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

_____ _____

Signature Date

Crosslinking the 3′ End of a VLAT to RNA Polymerase: A Method to Elucidate Forward Hyper-Translocation

by

Rebekah Wieland

A Paper Presented to the

Faculty of Mount Holyoke College in

Partial Fulfillment of the Requirements for

the Degree of Bachelors of Arts with

Honor

Department of Biochemistry

South Hadley, MA 01075

May, 2012

This thesis was prepared under the direction of:

Dr. Lilian M. Hsu

Professor of Biochemistry

for eight credits

This research is funded by an NSF-RUI Grant (MCB 0841452)

awarded to Professor Lilian M. Hsu

ACKNOWLEDGMENTS

To Professor Lilian Hsu, thank you for always supporting me, providing advice, and challenging me on the road of scientific discovery. You have been an amazing mentor, professor, and friend. I am honored to have worked in your lab and proud to graduate a biochemistry major.

To my family and friends, thank you for your guidance, honesty, and encouragement. To my father, for giving me the strength to never quit. To my mother, for being my role model of a strong, independent woman. To my brother, for keeping me grounded.

To Professor Megan Nuñez and Professor Kathryn McMenimen thank you for being on my thesis committee, providing support in times of stress, and guiding me to being a better mentor to others.

To the Hsu Lab members, thank you for all the great memories-full of laughter, anxiety, and inspiration.

To Mount Holyoke College for providing a wonderful community of talented women that will truly make a difference in the world.

TABLE OF CONTENTS

Page

LIST OF FIGURES

ABSTRACT

Transcription is the process by which an organism's genetic material encoded in DNA is arranged into a messenger molecule called RNA. Transcription is divided into three sequential phases: initiation, elongation, and termination. Of these three phases, initiation is the most highly regulated. In our model organism *Escherichia coli*, it begins with the specific binding of the RNA polymerase (RNAP) holoenzyme ($\alpha_2\beta\beta'\omega\sigma^{70}$) to two highly conserved hexamers (-35 and -10 elements) on the promoter within a gene encoding sequence of DNA. This initial binding of RNAP to the double-stranded DNA forms a complex referred to as the closed complex (RP_c). RNAP then proceeds to melt open a 13-14 base-pair (bp) stretch of DNA (deHaseth *et al.*, 1998; Naryshkin *et al*, 2000). This melting process produces a bubble in the DNA structure where the two strands have locally separated from each other. This structure is commonly called the open complex (RP_o). Following the formation of the open complex, *de novo* RNA synthesis can begin. As the nascent RNA begins to grow, the transcription bubble melts open an additional 6-8 bp. The bulk of the growing nascent RNA and transcription bubble builds tension within the RNAP producing a phenomenon known as scrunching (Revyakin *et al*., 2006). Many attempts are made to grow the nascent RNA to 8-15 nucleotides (nt), causing the production of small abortive RNA transcripts. After the RNA molecule has grown to 8-15 nt,

promoter escape occurs, disrupting the initial DNA promoter contacts, so the elongation phase can ensue.

Nevertheless, the process of promoter escape is a difficultly orchestrated accomplishment in which problems can occur. On some promoters, extremely strong RNAP and DNA promoter contacts require excessive scrunching to bring about promoter escape. The high stress accumulated inside the RNAP, when it succeeds to disrupt the promoter-RNAP contacts, releases so much force that the RNAP is propelled downstream by several nucleotides (Chander *et al.*, 2007). This process called forward hyper-translocation causes the growing end of the nascent RNA, which has now reached lengths of 16-19 nt, to move upstream and out of register with the active site. Thus, the RNA molecule cannot be elongated further, transcription is halted, and the production of **v**ery **l**ong **a**bortive **t**ranscripts (VLATs) occurs when a VLAT leaves the RNAP via the RNA exit channel.

To prove that the VLAT production is the result of forward hyper-translocation, I conducted photocrosslinking experiments, utilizing a photoactivatable thiol group attached to the 3′-end of a VLAT, to locate its position on a VLAT-producing RNAP complex. The crosslinking method is a very advantageous procedure that has been successfully used in the determination of the backtracking mechanism of abortive RNA 2-15 nt production, locating the 3′ end of backtracked abortive RNA molecules to the secondary channel of RNAP (Borukhov *et al.*, 1996; Opalka *et al*., 2003). For VLATs that are the products of

forward hyper-translocation, their 3' end is predicted to map to the vicinity of the RNA exit channel.

To undertake this crosslinking study, unique transcription reaction features are adopted that include an initiating ApU dinucleotide and the incorporation of a photocrosslinkable nucleotide (S_4UTP) at the 3′ end. Having found transcription reaction conditions that can specifically label the 3′-most residue with S_4UTP on VLAT RNA U19, U18, or U17, I began the photocrosslinking investigations and obtained UV-dependent and S_4UTP-dependent crosslinking of U19 and U18 RNA to the β/β′ subunits as shown by SDS-PAGE gel fractionation. After showing that specific crosslinking occurs, the next goal was to prepare a sample for MALDI-TOF mass spectrometry analysis to identify the exact amino acids crosslinked. To obtain a higher yield of sample, I investigated several methods for enriching the crosslinked product, including increasing the length of UV irradiation, changing the photocrosslinkable nucleotide, changing the UV irradiation wavelength, and increasing reaction volume followed by concentration through precipitation. I ultimately decided on the following reaction conditions: a 100-μL reaction containing 30 nM DNA DG203/*SPfullcon-U2/U19* or *U2/U18*, 1X transcription buffer, 50 mM KCl, 500 μM ApU, 10 μM G/C/ATP, 150 nM RNAP or 150 nM RNAP + 1.5 mM GreB, and 100 μM S_4UTP incubated at 37 °C with 365 nm UV irradiation for one hour; afterwards, reaction products are concentrated by TCA precipitation and fractionated as one sample in SDS-PAGE which can separate the crosslinked product from unmodified β/β’ subunits.

After excising the crosslinked product band, I performed in-gel digestion to remove the radioactivity associated with the RNA and to reduce the size of the crosslinked product through the use of nucleases and proteases. I was able to perform the technique of in-gel digestion with micrococcal nuclease (MNase), so far only with partial success. I suggest, as future directions, the use of both MNase and RNase T1 to remove all radioactivity prior to proteolytic treatment with trypsin to reduce the size of the crosslinked product.

INTRODUCTION

The process of transcription in *E. coli* is mediated catalytically by RNA polymerase (RNAP). The major RNAP holoenzyme we work with is comprised of five subunits, including: β (~150 kDa), β′ (~155/60 kDa), two identical α subunits (~36 kDa each) forming a dimer, ω (~6 kDa), and σ^{70} (~70 kDa) (Naryshkin *et al.*, 2000). The core enzyme, which lacks the σ subunit, resembles a crab-claw shape with the β and β′ subunits forming the pincer and the α_2 and ω surrounding the base. This pincer forms a narrow channel 27 Å in diameter, surrounded by positively charged amino acids (Zhang *et al.*, 1999). These amino acid residues ensure its ability to easily encompass the negatively charged double-stranded DNA. At the floor of the channel lies the active site, comprised of two Mg^{2+} ions. Three aspartate residues chelate these two Mg^{2+} (Nelson and Cox, 2008). One of the two Mg^{2+} called cMg2 stabilizes the pyrophosphate (PP_i) on the incoming nucleotide. Meanwhile, the other Mg^{2+} called cMg1 facilitates the nucleophilic attack of the 3′-hydroxyl group of the nascent RNA to the α-phosphate of the incoming ribonucleotide, with the subsequent cleavage and displacement of the pyrophosphate (Steitz and Steitz, 1993; Vassylyey *et al*., 2005).

Transcription initiation is the most regulated stage of transcription. To begin, transcription initiation starts with *E. coli* RNAP holoenzyme ($\alpha_2\beta\beta'\omega\sigma$) binding to the promoter within a gene encoding sequence of DNA. The *E. coli*

σ^{70} ($E\sigma^{70}$) promoters show a consensus sequence consisting of an UP element, the -35 element 5′-TTGACA-3′, a 17-bp spacer, the -10 element 5′-TATAAT-3′, connected via a DIS region to the +1 site of transcription (McClure, 1985). As bound, RNAP sigma subunit interacts with the promoter DNA: domain σ_4 contacts the -35 element and domain σ_2 contacts the -10 element. The strength of binding directly correlates with the consensus nature of a promoter, and the higher the binding affinity, the higher is the frequency of initiation from a promoter (McClure, 1980; Mulligan *et al*., 1984). However, with the tight-binding promoters, abortive initiation becomes a central issue during the transcript initiation process (McClure *et al.*, 1978).

Abortive initiation refers to the repetitive synthesis and release of short nascent RNAs by RNAP. It is a key process during initiation and its amplification can significantly decrease the amount of productive mRNA. Abortive initiation occurs repeatedly as reflection of failed attempts of the polymerase to transition to the elongation phase. The initiation to elongation transition by RNAP has been dubbed the promoter escape step. Both the promoter recognition region (PRR) spanning -60 to -1 and the downstream region (DSR) spanning +1 to +20 have been shown to control the amount and length of abortive RNAs (Carpousis *et al.*, 1982; Knaus and Bujard, 1990; Kammerer *et al.*, 1986). Early work on the *lac*UV5 promoter by Carpousis *et al.* showed that the closer the -35 and -10 hexamers are to the consensus sequence, the worse is promoter escape to synthesize full-length RNA (Carpousis *et al.*, 1982).

Additional work with this promoter showed that it undergoes more abortive synthesis than the highly productive T7 A1 promoter. Thus, this study gave the first evidence that the strength of the DNA promoter: RNAP contacts is involved in promoter escape and synthesis of abortive transcripts.

Work by Knaus and Bujard attempted to uncover why the T5 *N25* promoter was so successful in transitioning to the elongation phase. Changing several regions of the promoter allowed them to discover that differences in the downstream region (DSR), which spans from +1 to +20, rendered the promoter weaker in escape (Knaus and Bujard, 1990; Kammerer *et el.*, 1986). Specifically, creating the "anti" version of the DSR (changing A↔C and G↔T) made the promoter 10 times weaker in the synthesis of full-length mRNA without affecting the ability of the initiating DNA promoter: RNAP contacts. Additional work showed that even though T5 *N25* can escape efficiently, it nevertheless produces a very robust level of abortive RNAs ranging in size from 2-11 nt prior to escape. With T5 $N25_{antiDSR}$ promoter, the abortive ladder lengthened to 15 nt and the level of full-length RNA was greatly diminished (Hsu *et al.*, 2003). Never before had abortive initiation been found to produce long transcripts beyond 10 nt. Both work performed by Carpousis *et al.* (1982) and Knaus and Bujard (1990), and later Vo *et al*. (2003), demonstrated that both the upstream promoter recognition region (PRR) and the downstream initial transcribed sequence (ITS) region play key roles in controlling abortive RNA synthesis and promoter escape. The PRR

sets the initiation frequency and the ITS determines the abortive pattern as well as the extent of productive versus abortive synthesis (Hsu *et al*., 2006).

The enigmatic nature of the abortive initiation phenomenon prompted many investigations into a possible mechanism for the synthesis and release of abortive RNAs. Straney and Crothers (1987) proposed that RNAP bound at the *lac*UV5 promoter, upon initiating transcription but unable to escape, forms a highly stressed intermediate which they called the initial transcribing complex (ITC). To relieve the stress, the ITCs can collapse back to the open complex conformation, release the abortive RNAs, and reinitiate, leading to repetitive cycles of abortive initiation (Carpousis and Gralla, 1985). The collapse of the ITC was found to stem from RNAP backward translocation (i.e. backtracking), resulting in transcriptional arrest (Kommissarova and Kashlev, 1997). In an arrested complex, the 3′-OH of the nascent RNA has become disengaged from the RNAP active site and extruded into the RNAP secondary channel (Borukhov *et al.*, 1991; Opalka *et al.*, 2003). Interestingly, it was found that an arrested complex can be reactivated through transcript cleavage and subsequent re-elongation of the 5′ nascent RNA (Surratt *et al*., 1991). Here, two types of cleavages were discovered at different distances from the 3′ end; rescue from 2-3 nt and 10-11 nt in the 3′→5′ direction. This rescue was discovered to be mediated by two different transcription factors. The 2-3 nt 3′ terminal cleavage was found to be mediated by a factor called GreA (Borukhov *et al.*, 1992). Later, rescue of

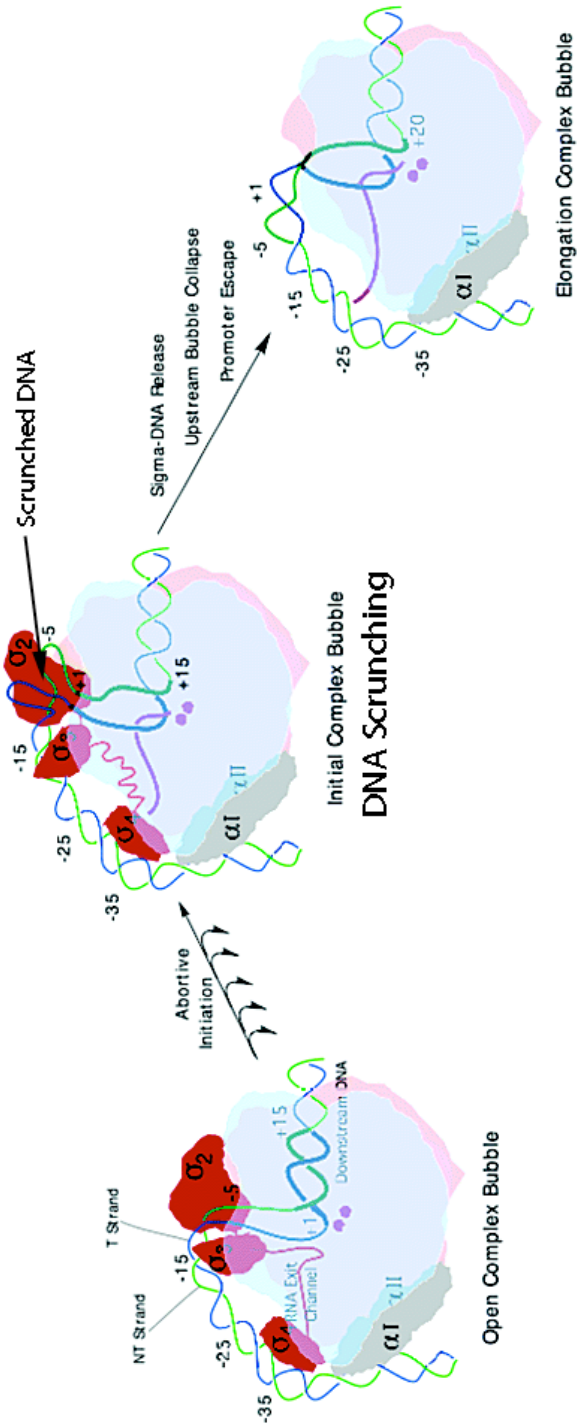
transcripts responsible for the more distal cleavages were found to be due to factor GreB (Borukhov *et al.*, 1993).

Overall, the importance of these factors is that the amount of abortive transcripts decreases significantly in the presence of Gre factors such as GreB and GreA (Feng *et al.*, 1994; Hsu *et al.*, 1995). These factors are transcription factors that augment the intrinsically low level of cleavage activity in RNAP and share a high degree of homology (Orlova *et al.*, 1995). GreA has been found to be more highly conserved and perhaps more biologically important, while GreB is more successful in rescuing long transcripts (Stepanova *et al.*, 2007; Hatoum and Roberts, 2008). GreB can directly attack an arrested complex very efficiently. Meanwhile, GreA does not easily cleave a transcript in an arrested complex, but instead prevents the arrest. The mechanism of Gre action involves a Gre factor binding in the secondary channel, stimulating the intrinsic hydrolytic activity of RNAP to cleave off the 3′-end that has extruded into the channel (Laptenko *et al.*, 2003; Opalka *et al.*, 2003). Their removal allows for the reintroduction of a new 3′-hydroxyl group at the active site, allowing the resumption of RNA transcript growth. Due to the requirement that the abortive transcripts have their 3′-hydroxyl group within the secondary channel, this fact also indicates that the transcripts arose through a process known as backtracking.

As previously mentioned abortive transcripts are the product of stressed intermediates from the initiation to elongation transition. During initiation, the melting open and expansion of the transcription bubble and subsequent growth of

the nascent RNA produces an unstable initial transcribing complex (RP_{itc}). Previous footprinting experiments identified that the upstream boundary of DNA was protected by RNAP in both RP_o and RP_{itc} (Straney and Crothers, 1987). Further experimentation revealed that during initiation RNAP remains stationary, maintaining promoter contacts, as it unwinds and pulls DNA inside and begins the growth of the nascent RNA (Revyakin *et al.*, 2006). This process is known as DNA scrunching and creates a lot of stress energy. If this stress energy is managed properly, the upstream edge of the transcription bubble rewinds, the RNAP loses contacts with the promoter, and transition to the elongation phase ensues (Figure 1). If this stress is not managed properly the downstream bubble can collapse, lodging the 3′-end of the abortive transcript in the secondary channel, through a process known as backtracking. Due to the location of the 3′-end in the secondary channel, the abortive transcript can be rescued by transcription factor GreB. Overall, DNA scrunching is the process that can account for the stress energy to bring about abortive initiation or the transition to the elongation phase.

Figure 1. *Transcription Initiation into Elongation: Open Complex (RP_o), Initial Transcribing Complex (RP_{itc}), Elongation Complex (TEC).* Following the formation of the open complex, the transcription bubble melts open an additional 6-8 bp and the nascent RNA begins to grow The bulk of the growing nascent RNA and transcription bubble builds tension within RNAP producing a phenomenon known as scrunching. Many attempts are made by RNAP to grow the nascent RNA to 8-10 nt, causing the production of small abortive RNA transcripts. After the RNA molecule has grown to 8-10 nucleotides (nt), promoter escape occurs, disrupting the initial DNA promoter contacts, so the elongation phase can ensue.


Scrunched DNA
Sigma-DNA Release
Upstream Bubble Collapse
Promoter Escape
Elongation Complex Bubble
Initial Complex Bubble
DNA Scrunching
Abortive Initiation
T Strand
NT Strand
RNA Exit Channel
Downstream DNA
Open Complex Bubble

Previous work in the Hsu laboratory attempted to characterize other T5 *N25* promoters that vary in the downstream region. They created about 10 different T5 *N25* promoters that only vary in the +3 to +10 region of DSR (Chander *et al.*, 2007). These studies led to the DG200-series promoters, which produce even longer abortive transcripts. Due to their unusual length, spanning 16-19 nt, they were named **v**ery **l**ong **a**bortive **t**ranscripts (VLATs). In particular promoter DG203 is most robust in producing VLATs. By replacing the 17-bp spacer between the -35 and -10 elements with a spacer sequence associated with a full consensus sequence obtained through *in vitro* selection (Gaal *et al*., 2001), no full-length transcripts are produced and only abortive transcripts are made. This ideal VLAT-producing promoter is called *DG203*/*SPfullcon-U2/U19* (Figure 2).

Previous to work done in the Hsu lab, backtracking was thought to be the only mechanism of producing abortive transcripts. Nevertheless, striking transcription reactions on the DG203 promoter in the presence of GreB indicate that transcripts less than 15 nt are rescued, but those 16-19 nt are not (Chander *et al.*, 2007). This evidence indicated that the 3′-hydroxyl group of VLATs were not ending up in the secondary channel, but instead, it was hypothesized to be in the RNA exit channel. Therefore, the production of VLATs had to be created through a different mechanism other than backtracking. Forward hyper-translocation has been postulated to be the mechanism of VLAT formation. This mechanism proposes that extremely tight promoter: RNAP contacts cause even

more DNA scrunching and stress-energy so that the RNAP is propelled downstream by several nucleotides when the upstream edge of the bubble rewinds and promoter escape occurs. This process causes the 3′-hydroxyl group of the growing transcript to be out of register with the active site. The 3′-hydroxyl group is now presume to lie within the RNA exit channel and cannot be rescued by GreB (Figure 3).

Although VLATs were shown to be GreB resistant, previous work in the Hsu lab has determined that the intensity of the VLATs in the presence of GreB decreased, suggesting some VLATs are produced through backtracking and lodging in the secondary channel. After quantifying this data they discovered that of the VLAT producing promoters: 17% of VLATs produced from DG203 hyperforward translocate; meanwhile 55% of VLATs produced from *DG203*/*SPfullcon* hyperforward translocate (Chander and Hsu, unpublished results). The remaining VLATs backtrack, thus GreB should be present in order to ensure RNAP only produces VLATs produced from forward hyper-translocation. The development of backtracking-resistant RNAP mutant RNAP rpoβ*35 (βH1244Q) also provided us an alternative RNAP to study the forward hyper-translocation mechanism (Dutta *et al.*, 2011). This RNAP is characterized by an amino acid replacement of histidine to a glutamine on the 1244th amino acid of the β subunit. This mutant RNAP could be studied without GreB since it intrinsically does not produce backtracked transcripts.

Figure 2. *Consensus, T5 N25, N25$_{antiDSR}$, DG203, DG203/SPfullcon Promoters.* Consensus promoter and T5 *N25* differ in the -35 element. *N25$_{antiDSR}$,* and DG203 differs from T5 *N25* in the initial transcribed sequence (ITS). DG203 differs from DG203/SPfullcon at the 17-bp spacer. RNAP makes contacts with the promoter at -35 element and -10 element with σ_4 and σ_2 domains, respectively.

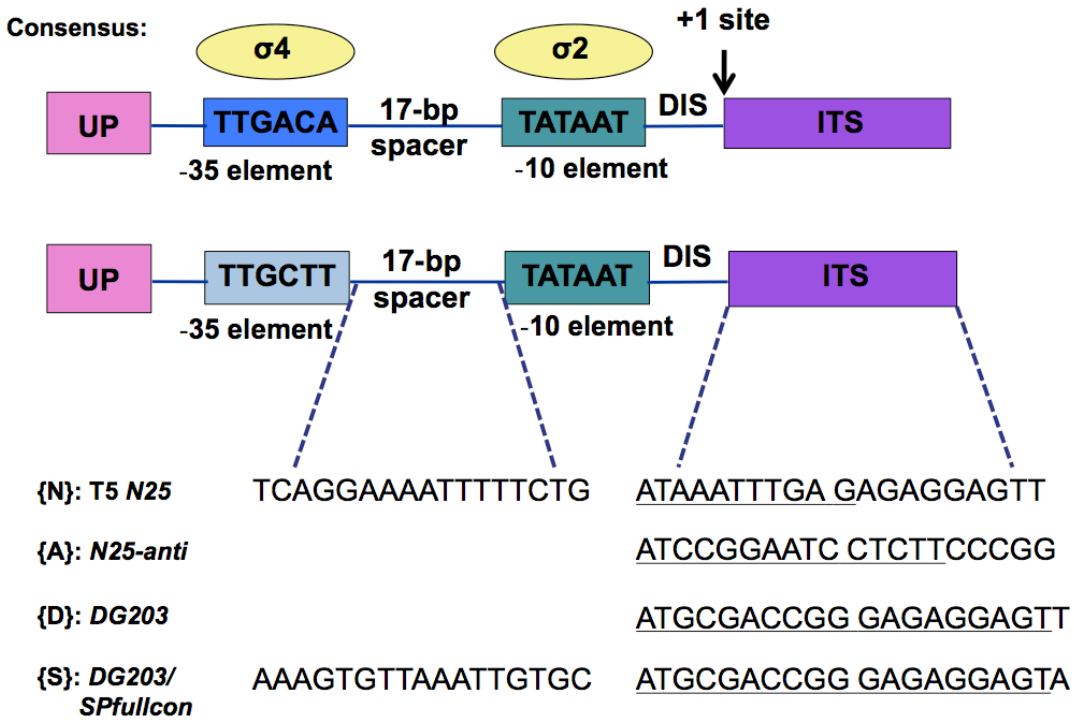
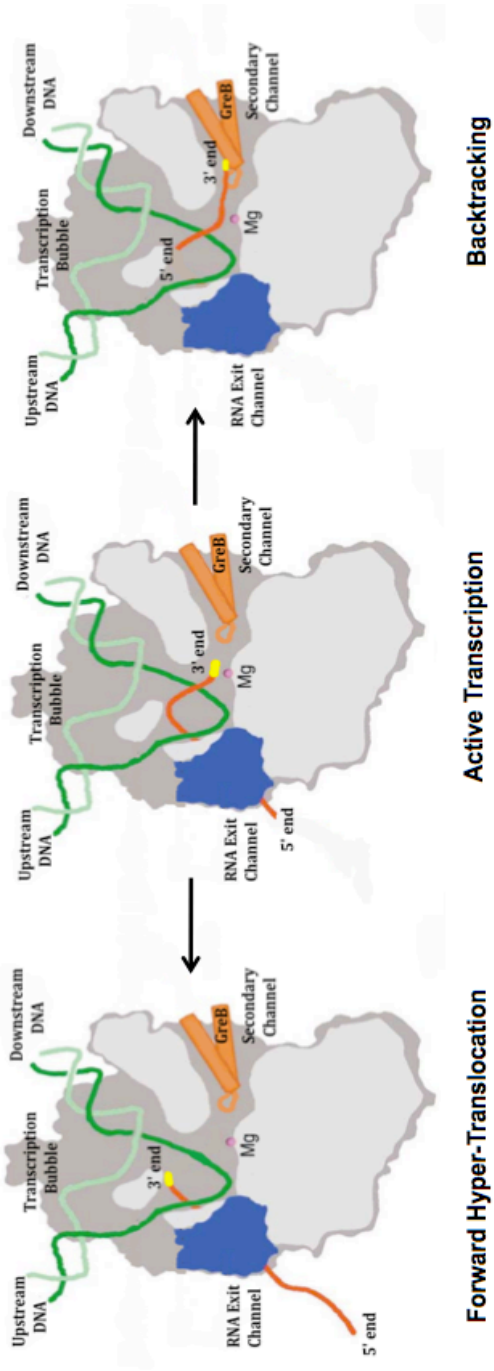

Consensus:
σ4
σ2
+1 site
UP
TTGACA
17-bp spacer
TATAAT
DIS
ITS
-35 element
-10 element
UP
TTGCTT
17-bp spacer
TATAAT
DIS
ITS
-35 element
-10 element
{N}: T5 N25
TCAGGAAAATTTTTCTG
ATAAATTTGA GAGAGGAGTT
{A}: N25-anti
ATCCGGAATC CTCTTCCCGG
{D}: DG203
ATGCGACCGG GAGAGGAGTT
{S}: DG203/ SPfullcon
AAAGTGTTAAATTGTGC
ATGCGACCGG GAGAGGAGTA

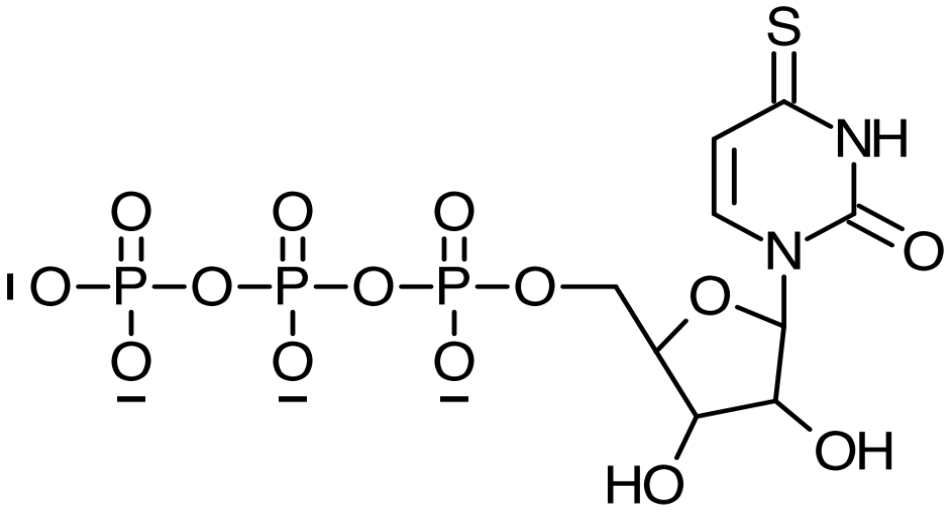
Figure 3. *Mechanisms of Abortive Initiation.* Active transcription is characterized by a 3′-OH end aligned at the active site, growing the nascent RNA. Backtracking occurs when the downstream bubble collapses, lodging the 3′ end of the nascent RNA into the secondary channel. Forward hyper-translocation is characterized by the lodging of the 3′-OH end in the RNA exit channel. Image adapted from Opalka *et al.* (2003).


Downstream DNA
Secondary Channel
GreB
3' end
Transcription Bubble
Mg
5' end
Upstream DNA
RNA Exit Channel
Backtracking
Active Transcription
Forward Hyper-Translocation

In order to demonstrate that the 3′-hydroxyl group of GreB-resistant VLATs ends up in the RNA exit channel, I plan on conducting a crosslinking study. In this study I plan to substitute the very last nucleotide on a VLAT with the photocrosslinkable nucleotide 4-thiouridine (S_4U) (Figure 4). In the presence of longwavelength UV (365 nm), this nucleotide will form a zero angstrom crosslink with the closest amino acid(s). In order to ensure that the S_4U will attach at either the 17^{th},18^{th} or 19^{th} position of VLATs, special transcription features were adopted. All of the *DG203/SPfullcon* templates we are working with have only two U's in the ITS, one at the 2^{nd} position and another at the 3′-end of the VLAT. To ensure that the S_4U is added at the 3′-most nucleotide of the VLAT, I will initiate the transcription reaction with an ApU dinucleotide. The concentration of ApU will be set 10 times higher than the concentration of regular ATP, thus ensuring that the majority of the VLATs are initiated with ApU, which in turn, allows the photocrosslinkable S_4U to be incorporated at the 3′ end.

The goal of my research will be to first demonstrate the occurrence of crosslinking and, if it exists, characterize its specificity. Next, I shall localize the VLAT crosslink to the RNAP subunit(s) through SDS-PAGE fractionation. Finally, I will perform in-gel digestion using RNases and trypsin to cut down the crosslinked product and mass spectrometry analysis to determine the exact amino acid the VLAT crosslinks to.

Figure 4. *Photocrosslinkable Nucleotide 4-Thiouridine Triphosphate (S_4UTP).* This S_4UTP is incorporated at the 3′ end of a VLAT. Under longwavelength UV irradiation (365 nm) it forms a zero-angstrom crosslink to the closest amino acid.


S
NH
O
N
O
O
O
O
O
O
O
O
O
O
O
O
O
HO
OH

MATERIALS AND METHODS

MATERIALS

Oligonucleotide primers and promoters

The primers for the DG203/SPfullcon templates were obtained from Integrated DNA Technologies (IDT). The DG203/*SPfullcon* upstream primer, designated as N25-u SPfullcon (SSX1), extends from -75 to -1 and has sequence 5′-AAG GCC ACC TAG GCC TCG AGG GAA ATC ATA AAA AAT TTA TTT GCT TAA AGT GTT AAA TTG TGC TAT AAT AGA TTC-3′. The DG203/*SPfullcon-U2/U19* downstream primer extends 5’ to 3’ from +57 to -26 and has the sequence 5′-TTC TGC GGC CGC TTT CTG CGT GGT CCT GCC TTG TTT GTA CTC CTC TCC CGG TCG CAT GAA TCT ATT ATA GCA CAA TTT AAC AC-3′; the letter highlighted in red corresponds to the +19 position. Likewise, the DG203/*SPfullcon-U2/U18* downstream primer has the sequence 5′-TTC TGC GGC CGC TTT CTG CGT GGT CCT GCC TTG TTT GTC ATC CTC TCC CGG TCG CAT GAA TCT ATT ATA GCA CAA TTT AAC AC-3′; and the DG203/*SPfullcon-U2/U17* downstream primer has the sequence 5′-TTC TGC GGC CGC TTT CTG CGT GGT CCT GCC TTG TTT GTC CAC CTC TCC CGG TCG CAT GAA TCT ATT ATA GCA CAA TTT AAC AC-3′. Promoters T5 N25 (FL=57), T5 $N25_{antiDSR}$ (FL=57), P_L (FL=68),

*lac*UV5 (FL=58), T7 A1 (FL=67), and *rrn*BP1$_{dis}$ (FL= 57) were provided by L. M. Hsu.

Reagents

ApU dinucleotide used for initiating the transcription reactions was obtained from Ribomed. The modified nucleotides 4-thiouridine triphosphate (S_4UTP), 2-thiouridine triphosphate (S_2UTP), 5-bromouridine triphophate (Br_5UTP), and 5-iodouridine triphosphate (I_5UTP) were obtained from TriLink BioTechnologies. Radioactive nucleotides [α- or γ-^{32}P]-NTPs were obtained from Perkin Elmer.

In primer extension, the agarose was purchased from Sigma-Aldrich, the 100 bp DNA ladder was purchased from New England Biolab (NEB), isopropanol, chloroform, isoamyl alcohol, and ethidium bromide (EtBr) were obtained from Fischer Scientific. Phenol/chloroform/isoamyl alcohol (25:24:1) mix and ultrapure dNTPs were purchased from US Biochemicals/Affymetrix. Chloroform/isoamyl alcohol (24:1) mix was prepared in the laboratory.

In transcription experiments, the 40% acrylamide-bisacrylamide (10:1) stock was prepared by L. Hsu. For SDS-PAGE gels, 40% acrylamide-bisacrylamide (37:5:1) stock, and N, N, N′, N′–tetramethylethylene diamine (TEMED) were obtained from Fisher Scientific. Ammonium presulfate (APS) was purchased from US Biochemicals/Affymetrix. 3X SDS-PAGE red loading

buffer and ColorPlus protein ladder (10-230 kDa) were obtained from NEB. Coomassie brilliant blue dye R-250 was purchased from Sigma-Aldrich.

In trichloroacetic acid (TCA) precipitation, TCA and the yeast tRNA carrier nucleic acid were obtained from Sigma-Aldrich. In-Gel Tryptic Digestion Kit was purchased from Pierce through Thermo Scientific.

Buffers

For primer extension, we used NEBuffer 2, which consists of 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT) at pH 7.9. Oligonucleotides were suspended in TE buffer, which consists of 10 mM Tris-HCl, 1 mM Na_2EDTA at pH 8.0. Transcription reactions were routinely performed in transcription buffer III (10), which consists of 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 10 mM β-mercaptoethanol (β-ME), 10 μg/mL acetylated bovine serum albumin (BSA). GES mix used to terminate a transcription reaction contains 1 mg/mL nuclease-free glycogen, 10 mM Na_2EDTA, and 0.3 M NaAc in DEPC-treated water. FLB (formamide loading buffer) is composed of 80% deionized formamide, 1X TBE buffer, 10 mM Na_2EDTA, 0.2% xylene cyanol, and 0.2% amaranth. Transcription gel electrophoresis requires different top- and bottom-reservoir buffers; the top buffer is 1X TBE and the bottom buffer, 1X TBE containing 0.3 M NaAc. 10X TBE buffer stock is made up, per liter, by dissolving 108 g Trisma Base, 55 g boric acid, 8.8 g Na_2EDTA in warmed deionized distilled water (ddH_2O). SDS-PAGE buffer contains, per liter, 14.4 g

glycine, 3.03 g Trisma Base, 1 g SDS in ddH_2O. For staining and destaining the SDS-PAGE gels, the Coomassie Staining Solution contains 0.5 g of Coomassie Blue R-250 (Sigma), 250 mL of isopropanol, 100 mL of glacial acetic acid, brought up to 1 L with ddH_2O. The protein gel destaining solution is 7.5% acetic acid in 45% methanol. It is made, per liter, by mixing 450 mL methanol, 75 mL glacial acetic acid, brought up to 1 L with ddH_2O. Agarose gel electrophoresis was carried out in 1X TAE buffer containing 0.5 μg/mL EtBr. 50X TAE buffer stock contains, per liter, 242 g Tris Base, 54.1 mL glacial acetic acid, 18.6 g Na_2EDTA in ddH_2O. RNAP storage buffer consists of 50% glycerol, 40 mM KPO_4 (pH 8), 0.1 mM Na_2EDTA (pH 8), 1 mM DTT and 200 mM KCl. For in-gel digestion, destaining solution consists of 80 mg ammonium bicarbonate, 20 mL acetonitrile, and 20 mL ddH_2O; digestion buffer consists of 10 mg ammonium bicarbonate and 5 mL ddH_2O (final concentration ~25 mM in ammonium bicarbonate); reducing buffer consists of 3.3 μL tris(2-carboxyethyl)phosphine (TCEP) and 30 μL digestion buffer per digest performed (final concentration ~50 mM); 5X alkylation buffer consists of 7 mg iodoacetamide (IAA) and 70 μL ddH_2O (final concentration ~500 mM in IAA); 1X alkylation buffer consists of 7 μL 5X alkylation buffer and 28 μL digestion buffer, per digest; trypsin working solution consists of trypsin stock and 45 μL ddH_2O (~ 0.44 μg/μL final concentration), activated trypsin solution consists of 1 μL trypsin working solution and 9 μL digestion buffer for each sample processed (~10 ng/μL final concentration).

Enzymes

E. coli RNA polymerase holoenzyme PC-16 (3.4 μM) and PC-17 (3.2 μM) were prepared by L. Hsu, with the assistance of T. Vallery and M. Thandar, during the summer of 2010. Mutant *E. coli* RNA polymerase rpoB*35 (βH1244Q) (5 mg/mL) was generously provided by Evgeny Nudler (New York University School of Medicine). Klenow polymerase (5,000 U/mL) used in primer extension, micrococcal nuclease (MNase) (2,000 gel units/μL), and TPCK-treated trypsin (20 μg) were obtained from NEB.

METHODS

Primer Extension

Primer extension was performed to make all single-promoter DNA templates. For each promoter preparation, two 100-μL reactions were initially created. First the annealing mix was made, consisting of 10 μL upstream primer (10 μM), 10 μL downstream primer (10 μM), 5 μL 10X NEBuffer 2, and 25 μL of deionized distilled water (ddH_2O), totaling 50 μL. Both 50 μL reactions were placed in the annealing cycle of 20 min at 70 °C, 10 min at 55 °C, 10 min at 42 °C, and 10 min at 37 °C. Meanwhile, the dNTP mixture was prepared, consisting of 5 μL 2.5 mM dNTP, 5 μL 10X NEBuffer 2, and 40 μL ddH_2O, totaling 50 μL.

This mixture was added to the annealing reaction, and 3 μL of Klenow polymerase (5,000 units/mL) was added to each 100-μL reactions and incubated for 30 min at 37 °C. The enzyme was heat killed for 20 min at 75 °C. Then, 3 μL of each 100-μL reaction was analyzed in 2% agarose gel to see if the product was made. After ensuring the product was present, the two 100-μL reactions were pooled, mixed with 25 μL 3 M NaAc and 700 μL EtOH to concentrate the DNA by ethanol precipitation at -20 °C overnight.

The ethanol pellet was collected by microcentrifugation at 13,400 rpm for 20 min at 4 °C, redissolved in 100 μL of TE buffer, mixed with 100 μL 10 M NH_4Ac and 600 μL EtOH. This mixture was left for 2 hours at room temperature, then spun for 25 min at room temperature. The supernatant was removed and saved in case not enough DNA was recovered. The pellet was dried in a SpeedVac rotary desiccator for 15 min, redissolved in 200 μL TE buffer, and extracted with 200 μL phenol-chloroform-isoamyl alcohol (25:24:1) mixture for 1 min and spun for 2 min to separate the phases. The aqueous top layer was transferred to a new tube and extraction with phenol-chloroform isoamyl alcohol (25:24:1) was repeated one more time as described above. The aqueous top layer was next extracted with 200 μL chloroform-isoamyl alcohol (24:1) to remove any residual phenol. After recovering the final top aqueous layer, 25 μL 3 M NaAc and 700 μL EtOH were added and mixed to precipitate the DNA. The contents were left overnight at -20 °C and spun for 20 min at 4 °C to collect the pellet. The pellet was washed with 70% EtOH, spun for 5 min and the EtOH decanted. The

pellet was dried using a SpeedVac rotary vacuum desiccator for 15 min and dissolved in 200 μL TE buffer. A 2% agarose gel was performed with 3 μL of the DNA to verify the size of the product. The concentration of the DNA was determined by 260 nm absorbance in a Nanodrop spectrophotometer.

Transcription Reactions

The transcription reactions were performed at 37 °C. The final volume of the reactions was 10 μL. A typical reaction contained 30 nM template DNA, 1X transcription buffer III(10), 50 mM KCl, various concentrations of NTPs (with either [γ-^{32}P]-ATP or [α-^{32}P]-NTP label added to a specific activity of ~10 cpm/fmol), 50 nM RNAP, and if supplemented, 500 nM GreB. To carry out the reactions, all of the reagents were assembled first. Then, reaction was started by adding either RNAP alone or RNAP+GreB (1:10) mix in a 30-second stagger and incubated in a 37 ^{o}C water bath for a specified length of time. The reactions were next terminated by the addition, in 30-sec stagger, of 100 μL GES mix and 330 μL EtOH. The mixture was pipetted up and down a few times to mix, then placed in the -80 °C for 2 hours or -20 °C overnight to precipitate the nucleic acids. The ethanol pellets were recovered by microcentrifugation (at 13,400 rpm) for 20 min at 4 °C. The (radioactive) EtOH supernatant was carefully removed and the pellet was dried in the speedvac for 15 min. Finally, each pellet was resuspended in 10 μL of freshly prepared FLB gel loading dye, vortexed and spun 3 times to

thoroughly redissolve the samples, and 4 μL was loaded onto the transcription gels for analysis. The remaining samples were stored in a -20 °C freezer.

The transcription gels are 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea. The preparation of the gels consists of adding the urea, 10x TBE, and 40% (10:1) acrylamide-bis stock together and placing the contents at 37 °C overnight or in a 37 °C water bath until clear. The content was cooled to room temperature and the 10% APS and TEMED were added right before pouring to initiate the polymerization reaction. Narrow gels are poured from a 40-mL mixture made by dissolving 16.8 g urea in 4 mL 10X TBE and 23 mL 40% (10:1) acrylamide-bisacrylamide stock; polymerization was initiated by adding 200 μL 10% APS and 20 μL TEMED. The wide gels require doubling the narrow gel recipe. The gel was poured quickly and allowed to polymerize for 30-40 min at room temperature. After setting the polymerized gel into an electrophoresis apparatus and outfitting the top and bottom reservoir with buffer, the gel is usually pre-run with fresh FLB in all of the wells for 10-30 min before the samples were loaded. The actual gel electrophoresis at 35W required 4 hours for the narrow gels and 5 hours for the wide gels.

Photocrosslinking Reactions

All photocrosslinking experiments were performed in microtiter plates placed on a 37 °C aluminum block. The first reactions were performed in a 10 μL volume. Later reactions were performed in 40- and 100-μL volumes for obtaining

more of the crosslinked material. Overall, final concentrations of the reactions consisted of 30 nM DNA, 1X transcription buffer III(10), 50 mM KCl, 500 μM ApU, 10 μM each of GTP, CTP, and ATP (G/C/ATP mix), 150 nM RNAP or 150 nM RNAP premixed with 1.5 mM GreB, and 100 μM S_4UTP, or S_2UTP, or Br_5UTP, or I_5UTP. Depending on the experiment, the transcription reaction proceeded for various times including 20 minutes or an hour either in subdued light or under long wavelength UV irradiation (365 nm or 302 nm). Following the reactions, 10 μL or 100 μL aliquots were taken from the microtiter plates and transferred to eppendorf tubes for processing.

After photocrosslinking, the reactions were to be analyzed in SDS-PAGE gels where each well can accommodate only 15 μL of sample volume. With 10 μL reactions, they can be loaded directly onto the SDS-PAGE gel after mixing with 5 μL of a 3X SDS-PAGE sample buffer containing DTT. However, samples with large reaction volumes (~100 μL) require prior concentration before the products can be loaded onto a SDS-PAGE gel for analysis. Here, two precipitation procedures were compared to determine 1) if a crosslinked product was produced, and 2) whether the crosslinked product is stable to the precipitation treatment. The more successful method I tried involves TCA precipitation. The reactions on the 37 °C plates were transferred to new eppendorf tubes on ice, each added with 50 μL YEP (50 mM NaPPi, 50 mM Na_2EDTA, 0.5 mg/mL yeast tRNA) and 300 μL 10% TCA. This solution is mixed, kept on ice for 10 min, and spun for 15 min at 4 °C to collect the precipitate. After removing the supernatant,

the pellet is washed with 300 μL acetone to remove residual TCA. The pellet is allowed to dry in a SpeedVac rotary vacuum desiccator for 15 min, redissolved in 15 μL 1X SDS-PAGE sample buffer with DTT, gently tapped and spun down.

Another precipitation protocol utilizes the GES/EtOH mix we routinely use to precipitate small RNAs. In this procedure, the reactions on the 37 °C plates were transfered to new eppendorf tubes, each added with 100 μL GES mix and 330 μL EtOH, and after mixing, were allowed to precipitate at -80 °C for 2 hours. The pellets were collected by microcentrifugation at 4 ^{o}C for 15 min, and dried in a speedvac for 15 min. Each pellet was resuspended in 15 μL 1X SDS-PAGE sample buffer with DTT, vortexed and spun 3X for complete resuspension.

Following the precipitation treatment but before loading onto the SDS-PAGE gels, the samples were heated at 95 °C for 8 min to denature the proteins. These samples were then loaded onto a SDS-PAGE gel along with two molecular size standards, one containing RNAP only and the other, 15 μL of the ColorPlus protein ladder from NEB. The RNAP-only standard was prepared using 1 μL *E. coli* RNAP (3.4 μM), 5 μL 3X SDS-PAGE sample buffer with DTT, and 9 μL DEPC H_2O.

The SDS-PAGE gel consists of an 8% resolving bottom gel and a 4% stacking top gel. The 8% resolving gel solution was made by mixing 2 mL of 40% (37.5:1) acrylamide-bis stock, 2.5 mL of 4X resolving gel buffer (1.5 M Tris-HCl, pH 8.8, 0.4% SDS), 5.5 mL ddH_2O, 50 μL 10% APS, and 10 μL TEMED. The 4% stacking gel solution was made by mixing 0.5 mL of 40% (37.5:1)

acrylamide-bis stock, 1.25 mL of 4X stacking gel buffer (0.5 M Tris-HCl, pH 6.8, 0.4% SDS), 3.25 mL ddH_2O, 37.5 μL 10% APS, and 7.5 μL TEMED. During the SDS-PAGE gel run, the samples were electrophoresed through the stacking gel at 100 V for 10 min, then through the resolving gel at 150 V for 30 min until the 10 kDa band in the NEB ColorPlus protein ladder reached the bottom of the gel. The gel was removed and stained in Coomassie Blue dye for 10 min and destained in destaining buffer. Many washes of the destaining buffer were made until the protein bands show up distinctly. The gel was then placed onto Whatman 3MM paper and dried at 80 °C under vacuum for 40 min in a BioRad Model 583 gel dryer. The RNAP control was marked with 0.2 μL of two-month-old radioactive nucleotides and allowed to air-dry. The dried gel was then exposed to a phosphor screen overnight and scanned the next morning in a GE Storm 820 phosphorimager.

In Gel Digestion and Purification

In gel digestion was carried out to reduce the size of the crosslinked products and remove the radioactivity associated with the RNA at the same time, for mass spectrometry analysis. In this process the small molecular weight (~13 kDa) endo-exonuclease micrococcal nuclease (MNase) and endonuclease ribonuclease RNase T1 (~11 kDa) were used to remove the radioactivity and truncating the VLAT RNA associated with the protein. Following this digestion, trypsin was used to reduce the size of the crosslinked β/β′ subunit.

In this procedure, after the SDS-PAGE gel has been destained so the protein bands are visible and read on a phosphoimager, the crosslinked product band of interest was excised using a razor blade and placed into 200 μL destaining buffer. The sample was incubated at 37 °C for 30 min with shaking. The destaining solution was removed and discarded. This destaining process was repeated once more.

This process was followed by reduction and alkylation procedure to remove the disulfide bonds irreversibly. The reduction buffer was prepared immediately prior to use; 30 μL was added to each sample and the samples were incubated at 60 °C for 10 min. After cooling, the reducing buffer was removed and discarded. Next, the alkylation buffer was prepared and 30 μL was added to each sample. The samples were incubated in the dark at room temperature for 1 hour, after which the alkylation buffer was removed and discarded. The samples were then washed with 200 μL destaining buffer and incubated at 37 °C for 15 min with shaking. Afterwards, the destaining buffer was removed and discarded, and the destaining procedure was repeated once more.

Following reduction-alkylation, VLAT digestion was investigated through the use of MNase. Here, the gel samples were dried using 50 μL of acetonitrile and incubated for 15 min at room temperature. After removing the acetonitrile, the samples were air dried for 10 min; upon drying, the gel pieces turned opaque. Next, 100 μL of freshly prepared reaction mixture (85 μL ddH_2O, 10 μL MNase buffer, 5 μL MNase stock) was added and the gels were incubated at 37 °C with

shaking overnight. The supernatant was removed and saved for scintillation counting.

The gel was then dried using 50 μL acetonitrile for 15 min at room temperature. After removing the acetonitrile and allowing the samples to air-dry for 10 min, 30 μL of the just-prepared activated trypsin solution (3 μL trypsin working solution and 27 μL digestion buffer [final trypsin concentration ~10 ng/μL]) was added per sample and the gel was allowed to swell for 15 min at room temperature. Then, 25 μL of the trypsin digestion buffer was added to each sample and the gel pieces were incubated at 37 °C for either 4 hours or at 30 °C overnight with shaking. The digestion mixture was removed for scintillation counting or for electrophoretic analysis in a narrow transcription gel.

RESULTS

The aim of this project is to photocrosslink the 3′-OH of a VLAT to the nearest amino acid(s) on RNAP and determine its location. In the presence of cleavage factor GreB, the GreB-resistant VLATs (16-19 nt) are hypothesized to have arisen from hyper forward translocation as promoter escape ensues. This translocation mechanism would move the 3′-OH end of the nascent RNA to the RNA exit channel which is lined by the β/β′ subunits (Murakami *et al*., 2002). Thus, we expect the photocrosslinking from the 3′-most residue of a VLAT to occur with the β/β′ residues.

To investigate this goal, it was necessary to demonstrate the feasibility of incorporating a single photocrosslinkable nucleotide at the 3′-most position of a VLAT RNA. Once that has been shown, we can pursue the photocrosslinking experiments. Several preliminary transcription reactions were performed to ensure that the incorporation of this photocrosslinkable nucleotide occurred and did not modify the process of transcription initiation.

Preliminary Transcription Reactions

Our study was performed with the VLAT-producing promoter DNA templates, including: DG203, DG203/*SPfullcon-U2/U19*, DG203/*SPfullcon-U2/U18*, and DG203/*SPfullcon-U2/U17*. DG203 and DG203/*SPfullcon-U2/U19*

have the identical initial transcribed sequence of A_{+1}UGCGACCGGGAGAG**GAGU**$_{+19}$; the VLAT positions (+16 to +19) are highlighted in bold. Because there are only two U-residues in this region, if we initiate transcription with the ApU dinucleotide, and transcribe in the presence of GTP, CTP and ATP (with one of them carrying the α-^{32}P label), we would obtain an 18-nt transcript. If we then supplement the reaction with S_4UTP, we can incorporate the photocrosslinkable probe specifically into the 19th position. Later, when we wished to investigate photocrosslinking from VLAT-18 and VLAT-17, we simply generated an ITS variant where the second U was moved to the 18th or 17th position, respectively.

With ApU initiation of in vitro transcription, two issues arise that can confound the transcript patterns we observe. First, the RNAs generated lack the 5′-triphosphate group of an ATP-initiated transcript and therefore migrate in denaturing polyacrylamide gels with an anomalous mobility compared to the ATP-initiated transcripts (L. Hsu, personal communication). Second, with ApU initiation, we have no choice but to label the transcripts with [α-^{32}P]-NMP; α-^{32}P labeling in turn highlights the existence of RNAP 3′-cleavage products (Surratt *et al*., 1991), which as 5′-monophosphorylated RNAs, also migrate differently than the 5′-triphosphorylated ATP-initiated transcripts (Hsu *et al*., 2006).

With these complexities in mind, I performed the first set of ApU initiated transcription reactions. The gel results (not shown) were indeed too difficult to interpret. Subsequently, to identify the bands in the ApU initated reaction, I

compared ApU- versus ATP-initiated transcription, with different [α-^{32}P]-NMP labeling, and the effect of the absence or presence of GreB in the reaction. The result is shown in Figure 5.

In Figure 5, the abortive RNAs of the size 19, 17 and 3 nt are labeled. It is immediately apparent that ApU-initiated 3-nt RNA (i.e. ApUpG) travels much more slowly through the gel than ATP-initiated 3-mer (i.e. pppApUpG). This slowness is due to the absence of four negative charges which gives rise to a much lower charge density for ApUpG compared to pppApUpG. This charge density differential persists through the shorter abortive RNAs until the length is around 17 nucleotides. The line across the gel indicates the point at which the mobility of the RNAs with different initiating nucleotides aligned. With this alignment, it was possible to work our way backwards to identify the smaller RNAs. Even with this knowledge, the identity of some bands, particularly the small abortive RNAs, were ambiguous. The identity of short RNAs is made especially difficult due to the presence of cleavage products—indicated by asterisks—that arise from the intrinsic cleavage activity of RNAP (Orlova *et al*., 1995) stimulated by the presence of GreB, and [α-^{32}P]-NMP labeling of the transcripts (Hsu *et al*., 2003).

In order to further resolve the identity of the small abortive RNA bands, several additional transcription reactions were performed with ApU initiation, but limited nucleoside triphosphates to allow partial elongation, and with differential [α-^{32}P] labeling. The gel results are shown in Figures 6 and 7.

Oddly in both gels, the 4-nt abortive RNA migrated faster than the 3-mer. The 5-nt spot seems to run just above the 3-mer, but not very well separated from it. Such an anomalous migration pattern is not common but had been observed before (L. Hsu, personal communication). Given these assignments, we were able to deduce the full ladder bands in lane 7 of Figure 6, containing labeled products from 6 nt up to 18 nt. However, even without GreB, we suspect the presence of some naturally occurring cleavage products (*). Having resolved the complex pattern of the abortive ladders, I could pursue the incorporation of a photocrosslinkable probe at the various VLAT positions.

For crosslinking studies, the incorporation of a photocrosslinkable nucleotide at the 3'-most position of a VLAT RNA is necessary. I performed a transcription reaction with *DG203/SPfullcon-U2U19* template by initiating with ApU, elongating in a G/C/ATP mix to the 18^{th} position, and then allowing the incorporation of either a regular UMP or a S_4UMP at the 19^{th} position. The result is shown in Figure 8. This gel shows that the incorporation of S_4UMP was successful and comparable to that of UMP. Thus, the 19^{th} position can be "labeled" specifically with a photocrosslinkable probe.

In order to recover more of the crosslinked material, I developed reaction conditions to amplify the amount of VLAT-19. I performed a reaction that varied the concentration of the photocrosslinkable nucleotide S_4UTP and the concentration of RNAP and RNAP/GreB, respectively (Figure 9). ImageQuant analysis showed that most VLAT-19 product was obtained when RNAP:DNA

ratio was tripled (final [RNAP] of 150 nM) and the S_4UTP concentration boosted from 10 to 100 μM. These conditions were used in all subsequent transcription-crosslinking experiments.

Although higher concentrations of S_4UTP and RNAP can raise the amount of VLAT-19 produced, its level is always lower than the amount of VLAT-18 and VLAT-17 synthesized in the same reaction (see GreB-containing lanes in Figure 9). This raised the possibility that we can perform photocrosslinking from the 18^{th} or 17^{th} position on a VLAT RNA with greater efficiency. To do this, I ordered new primers to construct the *DG203*/*SPfullcon-U2/U18* and -*U2/U17* templates and performed similar ApU-initiated transcription reactions to incorporate the S_4UMP at the new 3′-most positions. The results in Figure 10 indicate the experiment was successful, and that we can generate higher quantities of the 3′-S_4UMP labeled VLAT-18 and VLAT-17 RNA.

Figure 5: *DG203/SPfullcon Transcription Comparing ApU vs ATP Initiation, Differential [α-^{32}P]-NMP Labeling, and the Effect of GreB*. Each reaction contained the following final concentration of reagents: 30 nM DNA *DG203/SPfullcon-U2/19*, 1X transcription buffer III(10), 50 mM KCl, and a nucleotide mixture of either 500 μM ApU + 10 μM NTP or 100 μM ATP + 10 μM G/C/UTP. The radioactive labeling nucleotides were [α-^{32}P]-GTP, [α-^{32}P]-ATP, or [α-^{32}P]-CTP used individually at a specific activity of ~10 cpm/fmol. Reactions were started by the addition of either 50 nM RNAP or 50 nM RNAP + 500 nM GreB, The transcription reactions were incubated for 10 minutes at 37 °C. Reaction products were fractionated in a 23% (10:1) polyacrylamide gel in 1X TBE buffer and 7 M urea and electrophoresed in a gradient buffer set up for 3.5 hours at 35 W.

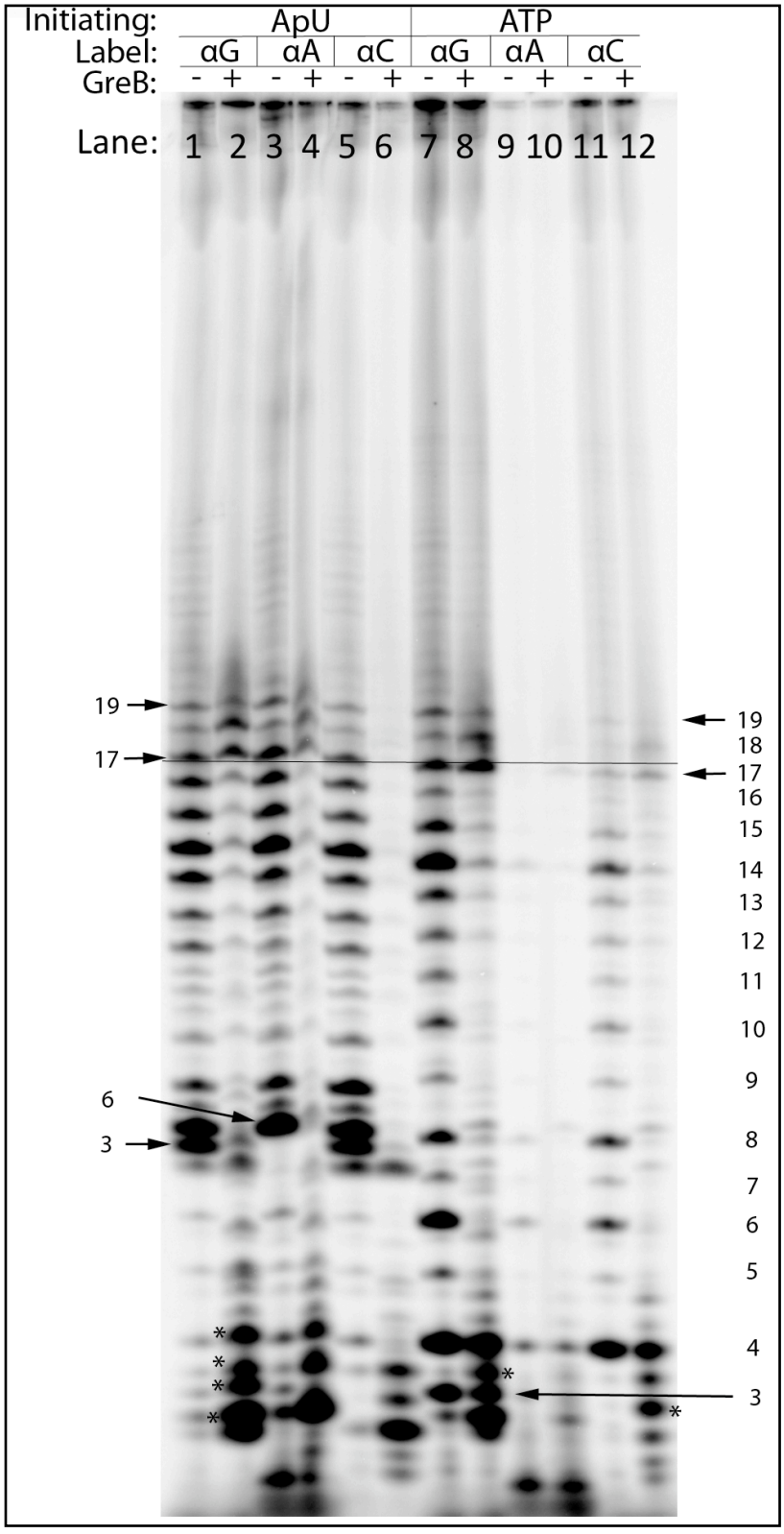

Initiating:
ApU
ATP
Label:
αG
αA
αC
αG
αA
αC
GreB:
- + - + - + - + - + - +
Lane:
1 2 3 4 5 6 7 8 9 10 11 12
19
17
6
3
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3

Figure 6: *Identification of Small ApU-Initiated Abortive RNAs: Attempt I*. All reactions were conducted without GreB. Varied radioactive labeling with [α-^{32}P]-GTP, [α-^{32}P]-ATP, or [α-^{32}P]-CTP are indicated. Final concentration of reagents included: 30 nM DNA DG203/*SPfullcon-U2/19*, 1X transcription buffer, 50 mM KCl, and a nucleotide mixture of either 500 μM ApU + 10 μM NTP or 100 μM ATP + 10 μM. Reactions were commenced by the addition of RNAP to a final concentration of 50 nM, and incubated for 10 minutes at 37 °C. Reaction products were fractionated in a 23% (10:1) polyacrylamide gel in 1X TBE buffer and 7 M urea and electrophoresed in a gradient buffer set up for 3.5 hours at 35 W.

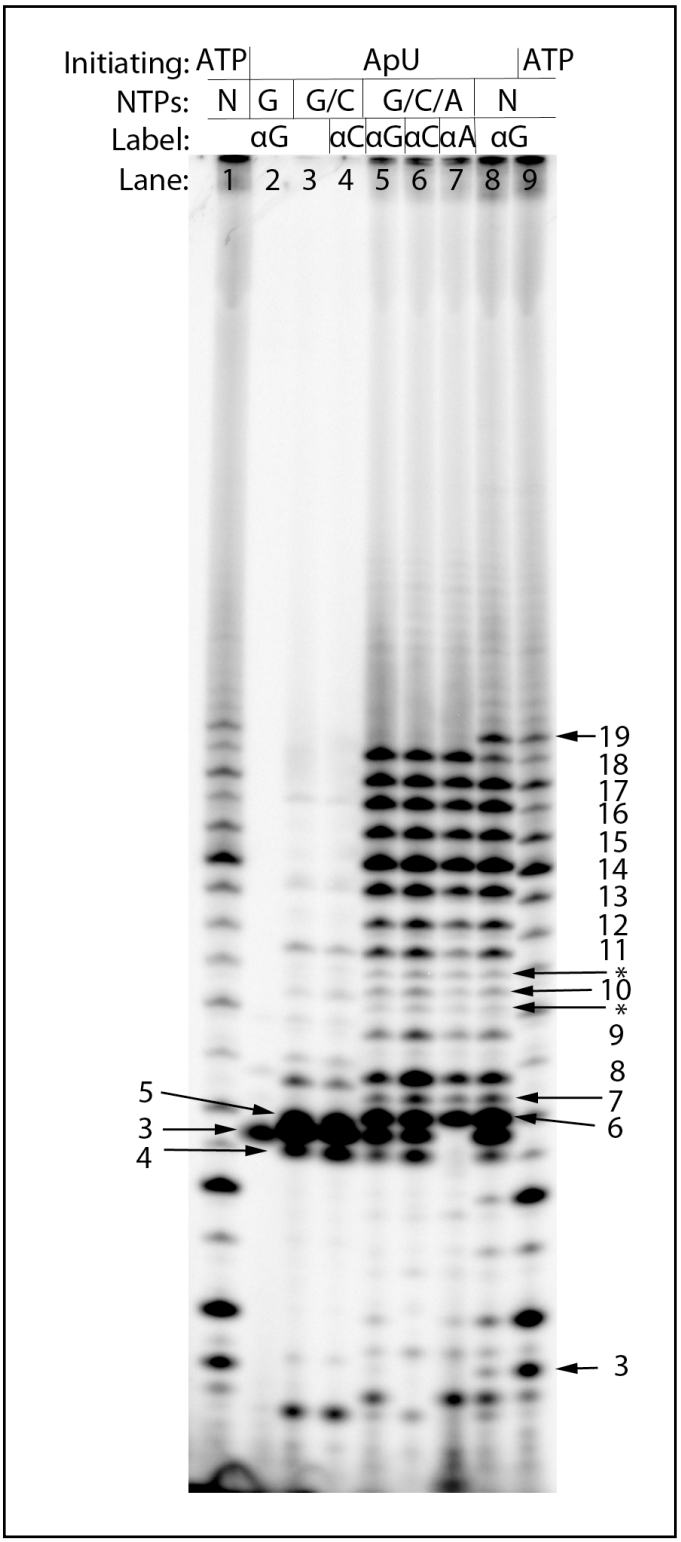

Initiating: ATP ApU ATP
NTPs: N G G/C G/C/A N
Label: αG αC αG αC αA αG
Lane: 1 2 3 4 5 6 7 8 9
19
18
17
16
15
14
13
12
11
*
10
*
9
8
7
6
5
3
4
3

Figure 7: *Identification of Small ApU-Initiated Abortive RNAs: Attempt II*. All reactions were conducted without GreB. Varied radioactive labeling with [α-^{32}P]-GTP, [α-^{32}P]-ATP, or [α-^{32}P]-CTP are indicated. Final concentration of reagents included: 30 nM DNA *DG203/SPfullcon-U2/19*, 1X transcription buffer, 50 mM KCl, and a nucleotide mixture of either 500 μM ApU + 10 μM NTP or 100 μM ATP + 10 μM. Reactions were commenced by the addition of RNAP to a final concentration of 50 nM, and incubated for 10 minutes at 37 °C. Reaction products were fractionated in a 23% (10:1) polyacrylamide gel in 1X TBE buffer and 7 M urea and electrophoresed in a gradient buffer set up for 3.5 hours at 35 W.

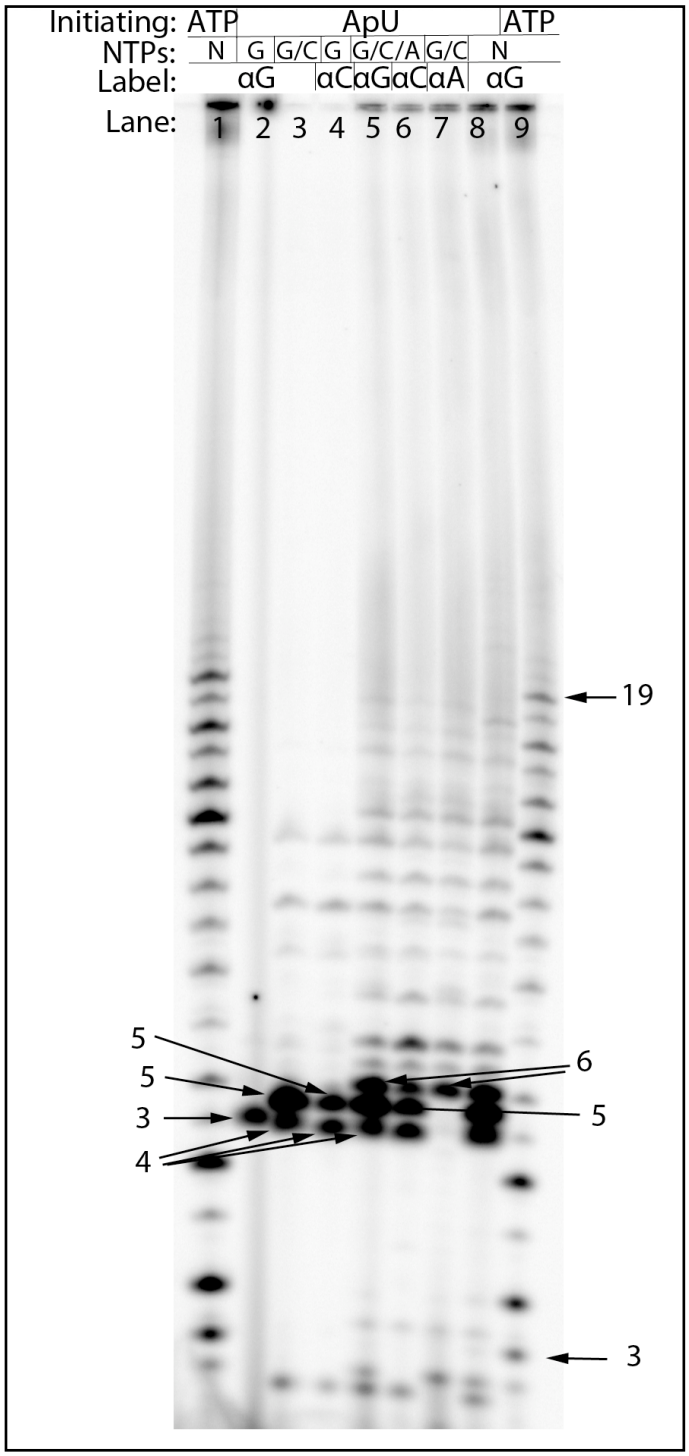

Initiating: ATP ApU ATP
NTPs: N G G/C G G/C/A G/C N
Label: αG αCαGαCαA αG
Lane: 1 2 3 4 5 6 7 8 9
19
5
5
3
4
6
5
3

Figure 8: *Incorporation of UTP vs S_4UTP at the 19th Position of VLAT-19*. Transcription reactions were initiated with ApU, elongated with G/C/ATP, and terminated with either UTP or S_4UTP. Alternate lanes contained GreB, and labeling nucleotides were [α-^{32}P]-ATP and [α-^{32}P]-CTP. Final concentration of reagents included 30 nM DNA DG203/*SPfullcon-U2/19*, 1X transcription buffer III(10), 50 mM KCl, and 500 μM ApU + 10 μM G/C/ATP. Reactions were commenced by the addition of either 50 nM RNAP (lanes 1, 3, 7, 9) or 50 nM RNAP:500 nM GreB (lanes 2, 4, 6, 8) and incubated for 10 minutes at 37 °C. Gel was 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea and ran for 4 hours at 35 W.

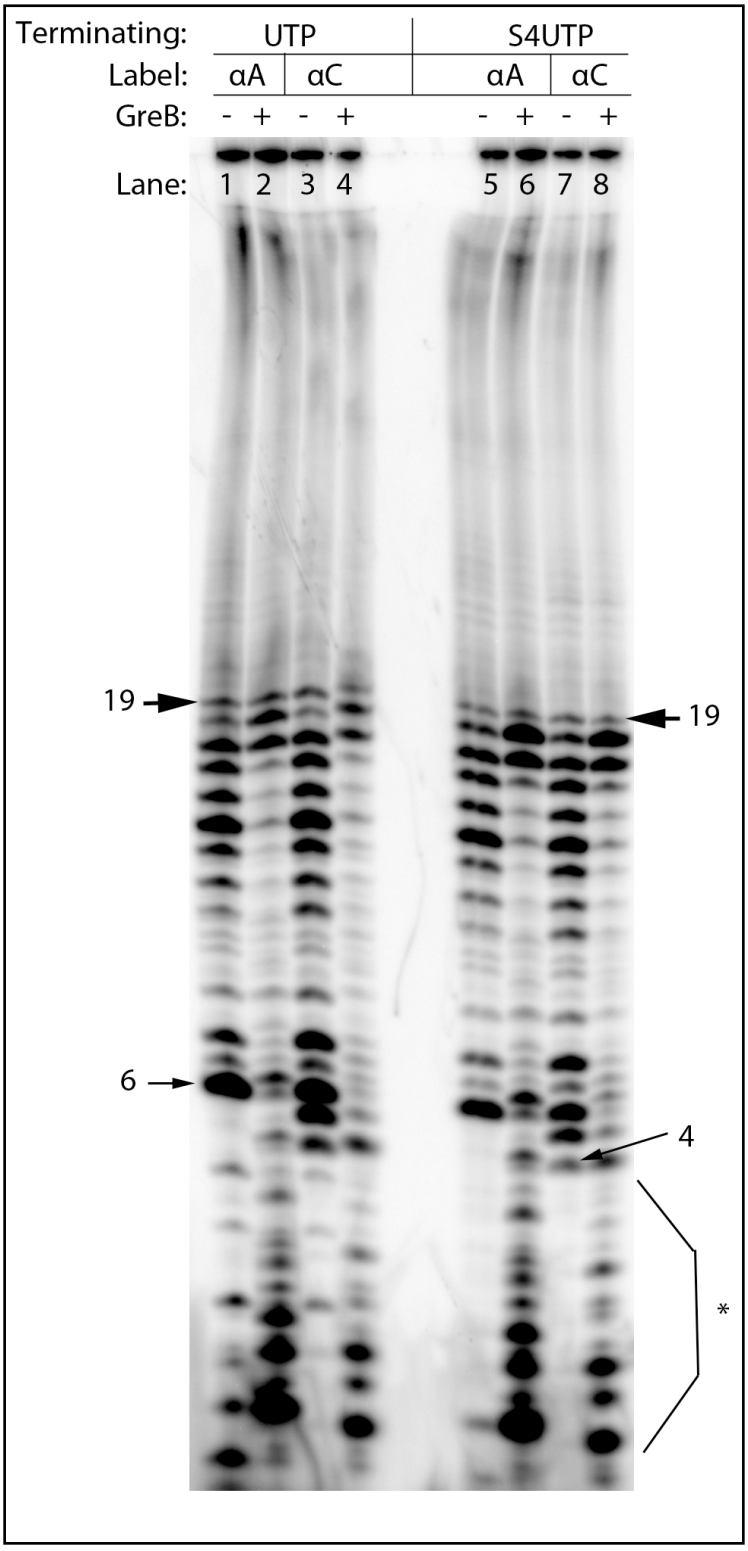

Terminating:
UTP
S4UTP
Label:
αA
αC
αA
αC
GreB:
- + - +
- + - +
Lane:
1 2 3 4
5 6 7 8
19
19
6
4
*

Figure 9: *Optimizing Reaction Conditions for the Production of VLAT-19.* To the basal transcription reaction containing 30 nM *DG203*/*SPfullcon-U2/19* DNA, 1X transcription buffer, 50 mM KCl, 500 μM ApU, and10 μM G/C/ATP was added different amounts of S_4UTP (10, 50, or 100 μM), RNAP (1X/2X/3X = 50/100/150 nM) or RNAP:GreB (1X/2X/3X = 50:500 nM/100 nM:1 mM/150 nM:1.5 mM). Reactions are incubated for 10 minutes at 37 °C in the dark. Gel was 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea and ran for 5 hours at 35 W.

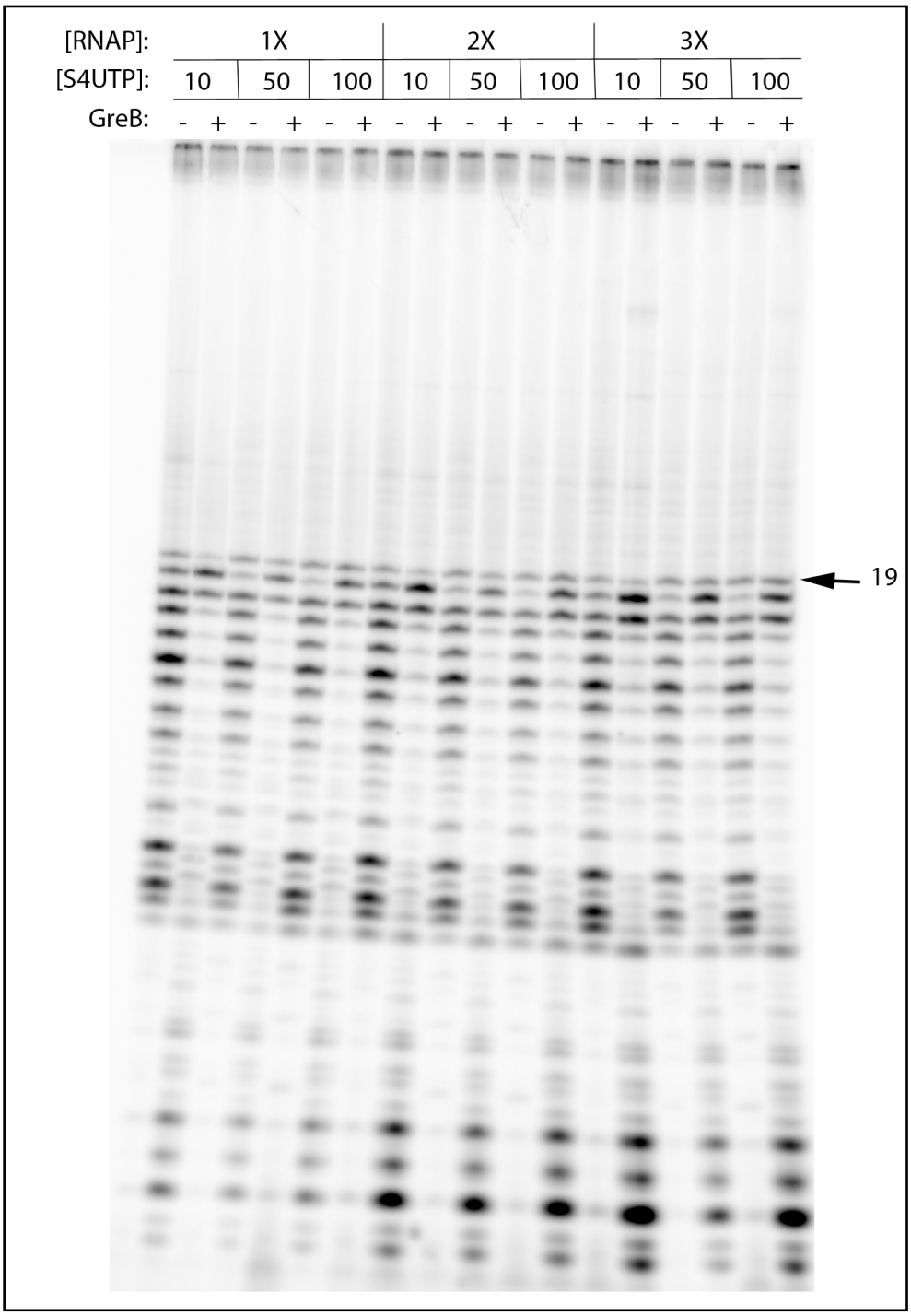
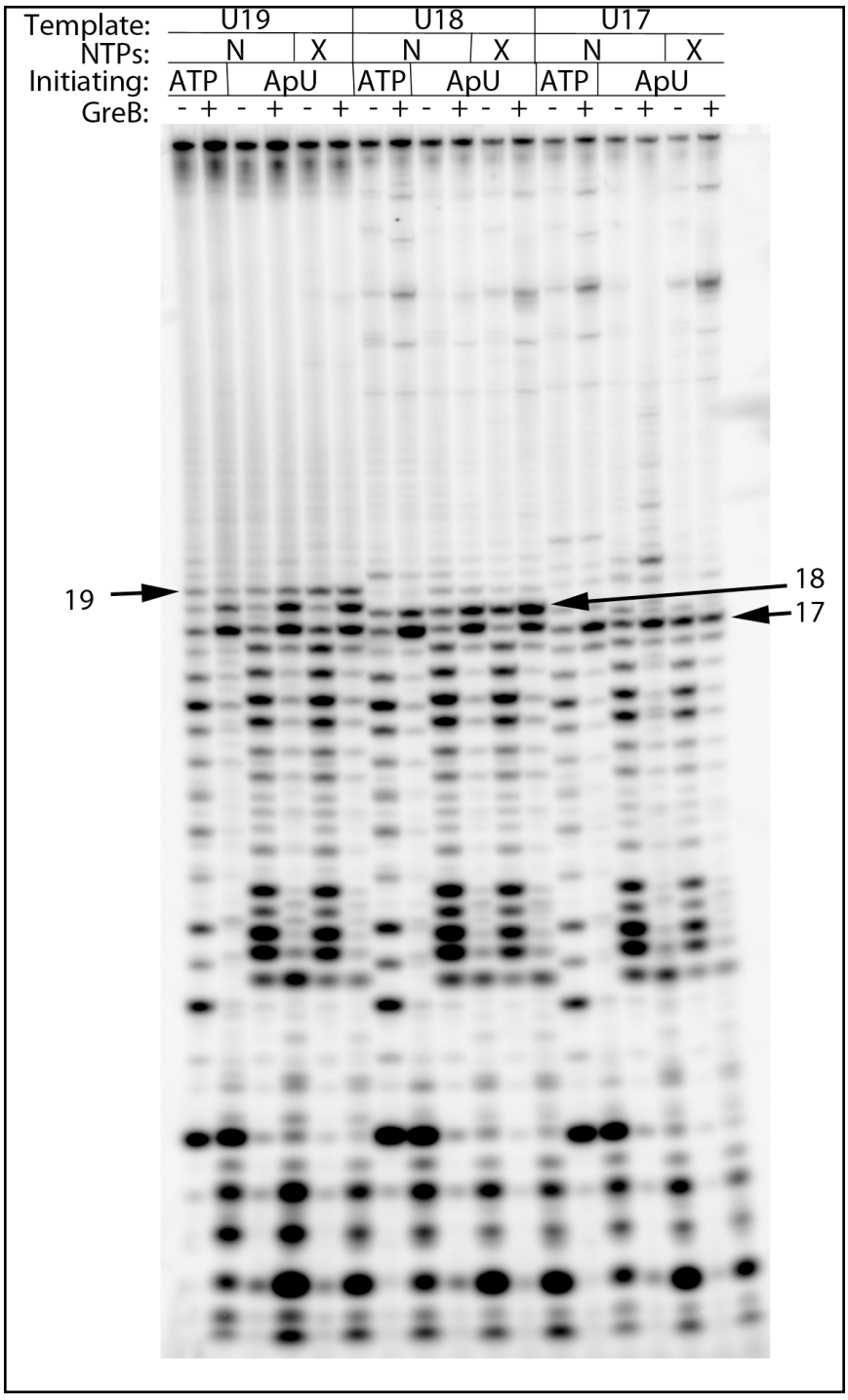

[RNAP]:
1X
2X
3X
[S4UTP]:
10
50
100
10
50
100
10
50
100
GreB:
- + - + - + - + - + - + - + - + - + - +
19

Figure 10: *Comparing the Level of VLAT-19, VLAT-18, and VLAT-17 Synthesis for Photocrosslinking Investigation.* Transcription reactions were performed, without or with GreB supplementation (in alternate lanes), with 30 nM DNA (either *DG203*/*SPfullcon*-*U2/19* or -*U2/U18* or -*U2/U17*) in 1X transcription buffer and 50 mM KCl. Different nucleotide mixture was added for ATP initiation (100 μM ATP + 10 μM NTP) or ApU initiation (500 μM ApU + 10 μM G/C/ATP + 100 μM S_4UTP); [α-^{32}P]-CTP was the labeling nucleotide. Reactions were commenced by the addition of either 150 nM RNAP or 150 nM RNAP + 1.5 mM GreB, and incubated for 10 minutes at 37 °C in the dark. Gel was 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea and ran for 5.5 hours at 35 W. Key: NTP (N) and G/C/ATP + S_4UTP (X).


Template:
U19
U18
U17
NTPs:
N
X
N
X
N
X
Initiating:
ATP
ApU
ATP
ApU
ATP
ApU
GreB:
- + - + - + - + - + - + - + - + - +
19
18
17

Photocrosslinking Reactions

Having demonstrated our ability to specifically incorporate a S_4U residue at the 3′ end of a VLAT, I embarked on a photocrosslinking experiment to determine a) whether the 3′-S_4U of a VLAT can crosslink to RNAP, and b) if so, how specific is the crosslinking reaction. To induce crosslinking, a transcription reaction mixture is transferred into a well in a microtiter plate and irradiated with UV light (365 nm) for a specified amount of time. Afterwards, the reaction mixture is withdrawn and analyzed by gel electrophoresis.

My first experiment with UV-crosslinking involved examining the dependence of crosslinking on UV irradiation. The reactions were analyzed on a high percentage polyacrylamide gel, and the result is shown in Figure 11. After discovering that the crosslinked protein was not entering the high percentage polyacrylamide gel, providing no information on crosslinking; I decided to analyze the photocrosslinking reactions in a SDS-PAGE gel, which has a much lower percentage of polyacrylamide and could distinctly separate the subunits of RNAP. The results are shown in Figure 12A and 12B for VLAT-19 and VLAT-18, respectively. Both VLAT-19 and VLAT-18 showed UV-dependent crosslinking to the β/β′ subunits; however, this crosslinking is not enhanced by the presence of GreB. In a separate experiment reported in Figure 13, VLAT-18 crosslinking to the β/β′ subunits is S_4U-dependent. Again, the level of crosslinking product is less in the presence of GreB. This observation is consistent with the fact that VLATs can arise by backtracking as well as forward

hyper-translocation mechanisms. The GreB-resistant VLATs, in the case of *DG203/SPfullcon* promoter, constitute 55% of the total VLATs made (Chander and Hsu, unpublished results). That we see fewer crosslinked products in the presence of GreB is an encouraging sign that we are capturing mostly the GreB-resistant VLATs.

The SDS-PAGE gel in Figure 12B showed the presence of several radioactive bands across the gel. It also contains size markers transcribed from several control promoters *DG203*, *N25* (full length 74 nt), and *N25* (full length 57 nt) to distinguish the source of the other radioactive bands.

I cut out several bands on the SDS-PAGE gel in Figure 12B and 15A, performed gel crush and re-ran the products on a transcription gel (see Figure 14). As indicated by the transcription gel result, only full-length RNA transcripts could cause these radioactive bands to appear. The various bands these full-length RNAs migrate to are hypothesized to be caused by the formation of secondary structures of the full-length transcripts. These secondary structures run slower in the SDS-PAGE gel, causing the appearance of these various bands.

Figure 11: *Polyacrylamide Gel Fractonation of Photocrosslinking Reaction Products*. 3'-S_4U containing VLAT-19, VLAT-18, and VLAT-17 were prepared from *DG203/SPfullcon-U2/19* or *U2/U18* or *U2/U17* templates in transcription reactions as described in the legend of Figure 10. RNAs were labeled with [α-^{32}P]-CMP incorporation. Reactions alternately lacks or contains GreB, incubated for 20 minutes at 37 °C either in the dark or under long wavelength (365 nm) UV irradiation. Reaction products were analyzed in a 23% (10:1) polyacrylamide/7 M urea gel as described in the legend to Figure 10.

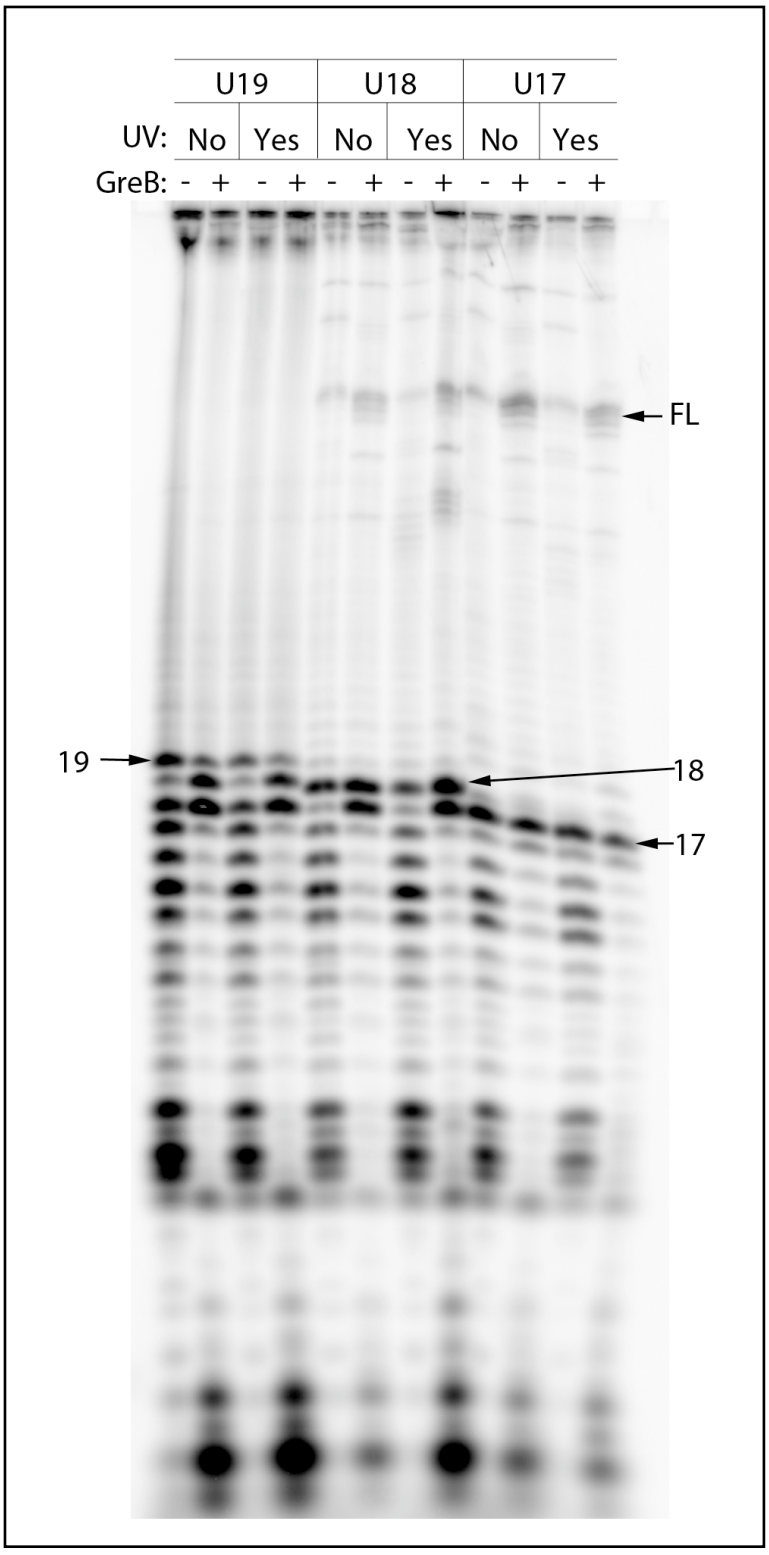
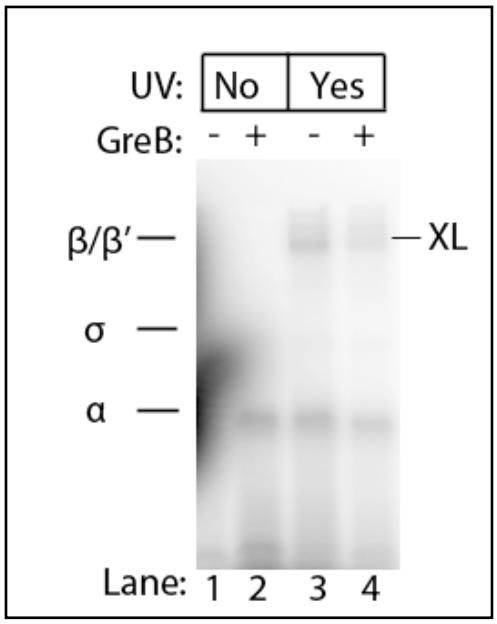
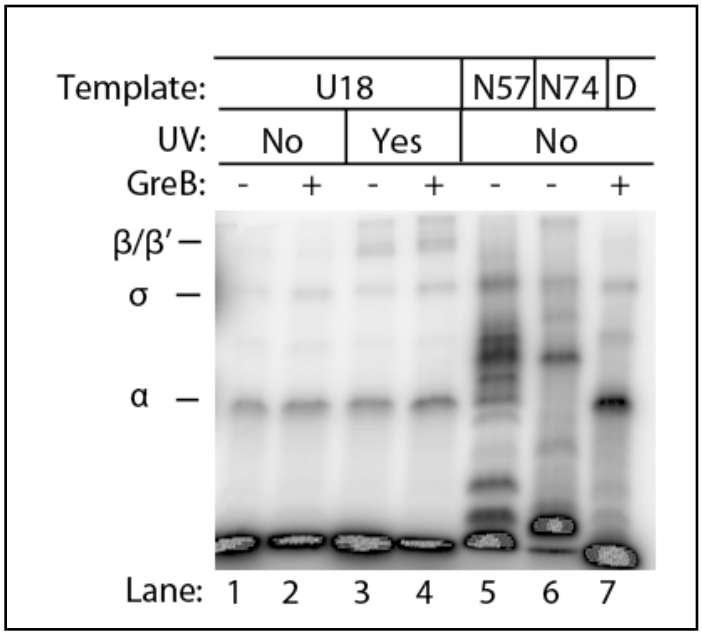

U19
U18
U17
UV: No Yes No Yes No Yes
GreB: - + - + - + - + - + - +
FL
19
18
17

Figure 12: *SDS-PAGE Analysis of VLAT-19 and VLAT-18 Photocrosslinking Reaction Products.* 3'-S_4U containing VLAT-19 and VLAT-18 were prepared from *DG203/SPfullcon-U2/19* or -*U2/U18* templates in transcription reactions as described in the legend of Figure 10. RNAs were labeled with [α-^{32}P]-CMP incorporation. Reactions alternately lacks or contains GreB, incubated for 20 minutes at 37 °C either in the dark or under long wavelength (365 nm) UV irradiation. Control promoters--*DG203* {labeled D}, *N25*(74) {labeled N74}, and *N25*(57) {labeled N57}--were transcribed under similar conditions to provide runoff RNAs (57 nt, 74 nt, and 57 nt, respectively) that can serve as size markers in the SDS-PAGE gel. SDS-PAGE was performed in a 4% stacking-8% resolving gel. A. Crosslinking reactions with VLAT-19 RNA. B. Crosslinking reactions with VLAT-18 RNA.

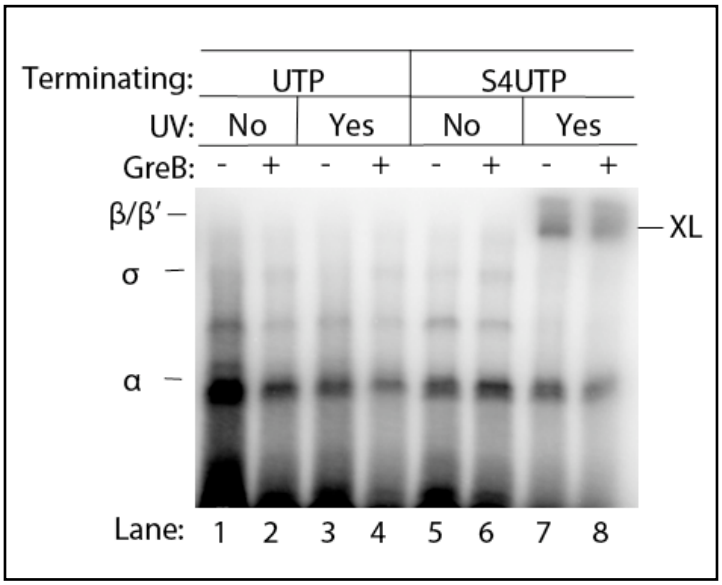


A. DG203/*SPfullcon-U2/U19*

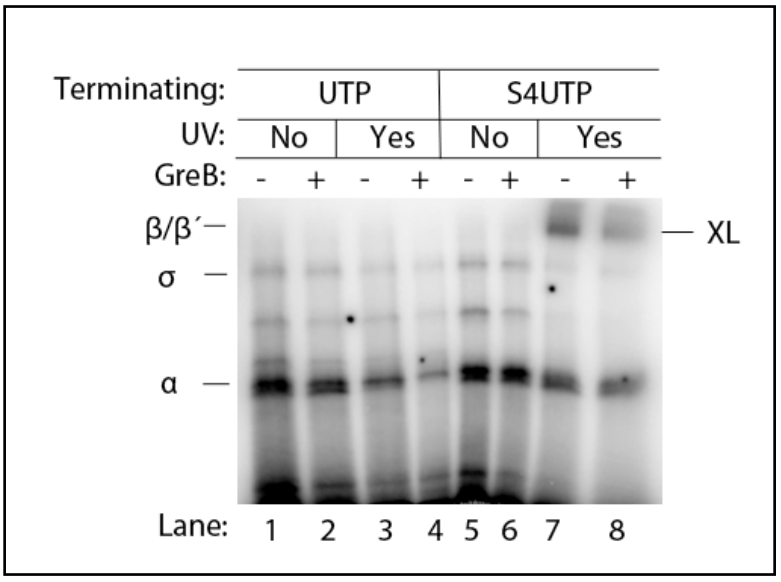


B. DG203/*SPfullcon-U2/U18*

Figure 13. *S_4U-Dependence of VLAT-19 and VLAT-18 Crosslinking*. 3'-S_4U-containing VLAT-19 / VLAT-18 and 3'-U containing VLAT-19 / VLAT-18 were made from the *DG203/SPfullcon-U2/U19* or *DG203/SPfullcon-U2/U18* template, respectively, in reactions as described in the legend of Figure 8. Transcripts were labeled with [α-^{32}P]-CMP incorporation. Reactions alternately lacks or contains GreB, subjected to 20 min incubation at 37 ^{o}C either in the dark or with longwave UV irradiation, and fractionated in a 4% stacking-8% resolving SDS-PAGE.

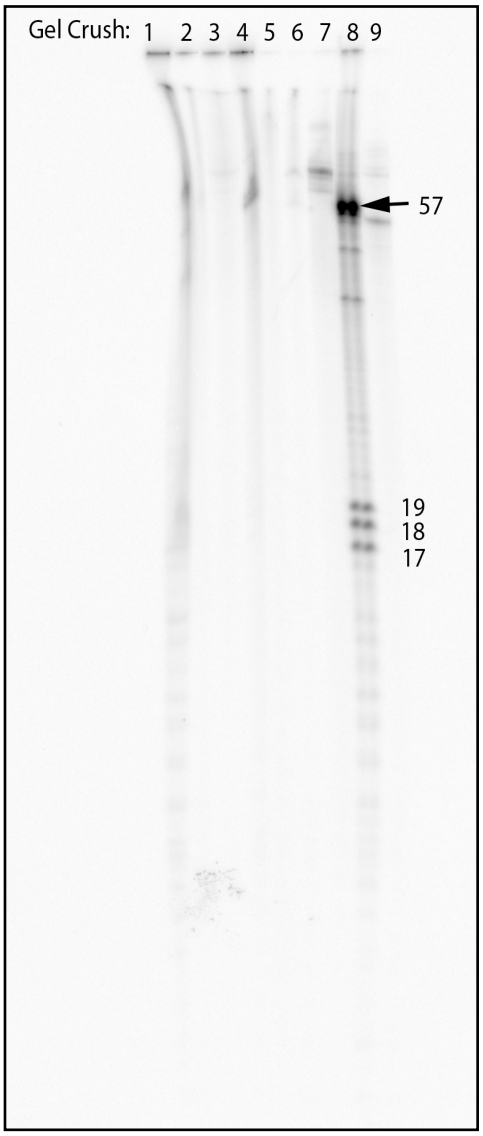


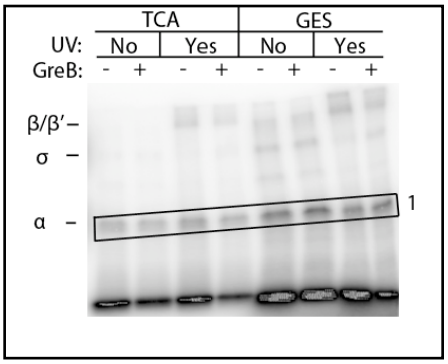
A. *DG203/SPfullcon-U2/U19*

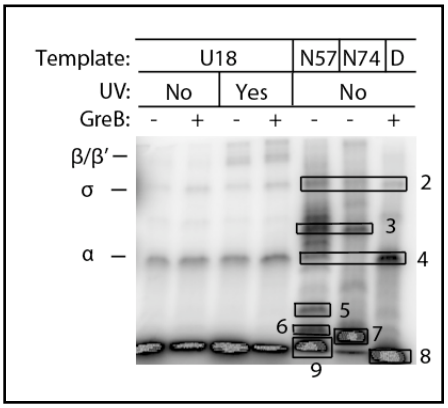


B. *DG203/SPfullcon-U2/U18*

Figure 14: *Gel crush of SDS-PAGE gel bands to deduce the size of omnipresent radioactively labeled bands*. Bands labeled 1-9 were excised, crushed and soaked in high salt buffer to extract the radioactive bands. Extracted material was fractionated in a 23% (10:1)/7 M urea polyacrylamide gel.


Gel Crush: 1 2 3 4 5 6 7 8 9
57
19
18
17


TCA
GES
UV: No Yes No Yes
GreB: - + - + - + - +
β/β′
σ
α
1


Template: U18 N57 N74 D
UV: No Yes No
GreB: - + - + - - +
β/β′
σ
α
2
3
4
5
6
7
8
9

Enhancing the Photocrosslinked Products

After ensuring that crosslinking is RNAP-, UV-, and S_4U- dependent and occurs to the β/β′ subunits, I performed several experiments to see whether I could enhance the intensity of the crosslinked product. These experiments included different procedures in the concentration-recovery process after the transcription reaction, different durations of the transcription reaction, different photocrosslinkable nucleotides, and different UV irradiation wavelengths.

Different concentration/recovery procedures were performed on the *DG203/SPfullcon-U2/U19* and -*U2/U18* templates. In my first experiments I varied TCA precipitation and GES precipitation. TCA precipitation provided clearer crosslinked bands and less free RNAs in the gel. Figure 15A,B show these results.

Furthermore, a comparison between direct addition of the transcription reaction to the SDS-PAGE gel and indirect addition (the reaction was first concentrated through TCA precipitation) was performed on both the *DG203/SPfullcon-U2/U19* and -*U2/U18* templates. The same transcription conditions were kept as the previous TCA/GES experiments. However, we also performed a time course, comparing reaction duration of 20 min or 1 hour. Figure 16 A, B and 17 A, B show these results.

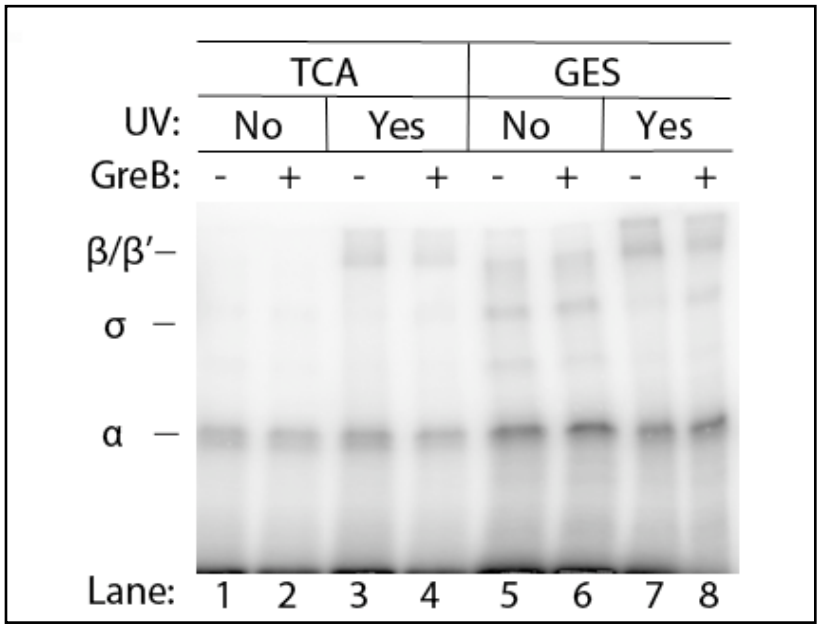
In order to recover more crosslinked product to prepare for mass spectrometry, it would be desirable to be able to run larger quantities of crosslinked product per lane in the SDS-PAGE gel. I performed different reaction

volumes of the crosslinking reactions and ran them on the SDS-PAGE gel after TCA precipitation to see whether there is a limit of the reaction volume and the amount of crosslinked product produced. I was able to successfully increase the reaction volume to 100 μL which showed a disproportionately large increase in crosslinked product. Figure 18 showed these results.

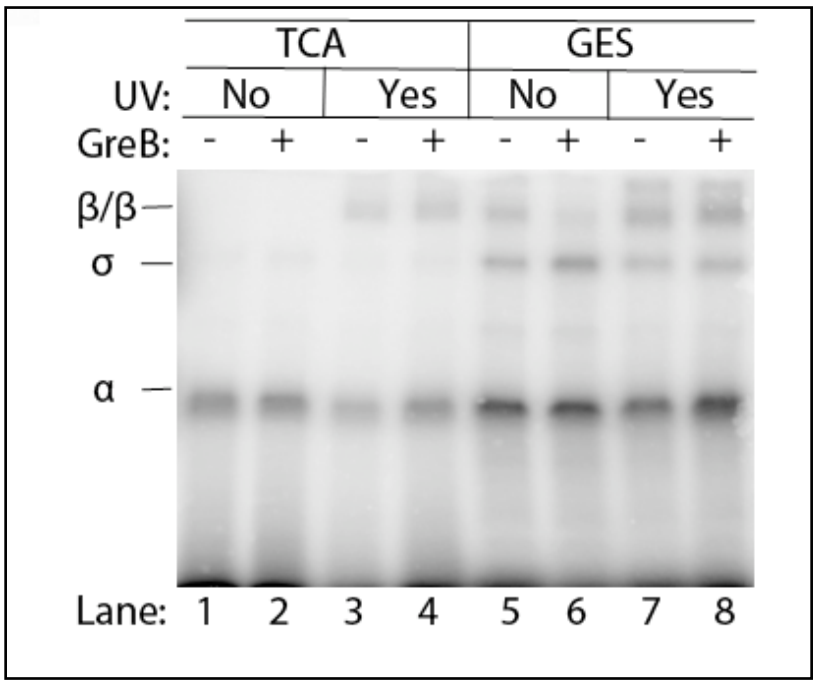
Additionally, I explored different photocrosslinkers including S_2UTP, Br_5UTP and I_5UTP. S_2UTP did not produce any photocrosslinking as seen in Figure 20. I attempted photocrosslinking with Br_5UTP and I_5UTP, but the gels produced were not very clear. I would recommend additional studies on these photocrosslinkers.

Finally, I tried 365 nm and 302 nm UV wavelengths. After doing more literature research, we found that the wavelength S_4UTP would be excited the most is at 312 nm. In the laboratory we do not have readily available a 312 nm lamp, but we do have a 302 nm lamp. The 302 lamp has a lambda max at 302 nm, but also more 312 nm wavelengths than the 365 nm lamp. These gel images were also not very clear and not shown. I would recommend additional studies comparing these different irradiation wavelengths.

Figure 15 A, B: *Comparing TCA vs GES Precipitation with Different Templates*. Reactions were alternated with and without longwave UV irradiation (365 nm) and with and without cleavage factor GreB, while radioactively labeled with nucleotide [α-^{32}P]-CTP. Overall, the volume of each reaction was 10 μL and the final concentration of reagents included 30 nM DNA *DG203/SPfullcon-U2/19* or -*U2/18*, 1X transcription buffer, 50 mM KCl, 500 μM ApU, 10 μM G/C/ATP, 150 nM RNAP or 150 nM RNAP + 1.5 mM GreB, and 100 μM S_4UTP. The transcription reaction ran for 20 minutes at 37°C either in the dark or under long wavelength UV irradiation.

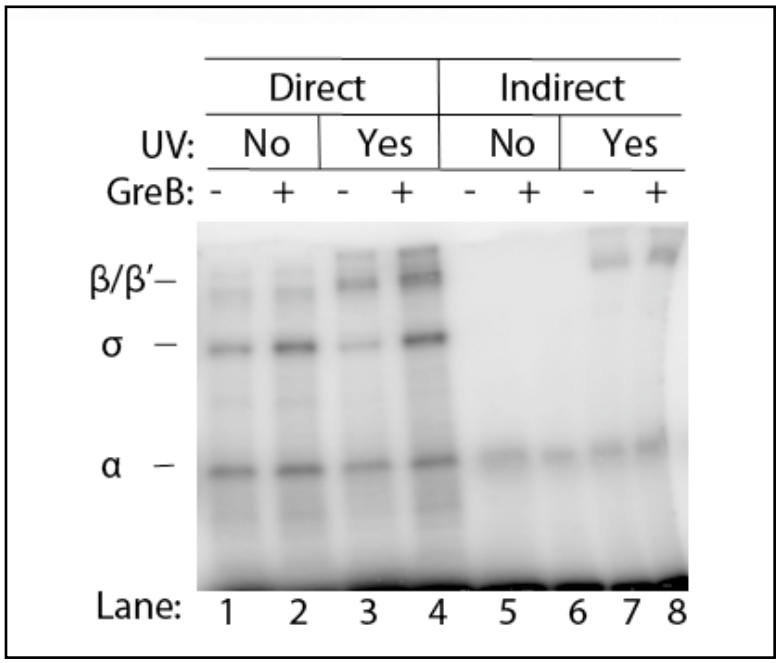


A. *DG203/SPfullcon-U2/U19*

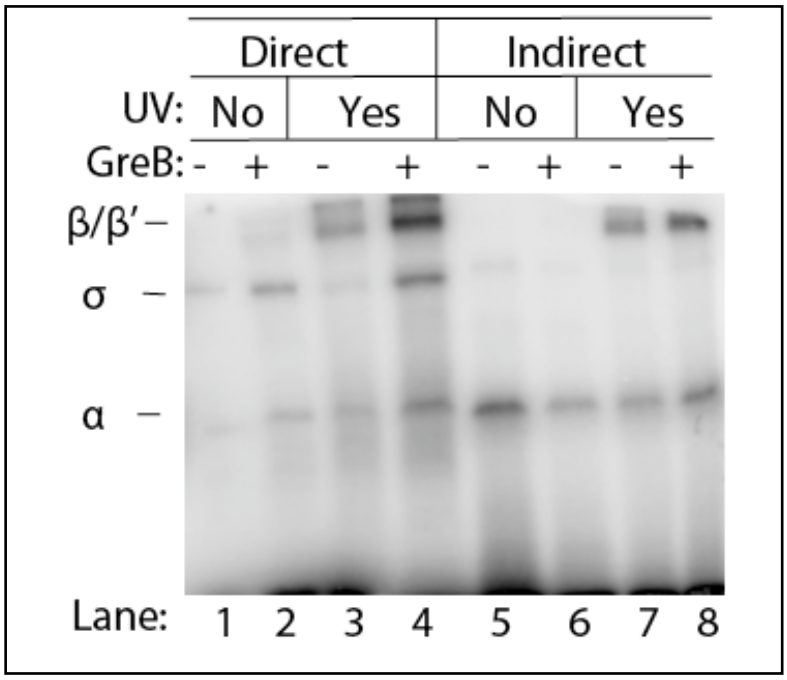


B. *DG203/SPfullcon-U2/U18*

Figures 16: A, B: *Comparing Direct and Indirect Addition and Time Course with U2/U18 Template*. Reactions alternated with and without cleavage factor GreB and UV irradiation. Transcripts were radioactively labeled with [α-^{32}P]-CTP. Indirect refers to TCA precipitation, while direct refers to no precipitation of transcription reaction.

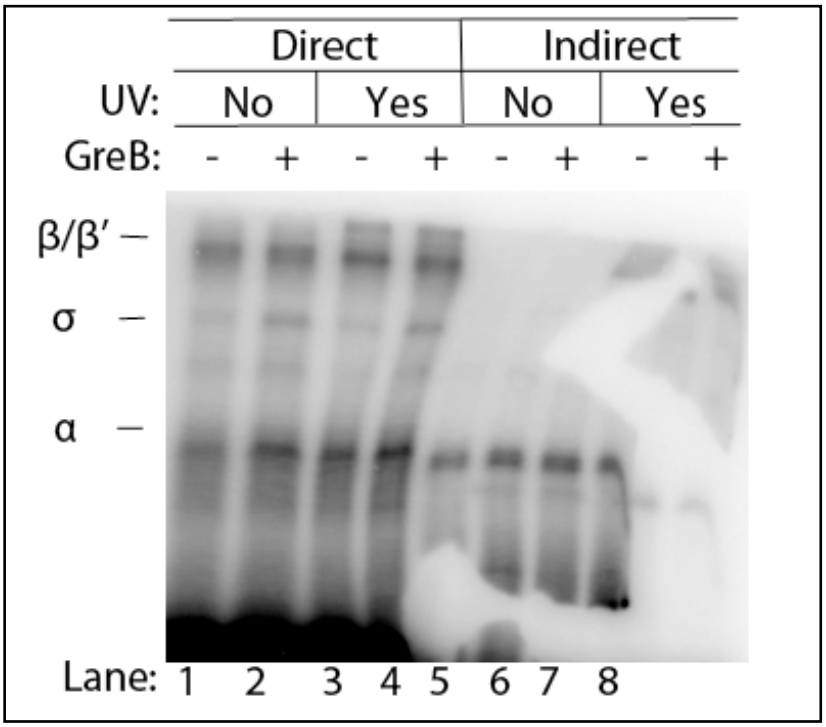


A. 20 Minute Reaction

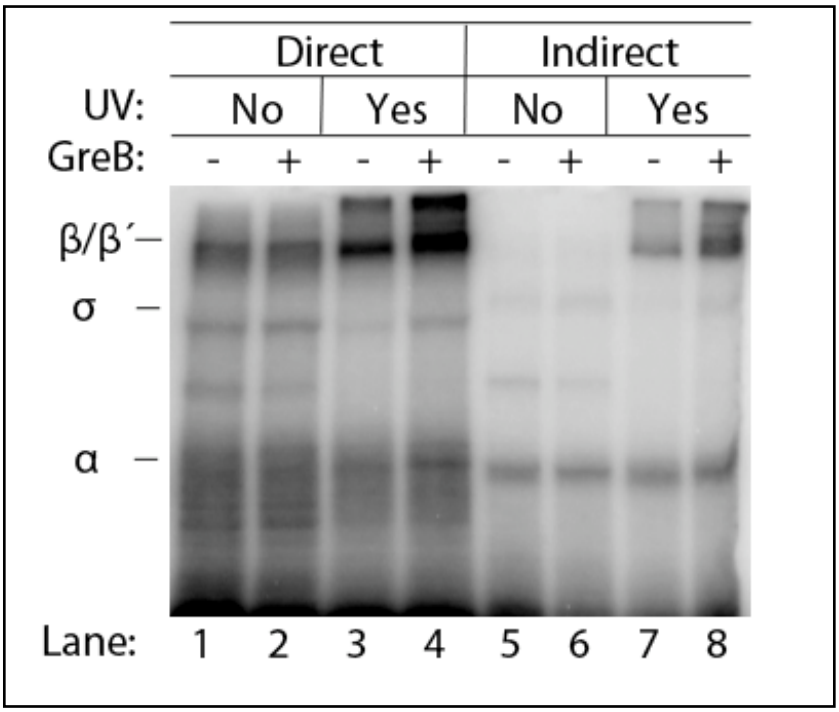


B. 1 Hour Reaction

Figure 17: A, B: *Comparing Direct and Indirect Addition to SDS-PAGE Gel and Time Course with U2/U19 Template*. Reactions alternated with and without cleavage factor GreB and UV irradiation. Transcripts were radioactively labeled with [α-^{32}P]-CTP. Indirect refers to TCA precipitation, while direct refers to no precipitation of transcription reaction.



A. 20 Minute Reaction



B. 1 Hour Reaction

Figure 18. *Different Reaction Volumes with DG203/SPfullcon-U2/U19 and DG203/SPfullcon-U2/U19 Templates with GreB.* Reaction volumes include 10 μL, 20 μL, 50 μL, and 100 μL. Transcripts were radioactively labeled with [α-^{32}P]-CTP. Crosslinking reaction ran for 1 hour. Crosslinked products were recovered using TCA precipitation.

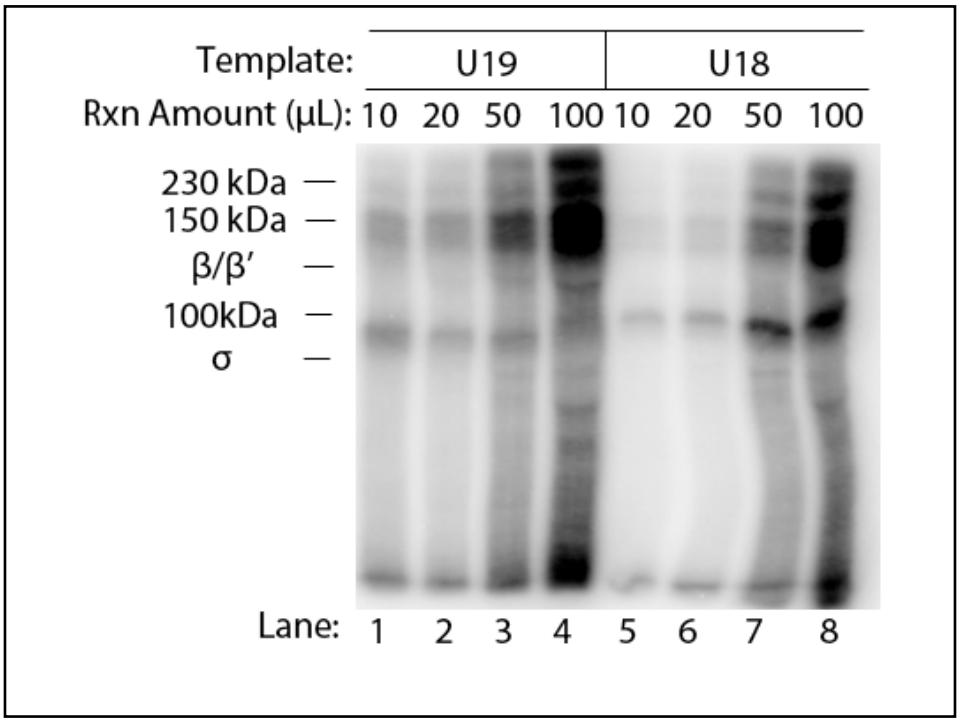
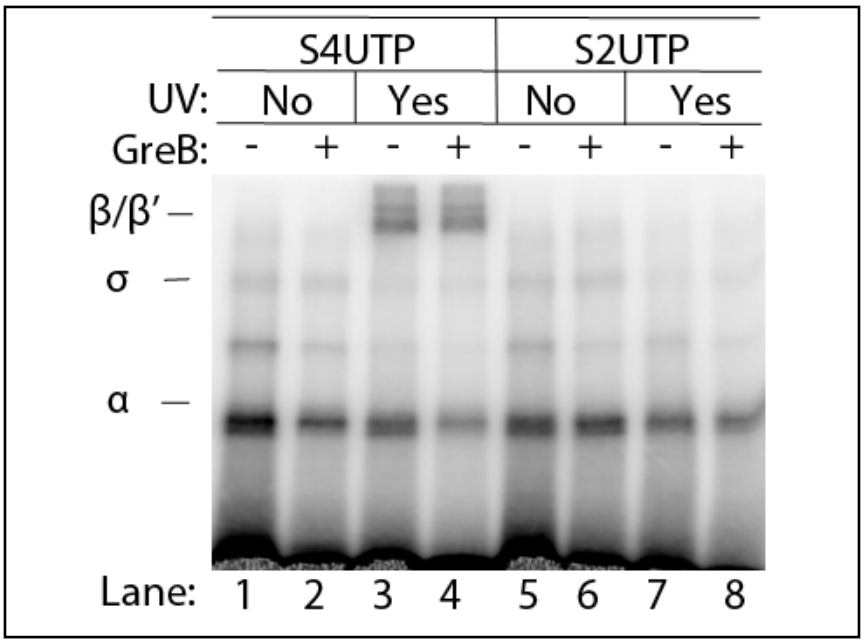
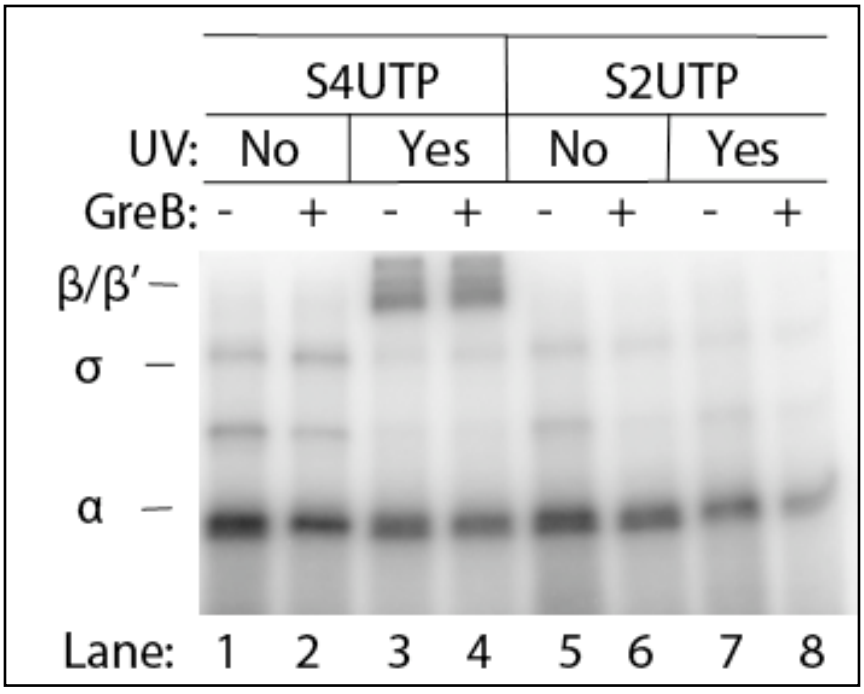

Template:
U19
U18
Rxn Amount (µL): 10 20 50 100 10 20 50 100
230 kDa
150 kDa
β/β'
100kDa
σ
Lane: 1 2 3 4 5 6 7 8

Figure 19. *Comparing S_4UTP to S_2UTP Crosslinking*. Reactions were performed with *DG203/SPfullcon-U2/U18* and *–U2/U19* templates, alternated with and without cleavage factor GreB and UV irradiation. Transcripts were radioactively labeled with [α-^{32}P]-CTP. UV irradiation at 365 nm was performed for 1 hour.



A. *DG203/SPfullcon-U2/U19*



B. *DG203/SPfullcon-U2/U18*

In Gel Digestion Reactions

Our crosslinking experiments showed that crosslinking is occurring to the β/β′ even in the presence of GreB, suggestive that the crosslink is within the RNA exit channel. Nevertheless, we want to understand to which exact amino acid(s) the crosslink is formed. MALDI-TOF mass spectrometry would be the best technique to elucidate where the crosslink(s) occur (Hillenkamp and Peter-Katalinic, 2007). To this end, we need to accomplish three goals: extract the crosslinked product from the SDS-PAGE gel, remove all radioactivity from the crosslinked product, and reduce the size of the crosslinked product. In-gel digestion is a technique used to both digest and extract proteins from an SDS-PAGE gel, thus optimal for our experiments. Radioactivity is contained within the VLAT, thus the use of nucleases to cleave the VLAT would be essential. Additionally, to identify to which amino acid the crosslink is formed, we want to reduce the size of the β/β′ subunits, allowing us to hone in on the crosslinked peptide fragment. The use of protease would be essential to reduce the size of the β/β′ subunit(s).

I began the in-gel digestion experiments trying to remove the radioactivity from the crosslinked product through the use of small nucleases that could penetrate into the polyacrylamide gel. I first tried micrococcal nuclease (MNase), which is only ~13 kDa in size. I performed a time course digestion of the excised bands for 2 hours and 14 hours digestion with MNase and a control under the same buffer conditions minus MNase and

incubated for 14 hours. I performed this digestion on two distinct bands of the crosslinked product, indicated in Figure 20. Following digestion I did a scintillation count of the digestion solution and the remaining gel. The scintillation counts and percentages within the solution and gel is presented in Figure 21. As this table indicates, I was not able to remove all the radioactivity from the gel. Additionally, the 2-hour and 14-hour incubation produced a similar amount of digestion. However, due to the length of the protocol, the 14 hour is much easier to perform because I can leave the reaction overnight and read the samples the following day.

To prepare for mass spectrometry, we want as little addition of foreign proteins in our sample as possible. Bovine serum albumin (BSA) is a protein used in conjunction with MNase to help with its activity. I performed in-gel digestion with and without BSA with the ~14 hour (overnight) digestion time. Both the samples with or without BSA left ~37% of the radioactivity within the gel. Thus, my future reactions with MNase can be performed without BSA.

MNase digestion procedure did not remove all of the radioactivity. Thus, future studies will look at a different nuclease RNase T1. This nuclease is 11 kDa, slightly smaller than MNase, and cleaves right after a G-residue in the RNA. Since the VLAT-19 RNA is labeled with [α-^{32}P]-CMP (indicated with *C) and S_4UMP at the crosslinking position (indicated with ^U), the RNA sequence (5'-AUG*CGA*C*CGGGAGAGGAG^U-3') is an excellent substrate for

RNase T1; any cuts at G-residues in the middle of the RNA should remove all of the *CMP signals. Future studies will investigate the use of RNase T1 alone or in conjunction with MNase. Finally, to reduce the β/β′ subunit(s), I will use trypsin. Trypsin cleaves at many sites on the β/β′ subunit(s) and will reduce the size to a more manageable and informative size for MALDI-TOF mass spectrometry analysis.

Figure 20. *In-Gel Digestion Excision Demonstrating Top and Bottom Crosslinking Bands.* A triplicate of 100-μL transcription reactions were performed with *DG203/SPfullcon-U2/U19* template in the presence of GreB, [α-^{32}P]-CMP labeling, and with UV irradiation for an hour at 37 ^{o}C. Samples were concentrated by TCA precipitation before loading into a SDS-PAGE. This gel was electrophoresed for 10 mins at 100 V through the stacking gel and 1 hour at 150 V until 20 kDa band has reached the bottom of the gel, allowing better separation of the crosslinked product from the β/β' subunits. Two pieces of gel, labeled as "top" and "bottom" are recovered from each lane, and subjected to control treatment (without enzymes) for 14 hours, with 10,000 units MNase for 2 hours, or with 10,000 units MNase for 14 hours. After the digestion reaction is over, the sample was spun to separate the supernatant and the gel pieces. Both fractions were counted in the scintillation counter to assess the completeness of the digestion reaction.

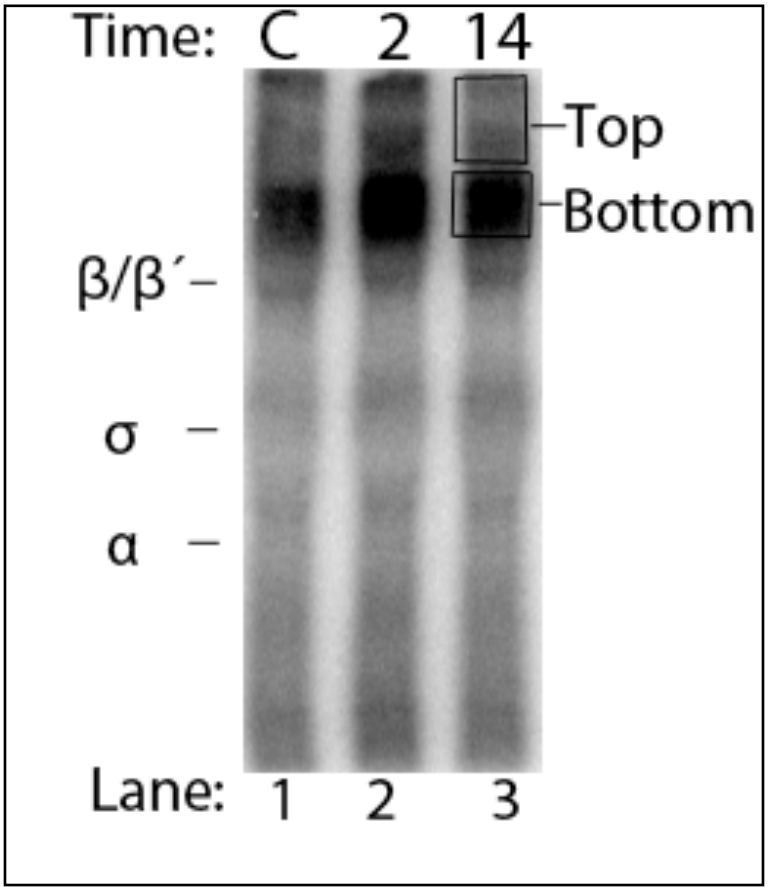

Time: C 2 14
Top
Bottom
β/β´
σ
α
Lane: 1 2 3

Figure 21. *Time Course of In-Gel Digestion with MNase and VLAT-19.* The in-gel digestion experiment was performed according to provided protocol. Incubation was at 37 °C for 2 hours or 14 hours (overnight) with shaking. Control was incubated without MNase for 14 hours. Following digestion, the gel piece and supernatant were separately analyzed by scintillation counting.

Figure 21: Time Course In-Gel Digestion with MNase and VLAT 19						
	Control (-MN)		**2 hr (+MN)**		**14 hr (+MN)**	
	Solution	**Gel**	**Solution**	**Gel**	**Solution**	**Gel**
Top:	4302	38639	22972	14570	16477	11889
	10%	90%	61%	39%	58%	42%
Bottom:	9501	18320	29820	11703	27067	8149
	33%	66%	72%	28%	77%	23%

Figure 22. *In-Gel Digestion of VLAT-18 and VLAT-19 with MNase With and Without BSA.* Followed protocol for in-gel digestion experiments. Incubation was with or without BSA at 37 °C for ~14 hours (overnight) with shaking. Control was incubated without MNase and BSA. Following digestion, the gel piece and supernatant were separately analyzed by scintillation counting.

Figure 22: With and Without BSA In-Gel Digestion with MNase and VLAT 18 and 19						
	Control (-MN)		**w/o BSA**		**w/ BSA**	
	Solution	**Gel**	**Solution**	**Gel**	**Solution**	**Gel**
VLAT 19	15513	28038	68883	35963	58741	35524
	36%	64%	66%	34%	63%	37%
VLAT 18	15616	34360	22407	14768	18849	12355
	32%	68%	61%	39%	61%	39%

Mutant RNAP Reactions

Beyond our normal wild-type (w.t.) RNAP, we investigated the activity of mutant RNAP rpoβ*35 (βH1244Q) generously donated by Evgeny Nudler, Professor of Biochemistry at New York University School of Medicine. This RNAP is characterized by an amino acid replacement of histidine to a glutamine on the 1244th amino acid of the β subunit. It has been deemed to be backtrack-resistant and could be experimentally insightful in our crosslinking studies.

I tested the hypothesized backtrack-resistant mutant RNAP rpoβ*35 (βH1244Q) on several templates to test the validity of its backtrack resistance. I tested templates *N25* (74) Anti, *DG203/SPfullcon-U2/U19* and *U2/U18*. The activity of the mutant RNAP is significantly lower than our w.t. RNAP, thus it was hard to make any conclusions from these results. However, many of the backtrack-induced abortive RNAs (i.e. abortive RNAs ≤ 15 nt) remained for the three promoters. Figure 23 shows these results.

After seeing that there may not be backtracking resistance with templates N25 (74) Anti, *DG203/SPfullcon-U2/U19* and *U2/U18;* I wanted to test if other, more common templates, produced backtracking resistance with the rpoβ*35 (βH1244Q) RNAP mutant. I tested templates T5 N25 (FL=57), T5 N25$_{antiDSR}$ (FL=57), P_L (FL=68), *lac*UV5 (FL=58), T7A1 (FL=67), and *rrn*BP1$_{dis}$ (FL= 57). Likewise with these promoters, the activity of the mutated RNAP was significantly decreased and the backtracking-resistant nature of this RNAP was inconclusive. Figure 24 shows these results.

Figure 23: *Comparing w.t. RNAP to Mutant rpoB*35 RNAP on Different Templates:I. Templates used were N25 (74) Anti* {A}*, DG203/SPfullcon-U2/U19 and -U2/U18.* Reactions were alternated with and without cleavage factor GreB and radioactively labeled with [α-^{32}P]-CTP. Reaction conditions included 30 nM DNA, 1X transcription buffer, 50 mM KCl, 100 μM NTP, 50 nM RNAP or 50 nM RNAP + 500 nM GreB. The transcription reaction ran for 10 minutes at 37 °C. Gel was 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea and ran for 4.5 hours at 35 W.

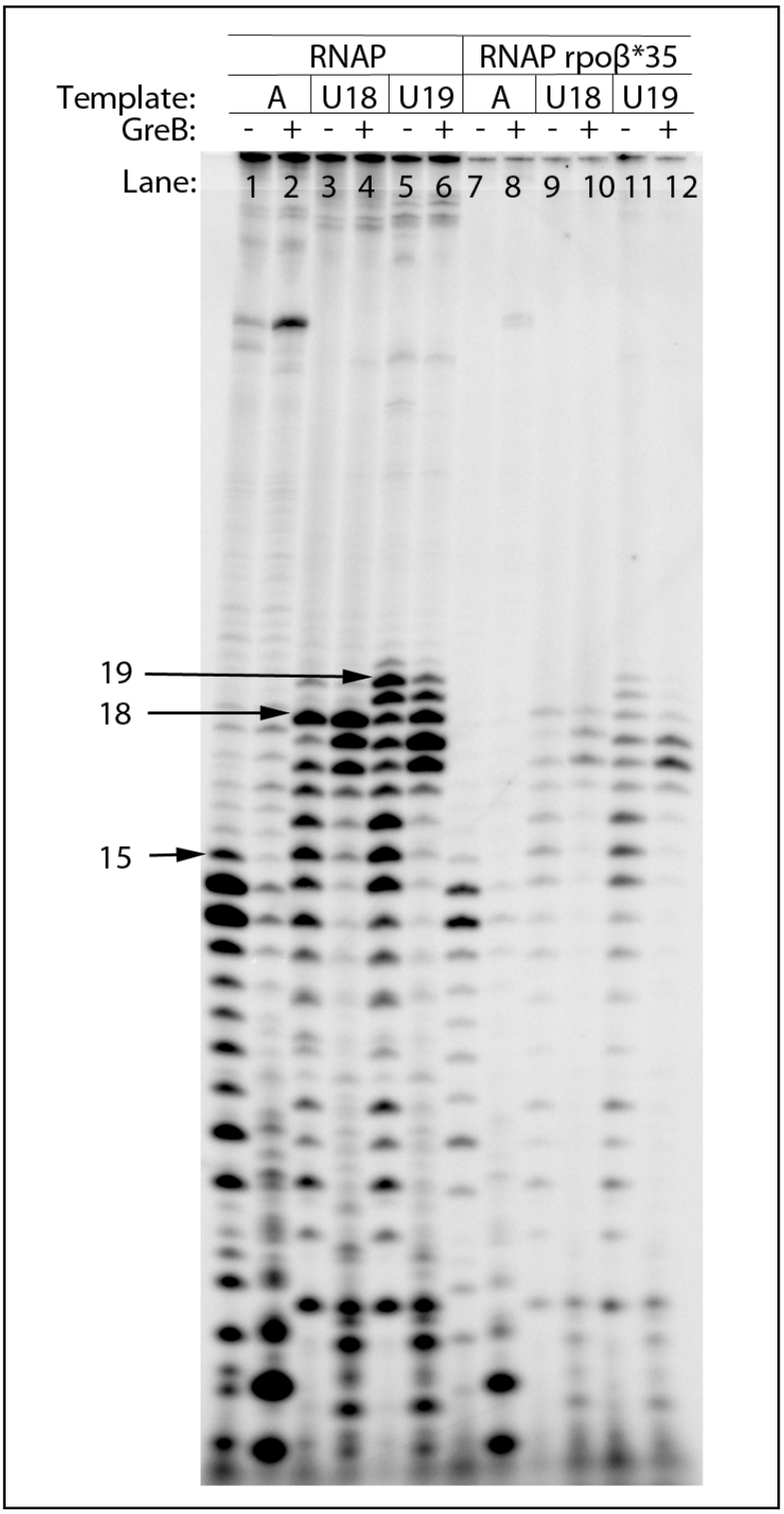
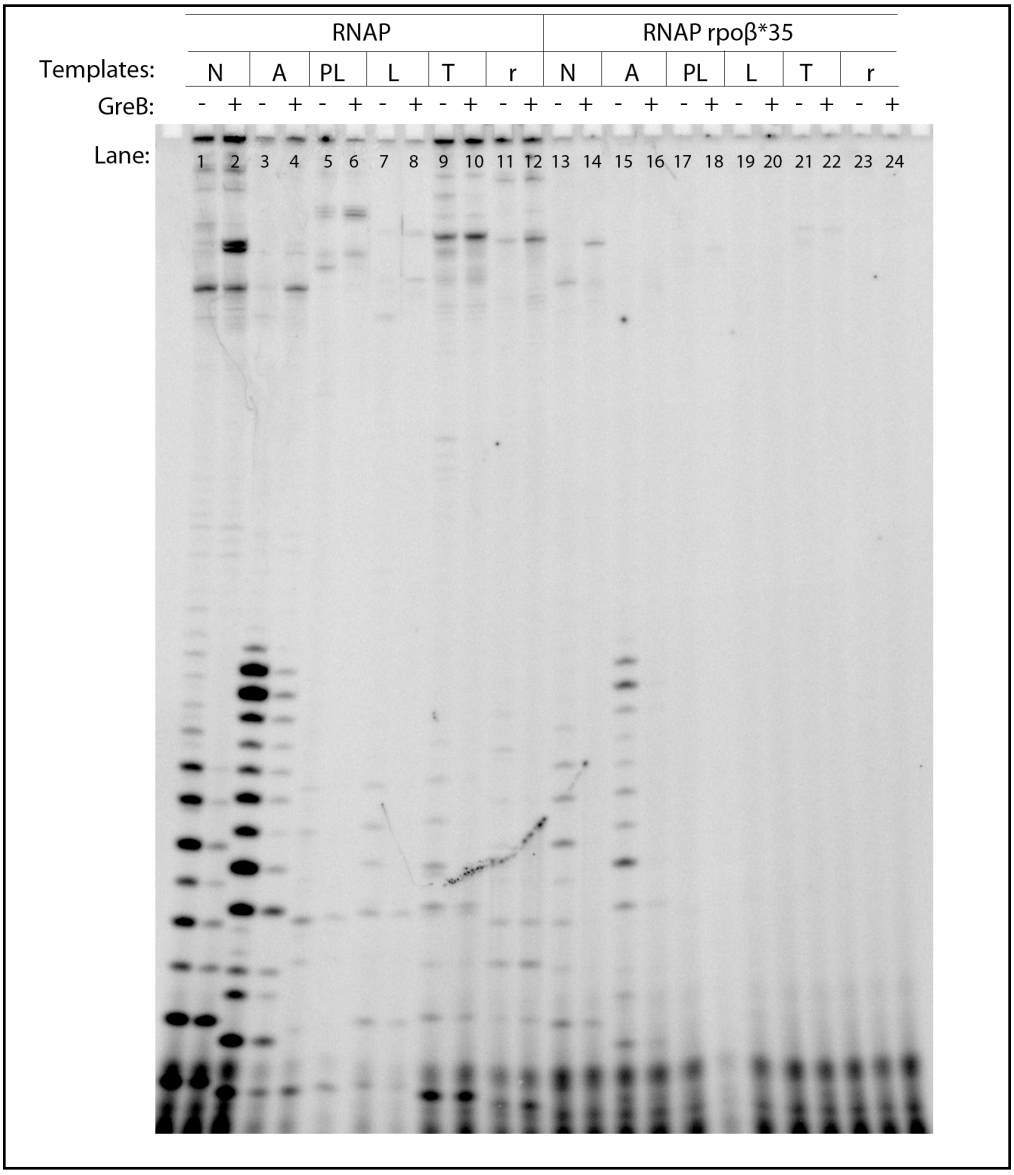

RNAP
RNAP rpoβ*35
Template:
A
U18
U19
A
U18
U19
GreB:
- + - + - + - + - + - +
Lane:
1 2 3 4 5 6 7 8 9 10 11 12
19
18
15

Figure 24. *Comparing w.t. RNAP to Mutant rpoβ*35 RNAP on Different Templates:II.* Templates used include: T5 *N25 (FL=57)* {N}*,* T5 $N25_{antiDSR}$ *(FL=57)* {A}*,* P_L *(FL=68)* {PL}*, lacUV5 (FL=58)* {L}*,* T7*A1 (FL=67)* {T}*, and* $rrnBP1_{dis}$ *(FL= 57)* {r}. Reactions were alternated with and without cleavage factor GreB and radioactively labeled with [α-^{32}P]-CTP. Reaction conditions included 30 nM DNA, 1X transcription buffer, 50 mM KCl, 100 μM NTP, 50 nM RNAP or 50 nM RNAP + 500 nM GreB. The transcription reaction ran for 10 minutes at 37°C. Gel was 23% (10:1) polyacrylamide in 1X TBE buffer and 7 M urea and ran for 6 hours at 35 W.


RNAP
RNAP rpoβ*35
Templates:
N A PL L T r N A PL L T r
GreB:
- + - + - + - + - + - + - + - + - + - + - + - +
Lane:
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

DISCUSSION

DNA scrunching is the process that accounts for the energy-stress necessary for RNAP's ability to lose promoter contacts and transition to the elongation phase. Work by Staney and Crothers (1987) first characterized through footprinting that the upstream edge of the DNA remained protected by RNAP between the open complex (RP_o) and the initial transcribing complex (RP_{itc}). This experiment provided the first evidence suggesting the creation of a stressed intermediate and that RNAP does not move down the DNA during initiation. Later work used single-molecule DNA nanomanipulation and demonstrated the process of DNA scrunching; where promoter escape involves scrunching of the DNA to provide force to disrupt promoter contacts and rewind the upstream bubble (Revyakin *et al.*, 2006).

Promoter escape is a highly regulated event and becomes more difficult with strong promoter contacts. Sometimes the mechanism of promoter escape fails. Instead of the upstream edge of the transcription bubble rewinding, the accumulated stress causes the DNA to rewind downstream (Revyakin *et al.*, 2006; Kapanidis *et al.*, 2006). This failure stops transcription and produces a small aborted transcript that backtracks into the secondary channel. Previous studies have found a specific molecule called GreB that plugs into the secondary channel to initiate the cleavage of the backtracked RNA (Opalka *et al.*, 2003).

Through the development of promoters like DG203, the RNAP needs to scrunch even more DNA, thus growing the nascent RNA to even longer lengths, to obtain the necessary strength to loosen promoter contacts. Nevertheless, when these contacts are lost and the upstream bubble rewinds, there is so much more energy released as to propel the RNAP downstream by several nucleotides. This process called forward hyper-translocation leaves the growing RNA 3′end out of register with the active site and located within the RNA exit channel. The RNA molecule cannot be elongated further and is released as a **v**ery **l**ong **a**bortive **t**ranscript (VLAT).

To prove that the VLAT is the result of forward hyper-translocation, I have conducted crosslinking experiments, utilizing a thiol group attached to the 3′ most nucleotide of the VLAT. The photocrosslinkable nucleotide S_4U is the preferred crosslinker because it is a natural analog to uridine, and unlike other nucleotide analogs, can be easily incorporated into RNA (Temiakov *et al.*, 2003). The crosslinking method is a very advantageous procedure that has been successful in the determination of the backtracking mechanism, locating the 3′ end of a backtracked transcript to the secondary channel (Borukhov *et al.*, 1991). The experiment covalently bonds the 3′ terminus of the VLAT to the nearest amino acid in the β′ subunit.

Previous studies from the Hsu lab have confirmed that the *DG203/SPfullcon* promoter would be ideal for a crosslinking investigation because if its high production of VLAT-19, its high function of GreB-resistant

VLATs, and its lack of productive transcription (Chander and Hsu, unpublished results). Its unique ITS: A_{+1}UGCGACCGGGAGAG**GAGU**$_{+19}$, contains only two U's with one at the 5′-end and the other at the 3′ most end. This sequence allows me to incorporate a photocrosslinkable nucleotide by initiating with ApU and terminating with S_4U. To ensure that the incorporation of the ApU and S_4U did not disrupt transcription, I did some preliminary transcription reactions. ApU initiation presented with two issues. First, the RNAs generated lacked the 5′-triphosphate group and traveled with anomalous mobility. Additionally, without the 5′-triphosphate group I had to label the abortive transcripts with [α-^{32}P] NTP. Due to this labeling, both the abortive transcripts and cleavage products are were visible on the transcription gels, making them more complicated to read. After multiple combinations of different NTP's and different [α-^{32}P] labels, I was able to figure out many of the bands and ensure the initiation with ApU did not present with any problems. Furthermore, I was able to tag the terminal VLAT position by the incorporation with the photocrosslinkable S_4U. This incorporation did not present with any problems.

After determining that the incorporation of ApU and S_4U did not disrupt transcription, I moved to techniques to increase the amount of VLATs to ensure the greatest yield after crosslinking. I changed the concentration of S_4U and RNAP and found through ImageQuant that a concentration of 100 μM S_4UTP and 150 nm RNAP and 1.5 mM GreB produced the highest yield.

Additionally, I noticed that VLATs-18 and -17 were higher in intensity than VLAT-19. Thus, I created new promoters that moved the U to the 18th and 17th positions respectively. I performed preliminary transcription reactions to ensure that these VLATs could also initiate with ApU and terminate with S_4U.

After finding the right conditions that produced the highest amount of VLATs, I moved to the crosslinking studies. My first study found that the crosslinked product could not be detected on a 23% polyacrylamide transcription gel. Instead, I analyzed my crosslinking experiments by denaturing SDS-PAGE gels. These gels are only 8% polyacrylamide, allowing for the entrance of the crosslinked product into the gel. I was able to see RNAP-, S_4U-, and UV- dependent crosslinking to the β/β′ subunit(s) with and without GreB. These subunit(s) line the RNAP exit channel, thus we expected crosslinking here. Crosslinking in the presence of GreB indicates that some portion of the VLATs are not crosslinked in the secondary channel and possibly ending up in the RNA exit channel, supporting the potential existence of forward hyper-translocation.

Although we suspect that the crosslink is to the RNA exit channel, knowing the exact amino acids of the crosslink would be definitive proof. Mass spectrometry has become an essential tool for biochemists in the postgenomic era. It can identify proteins, elucidate post-translational modifications, and read functional interactions (Chait, 2011). To this end I

wanted to perform matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) mass spectrometry (ms), which is the best mass spectrometry approach for biomolecules. Both MALDI and electrospray ionization (ESI) are soft ionizers, meaning ionization does not break peptide bonds. However, MALDI-TOF singly protonates the sample, producing a $(M+H)^+$ ion, where as other methods like ESI produces multiply protonated cations from +1 to +40 (Hillenkamp and Peter-Katalinic, 2007). After the cationic peptides are sent through the time of flight mass analyzer, the molecules hit a detector that reads the sample on its specific mass to charge ratio (*m/z*). By using MALDI, which singly protonates the peptides, the spectra is normalized, separating the peptides solely on mass. Thus, the use of MALDI as an ion source produces much simpler spectra to read and analyze.

To prepare my sample I wanted to enrich the amount of crosslinking to obtain the highest yield. Additionally, I wanted to extract the crosslinked product from the SDS-PAGE gel, remove any radioactivity and decrease the crosslink size. To enrich the crosslinked product, I first began by examining different crosslink-recovery techniques, comparing TCA precipitation (indirect), to direct loading of the reaction volume, to GES precipitation. I experimented with both the *DG203/SPfullcon-U2/U19* and *–U2/U18* templates. I found TCA precipitation was the most successful technique. Furthermore, there was little difference in the amount of crosslinking for either templates.

Beyond establishing the best conditions for recovery of the crosslinked product, I tested whether increasing the UV exposure for 1 hour would increase the amount of crosslinked product. I was successful in obtaining more crosslinked product. Additionally, I wanted to see if I could increase the reaction volume and obtain more crosslinked product. I was able to increase the reaction volume to 100 μL and saw a dramatic increase in crosslinked product.

Furthermore, I tested different photocrosslinkable nucleotides including S_2U, Br_5U, and I_5U. S_2U did not produce any photocrosslinking, the incorporation of Br_5U and I_5U gave inconclusive results. I would suggest repeating this study. Furthermore, I tested different UV irradiation wavelengths at 365 and 302 nm, which also gave inconclusive results. I would recommend repeating this study because the gel images were also not very clear.

After enriching the crosslinked product, I excised the crosslinked product from the gel, treated with nucleases to remove all radioactivity, and reduce the size of the crosslinked product. To accomplish these tasks I performed in-gel digestion. This technique with the combination of small nucleases like MNase and RNase T1 could penetrate into the gel and cleave to remove the radioactively labeled RNA containing [α-^{32}P]-CMP. Tryptic digest would then break the β/β′ into smaller peptides to identify where the crosslink is to. I only tested MNase digestion and was partially successful. I

was able to decrease the amount of radioactivity to about 23%. Nevertheless, we want to remove all radioactivity. Additional studies with RNase T1 or a combination of MNase with RNase T1 could remove all radioactivity. Tryptic digest would be the next logical step. If removing the radioactivity does not work, I could run the reaction without radioactivity and approximate the crosslinked product location. There is a distinctive shift in the SDS-PAGE gel between the uncrosslinked and crosslinked β/β′. I am able to see the uncrosslinked β/β′ through Coomassie Blue staining of an RNAP control and approximate the distance above this band.

The production of abortive RNAs is a key process during transcription initiation and can significantly decrease the amount of full-length transcripts. Promoters in this crosslinking study including DG203/*SPfullcon –U2/U19*, -*U2/U18*, and -*U2/U17* produce nearly no full length RNA. If such a promoter existed in a bacterium, the accumulation of abortive transcripts could have potential negative effects or even regulatory roles. It has been shown that abortive initiation occurs *in vitro* in all characteristic RNAPs, including: bacterial, archaeal, eukaryotic, and bacteriophage (Goldman *et al.*, 2009). Additionally, it had been shown that abortive initiation occurs *in vivo* in *E. coli.* Because abortive transcripts reach detectable levels in *E. coli*, Goldman *et al.* (2009) suggests that abortive transcripts could have a regulatory role. Recently, abortive transcripts have been shown to form antiterminator hairpin loops in the phage T7 (Lee *et al.* 2010). Here the abortive transcript

prevents the termination of transcription for a gene. Abortive transcripts play an integral part in the transcription initiation mechanism and new developments have shown their importance in other biological processes. Through understanding the mechanisms of abortive initiation, both backtracking and forward hyper-translocation, we gain greater insight into the machinery of RNAP and transcription initiation.

REFERENCES

Borukhov, S., Jookyung, L. and Goldfarb, A. (1991) Mapping of a contact for the RNA 3′ terminus in the largest subunit of RNA polymerase. *J. Biol. Chem.* **266**, 23932-23935.

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

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

Borukhov, S., Lee, J., and Goldfarb, A. (1996) Mapping of a contact for the RNA 3′ terminus in the largest subunit of RNA polymerase. *J. Biol. Chem*. **266**, 23932-23935.

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.

Carpousis, A. J., Stefano, J. E., and Gralla, J. D. (1982) 5′ nucleotide heterogeneity and altered initiation of transcription at mutant lac promoters. *J. Mol. Biol*. **157**, 619-633.

Chait, B. (2011) Mass spectrometry in the postgenomic era. *Annu. Rev. Biochem.* **80**, 239-246.

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

Chander, M. and Hsu, L. M. An AT-rich spacer enhances intrinsic and αCTD-dependent release of very long abortive transcripts (VLATs) during promoter escape, manuscript in submission.

deHaseth, P. L., Zupancic, M. L., and Record, M. T., Jr. (1998) RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *J. Bacteriol*. **180**, 3019-3025.

Dutta, D., Shatalin, K., Epstein, V., Gottesman, M, and Nudler, E. (2011) Linking RNA polymerase backtracking to genome instability in *E. coli*. *Cell* **146**, 533-543.

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., Ross, W., Estrem, S. T., Nguyen, L. H., Burgess, R. R., and Gourse, R. L. (2001) Promoter recognition and discrimination by EsigmaS RNA polymerase. *Mol. Microbiol*. **42**, 939-954.

Goldman, S., Ebright, R., and Nickels, B. (2009) Direct detection of abortive RNA transcripts in vivo. *Science* **324**, 927-928.

Hatoum, A., and Roberts, J. (2008) Prevalence of RNA polymerase stalling at *Escherichia coli* promoters after open complex formation. *Mol. Microbiol*. **68**, 17-28.

Hillenkamp, F., Peter-Katalinic, J. *MALDI MS: A Practical Guide to Instrumentation, Methods and Applications*. Weinheim, Germany: Wiley-VCH, 2007.

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., 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.

Kammerer, W., Deuschle, U., Gentz, R., and Bujard, H. (1986) Promoters of *Escherichia coli*: a hierarchy of in vivo strength indicates alternate structures. *EMBO J.* **5**, 2995-3000.

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

Knaus, R. and Bujard, H. in: Eckstein, F. and Lilley, D. M. (Eds.) (1990) Nucleic Acids and Molecular Biology. Springer-Verlag, Heidelberg, Germany. **4**, 789-800.

Kommissarova, 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.

Laptenko, O., Lee, J., Lomakin, I., and Borukhov, S. (2003) Transcript cleavage factors GreA and GreB act as transient catalytic components of RNA polymerase. *EMBO J.* **22,** 6322-6334.

Lee, S., Nguyen, H. M., and King, C. (2010) Tiny abortive initiation transcripts exert antitermination activity on an RNA hairpin-dependent intrinsic terminator. *Nucleic Acids Res.* **38**, 6045-6053.

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

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

McClure, W. R., Cech, C. L., and Johnston, D. E. (1978) A steady state assay for the RNA polymerase initiation reaction. *J. Biol. Chem.* **253**, 8941-8948.

Mulligan, M. E., Hawley, D. K., Entriken, R., and McClure, W. R. (1984) Escherichia coli promoter sequences predict in vitro RNA polymerase selectivity. *Nucleic Acids Res*. **12**, 789-800.

Murakami, K. S., Masuda, S., and Darst, S. A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. *Science* **296**, 1280-1284.

Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V., and Ebright, R. H. (2000) Structural organization of the RNA polymerase-promoter open complex, *Cell* **101**, 601-611.

Nelson, D. L. and Cox, M. M. *Principles of Biochemistry*. New York: W.H. Freeman and Company, 2008.

Opalka, N., Chlenov, M., Chacon, P., Rice, W. J., Wtiggers, 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., Liu, C., Ebright, R. H., and Strick, T. (2006) Abortive Initiation and Productive Initiation by RNA Polymerase Involve DNA Scrunching. *Science* **314**, 1139-1143.

Steitz, T. A. and Steitz, J. A. (1993) A general two-metal-ion mechanism for catalytic RNA. *Proc. Natl. Acad. Sci. USA* **90**, 6498-6502.

Stepanova E., Lee, J., Ozerova, M., Semenova, E., Datsenko, K., Wanner, B. L., Severinov, K., and Borukov, S. (2007) Analysis of promoter targets for *Escherichia coli* transcription elongation factor GreA in vivo and in vitro. *J. Bacteriol*. **189**, 8772-8785.

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.

Surratt, C. K., Milan, S. C., and Chamberlin, M. J. (1991) Spontaneous cleavage of RNA in ternary complexes of *Escherichia coli* RNA polymerase and its significance for the mechanism of transcription. *Proc. Natl. Acad. Sci. USA* **88**, 7983-7987.

Temiakov, D., Anikin, M., Ma, K., Jiang, M., and McAllister, W. (2003) Probing the organization of transcription complexes using photoreactive 4-thiol-substituted analogs of uracil and thymidine. *Methods Enzymol.* **371**, 133-143.

Vassylyev, D. G., Temiakov, D., Zenkin, N., Vassylyeva, M. M., Perederina, A., Tahirov, T. H., Kashkina, E., Savkina, M., Zorov, S., Nikiforov, V., Igarashi, N., Matsugaki, N., Wakatsuki, S., and Severinov, K. (2005) Structural basis for transcription inhibition by tagetitoxin. *Nat. Struct. Mol. Biol.* **12**, 1086-1093.

Vo, N., Hsu, L. M., Kane, C., 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.

Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K., and Darst, S. A. (1999) Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98**, 811-824.