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

Open Complex Contacts: An Analysis of the T5 *N25* Discriminator Region Interaction with *E. coli* RNA Polymerase

by

Cassandra White

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, 2010

This paper was prepared under the direction of
Professor Lilian Hsu for eight credits.

This research is funded by NSF Grant (RUI # MCB0841452)
Given to Professor Lilian Hsu

ACKNOWLEDGMENTS

To Professor Lilian Hsu, thank-you for your amazing support as a project advisor and as a professor. I have had a thoroughly awesome experience with you as my research mentor, guiding me through my years.

To my family, thanks for always believing in me and propping me up when I was falling, your helping hands throughout my research experience have been invaluable.

To my lab mates: Ahri Lee, Doris Tabassum, and Tenaya Vallery, it has been an enjoyable experience working with each of you, we've shared many laughters together, …beams.

To Professor Gary Gillis and Professor Sarah Bacon, thanks for joining my thesis committee and always being there to answer my unending questions and decoding my future.

To Steve Levery, thanks for agreeing to run samples for me on your mass spectrometers and aiding me through many research decisions.

To my friends, thanks for keeping me sane.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

E. coli RNA polymerase (RNAP) is known to contact a promoter via the UP element and its -35 and -10 hexameric sequences during the first stage of transcription to transition into a catalytically competent open complex. In this binding-activation process, strong promoter-RNAP interaction provides stability to the open complex, resulting in a high initiation rate. Bacteriophage T5 *N25* promoter is such an example where strong contacts at the open complex stage pose a high barrier for escape. As a result, RNAP undergoes repeated abortive cycling releasing short transcripts, prior to ultimately leaving the promoter and transitions into the elongation phase.

Recent research by Haugen *et al.* (2006, 2008) revealed that the discriminator (DIS) region makes an additional direct contact with RNAP through the nontemplate strand G residue positioned two nucleotides downstream of the -10 box, to further stabilize an open complex. Through photocrosslinking analysis of this G residue within the *rrnB* P1 $C_{-7}G$ and λ P_R promoters, their study found that contacts were made to the σ1.2 region of *E. coli* RNA polymerase.

Because the *N25* promoter is well know for its open complex stability, we wished to investigate whether this interaction contributes to its stabilization. Photocrosslinking and Western blot analysis completed by J. Wiwczar (2008) revealed that the *N25* promoter DIS region contacts the β' subunit of RNAP. To investigate the difference in the above results, research was initiated by the

construction of *N25* promoters that differ in the three nucleotides immediately downstream of the -10 box: AGA in CW1 (wildtype *N25*), GGA in CW2 (similar to GGT in λ P_R) and GGG in CW3 (matching GGG in $C_{-7}G$ *rrnB* P1). The goal was to determine whether contact to RNAP is a function of DIS region sequence composition. The constructed promoters were thereby subjected to UV-induced crosslinking to RNAP via their -5 S_6G residue.

Preliminary results show that all three promoters transcribe very similarly, indicating that the stability of the open complex is minimally affected by the DIS mutations. In crosslinking, all three promoters were found to contact the β' subunit. To further investigate which amino acids of the β' subunit of RNA polymerase are involved in making this contact, we strive to submit the complex for mass spectrometry analysis. To achieve this, the protein-nucleic acid conjugate must be shortened by using restriction enzymes and proteases. We have found that trypsin and *Apo* I digests together give a reasonably sized complex to be analyzed in the future.

In summary, our results indicate that the -5 G nontemplate base within the DIS region of the *N25* promoter was contacted by the β' subunit independent of the sequence of nucleotides bordering the G residue. To map which amino acids within β' are directly involved in contact, a shortened crosslinked complex can be applied for mass spectrometry analysis.

INTRODUCTION

RNA polymerase (RNAP) is responsible for gene transcription resulting in the production of RNA. Transcription occurs in a promoter-specified manner in three consecutive phases: initiation, elongation, and termination. In initiation, the RNAP comes into contact with the promoter via the promoter's recognition signals, which dictate the overall complex stability. RNAP then initiates de novo RNA chain synthesis, undergoes promoter escape by releasing the DNA contacts and transitions into the elongation phase where it synthesizes the RNA transcript as it moves down the gene. RNAP then reaches the terminator sequence that signals the enzyme to release the full-length transcript. A messenger RNA transcript is then translated by the ribosome into a protein.

Regulation of gene expression can occur in any of these three phases, with the initiation phase being the most highly regulated. Initiation is divided into two stages: promoter open complex formation, and abortive initiation and promoter escape. Open complex formation occurs when the promoter and RNAP first interact via the contact regions within the upstream promoter recognition region (PRR), from -60 to the -1 (relative to the +1 start site of transcription), to establish a closed complex (Murakami *et al*., 2002). This is followed by unwinding 14 basepairs (bp) of DNA from the -11 to the +3 positions to form an open complex, as illustrated in Figure 1 (Murakami *et al.*, 2002; Hsu, 2002). There are three well-studied regions within the PRR that control the stability of the open

complex: the UP element located between the -60 bp and -40 bp, -35 hexamer centered at ~35 bp upstream of the +1 start site, and -10 hexamer centered at ~10 bp upstream of the start site (Nelson and Cox, 2005). When the sequences within these signals are close-to-consensus, as shown in Figure 2, the open complex is highly stable with a long half-life. This stage sets the overall initiation frequency– the more stable the open complex, the higher the initiation frequency.

Upon initiation, RNAP may proceed into cycles of abortive initiation over the initial transcribed sequence (ITS) region from +1 to +20 as it attempts to move away from the promoter region to synthesize a full-length RNA (Hsu *et al*., 2003). Short abortive transcripts are produced when RNAP fails to escape from the promoter contacts. Promoters with tightly bound contact regions are correlated with high degree of abortive initiation. Once the promoter contacts are successfully released, the RNAP undergoes promoter escape and begins forming a full-length transcript in the elongation phase. This second stage of initiation determines the fraction of initiated chains that turns into productive synthesis. Clearly, understanding the structure and stability of open complexes is crucial for understanding the activity of a promoter.

Our group studies the highly abortive bacteriophage T5 *N25* promoter. This promoter contains a sequence in close agreement with the *E. coli* $E\sigma^{70}$ consensus, as shown in Figure 2, which is directly responsible for its high rate of initiation (Knaus & Bujard, 1990) and high rate of abortive initiation due to the stable contacts made with RNA polymerase (Vo et al., 2003). Research by

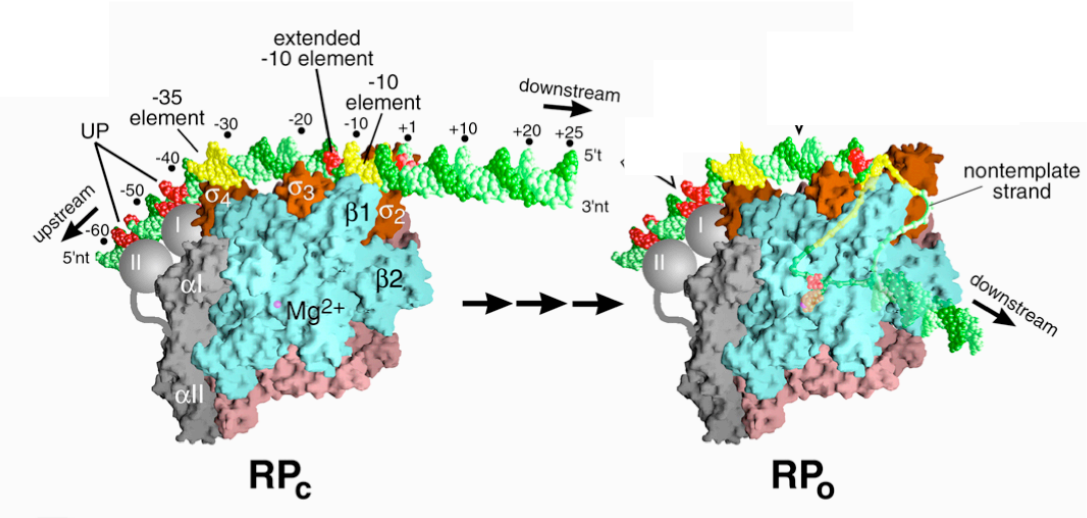


Figure 1. *RNA polymerase transitioning from closed complex to open complex during initiation*. The figure of the closed complex (RP_C) shows regions of the PRR that are in close contact with RNA polymerase. Melting then occurs starting at the -11 position of the -10 box. The unwound template and non-template strands are then pulled into the RNA polymerase to form an open complex (RP_O).

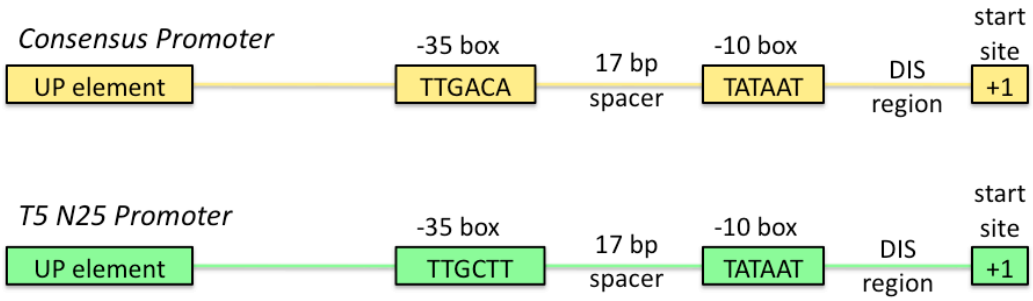


Figure 2. *Consensus sequence of the -35 box and -10 box in comparison with the T5 N25 promoter sequences*. The *E. coli* $E\sigma^{70}$ consensus promoter sequence (shown in yellow) contains -35 and -10 hexamers that form the tightest binding to RNA polymerase during initiation. Promoters with similar sequences to consensus, such as *N25* (shown in green), are known to bind to RNA polymerase strongly to result in stable open complexes with long half-lives; these promoters exhibit high degree of abortive initiation.

Knaus and Bujard (1990) additionally found that the open complex was long-lived with a half-life of 180 min. Because of these observations made about the stability of open complex of the *N25* promoter with RNAP, it was our interest to uncover the precise components of the T5 *N25* promoter that gives it such stability.

RNAP is a highly conserved multisubunit protein whose core enzyme consists of five subunits called α_2, β, β', and ω subunits. The additional σ subunit contains the specificity for directing the RNAP to the promoter and binds to core subunits to form the RNAP holoenzyme. The *E. coli* RNAP can form the holoenzyme with a variety of σ subunits, each specific for recognizing a different promoter signal controlling the expression of a group of genes. Our research focuses on the σ^{70} (M_r 70,000) initiation factor which recognizes the housekeeping genes responsible for maintaining growth and homeostasis within *E. coli*.

The sites of RNAP that are known to come into contact with the promoter regions during open complex formation are the α subunit C-terminal domain (α-CTD) with the UP element, the σ4 domain with the -35 box, and the σ2 domain with the -10 box (see Figure 1). It has been recently discovered through photocrosslinking that RNAP additionally contacts the promoter at the discriminator (DIS) region, referring to the 6-8 bp between the +1 start site and the -10 box.

Photocrosslinking is a commonly used method for determining close contacts in a protein-protein or protein-DNA complex. By using a thiol (–SH) group attached to one of the nucleotides within the promoter, as shown in Figure 3, we can explore amino acid groups on RNA polymerase that can form an –SR bond; thus, a thiol group is considered a 0-Å crosslinker (Suchanek *et al.*, 2005). In the cases studied below, the nucleotide chosen to be examined for contacts with RNAP is located at the second position downstream of the -10 box within the DIS region. UV-irradiation is required to activate the thiol group for reaction (Alexander and Moroson, 1962; Smith, 1962). The heteroconjugate of the promoter DNA covalently linked to RNA polymerase in the open complex conformation is the resulting product.

Haugen *et al.* (2006) first examined the open complex half life of the *E. coli rrnB* P1 promoter and found that it increased 37-fold after the -7 nontemplate cytosine (C) in the DIS region was mutated to a guanine (G) giving a -8/-7/-6 sequence of GGG. This research group also analyzed the λP_R promoter which has a -6/-5/-4 sequence of GGT (T = thymine). They found that when they mutated the wildtype -5 G non-template residue to a C, the open complex half-life decreased 14-fold. The importance of this base in open complex stabilization was further analyzed by crosslinking a 6-thio-deoxyguanine base at 2 nucleotides downstream from the -10 box. Their results of crosslinking the –7 G residue in the $C_{-7}G$ *rrnB* P1 promoter indicated that it made contact with RNAP via the σ1.2 domain. The thiol base was also placed at the -5 G nontemplate residue in the λ

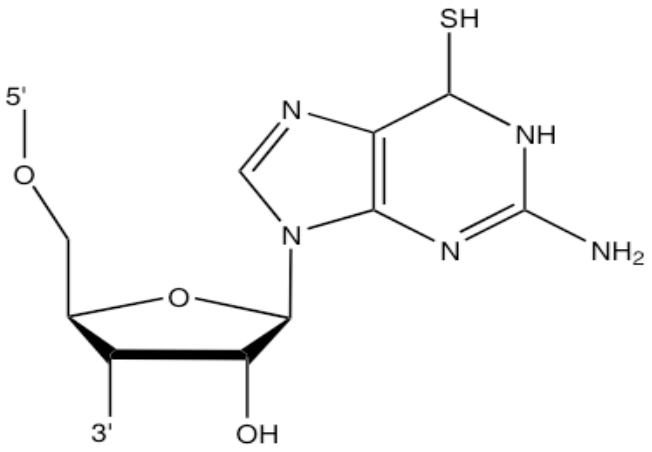


Figure 3. *6-thio-2'-deoxyguanine chemical structure*. This S_6G nucleotide is placed at the -5 nontemplate position within the constructed promoters. When the thiol group is irradiated with 365 nm UV light, it forms a covalent bond to the nearest element.

P_R promoter; when crosslinked, it likewise made contact to the σ1.2 subunit of RNAP (Haugen *et al.,* 2008). These researchers concluded that a G base positioned 2 nucleotides downstream from the -10 box forms a key interaction with RNAP.

The goal of my research is to examine this DIS region interaction in the bacteriophage T5 *N25* promoter open complex. In addition to the consensus-like -35 and -10 elements, the T5 *N25* promoter has a DIS region that contains a G residue at the -5 position (the 2nd base downstream from the -10 box). This G residue mirrors the G residues in the same position in the $C_{-7}G$ *rrnB* P1 and λ P_R promoters, enabling the long half-life of the stable open complex. Previous research in our lab found that the wild type *N25* promoter, containing a -6/-5/-4 sequence of AGA in its 6-bp long DIS region, crosslinks to the β' subunit through its -5 G residue (Wiwczar, 2008). DIS sequences and open complex half-lives of $C_{-7}G$ *rrnB* P1, λ P_R and *N25* promoters are shown in Figure 4.

Results from these previous investigations show that the G residue 2 positions down from the -10 box is key for the stability of the open complex due to its involvement in direct contact with a RNAP subunit. However, crosslinking of the -7/-5 G residue to different RNAP subunits was unexpected. To resolve the discrepancy of crosslinking between the *N25* promoters and *rrn*B P1 and λ P_R, we were interested in determining whether it's the 8-bp vs. 6-bp DIS length or whether it's the precise sequence of the DIS region that brings it into contact with different RNAP subunits.

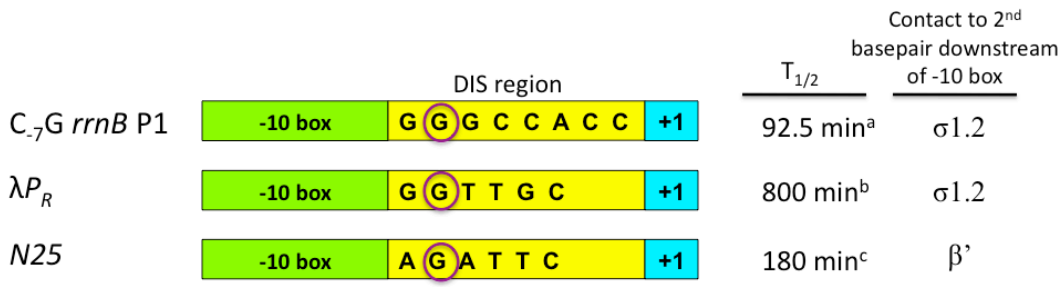


Figure 4. *Photocrosslinking results and half-lives of previously studied DIS regions.* Studies by Haugen *et al.* (2006, 2008) found that 2nd bp downstream from the -10 box in the C-7G *rrnB* P1 and λP_R promoters are both contacted by the σ1.2 subunit. Reasearch by Wiwczar (2008) found that the 2nd bp downstream of the -10 box in the *N25* promoter is contacted by the β' subunit.

[a] Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C., and Gourse, R. L. (2006) rRNA promoter regulation by nonoptimal binding of σ region 1.2: an additional recognition element for RNA polymerase. *Cell* **125**, 1069-1082.
[b] Haugen, S. P., Ross, W., Manrique, M., and Gourse, R. L. (2008) Fine structure of the promoter-σ region 1.2 interaction. *Proc. Natl. Acad. Sci. USA* **105**, 3292-3297.
[c] Knaus, R., and Bujard, H. (1990) Principles governing the activity of *E. coli* promoters. F. Eckstein and D. M. Lilley (eds.) *Nucleic Acids & Molecular Biology* **4**, 110-122. Springer-Verlag, Heidelberg, Germany.

This project began with the goal to examine the sequence composition of the bordering nucleotides around the -5 G residue on RNAP crosslinking by constructing *N25* promoter variants with mutations within the DIS region. The three promoters we studied-- *N25* (CW1), *$A_{-6}G$ N25* (CW2), and *$A_{-6}G/A_{-4}G$ N25* (CW3) --differ in the -6/-5/-4 nucleotides, changing from AGA in CW1 (which is identical to N25), to GGA in CW2 and GGG in CW3. The full sequences of these promoters are shown in Figure 5.

Our work reported here shows that the CW1 promoter as well as its modified CW2 and CW3 versions all crosslink to the β' subunit. Thus, it appears that the crosslinking of the N25 DIS region to the β' subunit is independent of the sequence composition and context. We decided to pursue experiments to pinpoint the exact location of the β' crosslink. Despite the enormous recent progress on solving RNAP structure, its interaction with the transcription bubble strands of promoter DNA has not been elucidated (Murakami et al., 2002). In aiming to decipher the β'-DNA crosslink, the size of the β' subunit poses an immediate challenge. The β' subunit is the largest of the RNAP subunits with a mass of 155 kDa. It is closely followed by the β subunit with a mass of 151 kDa. These two subunits together compose a clamp around the promoter, similar to two shells of a clam, to hold the template strand within the RNAP active site. The fact that their sizes are closely related pose difficulty in separation and recovery of the distinct subunits.

CW1

```
    -85
5′- GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT
3′- CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC CCTTTAGTAT TTTTTAAATA

    TTGCTTTCAG GAAAATTTTT CTGTATAATA GATTCATAAA TTTGAGAGAG
    AACGAAAGTC CTTTTAAAAA GACATATTAT CTAAGTATTT AAACTCTCTC

    GAGTTTAAAT ATGGCTGGTT CTCGCAGAAA GCTTCTGCAG CC -3′
    CTCAAATTTA TACCGACCAA GAGCGTCTTT CGAAGACGTC GG -5′
                                                 +57
```

CW2

```
-85
5′- GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT
3′- CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC CCTTTAGTAT TTTTTAAATA

    TTGCTTTCAG GAAAATTTTT CTGTATAATG GATTCATAAA TTTGAGAGAG
    AACGAAAGTC CTTTTAAAAA GACATATTAC CTAAGTATTT AAACTCTCTC

    GAGTTTAAAT ATGGCTGGTT CTCGCAGAAA GCTTCTGCAG CC -3′
    CTCAAATTTA TACCGACCAA GAGCGTCTTT CGAAGACGTC GG -5′
                                                 +57
```

CW3

```
-85
5′- GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT
3′- CCGAGCTCCT TAAGGGCCCC TAGGAAGCTC CCTTTAGTAT TTTTTAAATA

    TTGCTTTCAG GAAAATTTTT CTGTATAATG GGTTCATAAA TTTGAGAGAG
    AACGAAAGTC CTTTTAAAAA GACATATTAC CCAAGTATTT AAACTCTCTC

    GAGTTTAAAT ATGGCTGGTT CTCGCAGAAA GCTTCTGCAG CC -3′
    CTCAAATTTA TACCGACCAA GAGCGTCTTT CGAAGACGTC GG -5′
                                                 +57
```

Figure 5. *Sequences of CW1, CW2, and CW3 promoters constructed.* The red sequence within the promoters is the DIS region. This region was altered to examine the context around the -5 G nontemplate residue and its effect on contacts with RNAP.

Two approaches have been used in the past to define amino acid contacts between protein and DNA heterconjugate complexes: partial proteolytic digests of a tagged protein-nucleic acid conjugates (Severinov *et al.*, 1992) and mass spectrometry (MS) analysis (Jensen *et al*. 1996). The former method is cumbersome, especially where a large protein (i.e. β'subunit) is involved, and at best, it allows localization of a crosslinked probe to a proteolytic fragment of the protein. MS of protein-nucleic acid heterconjugates have been shown to be a more advanced and precise technique (Golden *et al.*, 1999). However, submission of crosslinked heteroconjugates introduces a few problems due to their conflicting requirements for ionization (Jensen *et al*. 1996) and large molecular weights. Sequencing of peptides by MS requires positive ionization, whereas sequencing of nucleic acids requires the negative ion mode (Golden *et al.*, 1999). On the other hand, heteroconjugates have been successfully analyzed by both MALDI-TOF and ESI on quadrupole MS after the molecular weight of the complex is reduced to below 50 kDa. Utilization of proteases and/or restriction enzymes on the crosslinked heterconjugate have been found to cut away non-crosslinked portions of both the protein and DNA sequence. This allows only a crucial contact region of the crosslinked DNA and protein to be prepared for submission for MS analysis.

A part of my research effort has been focused on testing reagents that can reduce the DNA-RNAP heteroconjugate to a suitable size for MS analysis. My

results indicate that the tryspin protease and *Apo* I restriction enzyme are possible candidates for decreasing the overall molecular weight of t

MATERIALS

Reagents

New England Biolabs (NEB) Buffer 2 [1X: 50 mM NaCl, 10 mM Tris-Cl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9], and dNTPs were used in primer extension reactions for constructing the promoters. The promoter's presence was confirmed by an agarose gel and ran with NEB 100-bp ladder (100-1517 bp) and stained with ethidium bromide (EtdBr) from Fischer Scientific. Various reagents including 3 M NaAc, 95% ethanol, 5 M NH_4Ac, phenol chloroform isoamyl alcohol (ϕOH/$CHCl_3$/IsoamylOH, 25:24:1), chloroform isoamyl alcohol ($CHCl_3$/isoamylOH, 24:1) and isopropanol were used in the ethanol precipitation procedures. TE buffer [10 mM Tris-HCl, 1 mM Na_2EDTA, pH 8.0] and STE [1X TE, 0.1 M NaCl] were used to (re)dissolve DNA. Transcription Buffer III [1X: 50 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, and 10 μg/mL acetylated BSA] and KCl were reagents used for transcription reactions. GES [1 mg/mL glycogen, 10 mM Na_2EDTA, 0.3 M NaAc] mix was used to stop the transcription reactions. Polyacrylamide gels were created using either 40% acrylamide-bisacrylamide (10:1) stock for denaturing transcription gels, or 40% acrylamide-bisacrylamide (19:1) stock (from Fisher) for native gels, or 40% acrylamide-bisacrylamide (37.5:1) stock (from Fisher) for SDS-PAGE, with N,N,N'N'- tetramethylethylene diamine (TEMED; Fisher), ammonium persulfate (APS), and TBE buffer [1X: 89 mM Trise Base, 89 mM Boric acid, 2.5 mM

Na_2EDTA, pH 8.3]. Prestained Protein Marker (10-230 kDa) and 3X Red Loading Buffer Pack [87.5 mM Tris-HCl (pH 6.8), 6% (w/v) SDS, 30% glycerol and 0.03% (w/v) phenol red] were both from NEB and used for SDS-PAGE analysis. Laemmli blue indicator buffer [0.1 M Tris-HCl, pH 6.8, 10% SDS, 20% glycerol, 0.2 g bromophenol blue (BPB) dye from Sigma] was also used for SDS-PAGE analysis. The SDS-PAGE gels were stained using Coomassie blue dye R-250 from Sigma and destained using destaining solution I [20% methanol, 7.5% glacial acetic acid]. NEBuffer 2 [1X: 50 mM NaCl, 10 mM Tris-Cl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9] and NEBuffer 3 [1X: 100 mM NaCl, 50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM Dithiothreitol, pH 7.9], and trypsin buffer [1X: 50 mM Tris-HCl, 20 mM $CaCl_2$, pH 8.0] were used in *Apo* I and trypsin digests, respectively.

Enzymes

Klenow polymerase from NEB was used in the primer extension procedure. RNA polymerase PC-45 (5.6 μM on 3/28/09) was purified by Monica Chander and Lilian Hsu and used in transcription reactions. Modified Trypsin (TPCK-treated) proteases were purchased from NEB and reconstituted with ddH_2O.

DNA

Oligonucleotide primers DT6-u (XE), CW1-d (PH), CW2-d (PH), CW3-d (PH), and KH1 were from Integrated DNA Technologies (IDT); their sequences are shown in Figure 6. The oligo primers KH2-S_6G, KH2-S_6G-1, and KH2-S_6G-2 were from TriLink Biotechnologies, with sequences included below. Individual dNTPs and dNTP mixes were from either NEB or USB. Radioactively labeled [γ-^{32}P]-ATP (NEG-002Z), [α-^{32}P]-ATP (NEG-003H), and [α-^{32}P]-dATP (NEG-512H) were from Perkin Elmer Life and Analytical Sciences.

DT6-u (XE), from -85 to -13: 5′- GGCTCGAGGA ATTCCCGGGG ATCCTTCGAG GGAAATCATA AAAAATTTAT TTGCTTTCAG GAAAATTTTT CTG -3′

CW1-d (PH), from +57 to -33: 5′- GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAA *TCT*ATTATAC AGAAAAATTT TCCTGAAAGC -3′

CW2-d (PH), from +57 to -33: 5′- GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAA *TCC*ATTATAC AGAAAAATTT TCCTGAAAGC -3′

CW3-d (PH), from +57 to -33: 5′- GGCTGCAGAA GCTTTCTGCG AGAACCAGCC ATATTTAAAC TCCTCTCTCA AATTTATGAA *CCC*ATTATAC AGAAAAATTT TCCTGAAAGC -3′

KH2-u (from -60 to -7): 5′- TCGAGGGAAA TCATAAAAAA TTTATTTGCT TTCAGGAAAA TTTTTCTGTA TAAT-3′

KH2-S_6G-u (from -60 to -2): 5′- TCGAGGGAAA TCATAAAAAA TTTATTTGCT TTCAGGAAAA TTTTTCTGTA TAATA**X**ATT -3′

KH2-S_6G-1-u (from -60 to -2): 5′— TCGAGGGAAA TCATAAAAAA TTTATTTGCT TTCAGGAAAA TTTTTCTGTA TAATG**X**ATT -3′

KH2-S_6G-2-u (from -60 to -2): 5′— TCGAGGGAAA TCATAAAAAA TTTATTTGCT TTCAGGAAAA TTTTTCTGTA TAATG**X**GTT -3′

KH1-1-d (from +57 to -33): 5′— GCGAGAACCA GCCATATTTA AACTCCTCTC TCAAATTTAT GAATCTATTA TACAGAAAAA TTTTCCTGAA AGCAAATAAA TTTTTTATGA TTTCCCTCGA -3′

Figure 6. *Sequences of primers used to construct promoters.* DT6-u was used as the upstream primer to be annealed with either KH1-d downstream primer to make the *N25* wildtype control promoter, or CW1-d to make the CW1 promoter, or CW2-d to make the CW2 promoter, or CW3-d to make the CW3 promoter. When the S_6G-marked promoters were constructed, the KH1-1-d was used as the downstream primer with either the KH2-u to make an unmarked *N25* promoter as control, or the KH2-S_6G-u to make a marked CW1-S_6G promoter, or KH2-S_6G-1-u to make the CW2-S_6G promoter, or KH2-S_6G-2-u to make the CW3-S_6G promoter.

METHODS

Formation of unmarked promoters CW1, CW2, and CW3

The wildtype *N25* promoter (also known as CW1 promoter) was constructed by first annealing two long overlapping primers DT6-u (the upstream non-template strand from -85 to -13) and CW1-d (the downstream template strand from +57 to -33). This was achieved by placing 10 μL of 10 μM of the primers in a 50 μL reaction in the presence of 1X NEBuffer 2 [50 mM NaCl, 10 mM Tris-Cl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9] and incubating for 20' at 70 °C, 10' at 55 °C, 10' at 42 °C, and 10' at 37 °C. The annealed primers were then subjected to the primer extension reaction by mixing with 50 μL of a 1X NEBuffer 2 solution containing 250 μM dNTPs and 3 μL Klenow Polymerase (10 units/μL) at 37 °C for 30'. This gave a complete double-stranded promoter from -85 to +57.

The CW2 promoter (using the DT6-u upstream primer with the CW2-d downstream primer) and CW3 promoter (using DT6-u with CW3-d primers) were constructed using the same primer annealing and extension method. Once all three promoters were obtained, 3 μL aliquots were mixed with 7 μL of 1X glycerol dye [10 mM Na_2EDTA, 0.02% bromophenol blue, 0.02% xyelene cyanol, and 4% glycerol] and then run on a 2% agarose gel with a 100-bp DNA ladder to confirm their presence and correct length.

The double-stranded full length promoters were then concentrated by ethanol precipitation where the promoter solution was adjusted to 0.3 M NaAc

and ethanol added to a final concentration of 70%. This reaction was left overnight at -20 °C. It was then centrifuged and the pellet resuspended in 200 μL TE. The samples were then subjected to phenol-chlorofom extraction where 200 μL of phenol-chloroform-isoamyl alcohol (25:24:1) was added, vortexed, and the aqueous layer transfered to a new tube. The extraction was repeated using 200 μL of chloroform-isoamyl alcohol (24:1). The final aqueous solution was made 0.3M NaAc and resuspended in an equal volume of isopropanol and left overnight at -20 °C. The solution was then centrifuged and pellet rinsed with 500 μL of ice-cold 70% ethanol. The supernatant was then removed and pellet dried for 15' in the SpeedVac rotatory vacuum desiccator. Finally, the pellet was resuspended in 200 μL of TE and concentration determined by the NanoDrop spectrophotometer.

Transcription

The promoters were first analyzed by their abortive-productive transcription properties through *in vitro* transcription reactions. A background control without RNA polymerase, and standard reactions with known templates (i.e. *N25* and $N25_{anti}$) were also prepared. Each reaction contained 30 nM of promoter DNA with final concentrations of 1X Transcription Buffer III, 200 mM KCl, 100 μM NTP containing [γ-^{32}P]-ATP at a specific activity of ~10 cpm/fmol, and DEPC H_2O, first made up to a volume of 9 μL. The reaction was initiated with the addition of 1 μL of 560 nM RNA polymerase (PC-45) which was freshly diluted with EcoRP diluent [10 mM Tris-HCl (pH 8), 0.2 mM β-mercaptoethanol,

10 mM KCl, 5% (v/v) glycerol, 0.1 mM Na_2EDTA, 0.4 mg/mL acetylated BSA, and 0.1% (v/v) Triton X-100] in a 30" stagger and incubated at 37 °C for 10'. The reactions were terminated with the addition of 100 μL of GES and precipitated with 330 μL ethanol overnight at -20 °C. The products were then centrifuged at 4 °C for 15', the supernatant was removed, and the pellet dried for 15' in the SpeedVac desiccator. Pellets was thoroughly resuspended in 10 μL of formamide loading buffer (FLB: 80% deionized formamide, 1x TBE, 10 mM Na_2EDTA, 0.08% xylene cyanol, 0.08% amaranth) and 4-μL aliquots were loaded into a thin 23% polyacrylamide (10:1) denaturing gel in the presence of 7 M urea and run at 35 watts until the amaranth dye has reached within 1 cm from the bottom of the gel. The gel was then exposed to a phosphorimager screen overnight. The screen was scanned, and product bands were quantified using ImageQuant.

Formation of ^{32}P-labeled promoters

The oligonucleotide KH1 was used as the downstream primer with either the oligonucleotide KH2 (for *N25*) or the oligonucleotide KH2-S_6G (for CW1-S_6G) or the KH2- S_6G-1-u (for CW2- S_6G), or the KH2- S_6G-2-u (for the CW3-S_6G) as upstream primers. Primer annealing was performed under the same conditions as described above for the CW1, CW2 and CW3 promoters, except the final reaction volume was 25 μL and the experiment (using S_6G- containing reagents) was carried out in dim light. This was followed by primer annealing by

mixing with a 25 μL reaction containing 0.2 mM dNTPs, 2.5 μL of 3.3 μM [α-^{32}P]-dATP, and 1X NEBuffer 2 and incubated with 1.5 μL of Klenow polymerase at 37 °C for 30'. An aliquot of the full length promoters were run on a 2% agarose gel with a 100-bp ladder to confirm the presence and correct length of the promoter DNA. The promoters then underwent ethanol precipitation with 0.3 M NaAc and left overnight at -20 °C. The solution was then centrifuged and dried in the SpeedVac desiccator. Pellets were resuspended in 45 μL of STE buffer. The samples were made 1X in glycerol loading dye and loaded onto a 6% (19:1) native polyacrylamide gel. The gel was electrophoresed at 200 volts until amaranth dye reached the bottom, and soaked in 1X TE with 0.5 μg/mL EtdBr. It was then visualized on a UV transilluminator. The major DNA bands were cut out and placed into a microfuge tube. The gel piece was crushed and soaked with 1X TE containing 0.3 M NaAc at 37 °C overnight. The supernatant containing leached DNA was recovered after centrifugation, and concentrated by isopropanol precipitation overnight with 50% isopropanol and 0.3 M NaAc at -20 °C. Samples were then centrifuged, the supernatant removed, and the pellet was dried in the SpeedVac desiccator. The pellets were then resuspended in 100 μL TE, followed by phenol-chloroform-isoamyl alcohol and chloroform-isoamyl alcohol extractions. The final aqueous layer was subjected to ethanol precipitation with 0.3 M NaAc and left overnight at -20 °C. After centrifugation, the supernatant was removed and the pellet rinsed with 500 μL of ice-cold 70% ethanol. This mixture was left on ice for 5', the 70% ethanol wash was removed, and the pellet

was dried in the SpeedVac desiccator and finally resuspended in 40 μL TE. The concentration was then determined using the NanoDrop spectrophotometer. Aliquots of the ^{32}P-labeled promoters were converted to STE buffer composition and 1X glycerol loading dye and run on an analytical native 8% (19:1) polyacrylamide gel to confirm their presence (amount), size, and purity. The gel was dried onto Whatman 3MM paper and then exposed to the phosphorimager overnight to be scanned and visualized.

Photocrosslinking ^{32}P-labeled promoter DNA to RNA polymerase

An open complex formation reaction in 10 μL contained 30 nM promoter DNA with 200 mM KCl, transcription buffer III and 50 nM RNAP. The reactions were incubated at 37 °C for 10'. The samples were then transferred to a 96-well plate and irradiated with 365 nm UV light (Spectroline Corporation, model ENF-280C) placed directly on the plate for 20' at RT. After exposure, the 10 μL reaction was mixed with 10 μL 2X Laemmli blue indicator buffer, boiled for ~7' at 95 °C, then loaded into a 4% stacking- 8% resolving SDS PAGE gel with both a protein marker (NEB Protein Ladder) and a RNAP marker. The RNAP marker was prepared by boiling 10 μL of 5 nM RNAP marked with 5μL 2X Laemmli buffer. Four controls were also run: 1) a thiol-marked promoter without UV irradiation, 2) a thiol-marked promoter without RNA polymerase, 3) a non-thiol-marked promoter with RNAP and UV irradiation, and 4) a thiol-marked promoter without UV nor RNA polymerase. The samples were electrophoresed in protein

gel buffer [43.2 g glycine, 9.1 g trisma base, 3.0 g SDS, brought up to 3 L with ddH_2O, pH 8.3] at 100 V while in the stacking gel, and 150 V while in the resolving gel for a total of 2 hrs where the 80 kDa pre-stained protein marker reached the mid-line of the gel. The gel was stained with coomassie blue dye [150 mL methanol, 30 mL glacial acetic acid, 120 mL ddH_2O, 0.75 g Coomassie blue dye R-250] and destained in destaining solution I [900 mL methanol, 150 mL glacial acetic acid, brought up to 2 L with ddH_2O] and dried onto Whatman 3 MM paper. The protein ladder within the gel was marked by radioactive spots. The marked gel was placed in the phosphorimager for overnight exposure to a phosphor screen and then scanned.

Gel shift of ^{32}P-labeled promoters

Open complex formation reactions in 10-μL volume contained 30 nM promoter DNA in 200 mM KCl and 1X Transcription Buffer III. RNA polymerase (1 μL of 560 nM) was next added in a 1' stagger and incubated for 10' at 37 °C. Afterwards, each reaction was added with 1 μL of 10X glycerol dye and immediately loaded into a thin native 4% (37.5:1) polyacrylamide gel running at 50 V. Promoter DNA-only control was also loaded. Once all samples were in the gel, the voltage was increased to 200 V and electrophoresis continued until the bromophenol blue dye reached the bottom of the gel. The gel was then dried and placed into the phosphorimager for overnight exposure and scan.

Enzymatic digestion of photocrosslinked open complexes

Enzymatic treatment of photocrosslinked open complexes was carried out to reduce the size of the protein-DNA conjugate for mass spectrometry analysis. We tried two cutting reagents: restriction enzyme *Apo* I and protease trypsin. Figure 7 shows the *Apo* I cleavage sites on the *N25* promoter, and Table 1 lists the trypsin cleavage sites of the β' subunit. All reactions were performed using the ^{32}P-labeled CW1-S_6G promoter previously crosslinked to RNA polymerase in a 10-μL reaction volume. Each 10-μL reaction contained 5 μL of 200 nM CW1-S_6G promoter in 50 mM KCl, 1X Transcription Buffer III, and 1 μL of RNA polymerase (a new tube of PC-45 at 2.1 μM), incubated at 37 °C for 10', and then irradiated with 365 nm UV light for 20' at RT. A control (uncut) aliquot was transferred to -20 °C upon completion of irradiation, for use in the SDS-PAGE analysis later. For *Apo* I cutting, the 10-μL reaction was adjusted with 1 μL 10X NEBuffer 3, 1 μL BSA (1 mg/mL), and 1 μL *Apo* I (NEB; 10 units/μL) and incubated at 37 °C for 2 hours. For trypsin cutting, the 10-μL aliquot was adjusted with 1 μL of 1X modified trypsin reaction buffer (NEB) [50 mM Tris-HCl, 20 mM $CaCl_2$, pH 8.0] and 1 μL of 210 μM TPCK-modified trypsin (TPCK-trypsin was supplied by NEB as lyophilized powder with specific activity of 2.1 μmol/min/mg and freshly dissolved in ddH_2O prior to use). This reaction was incubated at 25 °C overnight.

Upon completion of all reactions, each sample was mixed with 0.5 volume of 3X NEB Red Loading Buffer, loaded into a 8% SDS-PAGE and

electrophoresed with an RNA polymerase and protein ladder markers at 100 V in the stacking gel and 150 V in the resolving gel. The gel was (sometimes stained/destained prior to being) dried and placed into the phosphorimager for overnight exposure and scanned the following day. In some enzymatic cutting experiments, the samples were analyzed in native 8% (19:1) polyacrylamide gel. In these cases, the samples were adjusted with a tenth volume of 10X glycerol dye prior to loading. Electrophoresis was performed at 150 V until the xylene xyanol dye has migrated ¼ of the length into the gel. The gel was dried, placed into the phosphorimager for overnight exposure and scanned the next day.

Figure 7. *Apo I cleavage sites on the N25 promoter.* This figure shows that there are 4 cleavage sites. The red circled G residue is the location of the crosslink with RNAP, and orange A residues represent those that are radioactive. Complete *Apo* I digestion of the ^{32}P-labeled DNA (142 bp) would yield five fragments of 9, 34*, 20, 25, and 54* bp, of which only the 34-bp and 54-bp bands would be lightly labeled (indicated by asterisks). This figure reveals that there will be 25 bp remaining within the crosslinked construct after digest, with only 2 radioactive adenines.

Table 1. *Positions of trypsin cleavage sites on β' subunit*. Trypsin has a total of 174 cleavage sites. Most of the cleavage sites will result in peptides less than 10 aa in length.

# of cleavage sites	**β' amino acid position of cleavage site**
174	2 6 9 13 21 31 39 47 53 60 66 74 76 77 79 81 87 96 98 99 101 118 123 133 137 156 179 190 202 213 214 215 216 219 220 222 259 270 271 275 278 280 281 293 296 297 311 312 314 321 325 332 334 337 339 345 346 352 362 370 371 384 388 395 398 399 403 417 425 431 445 481 515 521 531 535 538 547 549 551 557 566 570 576 585 598 599 603 610 634 649 650 678 681 692 695 709 715 731 738 744 764 780 781 789 798 799 832 836 838 842 860 881 883 901 905 911 933 943 953 955 959 964 972 978 983 990 992 996 1005 1036 1048 1067 1072 1079 1104 1123 1132 1140 1148 1151 1167 1170 1172 1173 1174 1192 1194 1203 1206 1222 1224 1231 1242 1247 1251 1258 1262 1263 1284 1286 1290 1297 1304 1311 1330 1340 1341 1345 1348 1355 1369 1371 1373

RESULTS

Previous investigations by Haugen *et al.* (2006; 2008) have found that the DIS region of *rrn*B P1 and λ P_R promoters are contacted by RNA polymerase to contribute to the stability of their open complexes. Their results showed that the contact was between the nontemplate G base two nucleotides downstream from the -10 box with the σ1.2 domain of RNAP. These researchers further have pinpointed the exact amino acid residues of σ1.2 domain involved in the crosslinking (Haugen et al., 2008). Our work showed that when the T5 N25 promoter is crosslinked, it instead makes contact between the G base two nucleotides down from the -10 box and the β' subunit of RNAP (Wiwczar, 2008). To initiate investigation into which aspect of the DIS region dictates the precise amino acid contacts on RNAP, I constructed three variants of the *N25* promoter with different sequences at the 5' border of the DIS region. The ultimate goal of my investigation is to determine the precise amino acid contacts within RNAP that are directly involved in crosslinking to the -5 G residue on the *N25* promoter.

The CW1, CW2 and CW3 promoters were constructed by primer annealing and extension. Aliquots of the double-stranded full-length promoters were run on a 2% agarose gel to confirm their presence and correct length of 142 bp. The image is shown in Figure 8. After clean-up, the promoter DNAs were redissovled in 200 μL TE buffer. Concentration of the various DNA samples was

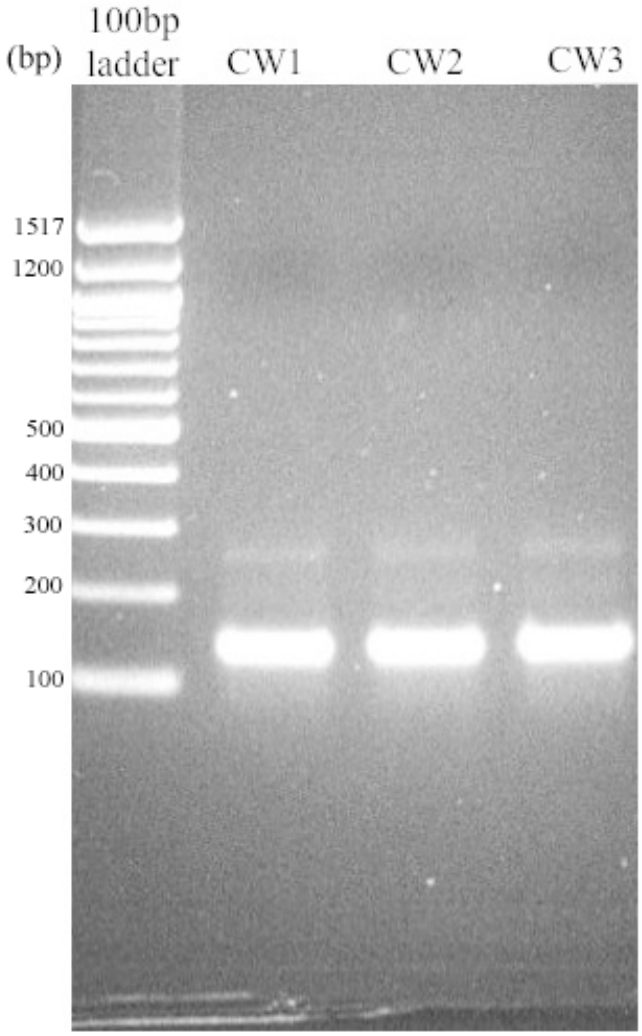


Figure 8. *2% agarose gel with the CW1, CW2 and CW3 promoters constructed by primer annealing.* This gel was run to confirm the correct 142 bp length of the promoters. It was run alongside the 100-bp DNA ladder from NEB.

determined by NanoDrop reading. The final concentration of CW1, CW2, and CW3 DNAs was 760 nM, 1041 nM, and 785 nM, respectively.

Despite the fact that these promoters have slightly different sequences within their DIS region, the -5 G residue is maintained between the three of them. Because this residue has been shown to provide open complex stability, the *N25* promoters with the same residue are expected to give similar transcription patterns. Therefore, once the correct length of the CW1, CW2 and CW3 promoters were confirmed, a transcription analysis was performed as a way to compare their relative open complex stability. This was achieved by a standard steady-state in vitro transcription reaction. The products were run on a denaturing 23% (10:1) polyacrylamide gel in the presence of 7 M urea. The gel image is shown in Figure 9. Abortive probability profiles were compiled from the intensity of each band quantified using ImageQuant for two sets of transcription reactions. The profile is shown in Figure 10. This plot compares the probability of abortive - productive RNA transcript formation between the promoters constructed. The profile reveals that the CW1, CW2 and CW3 promoters have similar abortive transcription patterns. This confirms the prediction that they form open complexes of equal relative stability.

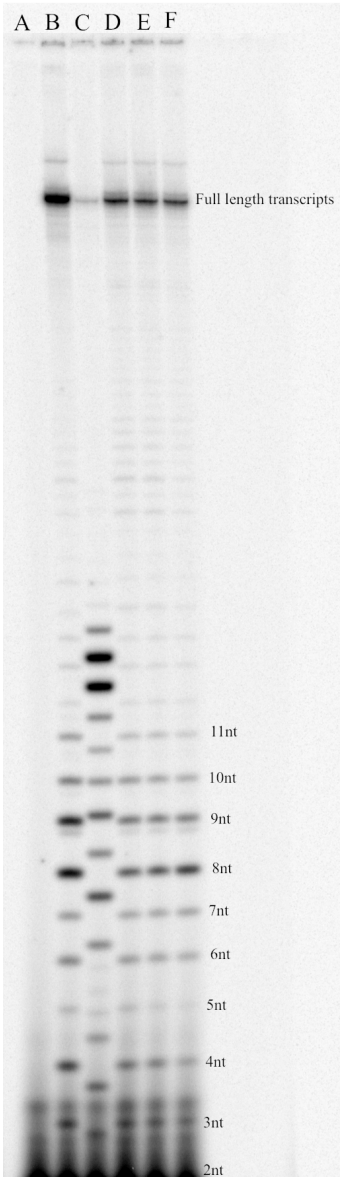


Figure 9. *Gel image of abortive transcripts and full-length transcripts from CW1, CW2 and CW3 promoters*. Lane A is the control without the RNA polymerase enzyme. Lane B contains transcripts formed from the *N25* promoter. Lane C is with the $N25_{anti}$ promoters. Lanes D-F are transcripts derived from the constructed CW1, CW2, and CW3 promoters, respectively.

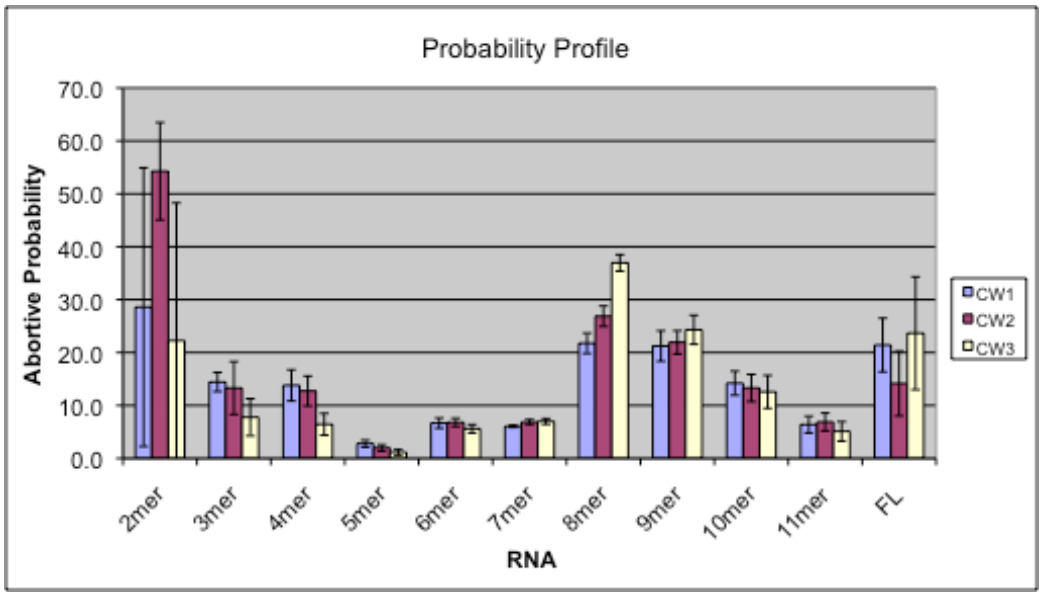


Figure 10. *Compiled abortive profiles of transcription products from CW1, CW2 and CW3 promoters*. Abortive probabilities are calculated according to the equation in Hsu (1996). The result represents the average from two sets of transcription reactions, followed by gel analysis, and ImageQuant quantitation.

Having shown that CW1, CW2 and CW3 promoters transcribe similarly, I was in position to investigate the crosslinking properties of these promoters. I began by constructing the *N25* and *N25*-S_6G pair of promoters studied by J. Wiwczar (2008) to see if I can 1) repeat her finding of the wildtype *N25*-S_6G promoter crosslinking to the β' subunit and 2) increase the crosslinking signal by labeling the DNA to a higher specific activity. To increase the degree of labeling per DNA, I performed the primer extension reaction in the presence of [α-^{32}P]-dATP where radioactive adenines are added into the promoter strands wherever a dAMP is required. In theory, this led to the incorporation of 29 [α-^{32}P]-dAMP residues in each promoter compared to one ^{32}P-phosphate group done by J. Wiwczar (2008) who used [γ-^{32}P]-ATP labeling instead.

Aliquots of the full length *N25* and *N25*-S_6G promoters (100 bp in length) were run on a 2% agarose gel with a 100-bp ladder after their construction, as shown in Figure 11. This gel image confirms their presence and correct length. Steps of purification using a 6% polyacrylamide native gel followed by phenol-chloroform extractions were used to separate the promoter away from side-products. The final concentrations after re-suspension in 40 μL TE were 183 nM and 448 nM of the *N25* and *N25*-S_6G promoters, respectively.

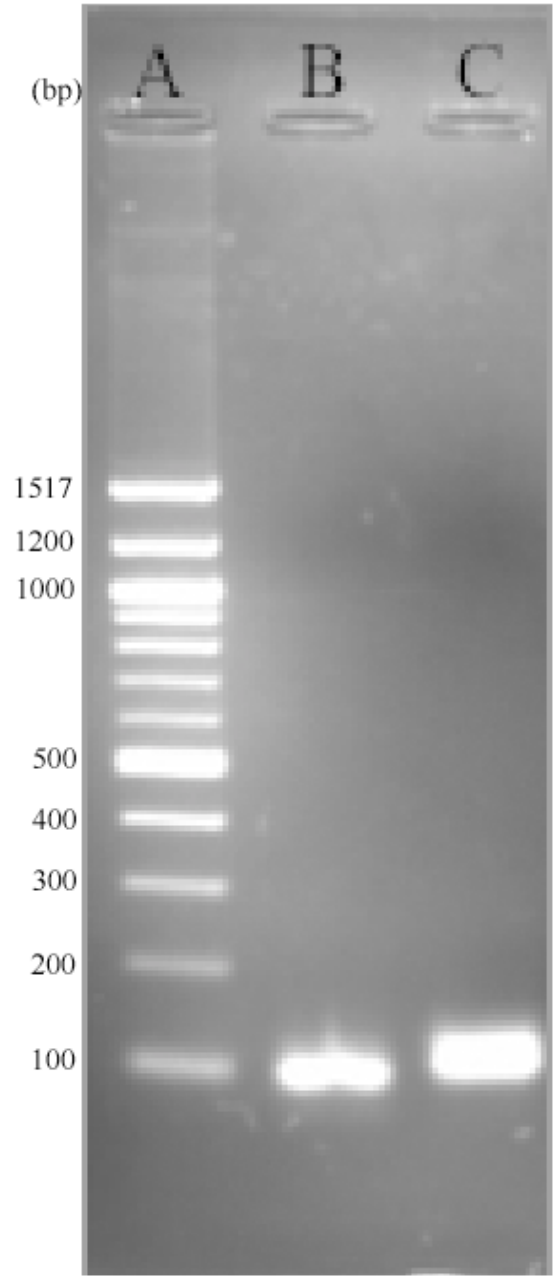


Figure 11. *Agarose gel image of the newly constructed N25 and N25-S_6G promoters.* This gel contains the 100-bp ladder (Lane A) to show that the promoters are of the correct size. Lane B: *N25*; lane C: *N25*-S_6G.

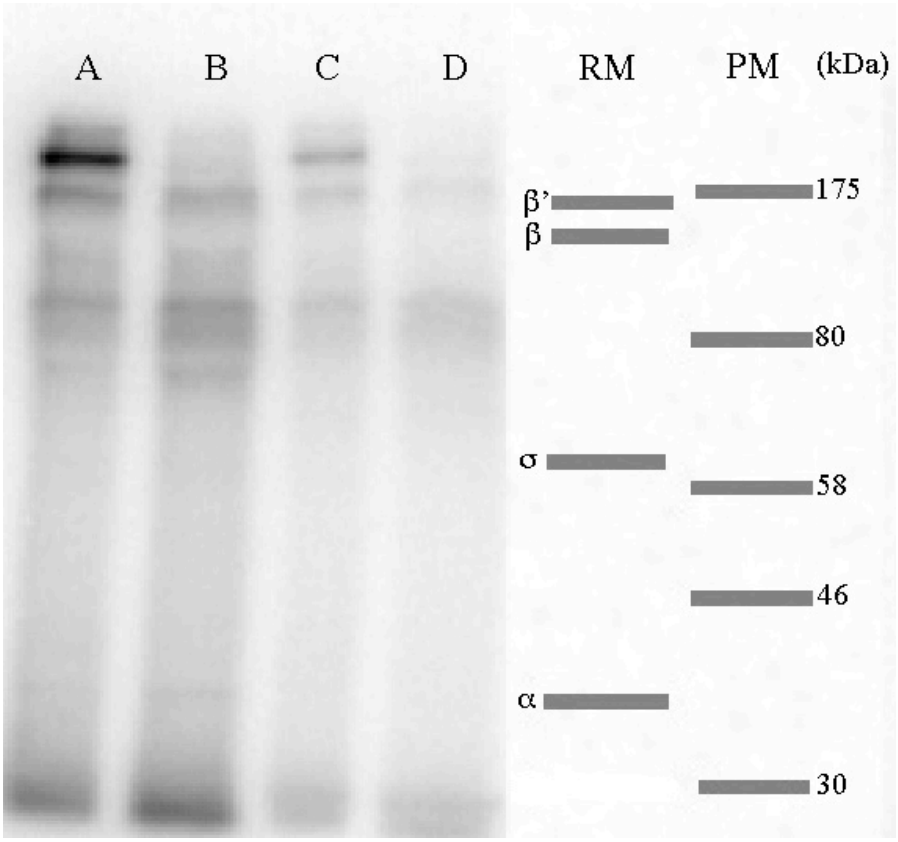
The newly labeled promoters were then run on an analytical native 8% (19:1) polyacrylamide gel to assess their purity. This image is given in Figure 12 and shows that both promoters are of equal length, and that the *N25* promoter is likely more radioactive than the *N25*-S_6G promoter due to the more intense band. The absence of other bands visible in the gel indicates that the promoters are isolated away from any side products.

Figure 12. *Image of 8% analytical native gel containing the ^{32}P-labeled N25 and N25-S_6G promoters.* This gel image indicates that the promoters were purified away from any side products, and that they are of equal lengths.

After the construction of the *N25* and *N25*-S_6G promoters, they could be applied to photocrosslinking analysis to confirm that contact is made with the β' subunit. The promoters were first set up as open complexes using 0.5 μM RNAP (1X normal amount in a 10 μL reaction) and then irradiated with UV light to photocrosslink to RNAP. The crosslinked constructs were then analyzed on an 8% SDS-PAGE with a set of controls (image not shown).

To increase the intensity of the photocrosslinked band for more precise analysis, the reaction was repeated using different concentrations of RNAP for comparative analysis. A higher concentration of undiluted RNAP (5.6 μM, 10X the normal amount in a 10 μL reaction) was tested to allow more open complex formation per reaction. The gel image of the crosslinked results is shown in Figure 13.

This gel shows that the wildtype *N25*-S_6G promoter is indeed contacted by the β' subunit, thereby confirming the work of J. Wiwczar (2008). Results additionally show that the intensity of the band increased with using 5.6 μM RNAP during open complex formation, indicating that a higher amount of open complexes are formed.



Figures 13. *Image of photocrosslinked N25 and N25-S_6G in 8% SDS-PAGE*. Lane A contains open complexes formed using 5.6 μM RNAP and *N25*-S_6G, with UV light for crosslinking. Lane B contains *N25*-S_6G and RNAP without UV light. Lane C contains *N25*-S_6G without RNAP, but with UV light. Lane D contains *N25* (missing thiol group) with RNAP and UV light. PM represents NEB protein marker where the bands, from top to bottom, correspond to m.w. of 175, 80, 58, 46, and 30 kDa. RM represents RNAP subunit markers; the bands from top to bottom are β', β, σ, and α, with corresponding m.w. of 155, 150, 70, and 36 kDa, respectively.

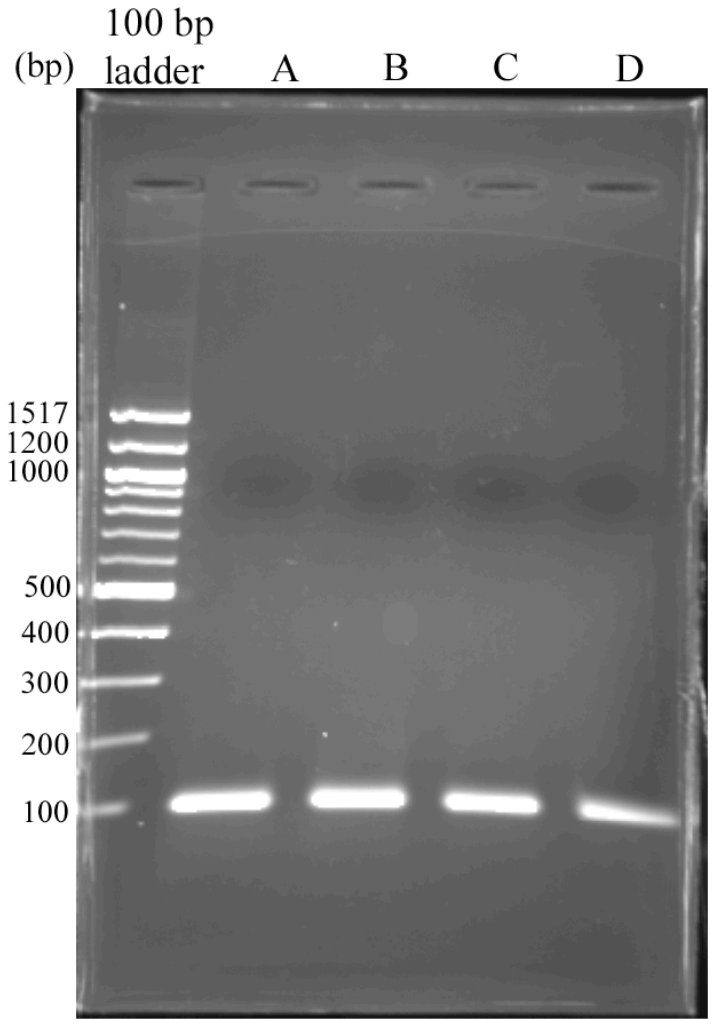


Figure 14. *Agarose gel image of crude full-length N25, CW1-S_6G, CW2-S_6G and CW3-S_6G promoters.* This gel is run with a 100 bp ladder to show that their lengths are approximately equal and correct. Lane A: *N25*; lane B: CW1-S_6G; lane C: CW2-S_6G; and lane D: CW3-S_6G.

Once I could repeat J. Wiwczar's findings, I was ready to examine the photocrosslinking behavior of the CW promoters. For this, I constructed ^{32}P-labeled *N25*, CW1-S_6G, CW2-S_6G and CW3-S_6G promoters using [α-^{32}P]-dATP in the primer extension reaction. The full length promoters were analyzed on a 2% agarose gel to confirm their presence and correct length (100 bp). This image is shown in Figure 14. This image shows that the promoters are of correct length around 100 bp.

These promoters were then recovered by ethanol precipitation, purified from a preparative 6% native polyacrylamide gel, and the extended DNA was cleaned up via phenol-chloroform extractions. The promoters were resuspended in 40 μL TE and concentrations then measured by the NanoDrop spectrophotometer which gave 457 nM of the *N25* promoter, 432 nM of CW1-S_6G, 461 nM of CW2-S_6G, and 463 nM of CW3-S_6G. The full-length promoters obtained were run on an analytical 8% (19:1) polyacrylamide gel to assess its purity. This gel image is shown in Figure 15 and reveals that the promoters are of equal length in comparison to one another. It additionally does not show any other bands, indicating that the purification process was successful.

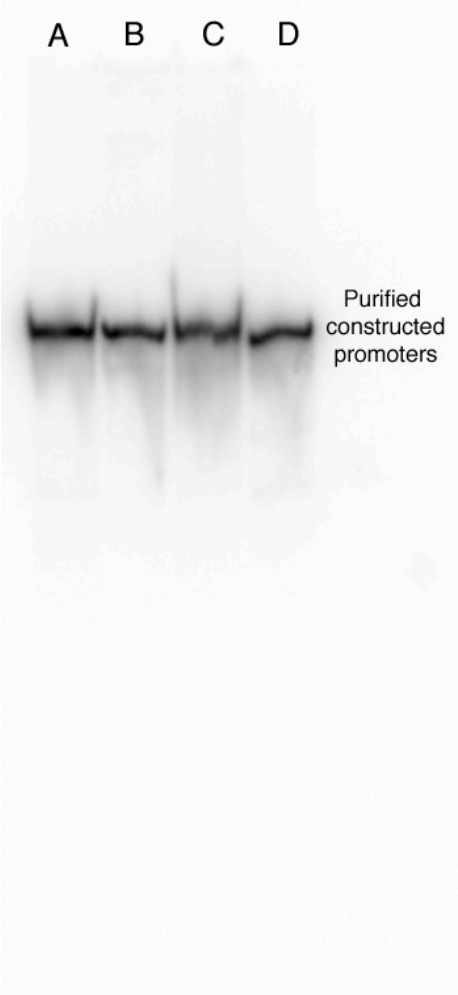


Figure 15. *Analytical 8% (19:1) polyacrylamide gel image of purified ^{32}P-labeled N25, CW1-S_6G, CW2-S_6G and CW3-S_6G promoters.* This image shows that the constructed promoters were the same length as each other and were separated from by-products by the purification process. Lane A: *N25*; lane B: CW1-S_6G; lane C: CW2-S_6G; and lane D: CW3-S_6G.

The purified ^{32}P-labeled *N25*, CW1-S_6G, CW2-S_6G and CW3-S_6G promoters could then be applied to photocrosslinking to give insight about contacts to RNAP with their varying DIS regions. Open complex formation was executed with 5.6 μM RNAP, the promoters underwent crosslinking with or without irradiation with UV light. The samples were then run together on a 8% SDS-PAGE with RNAP and protein m.w. markers. This image is shown in Figure 16. As expected, all of the controls worked. Without UV irradiation, no photocrosslinking was detected (see lanes B, D, F, and H). Without the S_6G probe, UV irradiation failed to generate crosslinks (on the N25 promoter; see lane A). CW1-S_6G and CW2-S_6G DNA gave rise to robust levels of photocrosslinking which is UV-dependent and S_6G-dependent (lanes C, and E). The fact that the CW2-crosslinked band migrates to the same position on the SDS-PAGE gel as the CW1 band indicates that the CW2-crosslink is also with the β' subunit. Thus, changing the A_{-6} residue of N25 to G did not significantly alter the open complex conformation and crosslinking.

Interestingly, CW3-S_6G promoter failed to crosslink with RNA polymerase (lane G). This result is intriguing and suggests two possible causes: either the open complex with CW3 promoter is highly unstable and falls apart before a crosslinking reaction can occur, or the open complex, due to the $A_{-4}G/A_{-6}G$ changes, has adopted a different conformation where no crosslinking is possible.

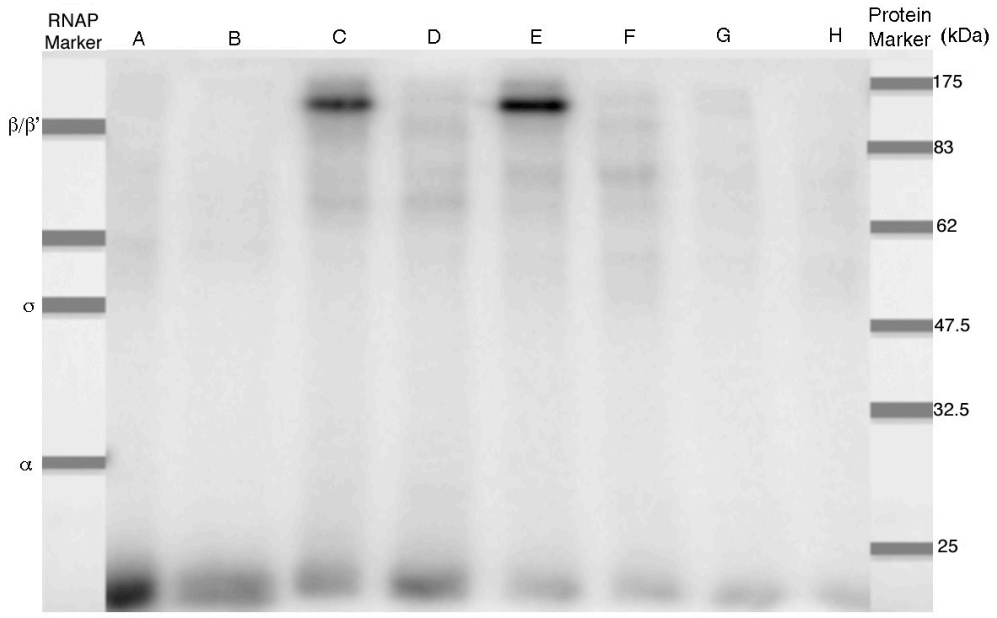


Figure 16. *SDS-PAGE image of photocrosslinking N25, CW1-S_6G, CW2-S_6G and CW3-S_6G promoters*. Lanes A, C, E, G: open complexes irradiated with 365 nm UV light; lanes B, D, F, H: open complexes without UV irradiation. Lanes A & B: *N25* promoter; lanes C & D: CW1-S_6G promoter; lanes E & F: CW2-S_6G promoter; and lanes G & H: CW3-S_6G promoter. RM represents RNA polymerase subunit markers (β', 155 kDa; β, 150 kDa, σ, 70 kDa, and α, 36 kDa) and PM represents protein marker of 175, 80, 58, 46, 30 and 25 kDa.

To ensure that the ^{32}P-labeled CW3-S_6G promoter was at least binding to RNAP to form an open complex, the complex was analyzed using a gel shift. The ^{32}P-labeled *N25*, CW1-S_6G, and CW2-S_6G promoters were included for comparison. The experiment was carried out and analyzed on a native 4% (37.5:1) polyacrylamide gel. Promoter DNA by itself and promoter in open complexes were loaded into a running gel to capture the bound promoter-RNAP complex. This gel image is shown in Figure 17 and reveals that the ^{32}P-labeled CW3-S_6G promoter does form an open complex with RNAP. However, the intensity of the CW3-S_6G band is minimal in comparison with the other promoters. This could be due to the promoter being less radioactively labeled in comparison to the others, or that the CW3-S_6G promoter does not form an open complex as efficiently as the other promoters.

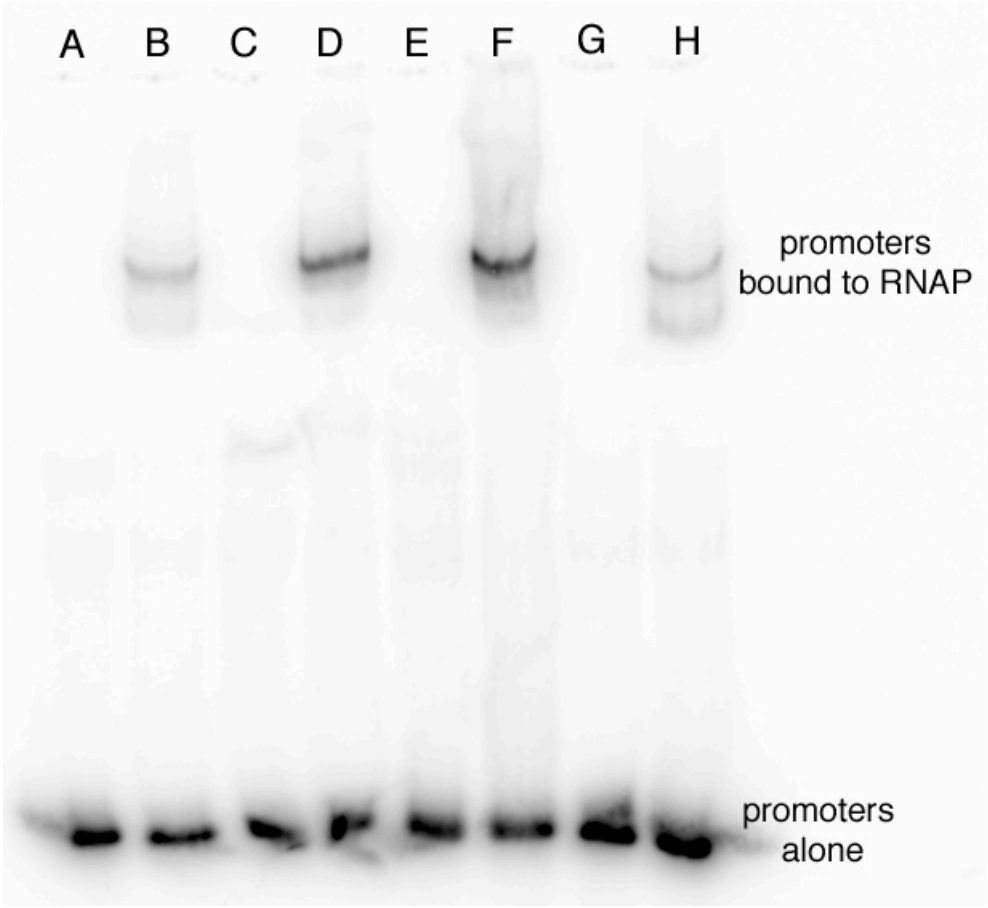


Figure 17. *Gel shift image of open complex formation*. Lanes A and B contain the *N25* promoter without and with RNA polymerase; lanes C and D contain CW1-S_6G promoter without and with RNA polymerase; lanes E and F contain CW2-S_6G promoter without and with RNA polymerase; and lanes G and H, contain CW3-S_6G promoter without and with RNA polymerase, respectively.

Open complex formation reactions are known to be sensitive to the concentration of KCl. Thus far in this study, the open complex reactions have been performed at 200 mM KCl. Such high KCl concentration can destabilize the weaker open complexes. Thus, I next examined the effect of KCl concentration on open complex photocrosslinking with CW3-S_6G promoter. KCl concentrations of 20, 50, 100, and 200 mM were tested both in open complex formation by a native gel shift analysis and in open complex photocrosslinking. The gel shift image shown in Figure 18 reveals that the 200 mM KCl results in the least amount of open complexes formed and that 50 mM KCl appears to be the optimum concentration. The bands within this gel indicate that the open complexes can be formed to a substantial degree when crosslinking is not involved. To test the effect of KCl concentrations on open complexes followed by crosslinking, the same concentrations of KCl were used during open complex formation followed by crosslinking, and results run on a 8% SDS-PAGE. The image of this gel is shown in Figure 19. The presence of visibly crosslinked bands was revealed in this image. These gels reveal that the CW3-S_6G promoter does appear to contact RNAP to form an open complex. To confirm this observation, the promoter can next be analyzed by repeating the photocrosslinking experiment with a full set of controls.

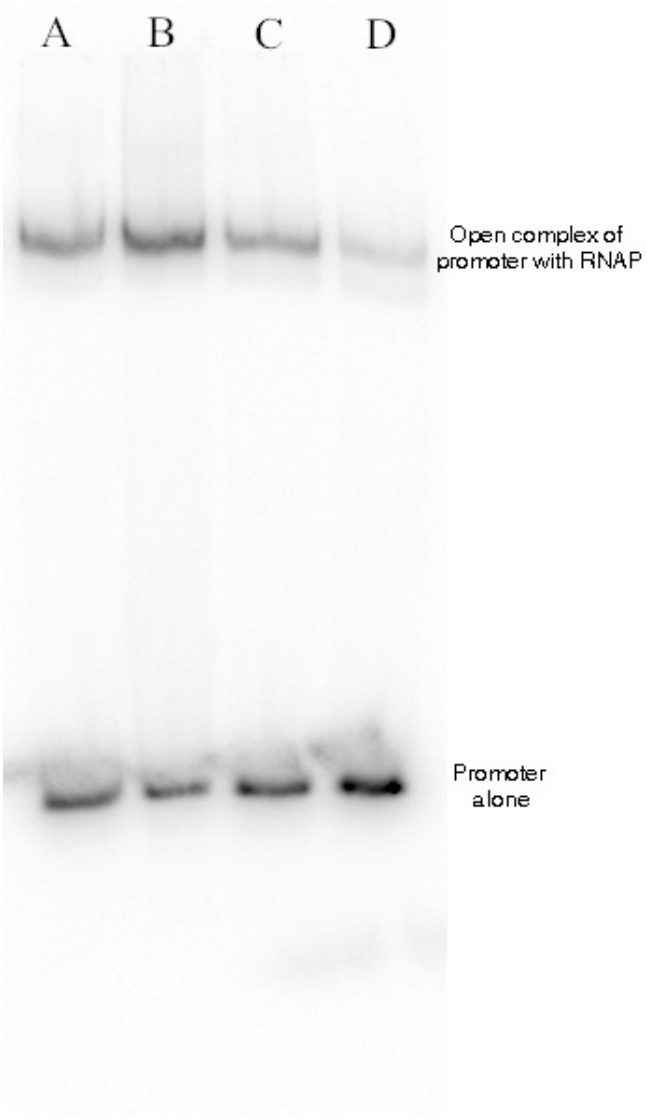


Figure 18. *Gel shift analysis of the CW3-S_6G promoter in an open complex with RNAP under varying KCl concentrations*. Lanes A-D differ in KCl concentration: from left to right, 20 mM, 50 mM, 100 mM, 200 mM were used in the open complex formation reaction.

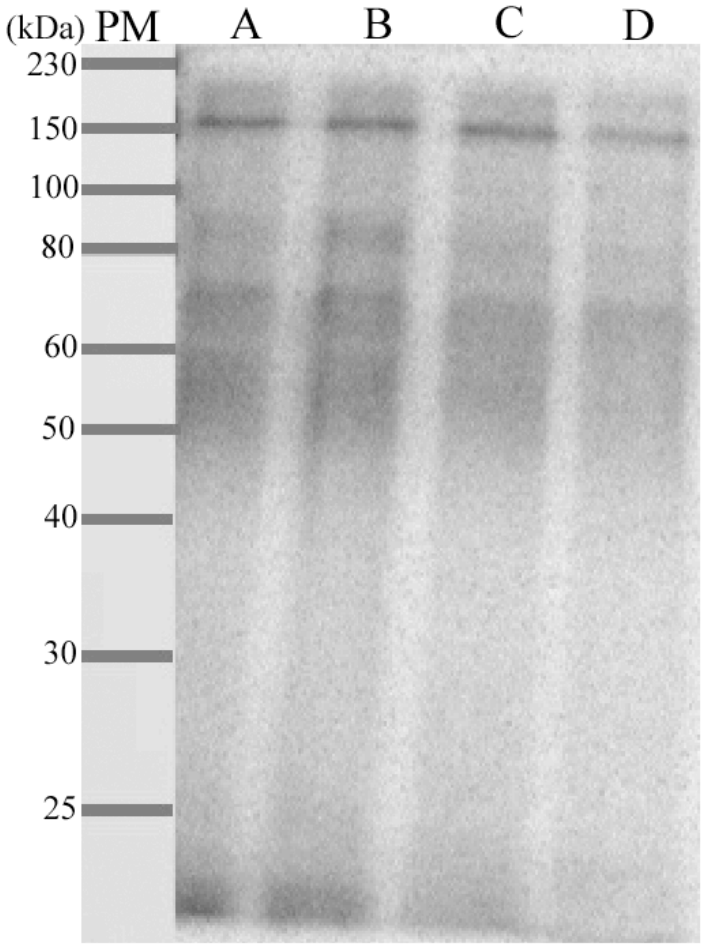
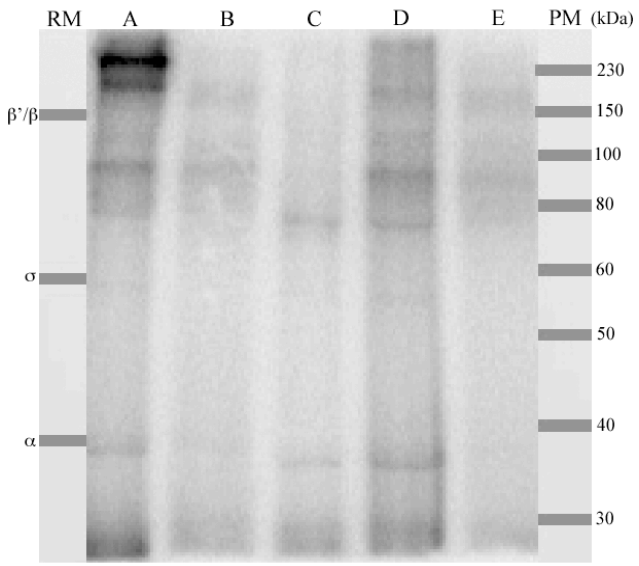


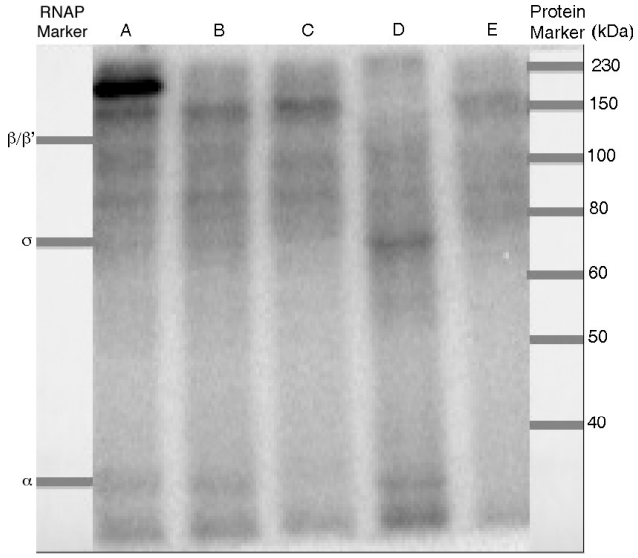
Figure 19. *SDS-PAGE analysis of CW3-S_6G promoter crosslinked to RNAP at varying KCl concentrations.* Lanes A-D differ in KCl concentration: from left to right, 20 mM, 50 mM, 100 mM, 200 mM. PM represents protein marker with the bands from top to bottom representing 230, 150, 100, 80 60, 50, 40, 30, and 25 kDa.

After discovering that the CW3-S_6G promoter, in addition to the CW1-S_6G and CW2-S_6G promoters, all disclose contacts to RNAP, I repeated individual photocrosslinking experiments using 50 mM KCl with a full set of controls to confirm this observation. Each promoter was set up for open complex formation with 5.6 μM RNAP and then irradiated with UV light. The crosslinked conjugate with four controls were run on a 8% SDS-PAGE. The images for the CW1-S_6G gel, CW2-S_6G gel and CW3-S_6G gel are shown in Figure 20. The CW3-S_6G gel did show some crosslinking, but not as robust as the CW1-S_6G and CW2-S_6G. However, the fact that the crosslinked band is missing from the control lanes gives insight that it is likewise contacted by the β' subunit. The CW1-S_6G gel shown is similar as expected to the previous gel in Figure 13. The controls for all three gels demonstrate how the crosslinked open complexes are RNAP-dependent, UV-dependent, and S_6G-dependent. These three gels compiled reveal that the -5 G residue within the DIS region of the *N25* promoter is contacted by the β' subunit. This indicates that the sequence composition around the -5 G is not important for determining the contacting subunit of RNAP.

CW1-S_6G



CW2-S_6G



CW3-S_6G

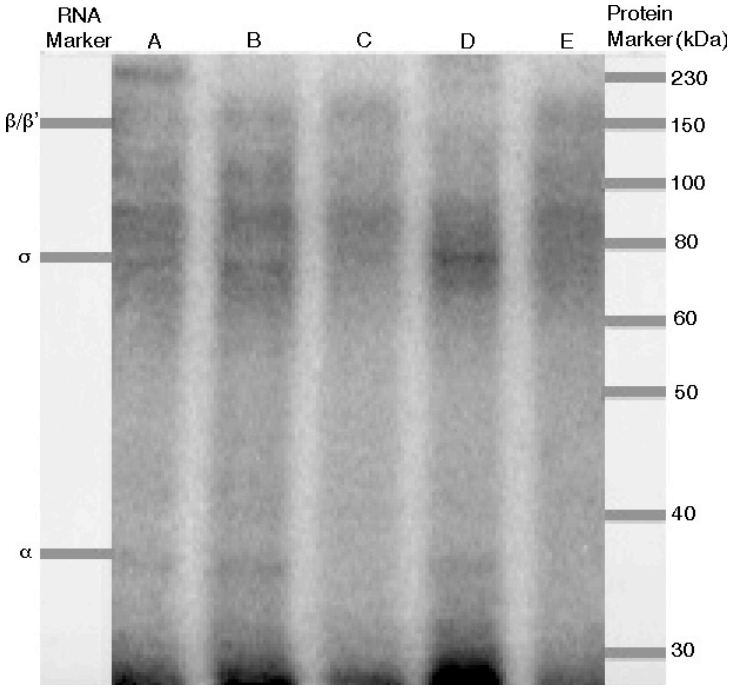


Figure 20. *Gel images of photocrosslinked CW1-S_6G, CW2-S_6G and CW3-S_6G with full set of controls*. In each of the gel images, lane A represents the promoter with RNAP and UV irradiation. Lanes B-E are controls where lane B is without UV irradiation, lane C without RNAP, and lane D without the S_6G thiol group. Lane E is the S_6G promoter without RNAP nor UV irradiation.

Exploring the feasibility of mass spectrometry analysis of the protein-DNA crosslink:

After it was confirmed that the β' subunit contacted the CW1-S_6G, CW2-S_6G and CW3-S_6G promoters, the next step in my investigation is to examine the precise context of contact to clarify the difference between our results and those from previous investigations (Wiwczar, 2008; Haugen *et al.,* 2006, 2008). It is my goal to map the exact amino acids of the β' subunit involved in contacting the -5 G residue. The traditional method of mapping contacts is by using a partial proteolytic digestion. This method is time-consuming and can only localize the crosslink to a segment of the β' subunit involved (Brodolin *et al*., 2000). Another approach is by mass spectrometry which would pinpoint the precise amino acids involved in crosslinking. This is a new field of research and therefore not fully developed (Shaw *et al.*, 1992; Shivanna *et al.*, 1993). I decided to advance forward with mass spectrometry analysis as it has been proven to work efficiently and accurately in discovering contacts between protein-nucleic acid heteroconjugates (Jensen *et al.*, 1994; Golden *et al.*, 1999; Steen & Jensen, 2002).

The first step to prepare the reaction for such analysis is to: 1) devise a method to maximize the amount of crosslinked heteroconjugates per lane in the SDS-PAGE and 2) apply the heteroconjugates to a gel that allows sufficient and straightforward extraction. To increase the amount of photocrosslinked heteroconjugates, an analysis was performed to vary the ratio of RNA polymerase:promoter during open complex formation using the CW1-S_6G and

CW2-S_6G promoter DNA. RNAP:promoter ratios of 4:1, 2:1 and 1:1 where the promoter DNA concentration is 150 nM were examined. A reaction with RNAP:DNA ratio of 1:1 where the [DNA] is 560 nM was likewise examined as an upper limit. These reactions were run on an 8% SDS-PAGE and the image is shown in Figure 21.

By visualizing the gel, one can see that the upper limit band in lane G appears to have the highest amount of crosslinked open complexes. The ImageQuant program was used to quantify the intensity associated with each band and give insight into the relative number of open complexes associated. The readings are given in Table 2. Both the appearance and readings reveal that the upper limit test in lane G containing the 1:1 ratio of [560 nM] RNAP:promoter gave the most concentrated band. Additionally, the CW1-S_6G promoter appears to give a higher amount of open complexes versus the CW2-S_6G promoter. However, both promoters maintain the G residue in the -5 position, indicating that their stability should be approximately equal. I therefore measured the counts per minute (cpm) associated with the promoters to investigate whether it was merely due to less radioactivity associated with the CW2-S_6G promoter. The cpm for each promoter is given in Table 3 and shows that the CW2-S_6G promoter is approximately half as radioactive as the CW1-S_6G promoter, which would result in the band appearing less intense. When the ratio between the bands quantified from the 4:1 RNAP:promoter readings in Lane A versus Lane D are taken

however, the CW1-S_6G band is approximately 7 times more intense than the CW2-S_6G band.

After finding how to produce a maximal amount of photocrosslinked complexes (Figure 21, lane G) in a single lane of a gel, the next step was to find a gel that would allow straightforward extraction. I ran a photocrosslinked open complex with a control non-crosslinked promoter on a 6% SDS-PAGE. This slightly lower percentage in the resolving gel was aimed to give greater separation between bands. An image of this gel is shown in Figure 22 and reveals that the crosslinked CW1-S_6G promoter is isolated away from the side products of the reaction to provide for easy extraction.

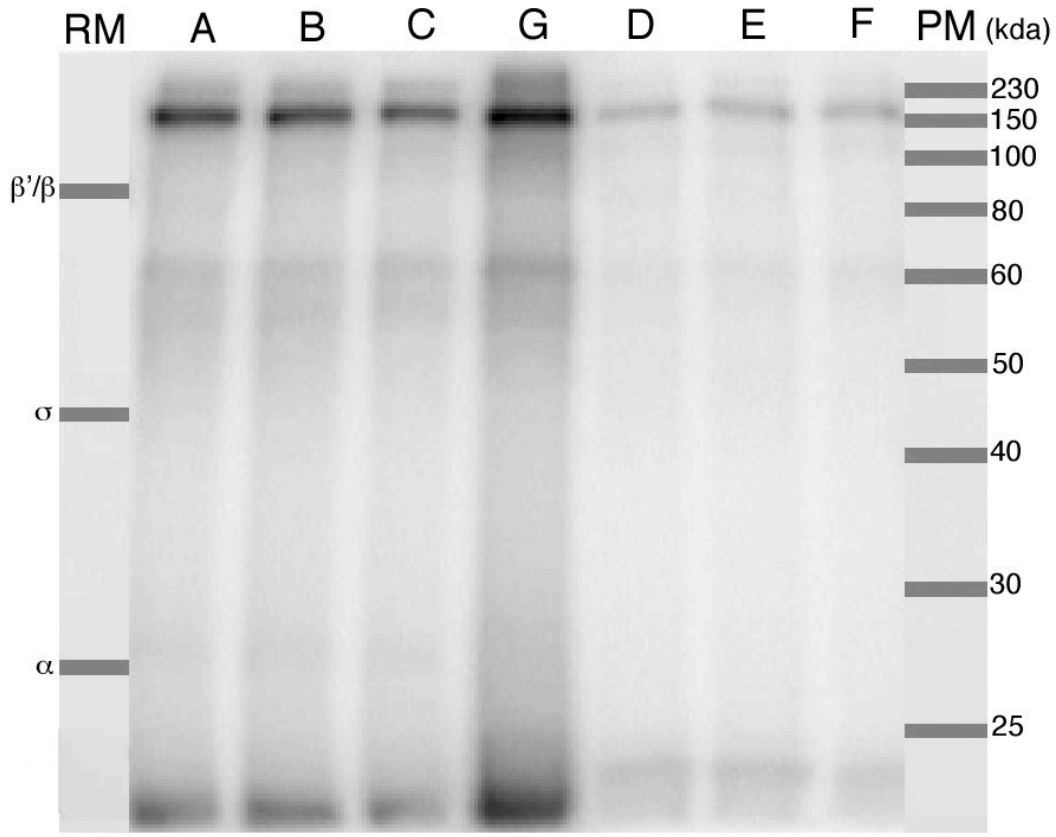


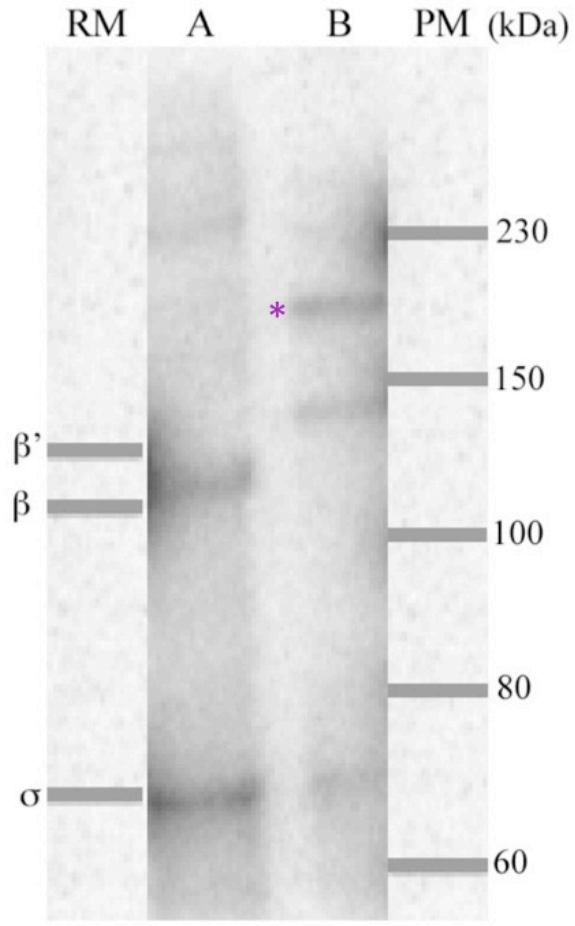
Figure 21. *Image of SDS-PAGE containing photocrosslinked open complexes at varying RNAP:promoter ratios*. Lanes A-C contain CW1-S_6G promoter in RNAP:promoter ratios of 4:1 in lane A, 2:1 in lane B, and 1:1 in lane C with the baseline [DNA] at 150 nM. Lanes D-F are corresponding ratios with CW2-S_6G promoter. Lane G contains concentrated CW1-S_6G promoter at 560 nM in a 1:1 ratio with RNA polymerase. RM represents RNA polymerase subunit markers. PM represents protein marker with each band from top to bottom representing 230, 150, 100, 80 60, 50, 40, 30, and 25 kDa.

Table 2. *Intensity of bands associated with the photocrosslink*. Lanes A-C, and G contain the CW1-S_6G promoter, lanes D-F contain the CW2-S_6G promoter. The baseline [DNA] is 150 nM, but in lane G, the [DNA] is 560 nM. These readings show that the 4:1 ratio of RNAP:promoter gives the highest intensity.

Band	**[RNAP]:[promoter]**	**Intensity**
A	4:1	802397
B	2:1	713409
C	1:1	493372
G	concentrated 1:1	1278370
D	4:1	108018
E	2:1	107412
F	1:1	104914

Table 3. *Counts per minute from 1 μL constructed promoters*. These values were measured after another set of promoters were constructed on 11/24/10.

Promoter	**Counts per minute (cpm)**
N25	39130
CW1-S_6G	30980
CW2-S_6G	15116
CW3-S_6G	29335



*- crosslinked open complex

Figure 22. *Separation of crosslinked bands on 6% SDS-PAGE*. This image contains the crosslinked CW1-S_6G promoter in Lane B, and a control of the CW1-S_6G promoter without crosslinking in Lane A. This 6% resolving gel was run for 2 hr to provide sufficient separation of the crosslinked band away from uncrosslinked β' subunit. PM represents protein marker, and RM represents the subunits of RNAP

The next step in preparing samples for MS analysis is to ensure they can be read in an accurate manner by the model of MS that they will be submitted to. Research by Jensen *et al.* (1996) examined a number of protein-nucleic acid heteroconjugates and came to the conclusion that MALDI-TOF and ESI on triple quadrupole spectrometers gave the most precise data. Results produced from these MS analyses are only accurate however when a heteroconjugae is no more than approximately 50 kDa. Because the *N25*-RNAP open complex is 187.5 kDa, I need to devise a method to decrease its mass. Studies by Steen & Jensen utilized a variety of proteases and restriction enzymes to cut away non-crosslinked protein and DNA (Steen & Jensen, 2002). Below I describe our various attempts at establishing conditions for trypsin and Apo I digestion.

I initiated the examination of reducing the promoter length by digesting the naked CW1-S_6G DNA strand with the *Apo* I restriction enzyme (cleavage sites shown in Figure 7 of methods). The digestion was carried out under varying reaction conditions to search for the optimum cleavage conditions. Products were run on an 8% (19:1) polyacrylamide gel with image shown in Figure 23. This gel shows that the digest is effective as there are a few bands migrating lower than the non-digested control band in Lane E. These bands represent potential fragments of 9, 34*, 20, 25, and 54* bp, of which only the 34-bp and 54-bp bands would be lightly labeled (indicated by astericks). From a qualitative examination of the gel, Lane B appears to have the optimal conditions as the bands are most intense signifying a greater amount digest products.

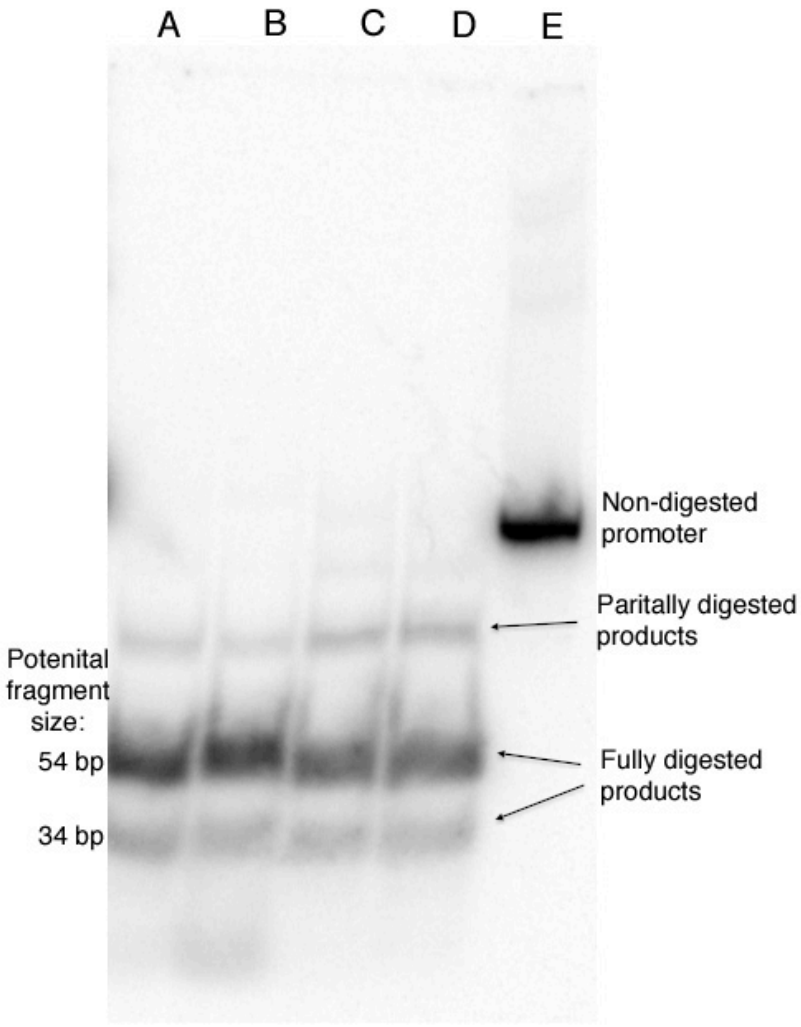
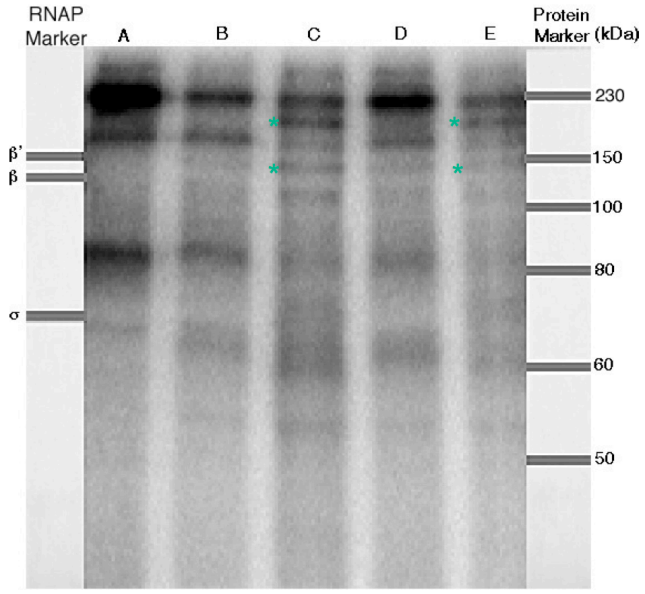


Figure 23. *ApoI digests of naked CW1-S_6G promoter*. Lane A is with a 1 µL *Apo* I digest done in NEBuffer 2 at 37 °C. Lane B is a 2 µL *Apo* I digest done with 2 µL NEBuffer 2 at 37 °C. Lane C has a 1 µL *Apo* I digets with 1 µL NEBuffer 3 and incubated at 50 °C. Lane D has a 2 µL *Apo* I digest with 2 µL NEBuffer 3 done at 50 °C. Lane E contains a control of the CW1-S_6G promoter with no digest. All reactions were completed with 5 µL of 200 nM promoter, giving a final reaction volume of 8 µL to Lanes A and C, and 15 µL to Lanes B and D.

Once it was confirmed that the *Apo* I restriction enzyme was functional, the promoter crosslinked within an open complex was next analyzed under varying digestion conditions. The products were run on an 8% SDS-PAGE with image shown in Figure 24. These results correlate with the results from the naked DNA digest where the condition in Lane C gave the greater amount of partial and full digest products. This is indicated by having a minimal non-digested band (migrating at the same distance as band in Lane A) and more intense newly-arisen bands. The conditions used by the digest run in Lane E also appear to have worked well. The lower asterisk-marked band is likely the full-digest product as it migrates at approximately the same length as the β', indicating a minimal amount of DNA (~25 bp) left bound the to the RNAP.



*- partial and full digest products

Figure 24. *SDS-PAGE image of Apo I digested photocrosslinked CW1-S_6G promoter*. Lane A contains the crosslink control of CW1-S_6G in an open complex. Lane B contains the crosslinked CW1-S_6G with a 1 µL *Apo* I digest with 1 µL NEBuffer 2 done at 37 °C. Lane C has the crosslinked CW1-S_6G with a 2 µL *Apo* I digest with 2 µL NEBuffer 2 done at 37 °C. Lane D has the crosslinked CW1-S_6G promoter with a 1 µL *Apo* I digest with 1 µL NEBuffer 3 done at 50 °C. Lane E has a 2 µL *Apo* I digest done with NEBuffer 3 at 50 °C of the CW1-S_6G crosslinked promoter to RNA polymerase. All reactions were incubated for 2 hours. The reaction volume was 13 µL for lanes B and D, and was 20 µL for Lanes C and E. Before loading completed digest products on the gel, all of the reaction were brought up to 20 µL using ddH_2O. RM is the RNA polymerase subunits and PM is the protein markers corresponding to m.w, from top to bottom, of 230, 150, 100, 80 60, and 50 kDa.

After recognizing that the *Apo* I digest decreased the amount of DNA left bound within the crosslinked heteroconjugate, I then looked into protease digestions of the β' subunit to cut away most of the protein portions not bound to the promoter. Trypsin is a protease known for being very effective, and widely used across the studies of preparing heterconjugates for MS (Golden *et al.*, 1999). There are 174 trypsin cut sites within the β' subunit (sites shown in Table 1 of methods), giving the possibility that only a few amino acid residues are left behind. I initiated this study in a similar manner as the *Apo* I study where I searched for optimal cleavage conditions. A series of crosslinked open complexes were first applied to a timecourse trypsin digest for 10', 30 ', 1 hr, and 2 hr. The products of these digests were then run on a native (19:1) polyacrylamide gel, with image shown in Figure 25. An *Apo* I digest using the conditions of sample C from Figure 24 was also completed and shown in Lane G. This gel shows that the 1 hr and 2 hr digestion times, shown in Lanes E and F gave bands of reduced size from the non-digested heteroconjugate. These lanes also gave more fully digested products in comparison to lanes B and C. The *Apo* I digest of the promoter crosslinked to an open complex is not visible in Lane G. This is potentially due to little signal (due to decay) associated with the attached fragment after digestion.

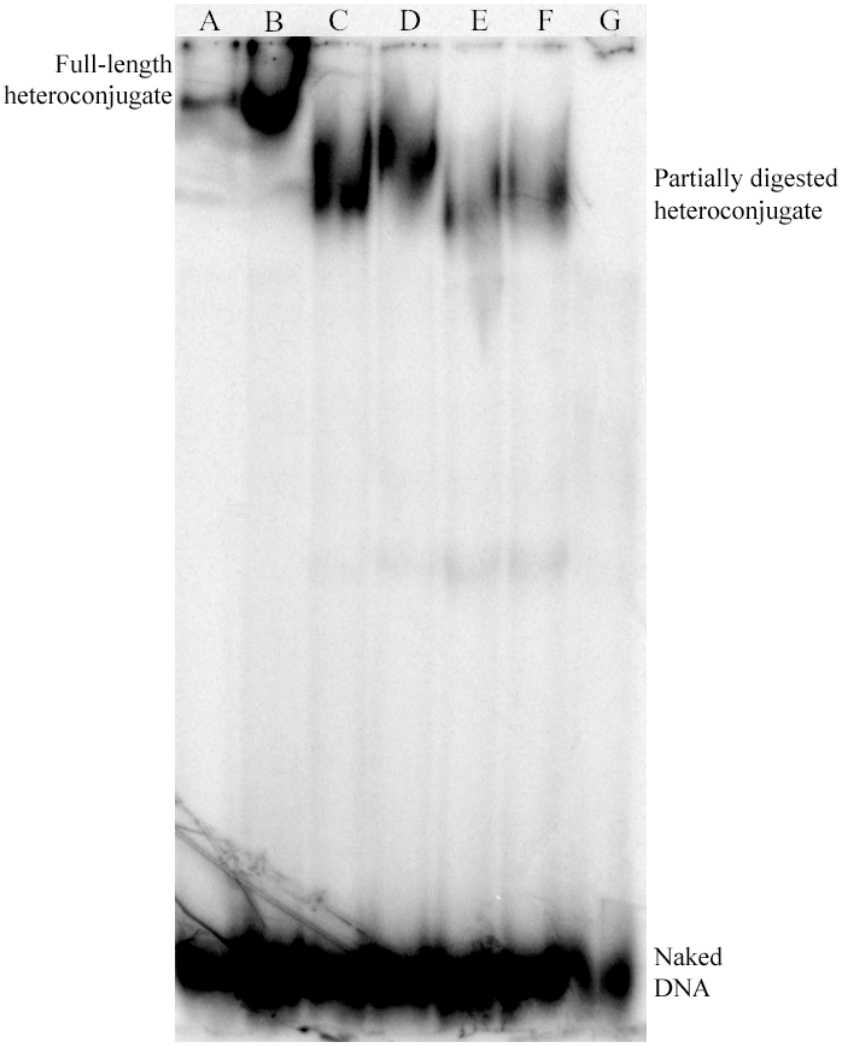


Figure 25. *Timecourse of trypsin digests and an Apo I digest of crosslinked open complexes*. Lane A is the CW1-S_6G without crosslinking and no digest. Lane B contains the crosslinked CW1-S_6G with no digest. Lanes C-F contains crosslinked CW1-S_6G digested with 200 μM trypsin for 10', 30', 1 h, and 2 h. Lane G contains an *Apo* I digest done under the same conditions as sample/lane C in Figure 24.

The previous results indicate that a 1-2 hr trypsin digest results in fewer amino acids left bound to the promoter. I next examined trypsin digest reagents to explore which conditions would be optimum for amino acid cleavage. Varying concentrations of tryspin were used in digests to either untreated or boiled crosslinked open complexes. The digests were carried out overnight at 37 ^{o}C. These samples were run in an 8% SDS-PAGE and image shown in Figure 26. The disappearance of the crosslinked open complex band (shown in Lane B) indicates that the trypsin digest worked efficiently across the board, with the exception of Lane F. This gel also reveals that the promoter is likely bound to a small segment of β' as the digest products are migrating around the same distance as naked DNA.

Analysis of the restriction enzyme and protease digestions have shown that both the *Apo* I and trypsin enzymes are functional on photocrosslinked open complexes. The digests results give insight into which conditions would give the greatest amount of digested conjugates.

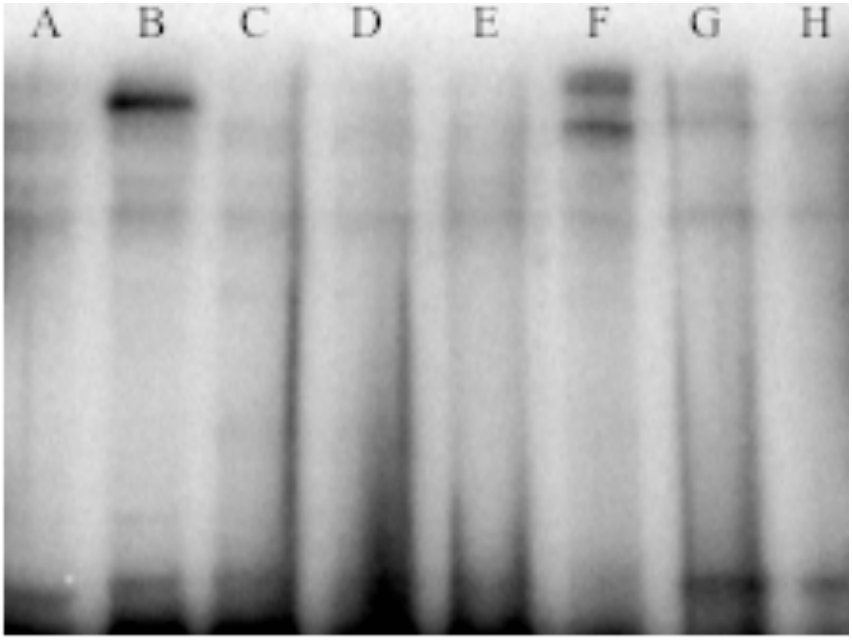


Figure 26. *8% SDS-PAGE image of trypsin digestion with untreated or boiled photocrosslinked CW1-S_6G promoter open complex*. Boiling treatment was for ~7 minutes. Lane A contains a control CW1-S_6G with no crosslink and no digest. Lane B contains a crosslinked CW1-S_6G complex with no digest. Lanes C-E involves treatment with trypsin at 200 nM, 400 nM, and 600 nM, respectively, and digested overnight at 37 ^{o}C. Lanes F-H are the boiled crosslinked CW1-S_6G promoter open complex digested with 200 nM (lane F), 400 nM (lane G), and 600 nM (lane H) of trypsin overnight at 37 ^{o}C.

DISCUSSION

The intention of this investigation was to examine the context of the interaction between the RNAP and -5 G nontemplate residue within the DIS region of the T5 *N25* promoter. The *N25* promoter is known for having close-to-consensus regions within its -35 box and -10 box. The strength of interaction between the *N25* promoter and RNAP has been shown to provide a long-lived open complex resulting in a high rate of initiation (Knaus & Bujard, 1990) and abortive initiation (Vo et al., 2003). We were piqued to investigate the newly uncovered DIS region interaction with RNAP to examine whether this was an additional feature of the *N25* promoter that contributed to its open complex stability.

Research by Haugen *et al*. (2006) suggested that a nontemplate strand G residue two nucleotides downstream of the -10 box enabled such stabilization. Through their photocrosslinking studies, they found that both G residues of *E. coli* $C_{-7}G$ *rrnB* P1 and λP_R promoters were contacted via the σ1.2 domain of RNAP. The *N25* promoter naturally contains a -5 G residue, and therefore the putative interaction with RNAP is a likely characteristic that contributes to its open complex stability. Photocrosslinking studies and Western blot analysis by J. Wiwczar (2008) examined the subunit of RNAP involved in contacting the *N25* promoter. Her results revealed that the -5 G residue indeed crosslinks to RNAP,

but was contacted by the β' subunit, thereby differing from results of Haugen *et al*. (2006; 2008).

It was our interest to investigate the difference in DIS region contacts to gain insight on factors of the DIS sequence that dictate its interaction with RNAP. Hence, after confirming the results of Wiwczar (2008), we set forth to examine the DIS region contacts of *N25* through mutational analysis. Two variants of the *N25* promoter were constructed along with the wildtype to yield CW1, CW2, and CW3 promoters. CW1 is identical to the wildtype N25 promoter with the three nucleotides immediately downstream of the -10 box being AGA, for this reason CW1 can also be designated N25-AGA. By this convention, CW2 is N25-GGA, and CW3, N25-GGG. These sequences were chosen because the $C_{-7}G$ *rrnB* P1 and λP_R promoters contain –GGG and –GGT, respectively, at the equivalent positions, so that we can investigate if a higher G-content resulted in positioning the DIS region closer to the σ subunit. DNA strands with a high G-content are known for introducing a slightly larger twist (Laird, 1971), which could potentially be accommodated for by placing the nontemplate strand into a different region of contact.

Another aspect of the DIS region that could potentially result in slightly different RNAP contact is length. The $C_{-7}G$ *rrnB* P1 promoter has an 8-bp DIS region, whereas *N25* has 6 bp. It is possible that a longer DIS connection between the -10 box and the +1 start site of transcription caused it to be pushed into a different pocket within RNAP, resulting in different subunit contact. This latter

speculation does not apply to λ P_R, which has a 6-bp DIS region but crosslinks to the σ subunit nevertheless.

Crystal structures of the RNAP-promoter open complex from the bacteria *Thermus thermophilus* show that the σ subunit lies in close proximity to the β' subunit with a short segment of the σ subunit (σ313-342) pulled into the core molecule (Vassylyev *et al.*, 2002). The observation that these two subunits are located close to one another can increase the feasibility that σ1.2 or β' subunits of the RNAP from *E. coli* could both be available for contact depending on DIS region promoter signals.

The constructed *N25* DIS region variants could thereby be studied with photocrosslinking to determine if the position of crosslink is a function of sequence composition. The CW1, CW2 and CW3 promoters were first analyzed by measuring their abortive initiation probability. Despite the fact that their DIS sequences differ, the -5 G residue which is known for adding open complex stability, is maintained between the three of them. The abortive probability profile shown in Figure 10 indicates that the CW1, CW2 and CW3 promoters have similar enough open complex stability such that their probability to undergo abortive initiation is approximately the same.

With the incorporation of a 6-thio-deoxyguanine base at the -5 position, these promoters were then subjected to photocrosslinking to determine which RNAP subunit it comes into contact with. The results show that all three variants of the *N25* DIS sequence are contacted by the β' subunit. The stability of the

open complex in contact with the CW3-S_6G was additionally found to be dependent on the concentration of KCl. Salts can affect the electrostatic interactions formed between the contacting amino acids and specific nucleic acid residues of the promoter. This leads to the possibility that the sequence composition of the CW3-S_6G DIS region affected its stability in open complex. However, the results are in support of the hypothesis that the sequence of nucleotides bordering the -5 G residue may not play a significant role in directing RNAP DIS contact due to the observed consistency of all three promoters interacting with the β' subunit.

To further analyze the discrepancy in the -5 G nontemplate residue interaction with RNAP, we were compelled to map the precise location of contact within the β' subunit. A variety of techniques are currently used to uncover regions of interaction between proteins and nucleic acids: partial proteolytic degradation (Severinov *et al.*, 1992), immunological methods (Lim *et al.*, 1991; Celander & Abelson, 2000a,b), biophysical techniques such as NMR spectroscopy and X-ray crystallographic data (Takeuchi & Wagner, 2006), and three-dimensional molecular modeling (Luger *et al.*, 1997). In recent years, the development of proteomics has made mass spectrometry a choice technique for protein identification and contact detection.

MS is evolving into a more commonly used method for detecting crosslinked conjugates and is known to have high sensitivity and speed for interaction analysis (Back *et al.*, 2003). Studies have shown that MS will give

information about the stoichiometry of the crosslinking reaction, identification of the binding domains, and localization of precise amino acids involved in contact (Steen and Jensen, 2002). A study by Steen and Jensen (2002) has found that photochemical crosslinks, especially those with minimal volume (such as a thiol group) are the most suitable for MS analysis. This is due to their crosslink being a more stable bond formed between a nucleobase and an amino acid, and results in a structure similar to the original conformation. The thiol group used for photocrosslinking the CW1, CW2, and CW3 promoters is therefore applicable for MS analysis.

Previous studies have shown that MALDI-TOF (matrix-assisted laser desorption/ionization on linear time-of flight) and ESI (electrospray ionization) on triple-quadrupole mass spectrometers are the most accurate when analyzing heteroconjugates. The MALDI-TOF was less hindered by sample impurities and more sensitive than ESI quadrupole, whereas ESI on the quadrupole produced greater mass accuracy and resolution than MALDI-TOF (Jensen *et al.*, 1996). Their requirements however introduce a few complications for sample preparation. One is that oligonucleotides generally require a negative ionization mode, whereas proteins require a positive ionization mode. Results are therefore more accurate when a heteroconjugate is more than 80% of one component versus the other. Another complication is the mass of the heteroconjugate. The upper limit is approximately 50 kDa, however results have shown that more accurate data is received from smaller heteroconjugates. For this reason, heteroconjugates

are digested with restriction enzymes and proteases to isolate only the crucial contacting components. Snake venom phosphodiesterase/alkaline phosphatase is another oligonucleotide degradation method (Golden *et al.*, 1999). However, the reaction is difficult to control and could leave no signals for detection. The truncated heteroconjugate can be purified from digestion side-products by using SDS-PAGE extraction or liquid chromatography (Steen and Jensen, 2002).

We chose to examine the *Apo* I restriction enzyme digest to cut away excess promoter from the crosslinked -5 G nontemplate residue. This enzyme in comparison to others has cleavage sites with closest proximity to the -5 G crosslink; at the -23 and +3 positions. This would leave behind 25 bp of promoter, which is a reasonable size for MS analysis. However, a complication is introduced by only 2 radioactively labeled adenines remaining after digestion. This makes detection/monitoring by radioactivity on SDS-PAGE or through the chromatographic steps difficult due to the little signal remaining. On the other hand, there is an advantage, since samples need to be nearly non-radioactive for MS analysis. *Apo* I digests were then analyzed by varying concentrations of reagents to find the optimal cleavage condition.

An analysis of potential proteases was initiated using a complementary cleavage mapping website provided by Swiss Institute of Bioinformatics. Theoretical cleavage sites of trypsin, caspase I, cyanogen bromide, and hydroxylamine were mapped out. Trypsin was found to have 174 cleavage sites on the β' subunit and would result in no more than 20 amino acids remaining

bound to the promoter. We therefore decided on using trypsin and analyzed the cutting efficiency under varying treatments to determine the optimal cleavage condition.

To conclude this investigation, we found that the wildtype *N25*, as well as the constructed ^{32}P-labeled CW1-S_6G, CW2-S_6G, and CW3-S_6G promoters all made 0 Å crosslinker contact to the β' subunit. Positive -5 G nontemplate residue crosslinking results confirm the conclusion made by Haugen *et al.* (2006) where RNAP contacts the 2nd base downstream from the -10 box. This gives insight that the -5 G nontemplate residue as well as close-to-consensus sequences are characteristic components of *N25* that cooperate to give a long lived open complex with high initiation and abortive initiation rates. Results also show that the sequence of nucleotides bordering the -5 G nontemplate residue of *N25* does not play a significant role in directing the nontemplate strand of the transcription bubble to surrounding RNAP segments for contact. For MS preparation, compiling *Apo* I and trypsin digests can give rise to a shortened oligonucleotide-protein heteroconjugate suitable for accurate MALDI-TOF and ESI on quadrupole analysis. Our next challenge is to recover sufficient amount of the truncated heteroconjugate for determining the β' amino acid contacts.

LITERATURE CITED

Alexander, P., and Moroson, H. (1962) Cross-linking of deoxyribonucleic acid to protein following ultra-violet irradiation of different cells. *Nature* **194**, 882–883.

Back, J. W., de Jong, L., Muijsers, A. O., de Koster, C. G. (2003) Chemical cross-linking and mass spectrometry for protein structural modeling. *J. Mol. Biol.* **331**, 303-313.

Brodolin, K., Zenkin, N., and Severinov, K. (2005) Remodeling of the σ^{70} subunit non-template strand contacts during the final step of transcription initiation. *J. Mol. Biol.* **350**, 930-937.

Celander, D. W., Abelson, J. N. (2000a) RNA-ligand interactions, Part A. *Methods Enzymol.* **317**, 276-291**.**

Celander, D. W., Abelson, J. N. (2000b) RNA-ligand interactions, Part B. *Methods Enzymol.* **318**, 163-182**.**

Golden, M. C., Resing, K. A., Collins, B. D., Willis, M. C., and Koch, T. H. (1999) Mass spectral characterization of a protein-nucleic acid photocrosslink. *Protein Sci.* **8**, 2806-2812.

Haugen, S. P., Berkmen, M. B., Ross, W., Gaal, T., Ward, C., and Gourse, R. L. (2006) rRNA promoter regulation by nonoptimal binding of σ region 1.2: an additional recognition element for RNA polymerase. *Cell* **125***,* 1069-1082.

Haugen, S. P., Ross, W., Manrique, M., and Gourse, R. L. (2008) Fine structure of the promoter-σ region 1.2 interaction. *Proc*. *Natl*. *Acad*. *Sci*. *USA* **105**, 3292-3297.

Hernandez, V. J., Hsu, L. M., and Cashel, M. (1996) Conserved region 3 of *Escherichia coli* σ^{70} is implicated in the process of abortive transcription. *J. Biol. Chem.* **271**, 18775-18779.

Hsu, L. M., (2002) Promoter clearance and escape in prokaryotes. *Biochem. Biophys. Acta.* **1577**, 191-207.

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.

Jensen, O. N., Kulkarni, S., Aldrich, J. V., and Barofsky, D. F. (1996) Characterization of peptide-oligonucleotide heteroconjugates by mass spectrometry. *Nucleic Acids Res.* **24***,* 3866-3872.

Knaus, R., and Bujard, H. (1990) Principles governing the activity of *E. coli* promoters. F. Eckstein and D. M. Lilley (eds.) *Nucleic Acids & Molecular Biology* **4**, 110-122. Springer-Verlag, Heidelberg, Germany.

Laird, C. (1971) Chromatid structure: relationship between DNA content and nucleotide sequence diversity. *Chromosoma* **32**, 378-406.

Lim, W. A., Sauer, R. T., Lander, A. D. (1991) Analysis of DNA-protein interactions by affinity coelectrophoresis. *Methods Enzymol.* **208***,* 196-201.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **289**, 251-260.

Murakami, K. S., Masuda, S., Campbell, E., Muzzin, O., Darst, S. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296**, 1285-1290.

Nelson, D. L., and Cox, M. M. (2005) Principles of Biochemistry, 4th Ed., W. H. Freeman and Company, New York.

Nudler, E., Gusarov, I., Avetissova, E., Kozlov, M., and Goldfarb, A. (1998) Spatial organization of transcription elongation complex in *Escherichia coli*. *Science* **281**, 424-428.

Severinov, K., Mustaev, A., Kashlev, M., Borukhov, S., Nikiforov, V., and Goldfarb, A. (1992) Dissection of the beta subunit in the Escherichia coli RNA polymerase into domains by proteolytic cleavage. *J. Biol. Chem.* **267**, 12813-12819.

Shaw, A. A., Falick, A. M., Shetlar, M. D. (1992) Photoreactions of thymine and thymidine with *N*-acetyltyrosine. *Biochemsitry* **31**, 10976-10983.

Shivanna, B. D., Mejillano, M. R., Williams, T. D., Himes, R. H. (1993) Exchangeable GTP binding site of beta-tubulin. Identification of cysteine 12 as the major site of cross-linking by direct photoaffinity labeling. *J. Biol. Chem.* **268**, 127-132.

Smith, K. C. (1962) Dose dependent decrease in extractability of DNA from bacteria following irradiation with ultraviolet light or with visible light plus dye. *Biochem. Biophy. Res. Comm.* **8**, 157-163.

Steen, H., and Jensen, O. N. (2002) Analysis of protein-nucleic acid interactions by photochemical cross-linking and mass spectrometery. *Mass Spectrometry Reviews* **21**, 163-182.

Suchanek, M., Radzikowska, A., and Thiele, C. (2005). Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nature Methods* **2**, 261–268.

Takeuchi, K., and Wagner, G. (2006) NMR studies of protein interactions. *Curr. Opin. Struct. Biol.* **16**, 109-117.

Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S., and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417**, 712-719.

Vo, N. V., Hsu, L. M., Kane, C. M., and Chamberlin, M. J. (2003) *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* **42**, 3798–3811.

Wiwczar, J. (2008) Probing an *E. coli* RNA polymerase open complex contact in the discriminator region of the T5 N25 promoter, Mount Holyoke College senior thesis.