming that the products of length 19 and 21 nts produced by DT-12 are not due to paused transcripts.** All the transcription reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP label and 200 mM KCl. NTP chase was performed in 1 mM NTP. Lanes 1-4 represent DT-2 (wt-*N25*), lanes 5-8 represent DT-12 in the absence of heparin and lanes 9-12 represent DT-12 in the presence of heparin. The first and third lanes for each group represent reactions that were chased with an excess of the NTP mixture after 3' and 10', respectively, and the second and fourth lanes represent the reactions that were terminated with GES after 3' and 10' without NTP chase, respectively. Since the adjacent lanes show no apparent change in the abortive ladder or the full length synthesis after the NTP chase even in the presence of heparin, we conclude that the 19-21 nt products observed with DT-12 are not paused transcripts but are indeed VLATs.

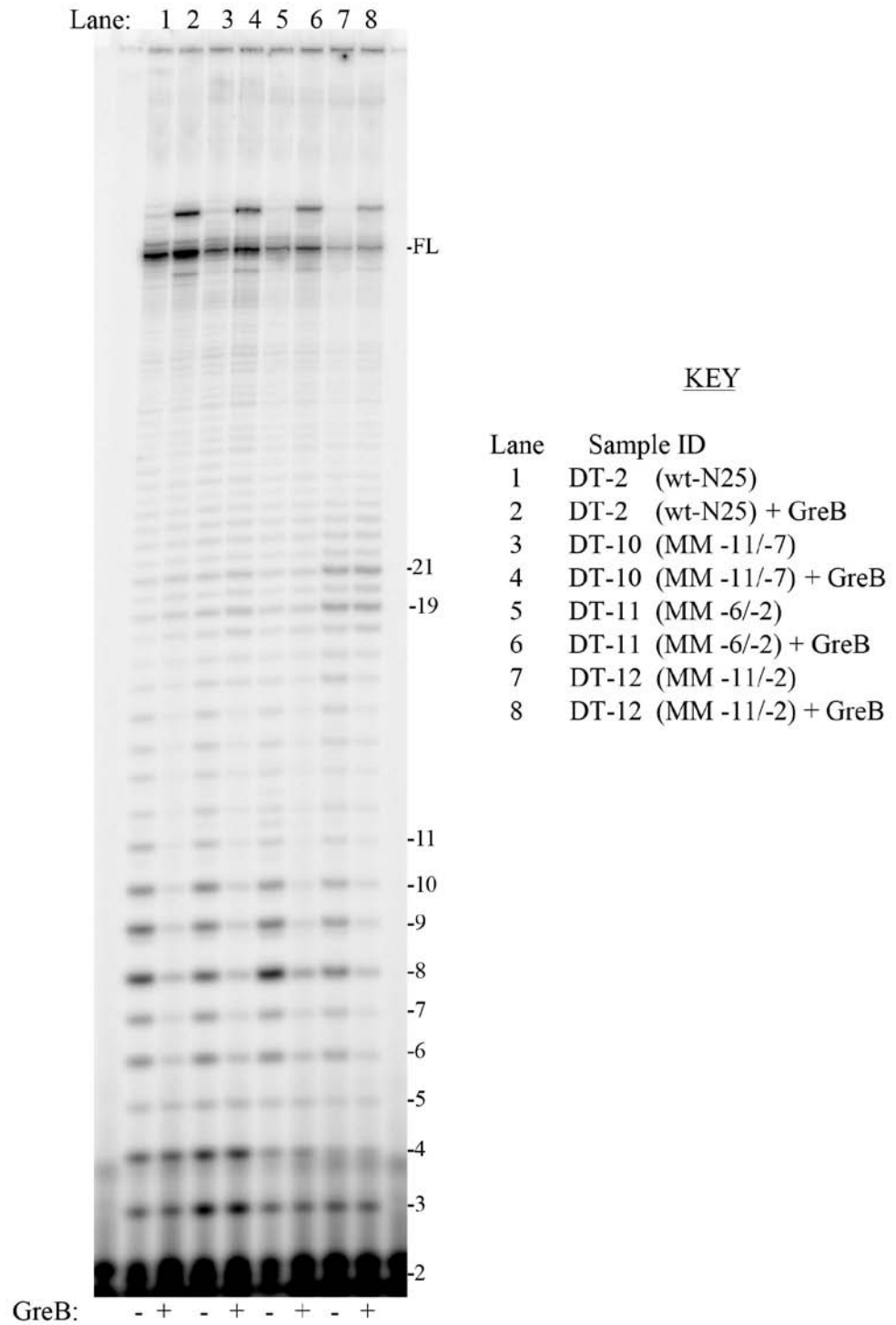


Another hallmark of VLATs is their resistance to GreB-mediated cleavage and rescue of abortive transcripts (Chander *et al.*, 2007). As previously mentioned, when RNAP backtracks during initial transcription, the RNA transcript 3'-OH end slips past the active site and becomes extruded into the secondary channel, and the transcription complex becomes arrested (Marr and Roberts, 2000). At this point, GreB bound in the secondary channel can rescue the abortive transcripts by inducing RNAP active site cleavage of the backtracked RNA (Opalka *et al.*, 2003; Roberts *et al.*, 1998; Orlova *et al.*, 1995). Thus repositioned, transcription elongation can proceed again. However, GreB-mediated rescue of abortive transcripts occurs only with RNAs 15 nts or shorter and not RNAs 16-21 nts in length (Chander *et al.*, 2007). The latter observation led to the speculation that VLATs are produced during the promoter escape maneuver when a fraction of RNAP molecules have undergone hyper forward translocation, putting the nascent RNA into the RNA exit channel, out of reach of the RNA cleavage activity (Chander *et al.*, 2007).

Based on the above information, we carried out a transcription experiment in the presence of GreB with DT-2 (control template) and DT-10 through 12. As demonstrated in Figure 13 and 14, this GreB mediated rescue of the shorter abortive transcripts can be observed in all four promoter constructs (DT-2 and 10 through 12). In the presence of GreB, the intensities of the bands due to the regular abortive ladder (up to 11 nts in length)

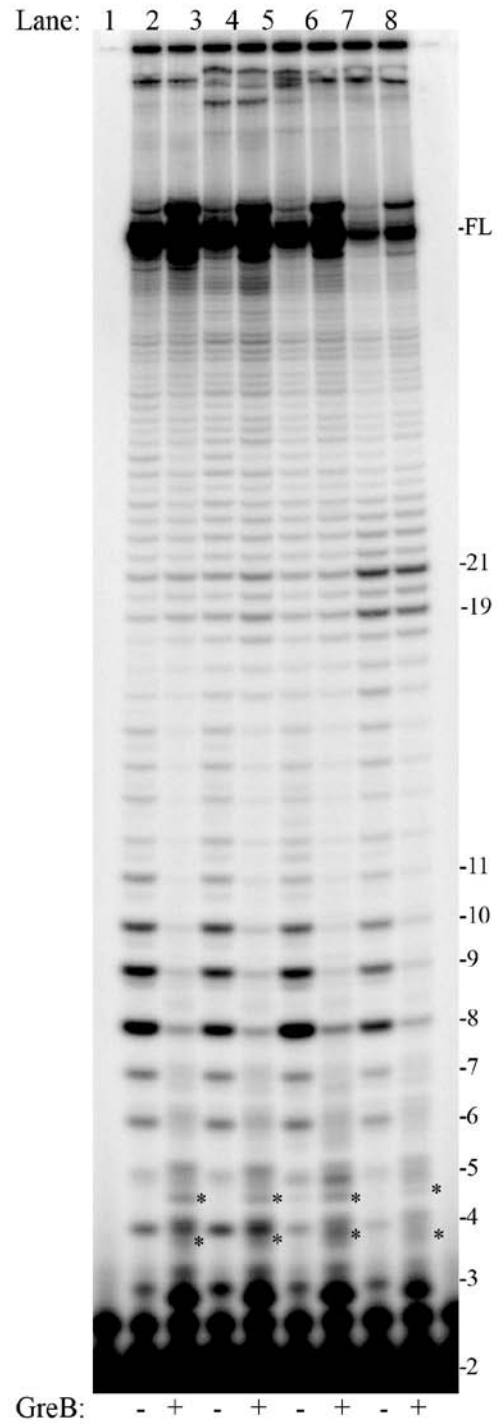
decreased drastically. However, the VLATs produced in DT-12 were unable to be rescued by GreB (compare lanes 7 and 8 on Figures 13 and 14), which is in accordance with results found in previous experiments looking at VLATs (Chander *et al.*, 2007). The same experiment was repeated with both [ $\gamma$ - $^{32}$ P]-ATP and [ $\alpha$ - $^{32}$ P]-UTP labels as illustrated in Figures 13 and 14 respectively. In fact, as shown in Figure 14, [ $\alpha$ - $^{32}$ P]-UTP labeling not only solidified the observation due to the greater intensities of the bands but also demonstrated the presence of the GreB mediated cleavage products (marked with asterisks). Thus we can be sure that indeed, in DT-12, VLATs are abortively released when the RNAP finally undergoes promoter escape at these long lengths (Chander *et al.*, 2007).

**Figure 13. Gel image with [ $\gamma$ - $^{32}$ P]-ATP label confirming that the VLATs produced by DT-12 are resistant to GreB rescue.** Transcription reactions carried out in the presence of 100  $\mu$ M NTP and 200 mM KCl. The image demonstrates that although the abundance of the shorter abortive products (5-11 nts in length) decreased in the presence of GreB for all the constructs, the VLATs in DT-12 show little or no change in their intensity. This corresponds to previous findings that VLATs cannot be rescued with GreB.





**Figure 14. Gel image with [ $\alpha$ - $^{32}$ P]-UTP label confirming that the VLATs produced by DT-12 are resistant to GreB rescue.** Transcription reactions carried out in the presence of 100  $\mu$ M NTP and 200 mM KCl. Alpha labeling highlights the presence of cleavage RNA (indicated by asterisks). The image once again demonstrates that even in the presence of GreB the VLATs in DT-12 show little or no change in their intensity. This reinforces our conclusion that DT-12 gives rise to delayed promoter escape due to compromised upstream rewinding that is reflected by the extended abortive ladder.



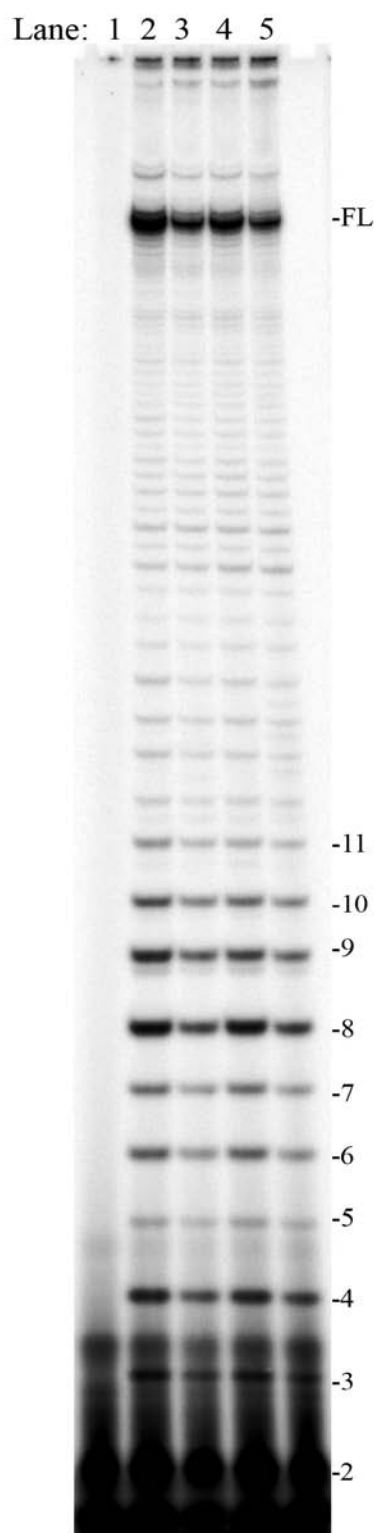
KEY

Lane	Sample ID
1	DT-2 (wt-N25)
2	DT-2 (wt-N25) + GreB
3	DT-10 (MM -11/-7)
4	DT-10 (MM -11/-7) + GreB
5	DT-11 (MM -6/-2)
6	DT-11 (MM -6/-2) + GreB
7	DT-12 (MM -11/-2)
8	DT-12 (MM -11/-2) + GreB

*EFFECT OF THE UPSTREAM REWINDING POTENTIAL ON THE ABORTIVE PROFILE OF N25 USING NON-TEMPLATE MISMATCHES*

To delve further into the matter of reduced rewinding potential, we continued making new upstream mismatch bubbles, but this time the required changes were introduced on the non-template strand, instead of the template strand. We took care to introduce the substitutions only at the DIS-region so that we would not change the escape kinetics by altering the contacts between the RNAP and the -10 box on the non-template strand (Marr and Roberts, 1997). This led to the construction of DT-14 through DT-16 where the mismatches were either 3-bp (DT-14 and DT-15) or 6-bp (DT-16) long. As can be seen in Figures 15 and 16, the mismatch substitutions on the DIS-region of the non-template strand did not elicit any significant effect, either in the abortive profile or in the full-length synthesis; this was true even when the entire DIS- region was mutated (i.e. DT-16). This was consistent with our previous results, showing that smaller mismatch bubbles are unable to affect the upstream bubble collapse significantly enough in order to alter the abortive profile of T5 N25.

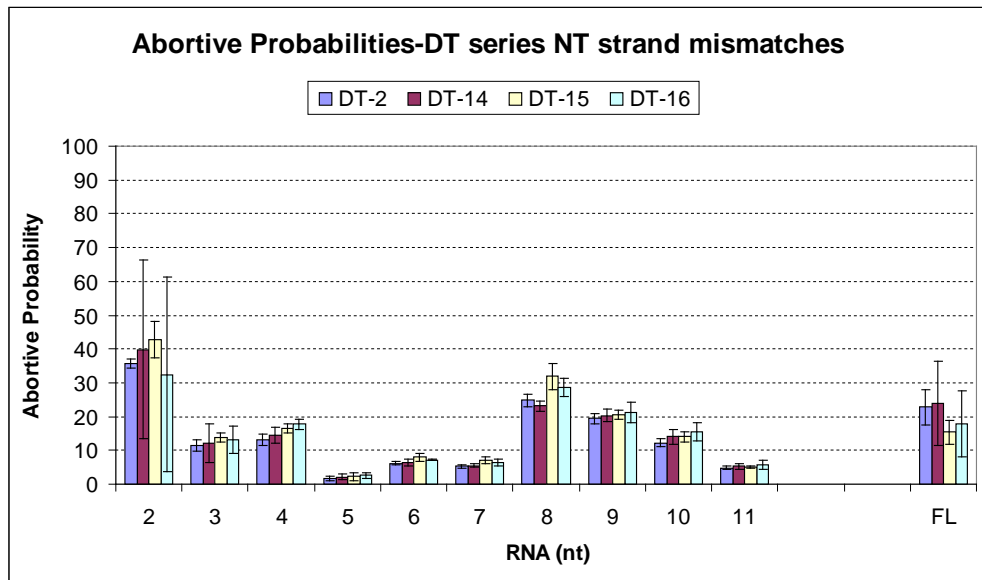
**Figure 15. Gel image of transcription reactions of the DT-series templates with mismatches on the DIS-region, non-template strand.** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP label and 200 mM KCl for 10 minutes. All the DTs produce a 57-nt long FL transcript. The lanes compare the full-length transcript and abortive ladder of DT-2 (wt-*N25*) with DT-14 (3-bp mm from -6 to -4), DT-15 (3-bp mm from -3 to -1) and DT-16 (6-bp mm from -6 to -1). The extent of the full-length synthesis appears to be consistent in all the constructs. The abortive ladders also do not show any remarkable difference between the wt-*N25* and the mismatched constructs.



KEY

Lane	Sample ID
1	-Enzyme Control
2	DT-2 (wt-N25)
3	DT-14 (MM -6/-4)
4	DT-15 (MM -3/-1)
5	DT-16 (MM -6/-1)

**Figure 16. Abortive profiles of the DT-series with mismatches on the DIS-region, non-template strand.** The figure compares the abortive probability along the ladder as well as the extent of full-length synthesis between DT-2 (wt-*N25*) with DT-14 (3-bp mm from -6 to -4), DT-15 (3-bp mm from -3 to -1) and DT-16 (6-bp mm from -6 to -1). The amounts of full-length synthesis as well as the abortive ladders do not show any remarkable differences between the wt-*N25* and the mismatched constructs.



*SUMMARY OF FINDINGS FROM THE DT-SERIES*

Our results thus far with the DT-series containing template strand mismatches in the -10/DIS region indicated that only the mismatched bubbles that were 5 bp or larger were able to produce a noticeable change in the abortive profile of *N25*. This result was further supported by the non-template strand DIS mismatches.

The first remarkable change in the abortive profile was observed with DT-11 (mm -6 to -2) where the abortive release of 8- and 9-nt fragments were significantly higher than that in wt-*N25*. The 8-nt abortive transcript is the most abundant one in wt-*N25*, and has been attributed to the inability of the initial transcribing complex at +8 to overcome the interaction between the  $\sigma_2$  domain of polymerase and the -10 box (Chan and Gross, 2001; Vo *et al.*, 2003). That the amount of the 8-nt abortive RNA is approximately doubled in the case of DT-11 suggests that the template stretch just upstream to the +1 site, when unable to re-anneal with the corresponding non-template stretch, has introduced an added barrier to promoter escape. DT-10 (mm -11 to -7), on the other hand, only shows a slight increase in abortive probability at the 3-, 4- and 10-nt positions relative to the wt-*N25* indicating the presence of some barrier to promoter escape in this case as well. In agreement, both DT-10 and DT-11 showed a significant decrease in full-length RNA synthesis to about half the amount of wt-*N25*.



The most dramatic change in the abortive profile was observed with DT-12, the construct with the largest region of mismatch (from -11 to -2). Interestingly, in this construct the distribution of the abortive probability up to the 11-nt RNA was quite comparable with that of wt-*N25*, but the presence of a considerable amount of VLATs (19- to 21-nt RNA) was indicative of compromised upstream rewinding to achieve escape. We speculate that, since the region of mismatch in this construct is so large, the RNAP continued to scrunch DNA and transcribe as far as the 21<sup>st</sup> nt before enough stress energy has been accrued to bring about promoter escape. However, when escape is finally achieved, the release of the large amount of stress energy caused a fraction of the enzyme to become arrested, this time in the forward hyper-translocated state where the RNA 3'-end was pushed far upstream into the RNA exit channel. These RNAs, now held by a few RNA-DNA hybrid basepairs and not susceptible to GreB-mediated cleavage and re-elongation, are soon released as very long abortive transcripts. This defective mode of escape was further reflected in a six-fold decrease in the amount of the full-length transcript in comparison with wt-*N25*.

Taken together, our results from the DT-series promoters allowed us to conclude that our initial hypothesis was too simplistic to describe the dependence of promoter escape on the upstream rewinding potential. Whereas we incorporated mismatches to decrease the upstream rewinding potential, all except the most extreme changes showed little or no effect on promoter escape, whether measured by the extent of full-length RNA

synthesis or, in the case of DT-12, by the delayed position of escape. Thus, our results indicate that the sequence of the upstream region, and how readily it rewinds, is not as consequential to bring about promoter escape as the initial transcribed sequence (ITS). Our results further indicate that it is the extent of DNA scrunching and not upstream rewinding which is a prerequisite to disrupting the  $\sigma_{2/-10}$  DNA contacts.

## THE GC-SERIES

### *RATIONALE*

The second set of promoters investigated in this work are the GC-series *N25* promoters (hereafter abbreviated as the GC promoters). These five promoters (GC-6, -7, -8, -8-1 and -8-2) all contain a very GC-rich DIS-region that would potentially make the upstream rewinding more favorable. The DIS region in these promoters varies in length from 6 to 8 bp. Figure 17 illustrates the open complex region in GC-promoters compared with wt-*N25* and the *rrnB* P1 promoter (See *Materials and Methods* for the full sequences of the GC-promoters).

The *rrnB* P1 promoter, which is normally associated with transcription of ribosomal RNA in *E. coli*, is similar in structure to the GC-promoters (especially GC-8 and its variants) in that it also has an 8 bp long, GC-rich DIS region (Mulligan *et al.*, 1984). These two features couple together to give the *rrnB* P1 promoter an open complex with a very short half life (~ few seconds or minutes) (Mulligan *et al.*, 1984; Barker and Gourse, 2001). Paradoxically, this relatively unstable open complex also gives the promoter its high efficiency of escape often approaching the maximum theoretical rate (Gourse, 1988). In addition to the sequence at the DIS region, the high rate of escape of *rrnB* P1 has also been attributed to the 16 bp long spacer region between the -35 and -10 boxes in this promoter (Barker and Gourse, 2001).

**Figure 17. Sequences of GC-promoters, *N25* mutants with GC-rich DIS regions compared with wt-*N25* and the *rrnB* P1 promoter.** The open complex region (from -11 to +3) for all the promoters is bold-typed and the +1 start sites are underlined. (A) Sequence of wt-*N25* showing the 6 bp DIS region in blue. (B) GC-promoters (GC-6, -7, -8, -8-1, and -8-2) showing the DIS region in green. The DIS-regions in these mutants are GC-rich and vary in length from 6 to 8 bp. (C) Sequences of *rrnB* P1 promoter showing the 8 bp DIS region in pink.

(A)

-10 box ← DIS → +1

wt-N25 5'-----CTGTATAATAGATTCATAAATT-----3'  
 3'-----GACA TATTATCTAAGTATTTAA-----5'

(B)

*Promoters with GC-rich DIS regions*

-10 box ← DIS → +1

GC-6 5'-----CTGTATAATGCGCTCATAAATT-----3'  
 3'-----GACA TATTACGCGAGTATTTAA-----5'

GC-7 5'-----CTGTATAATGCGCTCATAAATT-----3'  
 3'-----GACA TATTACGCGCAGTATTTAA-----5'

GC-8 5'-----CTGTATAATGCGGTCCATAAATT-----3'  
 3'-----GACA TATTACGCGCAGGTATTTAA-----5'

GC-8-1 5'-----CTGTATAATGCGGTCTATAAATT-----3'  
 3'-----GACA TATTACGCGCAGATATTTAA-----5'

GC-8-2 5'-----CTGTATAATGCGCTCTATAAATT-----3'  
 3'-----GACA TATTACGCGGAGATATTTAA-----5'

(C)

*rrnB P1 promoter sequence*

-10 box ← DIS → +1

*rrnB* P1 5'-----CCCTATAATGCGCACCACTGA-----3'  
 3'-----GGGA TATTACGCGGTGGTATTT-----5'

The GC-promoters were previously investigated by Katia Evans (1999) for her senior thesis. However, even after repeated attempts with various reaction conditions, she was unable to obtain good transcription initiation rates with the constructs. We thus tested several parameters in an attempt to optimize the transcription initiation rates. For that, we tested different NTP and KCl concentrations and determined the combination that worked best for all the constructs.

Evans (1999) had also observed that the GC-8 promoter has an alternative transcription start site at the -1 position which places a cytidine on the 5' terminus of the transcript instead of an adenine. Therefore the products arising from the alternative start site cannot be visualized with a [ $\gamma$ - $^{32}$ P]-ATP label. Hence, we used the [ $\alpha$ - $^{32}$ P]-UTP label to tag all the uridine residues along the transcript and compared the results from both types of assays.

The relatively higher upstream rewinding potential of the ITC in these constructs had the likelihood of engendering two opposing effects on open-complex formation: it could either limit the rate of formation of the open complex, or it could reduce the stability and hence the half-life of the open complex. Nevertheless, under ideal transcription conditions, we expected the GC-promoters to demonstrate escape properties similar to that of the *rrnB* P1 promoter, i.e. undergo an earlier escape accompanied by a short abortive ladder.

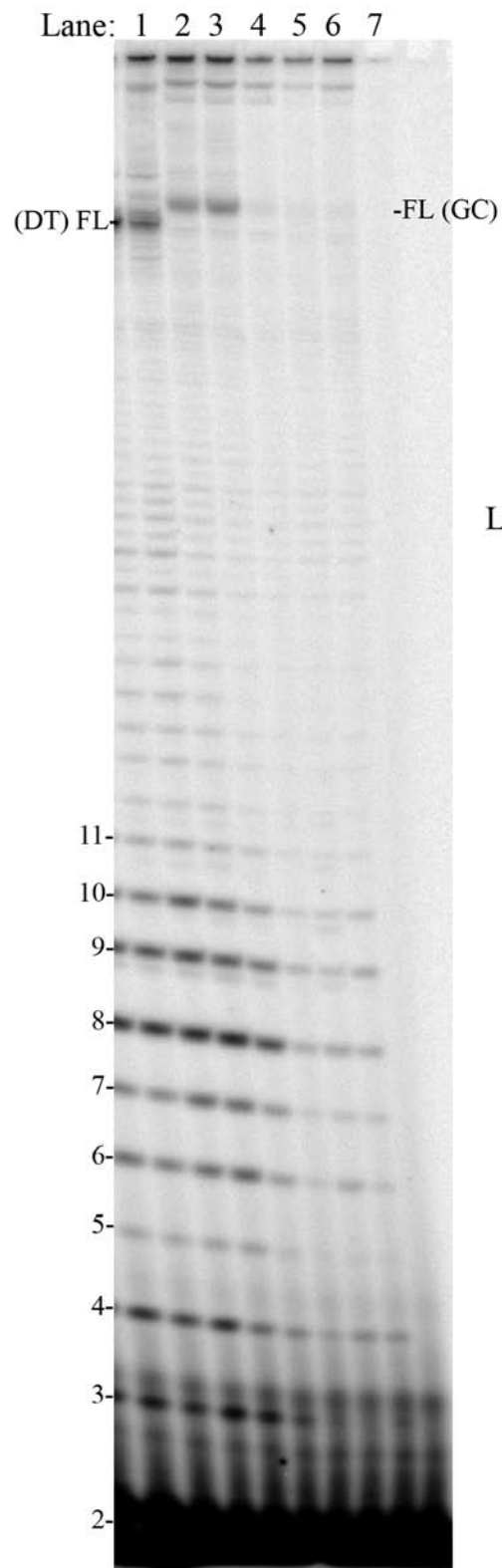
*TRANSCRIPTION OF THE GC-PROMOTERS WITH [ $\gamma$ -<sup>32</sup>P]-ATP LABEL*

It had been reported that the *rrnB* P1 promoter is sensitive to the salt concentration of the reaction mixture *in vitro*. The promoter was found to form the most stable open complexes at low salt concentrations of around 30 mM KCl (Gourse, 1988). Since the GC-promoters have similar DIS-regions as the *rrnB* P1, they were expected to show similar sensitivity to salt concentrations. In fact, Evans (1999) had shown that the GCs transcribed most efficiently at 20 mM KCl in contrast with the wt-*N25* which performs best at relatively high KCl concentrations ~200 mM (Sarah Barnes, senior thesis, 2008).

Following Evans's work, we carried out our initial experiments with the GC-promoters at 20 mM KCl. As can be seen in Figure 18, the bands for the abortive products as well as the full-length transcript for the GC promoters, especially GC-8 and its variants, at 20 mM KCl and 100  $\mu$ M NTP, were very faint making them difficult to quantitate and giving rise to large standard deviations. DT-2 (wt-*N25*) was transcribed under the same conditions for comparison and also seems to transcribe at relatively lower rates than those observed at higher salt concentrations as expected. As shown in Figure 19, despite the low signals, both GC-6 and -7 demonstrated a similar abortive profile to that of DT-2. On the other hand, we were unable to derive any particular conclusion about the GC-8s under these set of experimental conditions except that they produced little full-length products thereby indicating the possibility of rate limitation at the open complex formation step.

**Figure 18. Gel image comparing the GC promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt-N25).** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP label and 20 mM KCl for 10 minutes. DT-2 produces a 57-nt long FL transcript. All the GCs produce a 62-nt long FL transcript. Under the reaction conditions mentioned, the bands, especially for GC-8s were very faint indicating a possibility of low transcription initiation for these constructs.

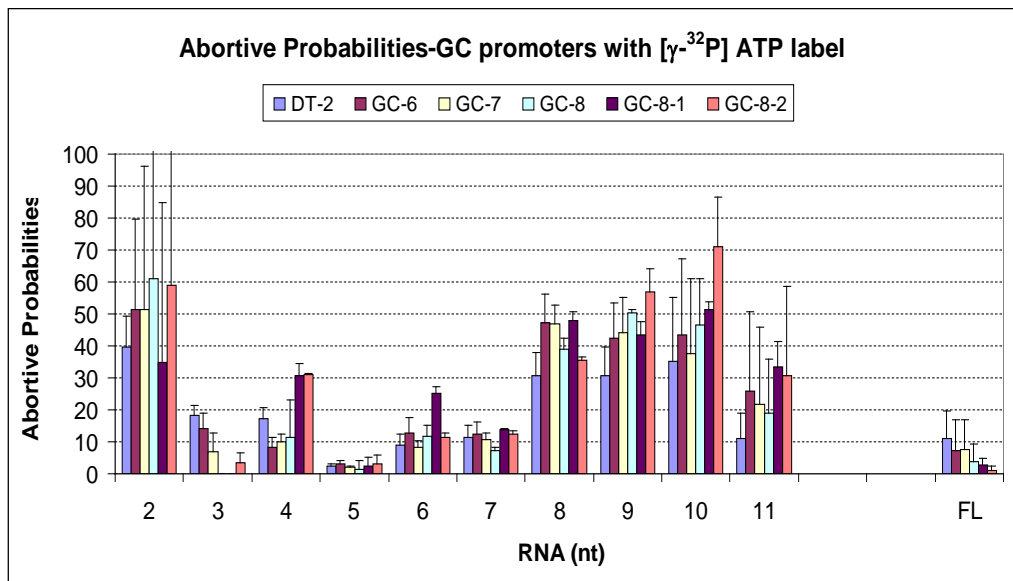




KEY

Lane	Sample ID
1	DT-2 (wt-N25)
2	GC-6
3	GC-7
4	GC-8
5	GC-8-1
6	GC-8-2
7	-Enzyme Control

**Figure 19. Comparison of the abortive profiles of GC promoters with DT-2 (wt-N25).** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP label and 20 mM KCl for 10 minutes. Overall, both GC-6 and 7 behave very similarly to DT-2. As seen from the faint banding patterns in Figure 18, the rates of transcription initiation for GC-8, GC-8-1 and GC-8-2 are very low thereby giving rise to large error bars. However, there seems to be no detectable alteration in the pattern of abortive release or promoter escape in the GC-8s under these reaction conditions.

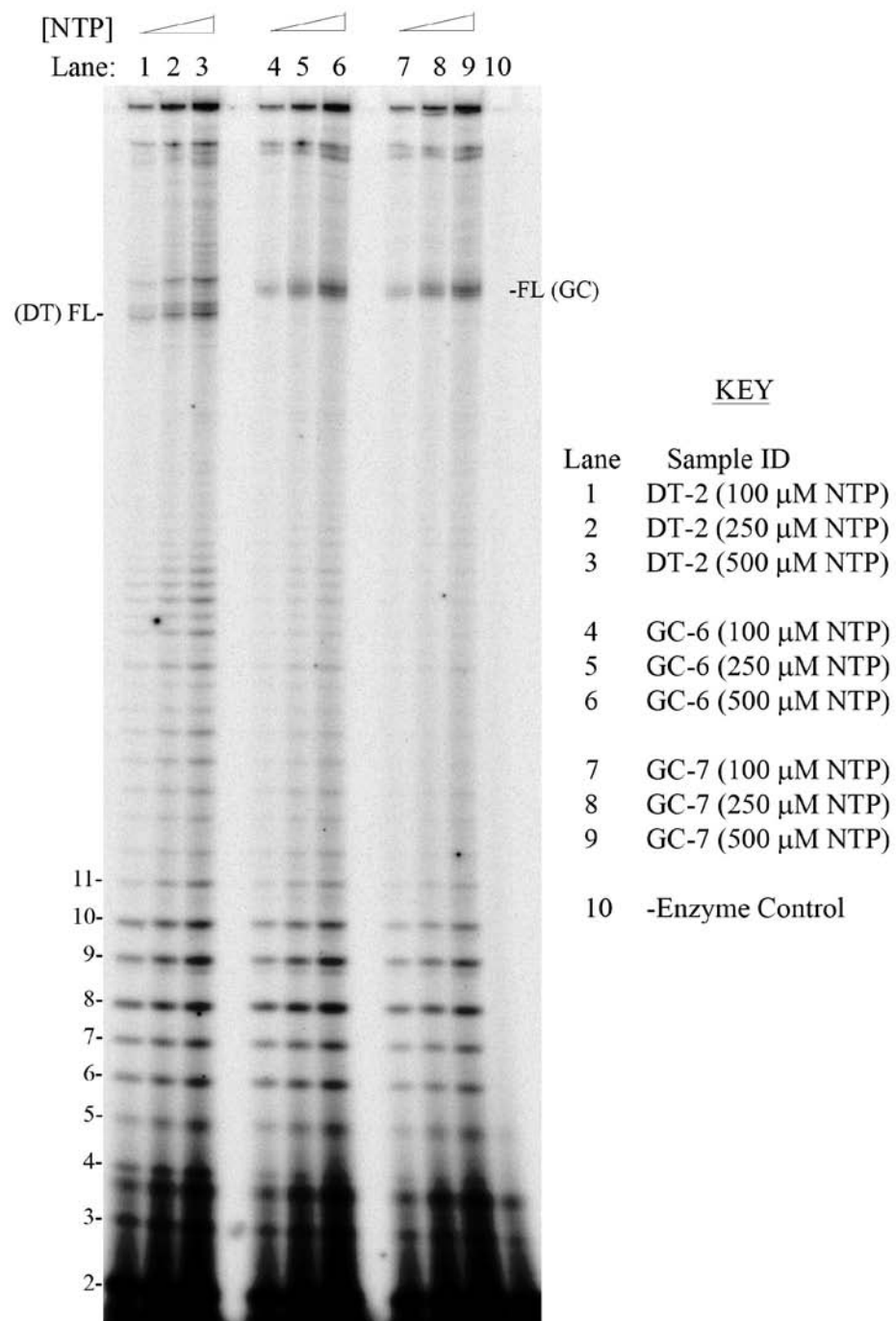


It had also been reported that the *rrnB* P1 promoters require a comparatively higher concentration of the initiating NTP than other promoters in order to ensue maximal transcription (Gaal *et al.*, 1997; Barker and Gourse, 2001). Thus, in an attempt to increase the intensity of the bands, we increased the NTP concentration of the reaction mixtures. At this point, we only focused our assays on GC-6 and -7 and compared them to DT-2 (wt-N25). We transcribed these promoters under three NTP concentrations: 100  $\mu$ M, 250  $\mu$ M, and 500  $\mu$ M and [ $\gamma$ -<sup>32</sup>P]-ATP label.

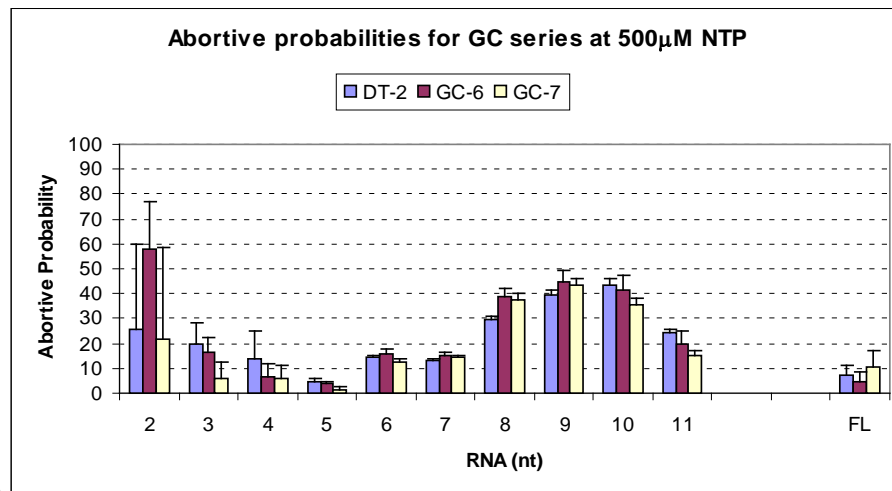
As demonstrated by the gel image and the bar graphs in Figures 20 and 21 respectively, the higher NTP concentrations indeed increased the intensities of the products although the bands were still not dark enough. The abortive profiles of the three constructs still did not show any significant difference regardless of the NTP concentration of the reaction.

Next, we compared the extent of initiation with the three promoters at the different NTP concentrations by measuring the total yield. We observed that at 250  $\mu$ M, there was about 6 fold increase in the overall initiation with GC-6 and -7 and ~2 fold increase in DT-2 in comparison with that at 100  $\mu$ M NTP. However, the use of 500  $\mu$ M NTP did not result in a proportional increase in the total initiation in the case of GC-7 although it had almost doubled with DT-2 and GC-6 (see Figure 22).

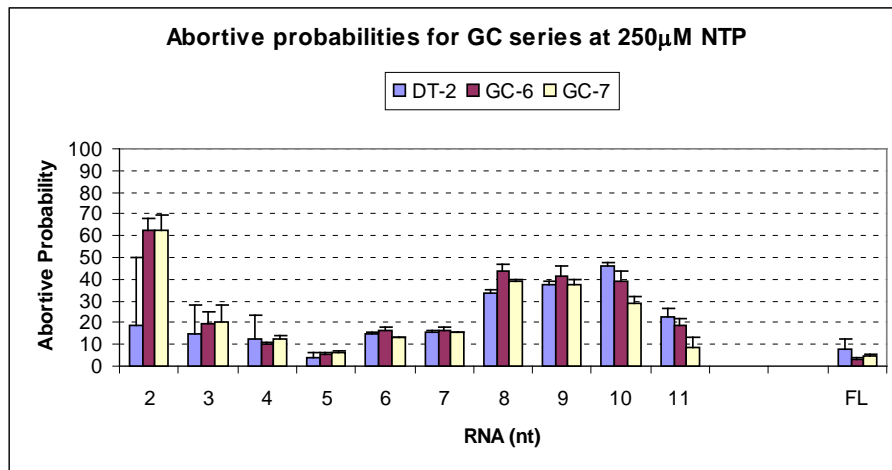
**Figure 20. Gel image comparing transcription of GC-6, -7, and DT-2 (wt-N25) under variable NTP concentrations (100, 250 and 500  $\mu$ M NTP) and [ $\gamma$ - $^{32}$ P]-ATP label.** All reactions were carried out in the presence of 20 mM KCl for 10 minutes. The intensities of the bands are higher for all the constructs at higher [NTP] indicating an increase in the occurrence of transcription initiation.



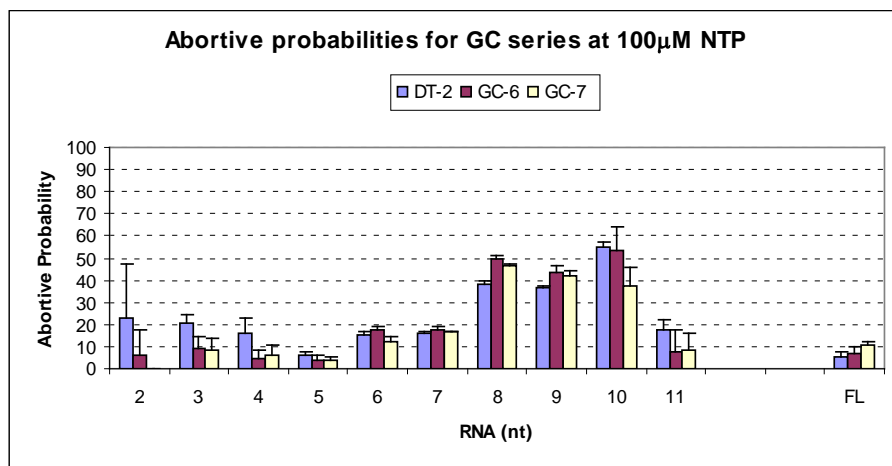
**Figure 21. Comparison of abortive profiles of GC-6 and -7 transcription with DT-2 (wt-N25) under variable NTP concentrations (100, 250 and 500  $\mu\text{M}$  NTP) and  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  label.** All reactions were carried out in the presence of 20 mM KCl for 10 minutes. (A) Abortive profile at 100  $\mu\text{M}$  NTP. (B) Abortive profile at 250  $\mu\text{M}$  NTP. (C) Abortive profile at 500  $\mu\text{M}$  NTP. There seems to be no significant difference between the abortive profiles of GC-6 and -7 in comparison with DT-2 at all NTP concentrations.



(A)



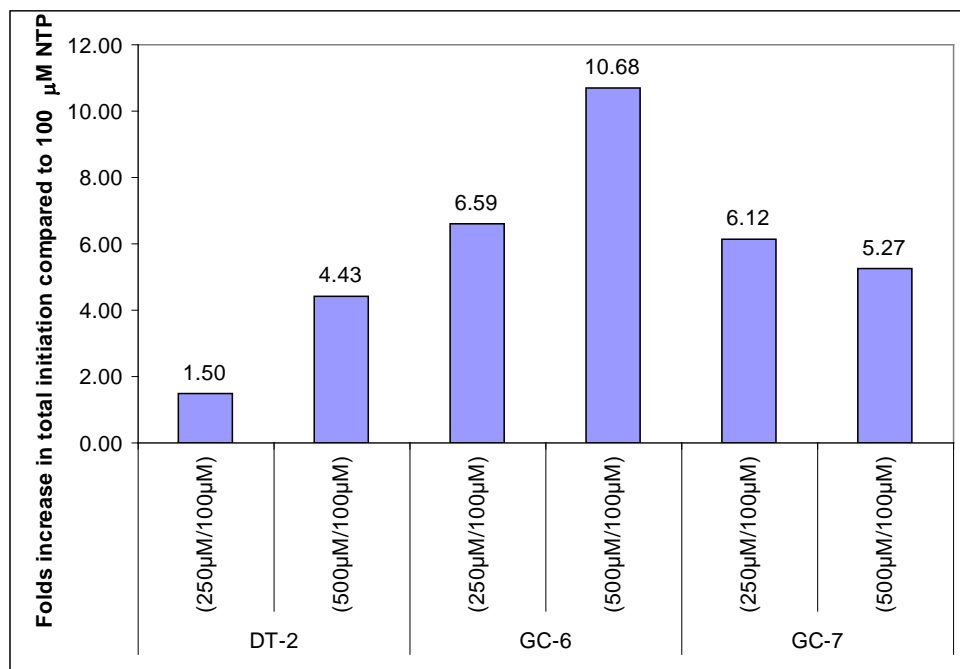
(B)



(C)

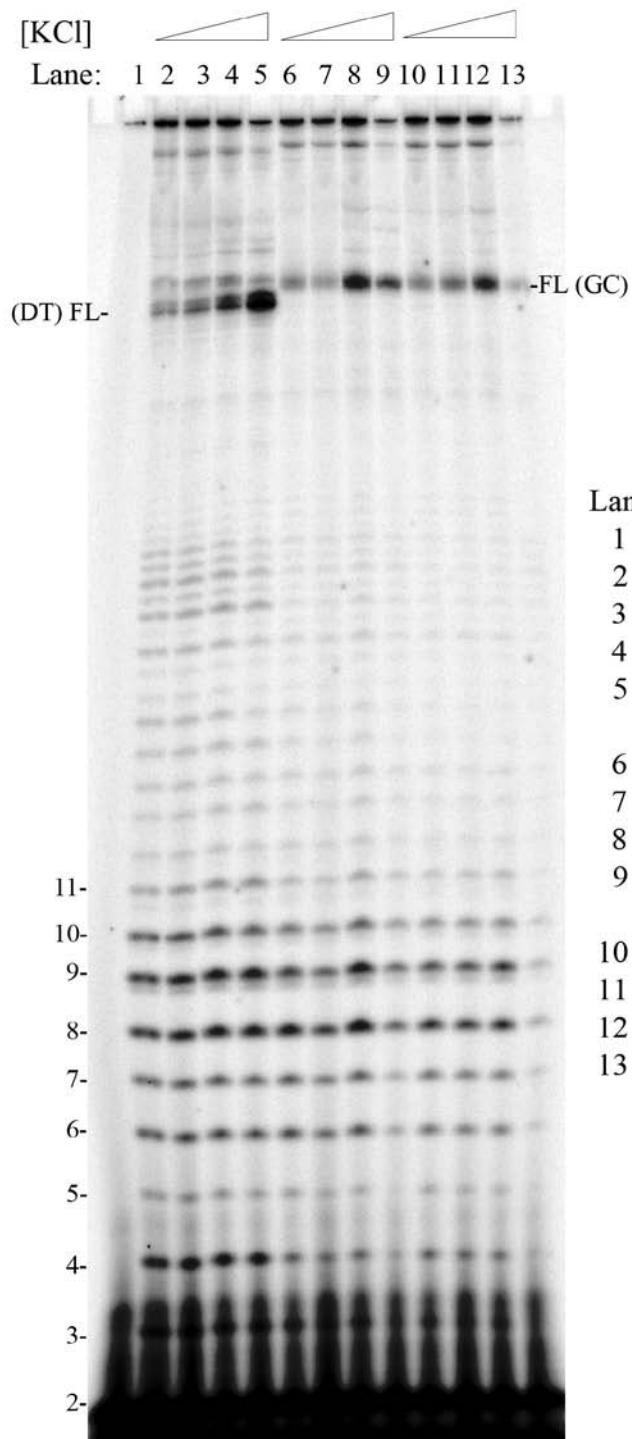


**Figure 22. Ratio of total transcription initiation events at higher [NTP] relative to that with 100  $\mu$ M NTP.** All three constructs show an increase in the amount of initiation at 250  $\mu$ M NTP, but only DT-2 and GC-6 show a similar increase when the NTP concentration is 500  $\mu$ M NTP.

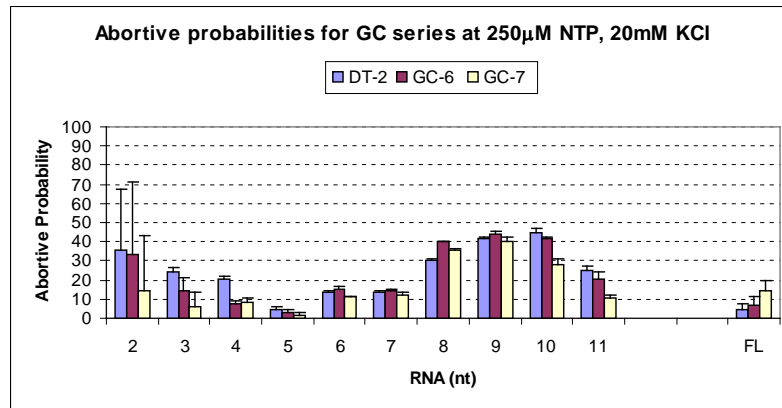


The higher incidence of transcription with a greater [NTP] prompted us to experiment with the KCl concentrations as well. Hereafter we used 250  $\mu$ M NTP and [ $\gamma$ - $^{32}$ P]-ATP for our experiments and transcribed DT-2, GC-6 and GC-7 at various KCl concentrations of 20 mM, 50 mM, 100 mM and 200 mM. As shown by the gel image in Figure 23, DT-2 synthesized the maximum amount of full-length product at 200 mM KCl, but both GC-6 and GC-7 transcribed maximally at 100 mM KCl. Also, as shown in Figure 24, as the salt concentration of the reaction was increased, the amount of abortive products decreased for all three promoters with a corresponding rise in the amount of full-length transcript. The patterns of the abortive profiles still remained the same for all three promoters at all salt concentrations. The latter observation is more profound in Figure 25. Although this was a surprising finding in the context of what is known for the *rrnB* P1 promoter (Gourse, 1988), the results made some sense because KCl has been shown to play a role in stabilizing the open complex and stimulating RNAP recycling for multiple rounds of steady state transcription (Sarah Barnes, senior thesis, 2008) especially in the presence of a high NTP concentration.

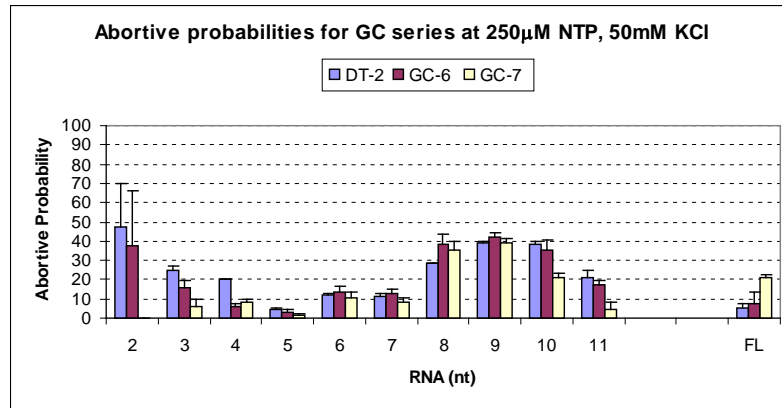
**Figure 23. Gel image comparing GC-6 and -7 transcriptions with DT-2 (wt- N25) under varying KCl concentrations (20, 50, 100 and 200 mM).** All reactions were carried out in the presence of 250  $\mu$ M NTP and [ $\gamma$ -<sup>32</sup>P]-ATP label for 10 minutes. There is an increase in the relative amount of transcription initiation with higher salt concentrations for all the constructs. GC-6 and -7 show the maximal transcription at 100 mM KCl whereas DT-2 shows maximal transcription at 200 mM KCl. Increasing salt concentrations also seem to favor an increased production of the FL transcript and a corresponding decrease in intensity of the abortive ladder probably due to enhanced RNAP recycling (Sarah Barnes, senior thesis, 2008).



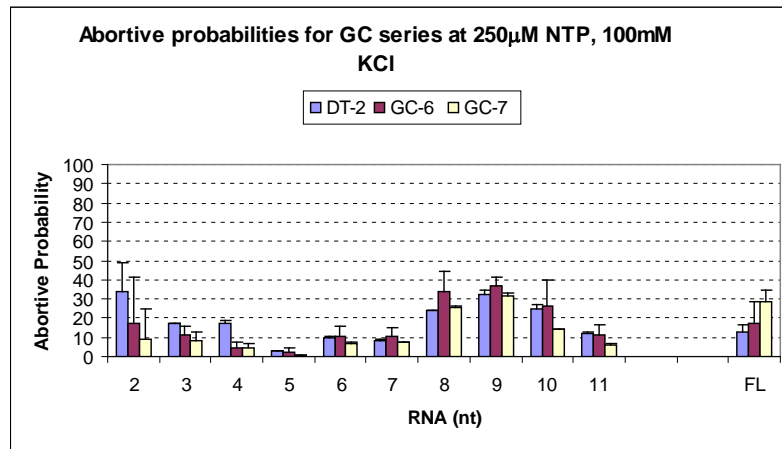
**Figure 24. Comparison of the abortive profiles of GC-6 and -7 with DT-2 (wt-N25) under varying KCl concentrations (20, 50, 100 and 200 mM).** All reactions were carried out in the presence of 250  $\mu$ M NTP and [ $\gamma$ -<sup>32</sup>P]-ATP label for 10 minutes. (A) Abortive profile at 20 mM KCl. (B) Abortive profile at 50 mM KCl. (C) Abortive profile at 100 mM KCl (D) Abortive profile at 200 mM KCl. There seems to be a general decrease in the abortive probability along the ladder with a corresponding increase in the amounts of FL for all the constructs as [KCl] increases.



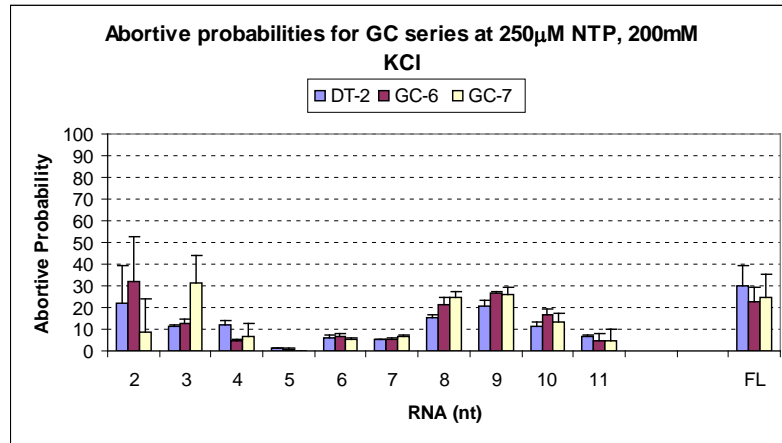
(A)



(B)



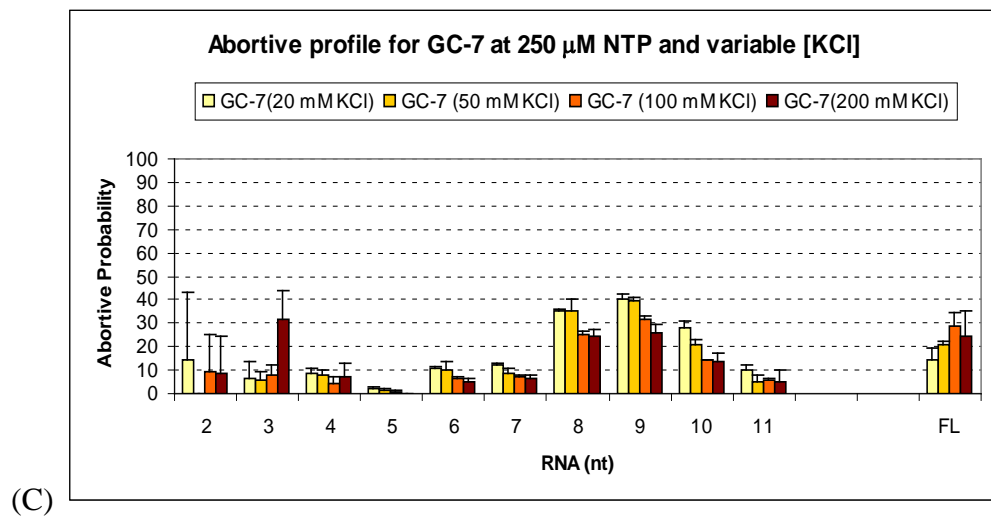
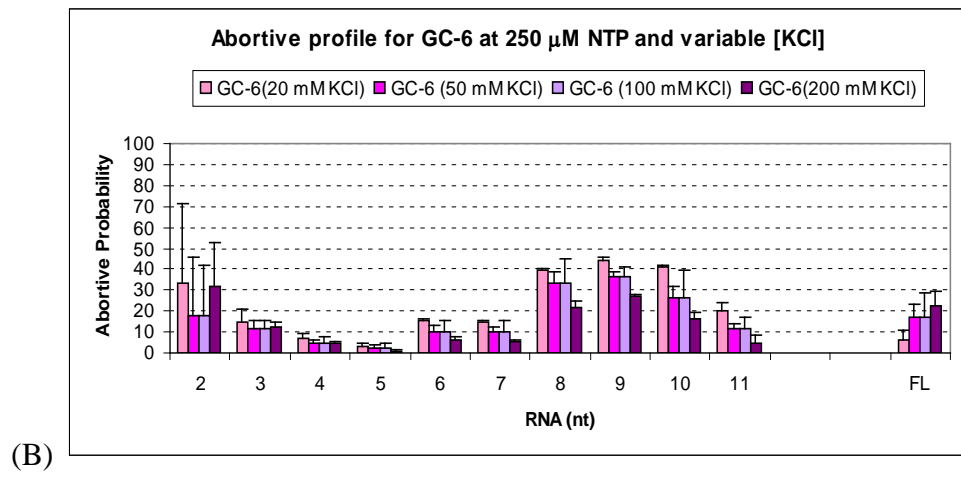
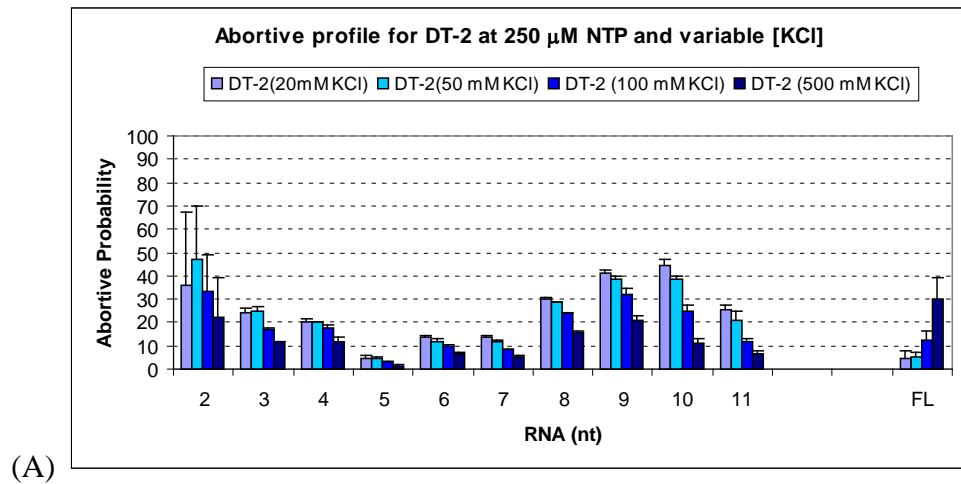
(C)



(D)

**Figure 25. Comparison of the abortive profiles of DT-2, GC-6, and -7 at different KCl concentrations (20, 50, 100 and 200 mM).** All reactions were carried out in the presence of 250  $\mu$ M NTP and  $[\gamma\text{-}^{32}\text{P}]\text{-ATP}$  label for 10 minutes. (A) Gel image comparing the different constructs with increasing [KCl]. (B) Abortive profile of DT-2. (C) Abortive profile of GC-6. (D) Abortive profile of GC-7. Increasing salt concentrations seem to favor escape for all the constructs as seen from an increase in the amounts of FL and a corresponding decrease in intensity of the abortive ladder. DT-2 shows maximum FL synthesis at 200 mM KCl whereas those for the GCs appear to max out at 100 mM KCl.





*TRANSCRIPTION OF THE GC-PROMOTERS WITH [ $\alpha$ - $^{32}$ P]-UTP LABEL*

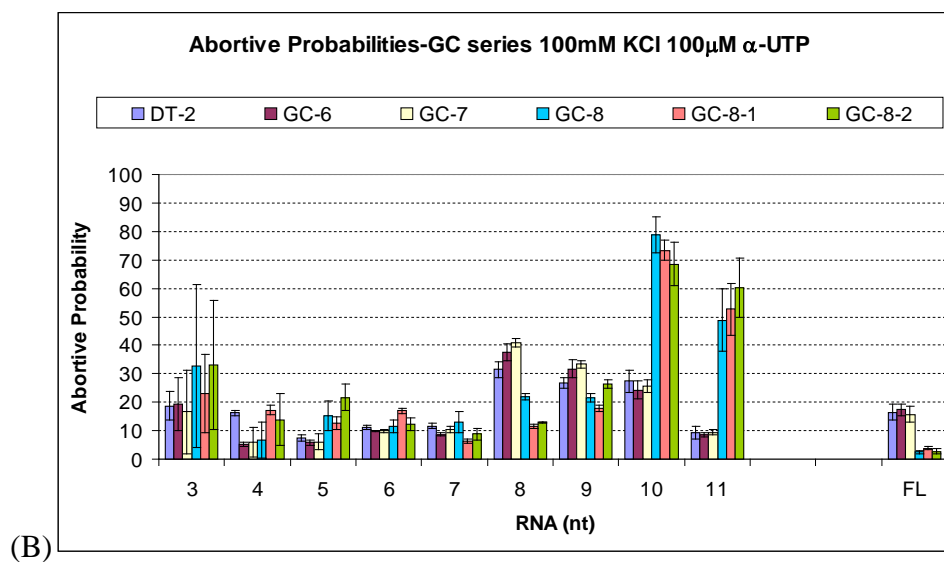
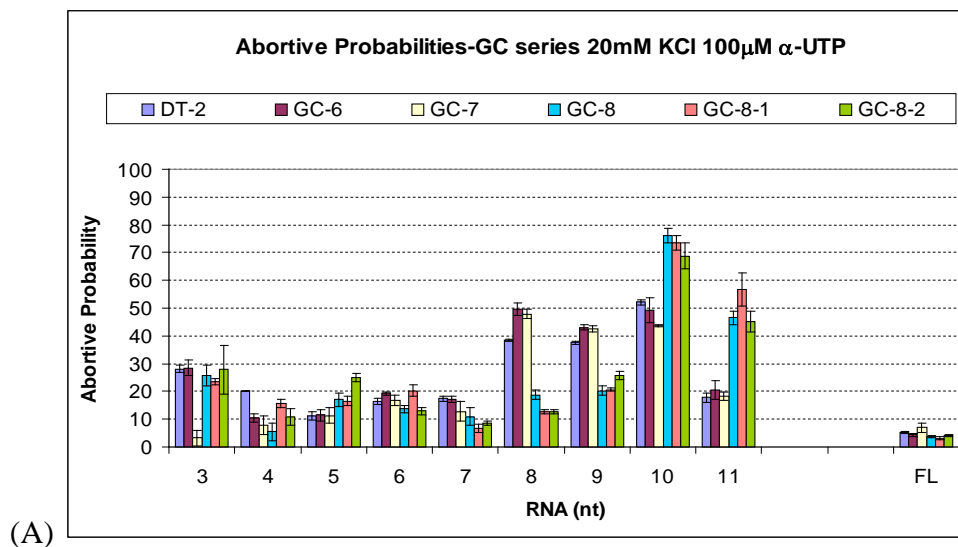
As previously mentioned, Evans (1999) had found that GC-8 has an alternative start site at the -1 position that puts a C at the 5' end of the transcript. This prompted us to pursue our investigations of variable [NTP] and [KCl] with the GC-promoters using [ $\alpha$ - $^{32}$ P]-UTP to label the transcript at every position where there is a U residue instead of just at the 5' end. Since we had found that GC-6 and -7 transcribed very efficiently at 100 mM KCl, we only used two different [KCl] (20 mM and 100 mM). For the first set of reactions, we used 100  $\mu$ M NTP.

A sample gel image from the experiments is illustrated in Figure 26. The image clearly shows the abortive products due to the alternative start site in GC-8, 8-1 and 8-2 as bands with a slightly higher rate of migration along the gel (lanes 5-7 and 12-14; labeled with -a). In fact, the abundance of the 10- and 11-nt products were seen to be much higher for the alternative start site than the regular start site with a corresponding drop in intensity of the 8- and 9 nt products. However, despite having different sequences, the products of length 6, 7 and 8 nts for both start sites of the GC-8s have the same nucleotide composition and thus migrated together to form a single band on the gel. Interestingly, in the presence of 100  $\mu$ M NTP, the intensities of the bands for GC-7 and the GC-8 variants were not higher at the greater KCl concentration of 100 mM contrary to our previous observation with GC-6 and -7. Hence we decided to repeat the experiment with 250  $\mu$ M NTP.

**Figure 26. Gel image comparing the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt-N25).** All reactions were carried out in the presence of 100  $\mu$ M NTP and [ $\alpha$ -<sup>32</sup>P]-UTP label with either 20 mM (lanes 1 through 7) or 100 mM KCl (lanes 8 through 14) for 10 minutes. DT-2 produces a 57-nt long FL transcript. All the GCs produce a 62-nt long FL transcript. The image shows that for DT-2 and GC-6, the intensities of the bands are much higher at the higher salt concentration. However, GC-7 and all the variants of GC-8 give better signal at the lower salt concentration. The GC-8 variants also have an alternative start site at -1 and this is indicated by the presence of additional bands with a slightly higher rate of migration for most of the abortive products. The bands arising from the alternative start site are numbered with an “-a”. The ladder arising from the alternative start site also shows higher intensities of the 9, 10, and 11 nt long abortive products.



**Figure 27. Comparison of the abortive profiles of the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt-N25).** All reactions were carried out for 10 minutes in the presence of 100  $\mu$ M NTP and [ $\alpha$ - $^{32}$ P]-UTP label with either (A) 20 mM, or (B) 100 mM KCl. The abortive probabilities for the GC-8s were calculated by adding the signals from both the regular and alternative start sites. The amount of FL synthesis for DT-2, and GC-6 and -7 almost triples at the higher salt concentration whereas those for the GC-8s remain very low (~5%) at both salt concentrations. Additionally, the abortive profiles of the GC-8s differ from those of DT-2, and GC-6 and 7, at positions 8 through 11.



As Figure 28 shows, in the presence of 250  $\mu\text{M}$  NTP, both the transcription rate as well as the amount of full-length synthesis were higher at the 100 mM KCl compared to the 20 mM concentration for all the GC-promoters which was in agreement with our previous findings. However, the abortive profiles, once again, did not deviate much from that obtained at 100  $\mu\text{M}$  NTP. At the higher [NTP], the only change was observed with the GC-8s which showed twice as much FL synthesis with 100 mM KCl compared to that with 20 mM KCl (compare Figures 29 A and B).

This differential sensitivity to [KCl] in the GC-8s compared to GC-6 or -7 is more clearly illustrated in Figure 30. As shown in Figures 30 A through C, DT-2, GC-6 and GC-7 all showed better escape at the higher salt concentration of KCl, indicated by a decrease in the abortive probabilities and an increase in the FL product. In fact, at the higher [KCl] the amounts of FL synthesis are almost tripled in all of these constructs. This trend was similar with both 100  $\mu\text{M}$  and 250  $\mu\text{M}$  NTP. On the contrary, the GC-8s transcribed maximally only when both the salt and the NTP concentrations were high. As demonstrated by panels D through F on Figure 30, at the lower [NTP] there was virtually no change in the amount of FL even at the higher [KCl]. But at 250  $\mu\text{M}$  NTP, the FL synthesis in all three of the GC-8s increased by about 1.5 fold. This suggests that stabilization of the open complex for the GC-8s with the higher [KCl] is only effective in promoting escape if there is a higher initiation frequency brought about by the higher [NTP]. This is yet another indication of a relatively unstable open complex in these constructs probably

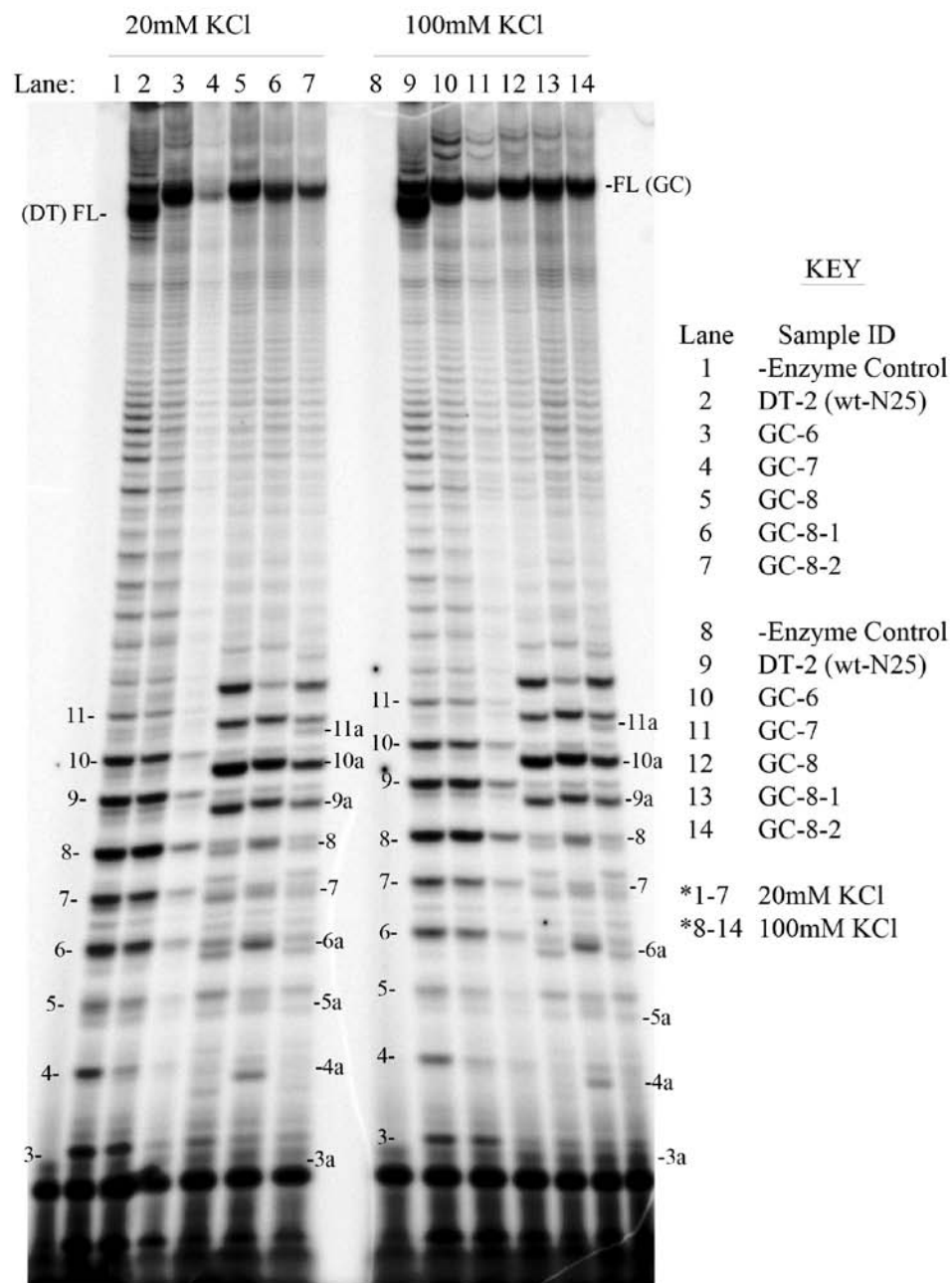
arising due to the longer GC-rich DIS region. Unfortunately, we were unable to elucidate any further conclusions from the results available. Perhaps additional investigations into the kinetics of each step of transcription initiation in these constructs, along with ppGpp (Barker *et al.* 2001a) and heparin analysis may be able provide some more insight into the dynamics of promoter escape in these mutant promoters.

It is very interesting to note that all the GC-8s showed an increase in the abortive probability of the longer fragments (10-11 nt) relative to the fragments of medium length (8-9 nt). This is not only opposite to our expectations but also quite different from the abortive profile of wt-*N25* where the 8 nt product is the most abundant. Additionally, this altered abortive pattern had resulted from the alternative rather than the regular start site. This indicated a preference of the *E. coli* RNAP for a 6-7 bp long DIS region over an 8 nt one.

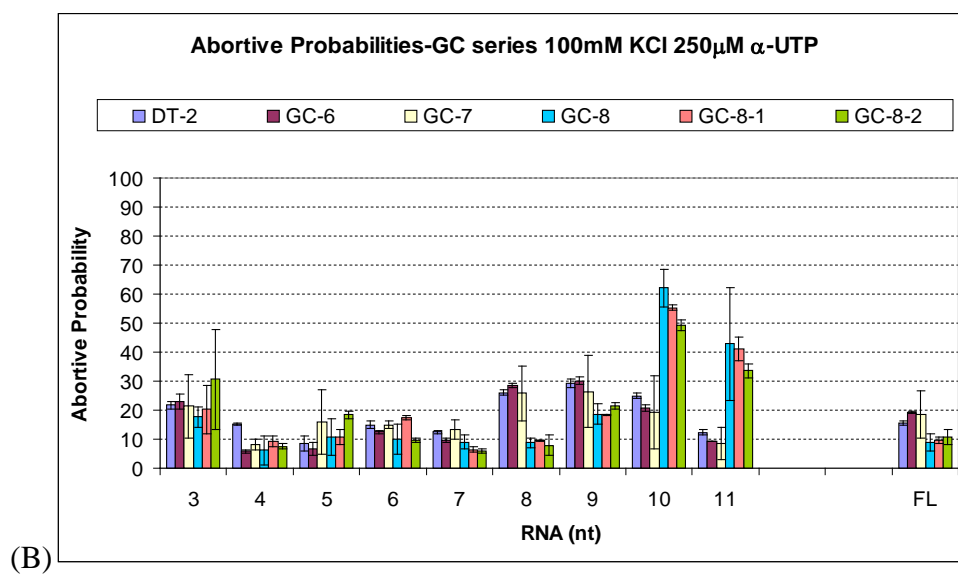
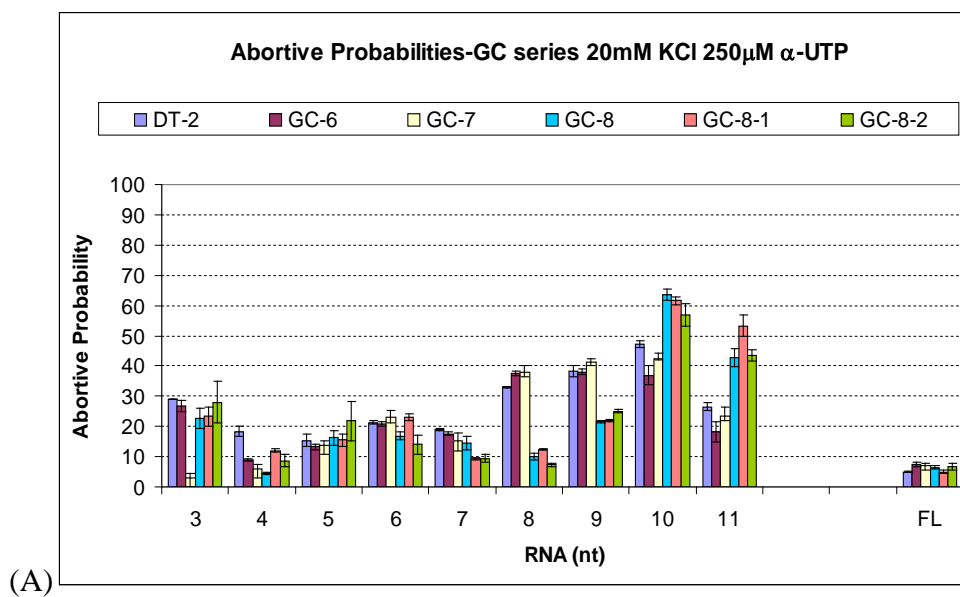
It can be speculated that, since the transcripts from the alternative start site have different sequences than those arising from the regular start site, they confer a different RNA-DNA heteroduplex stability leading to the dissimilarity in the abortive profiles. But, there is not enough evidence to support this conjecture. Despite the difference in the abortive ladder, the markedly low levels of FL product in the GC-8s show that most of the initiation events never undergo successful promoter escape and further indicate the rate limitation at the open complex formation step.



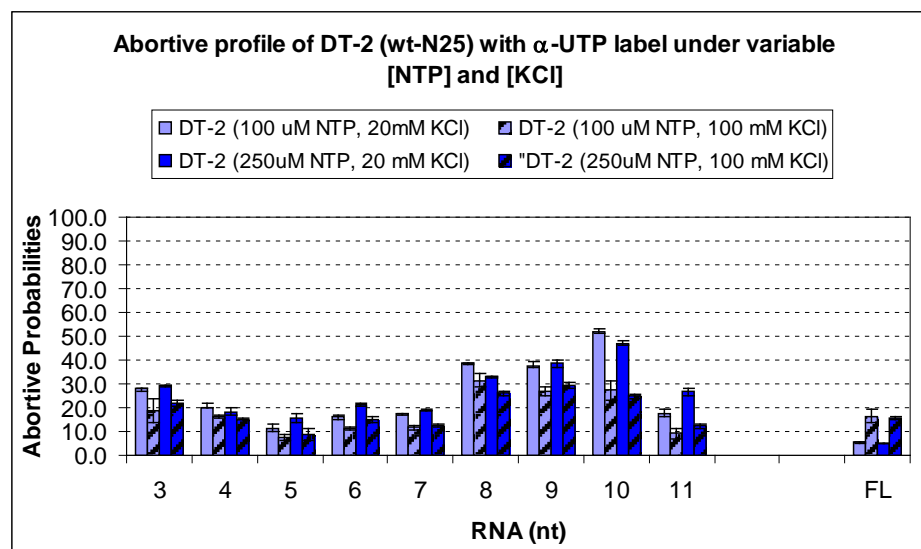
**Figure 28. Gel image comparing the GC-promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt-N25).** All reactions were carried out in the presence of 250  $\mu$ M NTP and [ $\alpha$ -<sup>32</sup>P]-UTP label with either 20 mM (lanes 1 through 7) or 100 mM KCl (lanes 8 through 14) for 10 minutes. DT-2 produces a 57-nt long FL transcript. All the GCs produce a 62-nt long FL transcript. The image shows that in all the promoters, including the GC-8s, the intensities of the bands are much higher at the higher salt concentration. The extent of transcription initiation, as well as full-length synthesis, seems to be higher for GC-8s at this [NTP]. The bands arising from the alternative start site in the GC-8s are numbered with an “-a”. Once again, the alternative start site shows a greater rate of abortion at the 9, 10, and 11 nt positions.



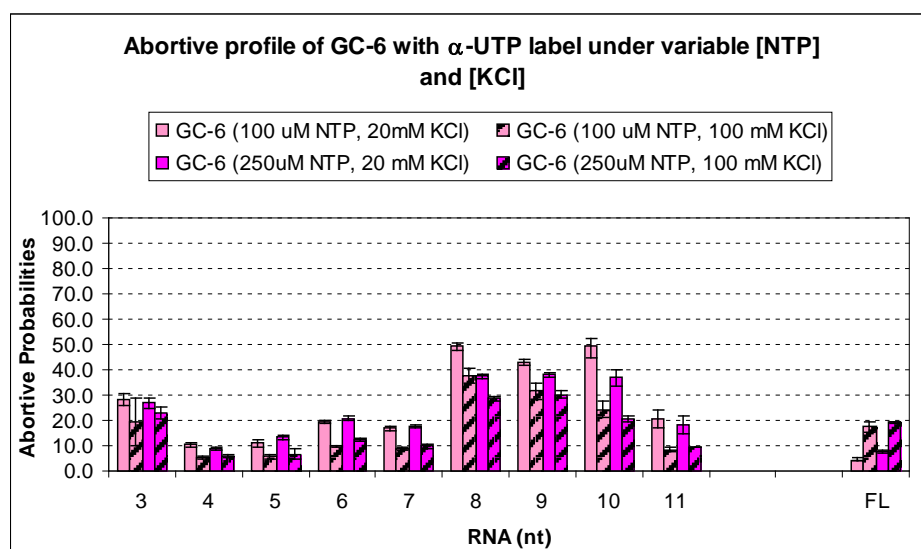
**Figure 29. Comparison of the abortive profiles of the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) with DT-2 (wt-N25).** All reactions were carried out for 10 minutes in the presence of 250  $\mu$ M NTP and [ $\alpha$ - $^{32}$ P]-UTP label with either (A) 20 mM, or (B) 100 mM KCl. The abortive probabilities for the GC-8s were calculated by adding the signals from both the regular and alternative start sites. The extent of FL synthesis for all the promoters is higher at the higher salt concentration. The abortive profiles of the GC-8s still show a difference from those of DT-2, and GC-6 and 7 at positions 8 through 11.



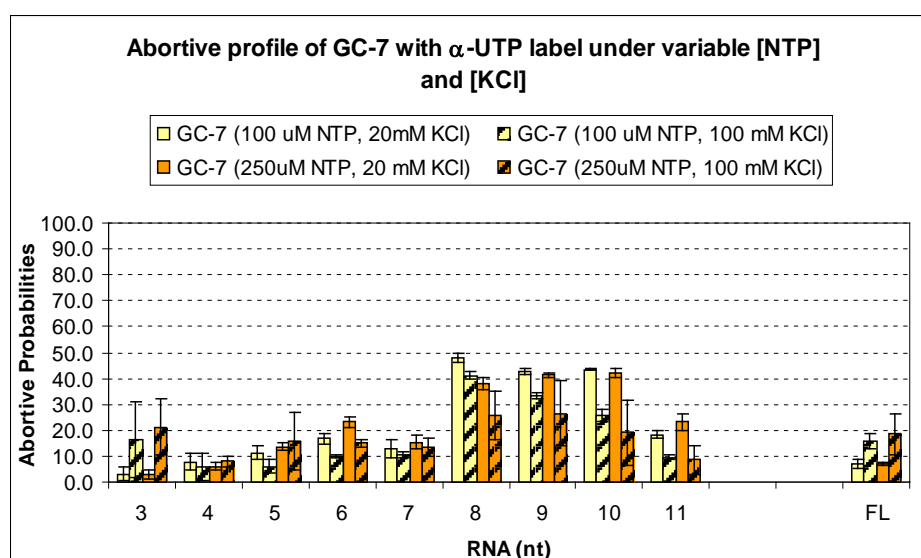
**Figure 30. Comparison of the abortive profiles of DT-2 (wt-*N25*) and the GC- promoters (GC-6, 7, 8, 8-1 and 8-2) under different [NTP] and [KCl] combinations.** All reactions were carried out for 10 minutes with [ $\alpha$ - $^{32}$ P]-UTP label. Abortive profiles of (A) DT-2; (B) GC-6; (C) GC-7; (D) GC-8; (E) GC-8-1; and (F) GC-8-2. Unlike, DT-2, GC-6 and GC-7, that show increase in FL synthesis with the higher [KCl] at both [NTP]'s, GC-8 and its variants show better escape at the higher [KCl] only at the higher [NTP]. The abortive probabilities of the GC-8s are different from those of DT-2 and GC-6, -7 at positions 8 through 11 nt.



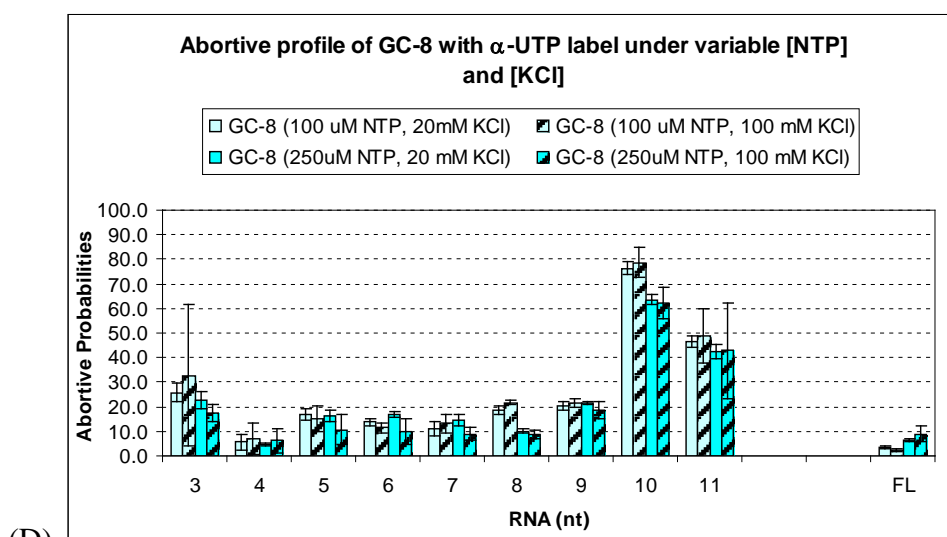
(A)



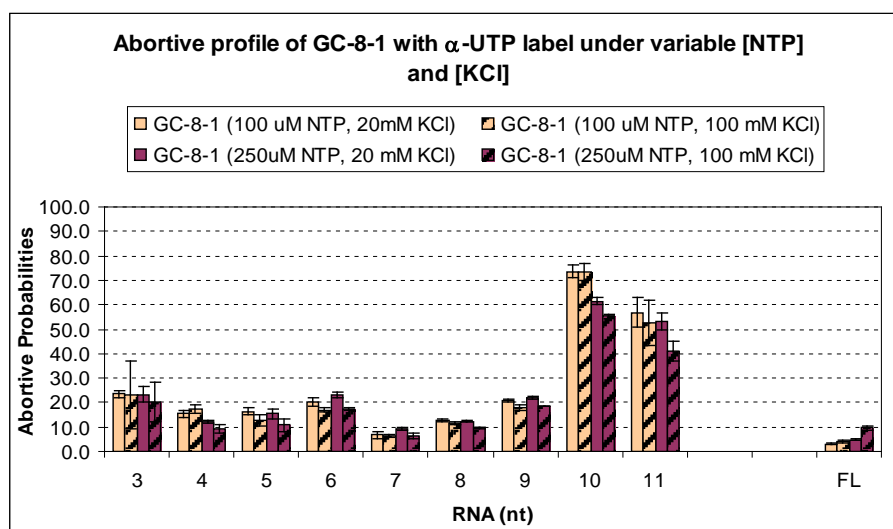
(B)



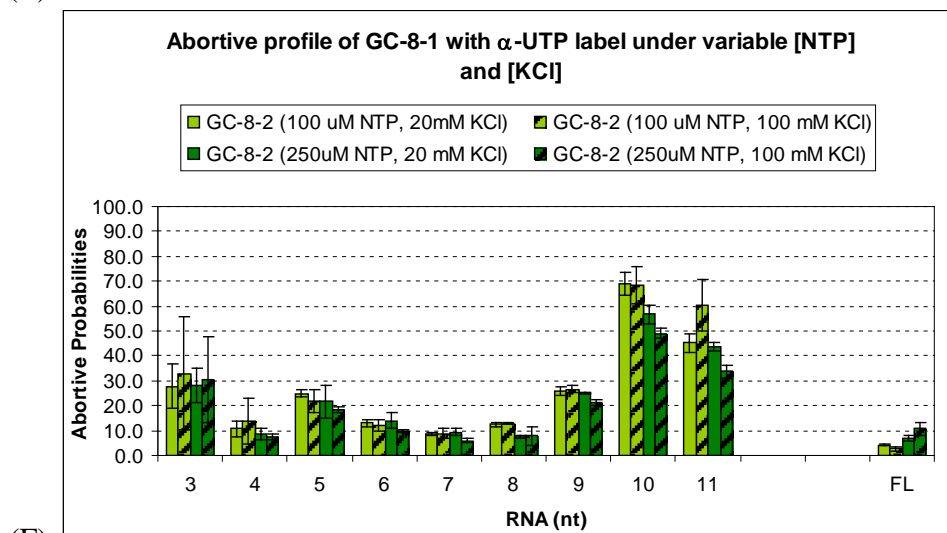
(C)



(D)



(E)



(F)

*SUMMARY OF FINDINGS FROM THE GC-SERIES*

Our results from the GC-series demonstrated that the GC-richness of the DIS-sequence does not alter the abortive profile of *N25* promoters if they are 6- or even 7-bp in length. The GC- promoters were found to transcribe better at higher [NTP] and are somewhat sensitive to [KCl] in the reaction mixture similar to *rrnB* P1 which also has a relatively GC rich DIS region (Gourse, 1988; Gaal *et al.*, 1997; Barker and Gourse, 2001).

Unlike GC-6 and -7, the longer and more GC-rich DIS sequences for GC-8s behaved quite differently from wt-*N25*. These promoters not only transcribed from an alternative start site at the -1 position but their abortive probabilities also differed from the wt-*N25* at the 8-11 nt positions. Although promoter escape still occurred at the +12 position in these promoters, it only did at very low frequencies. This observation further supports the hypothesis that they are rate limited at the open complex formation step rather than the promoter escape step, similar to the *rrnB* P1 promoter (Mulligan *et al.*, 1984). Hence, the results from the GC-series promoters, although intriguing, cannot be used to derive further associations between upstream ITC bubble collapse and promoter escape in the T5 *N25* mutants.



## DISCUSSION

Long before much of the mechanistic details of initiation to elongation transition during transcription were known, Straney and Crothers (1987) had suggested that the energy for breaking the contacts between the RNAP and the promoter, in order to cause successful escape, must come from a “stressed intermediate.” Indeed, just about 20 years later, their hypothesis was supported by the evidence of DNA scrunching revealed through studies using single molecule nanomanipulation (Revyakin *et al.*, 2004; Revyakin *et al.*, 2006) and FRET (Kapanidis *et al.*, 2006). However, this phenomenon of DNA scrunching and the formation of the stressed complex, although very useful in describing the energetics of promoter escape, cannot be used to address some of the sequence specific observations that also seemed to influence the process (Hsu *et al.*, 2006). The results of this investigation have been used to resolve and clarify some of these issues and to elucidate the roles of the DNA sequences within the ITC in promoter escape.

The dependence of promoter escape of T5 *N25* on the sequence-context, especially at the downstream end of the ITC bubble, was initially investigated by Hsu and colleagues (2006) when they compared the abortive profile of ~40 ITS mutants. Their motivation for undertaking such an enormous task had come from earlier observations with the *N25<sub>anti</sub>*. As previously mentioned, *N25<sub>anti</sub>* has the direct anti (A↔C and G↔T

substitutions) ITS sequence to that of the *N25* spanning from the +3 to +20. But, this difference is enough to render *N25<sub>anti</sub>* ~10 folds weaker in undergoing escape. Additionally, the position of escape in *N25<sub>anti</sub>* is delayed to +16 whereas that for *N25* is at +12 (Hsu *et al.*, 2003). This indicated that the phenomenon of promoter escape is strongly dependent on the ITS sequence, i.e. the region spanning the downstream edge of the ITC bubble.

The above hypothesis was indeed supported by results from the mutagenesis study. All the random ITS mutants (referred to as the *DG100* series) showed a delayed position of escape at the +15/+16 juncture. Not only that, the productive synthesis in the mutants varied ~25 folds relative to that of wt-*N25* (Hsu *et al.*, 2006). This finding was further supplemented by results from yet another subsequent investigation with a series of ITS mutants referred to as the *DG200s* (Chander *et al.*, 2007). Of particular interest in the latter series is *DG203*, which, despite being different from *N25* only at the +3 to +10 position of the ITS, exhibits an even greater delay in escape to the +20 position. This intriguing connection between the ITS and promoter escape eventually led to a closer examination of the sequence composition and the corresponding rewinding potential at the downstream end of the ITC and led to the proposal of the model represented earlier in Figure 2. Hence, the project described in this work was initiated in which we attempted to elucidate whether or not sequence changes in the upstream end of the ITC bubble (spanning the -10/DIS region) would have similar effects on promoter escape.

The effect of the relative rewinding or bubble collapse at the upstream and downstream ends of the initial transcription bubble have been investigated by Gong and colleagues with the T7 RNAP (Gong *et al.*, 2004; Gong and Martin, 2006). They altered the rewinding potential of different regions within the bubble by using several approaches, including mismatch incorporation, nicked promoters, partial single-stranded templates, etc. They used a consensus promoter sequence for T7 RNAP for all their analyses. Of particular interest to us were their investigations using mismatched stretches within the transcription bubble. T7 RNAP normally escapes at the +9 position (Martin *et al.*, 1988). However, the results from some of the mismatch experiments have indicated that the presence of a non-complementary stretch within the upstream region of the elongation bubble was able to increase the abundance of 11-13 nt abortive products with a corresponding drop in the amount of the FL product. Their data also revealed that only about 2-bp mismatch window was sufficient to demonstrate this change in the abortive pattern, provided that it was placed very close to the +1 start site. They further showed that the nucleotides from the -4 to -1 region of the promoter were the most involved in ensuing successful promoter escape (Gong *et al.*, 2004).

A similar study with T7 RNAP used mismatched constructs to monitor the translocation of the initiation bubble into the elongation phase (Zhou *et al.*, 2007). The results from the study have revealed that an impaired collapse at upstream end of the transcription bubble increased the lifetime of stalled

RNAP-promoter complex (T7 RNAP often pauses at the promoter just before it undergoes escape). This process of halting was due to the impaired collapse of the upstream end of the ITC bubble and it drastically reduced the likelihood of successful escape (Zhou *et al.*, 2007).

The findings with the T7 RNAP inspired us to use mismatches in our constructs to investigate a reduced upstream rewinding potential of the ITC and its effect on promoter escape of *E. coli* RNAP. However, we were faced with a major obstacle. The upstream boundary of the ITC includes the -10 box and the discriminator (DIS) region and so our conundrum was to find a way to alter this portion of the PRR without interfering with the contacts between the  $\sigma_2$ - domain of the RNAP and the -10 box, which could potentially affect the stability of the open complex altogether (Marr and Roberts, 1997). Fortunately, in the case of *E. coli* RNAP, the  $\sigma_2$ - domain only contacts the non-template strand of the -10 box sequence (Roberts and Roberts, 1996; Murakami *et al.*, 2002). Hence we resorted to changing the template strand sequence in our initial experiments in order to investigate our hypothesis and the DT-series promoters with smaller mismatches were designed.

Interestingly, our initial results with the DT-s containing small mismatch bubbles 1-3 bp in length were unlike what was obtained with the T7 RNAP. In the latter case, even a small 2-bp mismatch window, if placed close to the transcription start site, was enough to give rise to a noticeable change in the promoter escape pattern (Gong *et al.*, 2004). However, we failed to detect

any such effect on the abortive profiles in the case of the DT-promoters regardless of where the small mismatches were placed. These findings were later supported by the results from the non-template mismatches limited within the DIS-region (DT-14, 15 and 16).

We, therefore, continued our investigation with the DT-series containing template strand mismatches that were 5 bp or larger. We observed some changes in the abortive profile and the amount of FL synthesis between the wt and the mutants. First of all, the mutants were severely impaired in promoter escape as demonstrated by the marked decrease in the amount of the FL product (see Figure 11). Furthermore, DT-12, the construct with the largest region of mismatch (10 bp from -11 to -2), showed a considerable amount of VLATs (19- to 21-nt RNA) synthesis beyond the regular abortive ladder. These VLATs were also insensitive to GreB rescue indicating that they do not arise from backtracking (Chander *et al.*, 2007). As Revyakin *et al.* (2006) had suggested, *N25* promoter escape requires ~18 kcal/mol of energy in order to displace the promoter RNAP contacts and this usually requires about 9 bp of DNA to be scrunched within the enzyme. The fact that DT-12 is extremely defective in upstream rewinding suggests that it requires a greater amount of energy to undergo escape which necessitates more extensive DNA scrunching. This is probably why RNAP continues to transcribe until the 21<sup>st</sup> nt in order to build up enough stress for promoter release. Inevitably, with considerable added stress in the ITC, a fraction of these complexes may also undergo a process called hyper-forward translocation which eventually gives

rise to VLATs that are beyond the reach of GreB mediated cleavage and rescue (Chander *et al.*, 2007). Regardless of how VLATs arise, this delayed position of escape was the only supportive evidence to our initial hypothesis regarding the DTs. These latter findings were in agreement with what had been observed in the T7 RNAP (Gong *et al.* 2004).

Thus, our overall results from the DT-series indicated that compromised bubble collapse may not be the only factor limiting the escape from these templates. In fact, we saw with DT-12, that escape can occur without the possibility of upstream bubble rewinding! This is quite different from what has been seen in the case of T7 RNAP where the upstream transcription bubble occurs simultaneously with promoter escape (Liu and Martin, 2002; Gong *et al.* 2004). However, the discrepancy in the two enzymes may very well be due to the inherent structural differences between *E. coli* RNAP and T7 RNAP (Hsu, 2002a)

At the other end of the spectrum, our use of GC-rich DIS sequences to investigate the effect of an elevated upstream rewinding potential had the possibility to engender two effects on the RNAP-promoter interactions. On one hand, the GC-richness could limit the rate of formation of the open complex by making it harder to separate the two DNA strands. On the other, the open complex once formed, could collapse very readily and thus have a shorter half-life. Both effects together give rise to a low fraction of unstable open complexes, limiting the level of overall transcription initiation at these

promoters. This is, in fact, what we observed, especially with the GC-8 promoter variants.

The GC-rich DIS sequences are very similar to that of the *rrnB* P1 promoter, especially in the case of GC-8 and its variants where the DIS region is 8 bp long. Both these sets of promoters also have strong binding affinity with the RNAP due to consensus-like sequences at the -35 and -10-boxes (Barker and Gourse, 2001). Indeed, some of the aspects arising from these similarities were reflected in our data. For example, the *rrnB* P1 promoter is known to transcribe very efficiently at high NTP concentrations of about 500  $\mu$ M (Barker and Gourse, 2001). In the same way, the GC-promoters (most noticeably GC-6 and -7) also transcribed better at high NTP concentrations, as reported in Figures, 20, 22 and 30. This could be an indication that the open-complexes of the GC-promoters have a short half-life similar to that of the *rrnB* P1, hence requiring the abundance of NTPs in the reaction mixture to trap the complex in its open form by mass action and drive it towards initiation.

Nonetheless, some differences between the GC-promoters and *rrnB* P1 were also noted. As reported by Gourse (1988), the relatively short half-life of the open-complex of the *rrnB* P1 also makes it highly sensitive to the salt concentration in the reaction medium. Thus, the promoter transcribes most efficiently at very low salt concentrations,  $\sim$  30 mM. Our results, on the other hand, indicated a slightly higher tolerance of the GC-constructs to the

concentration of KCl. As illustrated most clearly by GC-6 and -7 (see Figures 25 and 30), comparable transcription efficiencies to that of wt-*N25* were obtained at KCl concentrations as high as 100 mM. In addition, the data indicated that the overall abortive profile of GC-6 and -7 are qualitatively very similar to that of wt-*N25* under all reaction conditions, which was very interesting indeed. The presence of the 11 nt long abortive ladder for all the GC-promoters can also be attributed to their 17 bp long spacer region in contrast with the 16 bp long spacer in *rrnB* P1 (Barker and Gourse, 2001; Vo *et al.*, 2003).

Of particular interest to us were the results with the GC-8 variants. In these promoters, the longer DIS connection between the -10 box and the +1 start site led to the utilization of both the regular and an alternative start site at the -1 position. The alternative abortive ladder was detected with [ $\alpha$ -<sup>32</sup>P]-UTP labeling. Unfortunately, quantitation in these constructs were confounded by the fact that some of the abortive products from the regular and alternative start site, despite having different sequences, had almost identical molecular weight and thus migrated as one band on the gel. Nevertheless, when the abortive products from both start sites were plotted together for the GC-8s, we noticed that the abundance of the larger abortive products (10 and 11 nt) was almost twice as that of the ones found in wt-*N25* and GC-6 and -7. This increase was supported by a corresponding drop in the abundance of abortive products that were 8 and 9 nts in length. However, whether this is related to the stability of the open complex or some other intrinsic factor



within the promoter, or it is caused by the difference in the sequence of the transcript is still unclear. But the fact that the ladder arising from the alternative start site was contributing more to the abortive profile of the GC-8s indicated that the *E. coli* RNAP must have a preference for a 6-7 bp DIS region over an 8-bp one with the *N25* promoter.

Yet another confounding discovery with the GC-8 variants is illustrated in Figure 30. Whereas GC-6 and GC-7 showed better escape at the higher salt concentration of KCl at both 100  $\mu$ M and 250  $\mu$ M NTP, the GC-8s transcribed best only when both the salt and the NTP concentrations were high. This suggests that the latter constructs must require both the stabilization of the open complex (due to the higher [KCl]) and a higher initiation frequency (brought about by the higher [NTP]) in order to undergo an elevated rate of promoter escape. This is probably attributable to the relatively longer DIS region that has the ability to hinder the open complex stability similar to that of the *rrnB* P1 promoter. Nevertheless, further experiment, especially kinetic assays and ppGpp and heparin challenges must be done in order to derive more concrete conclusions about the factors and/or rate constants that are governing the promoter escape in the GC-promoters. ppGpp assays will be especially insightful in the case of the GC-promoters because it has been shown to inhibit transcription at promoters with short-lived open complexes such as *rrnB* P1 (Barker *et al.*, 2001a and 2001b). Additionally, the promoters could be cloned into plasmids to probe the effect of DNA supercoiling on the transcription of the GC-rich constructs. Despite the

matter, since the GC-8s have been repeatedly indicated to be rate-limited at the open complex formation step we cannot use them for further analysis in the context of promoter escape.

Taken together, our results from the analysis of both the DT and GC-series promoters indicate that upstream rewinding potential has very little influence on the promoter escape of *E. coli* RNAP. The position of escape is not really affected regardless of whether the upstream rewinding is impaired (in the case of the DTs) or favored (in the case of GCs). Studies by Liu and Martin (2002) and Gong *et al.* (2004) have shown that, in the case of T7 RNA polymerase, initial bubble collapse and transition to elongation occurs in close proximity to each other suggesting that these two events must be taking place in unison. Contrary to this observation, our results provide evidence that energy for disrupting the  $\sigma_{2/-10}$  DNA contacts during promoter escape of *E. coli* RNAP is not necessarily provided by the collapse of the upstream bubble. Furthermore, in support of the hypothesis by Straney and Crothers (1987), the results of this work indicate that the energy for promoter escape is primarily derived from DNA scrunching events (Revyakin *et al.*, 2006) and less from the reannealing of the melted ITC bubble.

## CONCLUSION

Promoter escape usually requires the occurrence of two major events: (1) the rewinding of the upstream boundary of the ITC, leading to the forward translocation of the bubble; and (2) release of promoter contacts by the  $\sigma$ -subunit of the RNAP to move into elongation. The latter step may or may not be accompanied by a simultaneous liberation of the  $\sigma$ -subunit from the RNAP holoenzyme (Mukhopadhyay *et al.*, 2001; Hsu, 2002a; Mooney *et al.*, 2005). However, just like the “chicken-and-the-egg” problem, the exact sequence of the above events is yet to be established.

The results of this work have repeatedly demonstrated that even when upstream rewinding is impaired, the *E. coli* RNAP is still able to undergo promoter escape. Only in the case of a severely compromised upstream bubble collapse (DT-12) was the position of escape noticeably delayed. This was complemented by the abortive profiles of GC-6 and -7 where no apparent change in the position of escape was observed even though the upstream rewinding may have been facilitated by the GC-rich DIS region. The results further indicate that, unlike T7 RNAP, *E. coli* RNAP escape probably involves promoter release, i.e. disruption of the contacts between RNAP and promoter DNA, with or without prior collapse of the open complex bubble. Therefore, this work has the potential to open new avenues of investigation that may lead to the clarification of yet more mechanistic details of promoter escape and abortive initiation.

## REFERENCES

- Artsimovitch, I., and Landick, R. (2000) Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA* **97**, 7090-7095.
- Barker, M. M., Gaal, T., Josaitis, C. A., and Gourse, R. L. (2001a) Mechanism of Regulation of Transcription Regulation by ppGpp. I. Effects of ppGpp on Transcription Initiation *In Vivo* and *In Vitro*. *J. Mol. Biol.* **305**, 673-688.
- Barker, M. M., Gaal, T., and Gourse, R. L. (2001b) Mechanism of Regulation of Transcription Regulation by ppGpp. II. Models of Positive Control Based on Properties of RNAP Mutants and Competition for RNAP. *J. Mol. Biol.* **305**, 689-702.
- Barker, M. M., and Gourse, R. L. (2001) Regulation of rRNA Transcription Correlates with Nucleoside Triphosphate Sensing. *J. Bacteriol.* **183**, 6315-6323.
- Barnes, S. (2008) The Effect of Varying Potassium Chloride Concentrations on the Production of Full Length Transcripts in Prokaryotic Transcription Initiation [Senior Thesis]. Mount Holyoke College, South Hadley, MA-01075.
- 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**, 3245-3253.
- Carpousis, A. J., and Gralla, J. D. (1985) Interaction of RNA Polymerase with *lac* UV5 Promoter DNA during mRNA Initiation and Elongation. Footprinting, Methylation, and Rifampicin-sensitivity Changes Accompanying Transcription Initiation. *J. Mol. Biol.* **183**, 165-177.
- Chamberlin, M. J. (1974) The Selectivity of Transcription. *Annu. Rev. Biochem.* **43**, 721-775.
- Chan, C. L., and Gross, C. A. (2001) The Anti-Initial Transcribed Sequence, A Portable Sequence that Impedes Promoter Escape, Requires Sigma70 for Function. *J. Biol. Chem.* **276**, 38201-38209.
- Chander, M., Austin, K. M., Aye-han, N. -N., Sirkar, 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.

- Das, A. (1993) Control of Transcription Termination by RNA Binding Proteins. *Annu. Rev. Biochem.* **62**, 893-930.
- Evans, K. (1999) RNA Polymerase Open Complex Stability and Abortive Cycling: Role of the Discriminator Region [Senior Thesis]. Mount Holyoke College, South Hadley, MA-01075
- Feng, G., 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., Bartlett, M. S., Ross, W., Turnbough, Jr., C. L., and Gourse, R. L. (1997) Transcription Regulation by Initiating NTP Concentration: rRNA Synthesis in Bacteria. *Science* **278**, 2092-2097.
- Gong, P., Esposito, E. A., and Martin, C. T. (2004) Initial Bubble Collapse Plays a Key Role in the Transition to Elongation in T7 RNA Polymerase. *J. Biol. Chem.* **279**, 44277-44285.
- Gong, P., and Martin, C. T. (2006) Mechanism of Instability in Abortive Cycling by T7 RNA Polymerase. *J. Biol. Chem.* **281**, 23533-23544.
- Gourse, R. L., deBoer, H. A., and Nomura, M. (1986) DNA Determinants of rRNA Synthesis in *E. coli*: Growth Rate Dependent Regulation, Feedback Inhibition, Upstream Activation, Antitermination. *Cell* **44**, 197-203.
- Gourse, R. L. (1988) Visualization and Quantitative Analysis of Complex Formation between *E. coli* RNA Polymerase and the rRNA Promoter *In Vitro*. *Nucl. Acids Res.* **16**, 9789-9809.
- Hawley, D. K., and McClure, W. R. (1983) Compilation and Analysis of *Escherichia coli* Promoter DNA Sequences. *Nucl. Acids Res.* **11**, 2237-2355.
- 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. USA* **92**:11588-11592.
- Hsu, L. M. (2002a) Promoter Clearance and Escape in Prokaryotes. *Biochim. Biophys. Acta* **1577**, 191-207.
- Hsu, L. M. (2002b) Open Season on RNA Polymerase. *Nat. Struct. Biol.* **9**, 502-504.

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.

Hsu, L. M. (2009) Monitoring Abortive Initiation. *Methods* **47**, 25-36.

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.

Knaus, R., and Bujard, H. 1990. Principles governing the activity of *E. coli* promoters. In Eckstein, F., and Lilley D. M. J. [eds.] *Nucleic Acids and Molecular Biology* Vol. **4**, pp. 110-122. Springer-Verlag, Heidelberg.

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. USA* **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.

Krummel, B., and Chamberlin, M.J. (1992) Structural Analysis of Ternary Complexes of *Escherichia coli* RNA Polymerase: Deoxyribonuclease I Footprinting of Defined Complexes. *J. Mol. Biol.* **225**, 239-250.

Liu, C., and Martin, C. T. (2002) Promoter clearance by T7 RNA polymerase. Initial Bubble Collapse and Transcript Dissociation Monitored by Base Analog Fluorescence. *J Biol. Chem.* **277**, 2725-2731

Marr, M. T., and Roberts, J. W. (1997) Promoter Recognition As Measured by Binding of Polymerase to Nontemplate Strand Oligonucleotide. *Science* **276**, 1258-1260.

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

Martin, C. T., Muller, D. K., and Coleman, J. E. (1988) Processivity in Early Stages of Transcription by T7 RNA Polymerase. *Biochemistry* **27**, 3966-3974.

Mooney, R. A., Darst, S. A., Landick, R. (2005) Sigma and RNA polymerase: An On-again, Off-again Relationship? *Mol. Cell* **20**, 335-345.

- Mukhopadhyay, J., Kapanidis, A., Mekler, V., Kortkhonjia, E., Ebright, Y., and Ebright, R. (2001) Translocation of Sigma70 with RNA Polymerase during Transcription: Fluorescence Resonance Energy Transfer Assay for Movement Relative to DNA. *Cell* **106**, 453-46
- Munson, L. M., and Reznikoff, W. S. (1981) Abortive Initiation and Long Ribonucleic Acid Synthesis. *Biochemistry* **20**, 2081-2085.
- Mulligan, M. E., Hawley, D. K., Entriken, R., and McClure, W. R. (1984) *Escherichia coli* Promoter Sequences Predict *In Vitro* RNA Polymerase Selectivity. *Nucl. Acids Res.* **12**, 789-800.
- 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**, 1285-1290
- 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.
- Orlova, M., Newlands, J., Das, A., Goldfarb, A., and Borukhov S. (1995) Intrinsic Transcript Cleavage Activity of RNA polymerase. *Proc. Natl. Acad. Sci. USA* **92**, 4596-4600.
- Revyakin, A., Ebright, R. H., and Strick, T. R. (2004) Promoter Unwinding and Promoter Clearance by RNA Polymerase: Detection by Single-Molecule DNA Nanomanipulation. *Proc. Natl. Acad. Sci. USA* **101**, 4776-4780.
- 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, C. W., and Roberts, J. W. (1996) Base-Specific Recognition of the Non-template Strand of Promoter DNA by *E. coli* RNA Polymerase *Cell* **86**, 495-501.
- Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H., and Roberts, C. W. (1998) Antitermination by Bacteriophage  $\lambda$  Q Protein. *Cold Spring Harbor Symp. Quant. Biol.* **63**, 319-325.
- Roberts, J. W. (2006) RNA Polymerase, a Scrunching Machine. *Science* **314**, 1097-1098.
- Straney, D. C., and Crothers, D. M. (1987) A Stressed Intermediate in the Formation of Stably Initiated RNA Chains at the *Escherichia coli lac UV5* Promoter. *J. Mol. Biol.* **193**, 267-278.

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.

Zhou, Y., Navaroli, D. M., Enumeh, M. S., and Martin, C. T. (2007) Dissociation of the Halted T7 RNA Polymerase Elongation Complexes Proceeds via Forward-Translocation Mechanism. *Proc. Natl. Acad. Sci. USA* **104**, 10352-10357.