THE EFFECT OF PARTIAL PROMOTER SEQUENCES ON PRIMER LABELING AND *DE NOVO* INITIATION BY T7 RNA POLYMERASE

by

Ramesh Padmanabhan

APPROVED BY SUPERVISORY COMMITTEE:

Dennis L. Miller, PhD, Chair

Hyun-Joo Nam, PhD

John G. Burr, PhD

Kelli L. Palmer, PhD

To my Mother, and Mother Goddess Devi

THE EFFECT OF PARTIAL PROMOTER SEQUENCES ON PRIMER LABELING AND *DE NOVO* INITIATION BY T7 RNA POLYMERASE

by

RAMESH PADMANABHAN, BSC, MSC, MS

DISSERTATION

Presented to the Faculty of

The University of Texas at Dallas

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY IN

MOLECULAR AND CELL BIOLOGY

THE UNIVERSITY OF TEXAS AT DALLAS

May 2018

ACKNOWLEDGMENTS

My sincere thanks to the faculty of the Department of Biological Sciences, The University of Texas at Dallas for giving me a wonderful opportunity to conduct research. I feel highly privileged to express my deep sense of gratitude and profound thanks to my advisor, Dr. Dennis L. Miller, for his able guidance, timely advice, and mentoring throughout my research. I offer my profound thanks to my committee members, Dr. Hyun-Joo Nam, Dr. John G. Burr, and Dr. Kelli L. Palmer, for their continuous support and insightful comments. On my personal account, I extend special thanks to Dr. Stephen Spiro and Dr. Lawrence Reitzer for their timely help and advice. Many thanks to Dr. Ruben Ramirez, and Dr. John Kolar, for all their support during my period as graduate student.

On a personal note, I ardently express my respect and gratitude to my mother, Raji Padmanabhan, for the sacrifices and prayers she made for me to finish my academic research. I also express my sincere thanks to my wife, Gomathy Radhakrishna Iyer, for rendering emotional support during testing times. My sincere thanks to my sister, Priya Padmanabhan, and her family to help me throughout my studies. Finally, my prayers to God Almighty and Goddess Devi, and all my teachers for giving strength and blessings to complete my research successfully.

March 2018

THE EFFECT OF PARTIAL PROMOTER SEQUENCES ON PRIMER LABELING AND *DE NOVO* INITIATION BY T7 RNA POLYMERASE

Ramesh Padmanabhan, PhD
The University of Texas at Dallas, 2018

Supervising Professor: Dennis L. Miller, PhD

The bacteriophage T7 RNA polymerase (T7 RNAP) is the prototype of the family of single subunit RNA polymerases which includes the T3, SP6 and mitochondrial RNA polymerases. It is also the most well characterized enzyme of this family of polymerases. T7 RNAP is the primary choice in studying the mechanistic aspects of transcription and promoter evolution owing to its high specificity for its promoter, requirement of no additional cofactors, and high fidelity of initiation from a specific site in its promoter. Although the two groups of single subunit polymerases, bacteriophage and mitochondrial, display a remarkable structural conservation, they recognize quite dissimilar promoters. Functional domains involved in promoter recognition and transcription initiation have been well characterized by thorough mutational studies by systematic deletions within these domains. T7 RNAP recognizes a 23 nucleotide (nt) promoter which can be divided into several functional domains. T7 RNAP can recognize a range of promoter sequences which are closely related to this consensus sequence. The promoter-specific recognition is achieved by the 13 nucleotide promoter recognition region which extends from -5 to -17 and contains sites important for polymerase recognition and positioning.

Here, we have developed an *in vitro* transcription system to study the ability of T7 RNAP to use truncated promoters similar to mitochondrial promoters. In this system, we used oligonucleotides which are capable of forming intra- and inter-molecular base pairing to produce a recessed 3'end on an extended 5' template. We then systematically deleted nucleotides from the 5' end of the 20-nucleotide double stranded region on these oligonucleotides containing the T7 RNAP promoter to determine (1) If they would label at the 3' end of the oligonucleotide, (2) Initiate template-dependent *de novo* transcription, (3) Initiate template-independent and promoter-dependent *de novo* transcription or, (4) Fail to incorporate label when supplied with T7 RNAP and a single radiolabeled ribonucleotide triphosphate, which can base pair with the first unpaired base in the 5' extension of the template. Under this condition, when a complete or almost complete (20 to 16 nt) double stranded T7 RNAP promoter is present, small RNAs are produced through template-independent and promoter-dependent stuttering corresponding to abortive initiation, which is lost when supplied with a completely scrambled promoter sequence. When partial double stranded promoter sequences (10-12 nt) are provided, template dependent *de novo* initiation of RNA occurs but at a site different from the canonical +1 initiation site (-GGG). In situations where no promoter is present but the potential exists to form a transient duplex region with a recessed 3' end, the T7 RNAP adds a templated nucleotide to the 3' end (primer labeling). Understanding the mechanisms underlying these observations helps us to understand the roles played by promoter elements. The evolution of promoter sequences of the single subunit RNAPs and to use this as a technique to synthesize defined RNAs without 5'- sequence constraints are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 2

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

1.1 Bacteriophage T7 RNA polymerase

The bacteriophage T7 RNA polymerase (T7 RNAP), encoded by gene 1 of bacteriophage T7 is required for the transcription of the late phase of the phage life cycle (Tabor & Richardson, 1985). The T7 RNA polymerase is a prototype of the super family of single subunit nucleic acid polymerases, which includes the Phage T3, SP6, and mitochondrial RNA polymerases, is the best characterized and understood in its class, and thus is the best choice to study the fundamental process of transcription (McAllister, 1993; Sousa, 2004; Tunitskaya & Kochetkov, 2002). T7 RNA polymerase finds a wide range of application in molecular cloning and gene expression studies because of its simple structure and molecular size of ~100 kDa. One of the most important characteristics of T7 RNAP is the high specificity for its promoter (Rong, He, McAllister, & Durbin, 1998). Unlike the bacterial and eukaryotic multi-subunit RNA polymerases, which require a wide variety of molecular sub-units and accessory transcription factors to control the transcription process, and are structurally complex (Borukhov & Severinov, 2002; Borukhov & Nudler, 2003; Cramer, 2002; Murakami & Darst, 2003), T7 RNA polymerase does not require any accessory factors, and is a highly processive enzyme (*Processivity in early stages of transcription by T7 RNA polymerase*). Other than its DNA-dependent RNA polymerase activity, T7 RNA polymerase is also able to replicate RNA hairpin loops in an RNA-dependent fashion (Arnaud-Barbe, Cheynet-Sauvion, Oriol, Mandrand, & Mallet, 1998). Although T7 RNAP is structurally homologous to the Pol I family DNA polymerase (DNAP), there are many

differences in terms of their functions (Cheetham & Steitz, 1999a). Unlike the DNAPs, RNAPs do not require a primer to initiate transcription, but they do have a requirement of a specific promoter for initiation. The T7 RNAP catalyzes the formation of a phosphodiester bond with the incoming ribonucleoside triphosphates (rNTPs) in the presence of two Mg^{2+} ions in the active site (Steitz, Smerdon, Jager, & Joyce, 1994). The extensive investigation of the crystal structure of T7 RNAP has provided a detailed account on mechanistic aspects of transcription by T7 RNA polymerase (Sousa, Chung, Rose, & Wang, 1993; Sousa, Rose, Je Chung, Lafer, & Wang, 1989).

1.2 T7 RNA polymerase promoter structure

The requirement of a specific T7 RNAP promoter for transcription initiation is an important characteristic of this enzyme that provides advantages as a model to study DNA-protein interactions (Tang, Bandwar, & Patel, 2005). The T7 RNA polymerase can recognize a range of promoter sequences which are closely related to each other by a consensus sequence (Dunn, Studier, & Gottesman, 1983). The T7 RNAP promoter has a 4-partite structure (Figure 1.1**)**, which contains a binding region that spans from -17 to -5 base-pairs, which helps in sequence-specific recognition by the polymerase, and an initiation sequence that spans from -4 to -1 which helps in melting the promoter sequence. Base-pairs extending from +1 to +6 represent the initially transcribed sequence (Brieba, Padilla, & Sousa, 2002; Sousa, 2013). The recognition region can be further subdivided into the specificity loop region (-12 to -5) and the AT-rich recognition loop region (-17 to -13).

Whereas, class I genes are transcribed by *E.coli* RNA polymerase, two classes of promoters are used by T7 RNAP to transcribe class II and class III genes which encode proteins involved in T7

replication, and phage structural proteins, respectively (Ikeda, Lin, & Clarke, 1992). The use of these promoters also depends on the variables like salt concentration, temperature, and presence of denaturing agents, and the conditions that stabilize the DNA helix (McAllister & Carter, 1980). It has been shown that higher salt concentrations inhibit class II T7 promoters as a function of temperature. Measurements of relative efficiencies of class

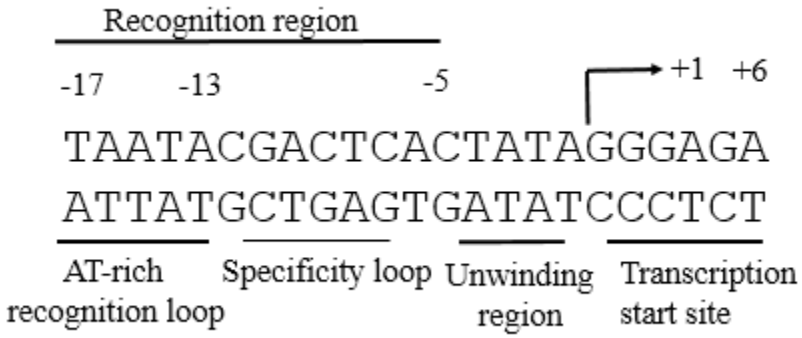


Figure 1.1. Modular organization of T7 RNA polymerase promoter**.**

II and class III promoters showed that Class III promoters are shown to be stronger than class II promoters, and class III promoters are relatively efficient in converting the initial transcription complex into a stable elongation complex (Ikeda, 1992). Other studies found no significant difference in the binding efficiency between class II and class III promoters, but the efficiency of initiation varied as a function of ionic concentration. Nitrocellulose filter binding assays using cloned T7 RNAP promoter were used to characterize the binding efficiency of T7 RNAP to the two classes of promoter (Smeekens & Romano, 1986). The results from this study indicated that the relative binding efficiencies by the two classes of promoters were identical based on their increase relative to ionic concentration. The initiation efficiencies, though, differed significantly in the presence of class II versus Class III promoters. The efficiency of initiation increased

initially followed by a decrease when the salt concentration was increased to 25 mM for the class III promoters. On the contrary, the efficiency of initiation from class II promoters continuously declined as the salt concentration was increased to 25 mM. Consistent with this observation, there was a significant increase in the binding efficiency of T7 RNAP to class III promoters in the presence of GTP using quantitative foot printing assay compared with class II promoters (Gunderson, Chapman, & Burgess, 1987). Transcription initiation using three different class II and class III promoters on linear templates showed an apparent contradiction of the earlier observed fact that class III promoters are qualitatively much stronger than class II promoters. This study showed that although, the class II promoters required a higher promoter concentration than that required for the class III promoters to achieve ½ V_{max}, there was no significant difference in promoter concentration between these promoters to achieve V_{max} (Ikeda, Lin, & Clarke, 1992). Further, this study also supported the relationship between the formation of a stable transcription complex and relative promoter strengths. Linear run-off templates from plasmids, containing the different sequences of the two promoter classes in the above study were used in run-off synthesis in the presence of excess amounts (200 nM) of T7 RNAP to study the relative efficiencies of formation of stable transcription complexes (Ikeda, 1992). The results from this study showed that the class III promoters were significantly stronger and able to produce strong transcription complexes when compared with class II promoters. Overall, the above studies indicate the high promoter binding capacity and efficiency of transcription of class III T7 RNAP promoters.

1.3 T7 RNA polymerase structure

The initial effort to clone and express the gene for the T7 RNAP was taken by Stahl and Zinn (Stahl & Zinn, 1981). They cloned the gene encoding the T7 RNAP extending from nucleotides 3127-5821 into the *PstI* site of the plasmid pBR322. This clone lacked a single nucleotide from the termination codon which would produce an aberrant protein with extra amino acids in the carboxy-terminus. Subsequently, the complete coding sequence of the T7 RNAP was cloned in the *BamHI* site of pBR322, and expressed abundantly from lac UV5 promoter in *Escherichia coli (E. coli)* (Davanloo, Rosenberg, Dunn, & Studier, 1984). Extensive studies have been conducted to determine the three dimensional structure of T7 RNAP both with single crystals at 3.3Å (Sousa, Rose, Chung, Lafer, & Wang, 1989), as well as in conjugation with initiation complex at 2.4 Å (Cheetham & Steitz, 1999), and elongation complex at 2.9 Å and 3 Å respectively (Durniak, Bailey, & Steitz, 2008; Tahirov et al., 2002), and T7 RNAP inhibitors like the T7 lysozyme (Jeruzalmi & Steitz, 1998a). These studies have provided an immense amount of information not only on the structure of T7 RNAP, but also have enabled an understanding the process of transcription initiation, elongation, and transition from initiation complex (IC) to a stable elongation complex (EC) (Cheetham, Jeruzalmi, & Steitz, 1999; T. A. Steitz, 2004).

T7 RNA polymerase has a modular organization with different functions of the enzyme controlled by specific domains (Figure 1.2). T7 RNAP shares sequence homology with single-subunit polymerases, like the eukaryotic mitochondrial RNA polymerases (Cermakian, Ikeda, Cedergren, & Gray, 1996; Chen, Kubelik, Mohr, & Breitenberger, 1996), and structural homology with many evolutionarily distant enzymes like the DNA pol I klenow (Kim et al.,

1995; T. A. Steitz, 1999) fragment, RNA-dependent DNA polymerase or reverse transcriptase (Kohlstaedt, Wang, Friedman, Rice, & Steitz, 1992; T. A. Steitz, 1997), and *Thermus aquaticus* DNA polymerase (Beese, Derbyshire, & Steitz, 1993).

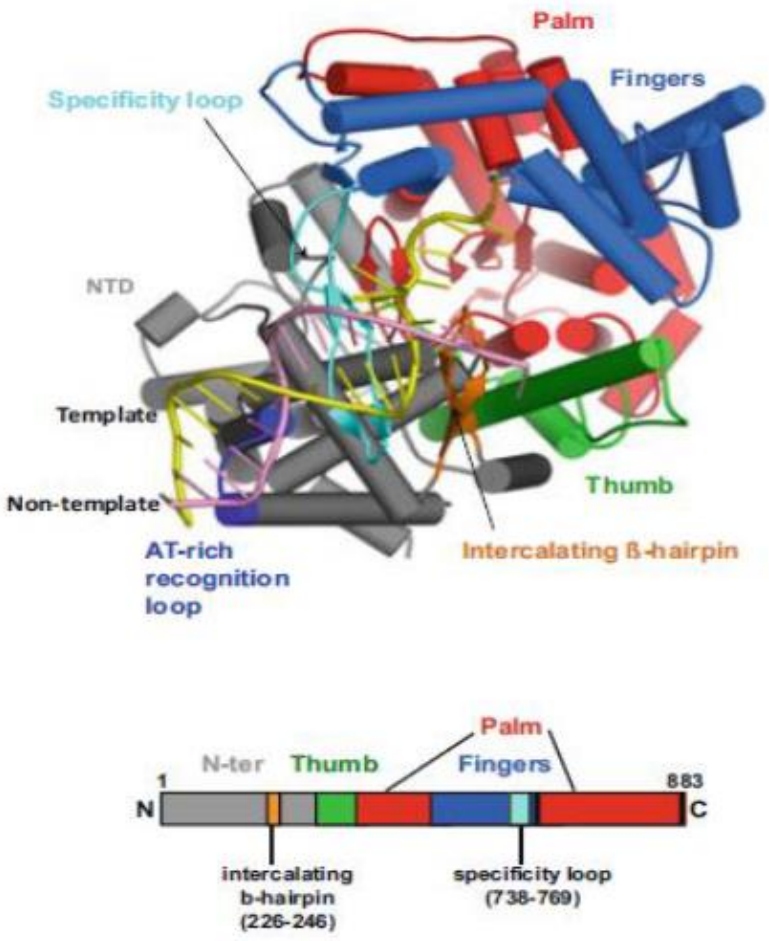


Figure 1.2. Structure of the T7 RNA polymerase with representative domains in complex with promoter DNA (template strand in yellow and non-template strand in pink). T7 RNA polymerase has a cupped right hand-like structure. The N-terminal domain is represented in grey, palm sub-domain in red, Thumb sub-domain in green, Fingers sub-domain in blue. The modular organization is also shown. [Adapted from: Basu, R.S. and Murakami, K.S. (2014) In Murakami, K. S. and Trakselis, M. A. (eds.), *Nucleic Acid Polymerases*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 237-250.]

Overall, the T7 RNAP consists of two distinct domains, the N-terminal domain (NTD) extending from 1-312 amino acids, and a polymerase domain, also known as the C-terminal domain (CTD), extending from 313-883 amino acids. The polymerase domain can be again sub-divided into the "fingers", the "palm", and the "thumb" subdomains; because they exhibit structural analogy to a cupped right hand (Jeruzalmi & Steitz, 1998a; Murakami, Trakselis, & SpringerLink, 2014). The "palm" sub-domain enclose a deep polynucleotide-binding cleft of 60Å length, 15-25 Å width and 25-40 Å depth which can hold two complete turns of the DNA double helix with the "fingers" and the thumb sub-domains forming the side walls of the cleft, containing charged amino acid residues. The walls are formed by α-helices and the back of the cleft is formed by β-strands (Tunitskaya & Kochetkov, 2002). This cleft contains the conserved active site aspartate residues, D537 and D812 that make coordinate bonds with two divalent Mg^{2+} ions for the formation of a phosphodiester bond during enzyme catalysis (Cheetham et al., 1999). The two Mg^{2+} ions are required to produce the 3' end for nucleophilic transfer reaction, and to equalize the charge distribution of the incoming nucleotide during the transition state (Murakami et al., 2014). The active site has been characterized in detail by generation of T7 RNA polymerase with mutations in the active site residues (D537A, D537S, D812A, D812S, D812N). These mutations caused decreased activity and processivity during both the abortive initiation and processive elongation phases, as well as an increase in the slippage-dependent poly (G) transcript synthesis during initiation (Bonner, Patra, Lafer, & Sousa, 1992; Bonner, Lafer, & Sousa, 1994). Targeted mutation of conserved residues in the A, B and C motifs of T7 RNA polymerase and DNA pol I Klenow fragment confirmed the proposal that the active sites of these nucleic acid polymerases shared similar structural characteristics (Bonner et al., 1992). Using electron paramagnetic

resonance (EPR) spectroscopy, the active site mutants D537N and D812N showed no activity at all, whereas D537E and D812E had very low activity indicating that the carboxyl groups of these aspartate residues are essential to coordinate the divalent metal ion during transcription (Woody, Eaton, Osumi-Davis, & Woody, 1996). A sequence alignment study using bacteriophage SPO2 polymerase, *E. coli* pol I klenow fragment, T7 and T5 RNA polymerases found that these aspartate residues are invariably conserved in all these polymerases and confirmed that these active site residues are required for polymerase activity (Delarue, Poch, Tordo, Moras, & Argos, 1990). These studies support the proposal that the active site provides correct geometry for the two-metal ion catalysis. Accessory domains of the T7 RNA polymerase N-terminal domain include amino acid residues 93-101 of the N-terminal domain that interact specifically with the AT-rich region of the promoter, and residues 232-242 which form the intercalating loop which inserts into the promoter DNA holds it in an open conformation. Accessory domains of the CTD include amino acid residues 739-770 (extended region of promoter recognition loop), amino acid residues 839-883 form a C-terminal loop ("fingers" sub-domain) which interacts with T7 lysozyme during the phage infection which shuts off transcription from the class II promoters (Jeruzalmi & Steitz, 1998; Sousa, 2004).

1.4 Mechanism of T7 RNA polymerase transcription

T7 RNA polymerase transcribes a template DNA in three distinct steps- (1) initiation, (2) elongation, and (3) termination mediated by the interaction of the promoter and the template DNA with different domains of the enzyme (Durniak, Bailey, & Steitz, 2008).

1.4.1 Initiation of transcription

Initiation is invariably the first step of the multi-step transcription, which is also known as the *de novo* RNA synthesis (Kennedy, Momand, & Yin, 2007). The general mechanism of initiation of transcription can be outlined as three distinct steps (Cheetham & Steitz, 1999, Ikeda & Richardson, 1986).

1. RNA polymerase binds to the promoter sequence during the promoter recognition step to form a closed complex (refer above for a detailed description of promoter).
2. Formation of an initiation-competent open promoter complex.
3. Initiation of RNA synthesis (de novo initiation and abortive cycling) and isomerization into a transcriptionally competent complex.

Promoter interaction has been extensively studied employing a range of biochemical methods including footprinting using the intercalating agent methidium-propyl-EDTA Fe(II) (Ikeda & Richardson, 1986), indicating that T7 RNA polymerase protects a region of -17 to -4 of the promoter. Methidium-propyl- EDTA Fe(II) combines the advantage of a metal chelator and an intercalating agent that cleaves the DNA thus providing a good agent for footprinting assays (Hertzberg & Dervan, 1984). Methylation and ethylation interference technique was used to study the polymerase-promoter interaction (Jorgensen, Durbin, Risman, & McAllister, 1991). This study demonstrated that specific residues in the major groove region -12 to -5 are important for promoter binding, and a close approach of RNAP to one side of the double helix (Diaz, Raskin, & McAllister, 1993).

During initiation, the NTD binds to the thirteen base pair sequence -17 to -5 (binding domain) of the promoter, followed by the melting of the AT-rich region of the promoter leading to the direction of the template strand into the catalytic cleft of the enzyme (Cheetham, Jeruzalmi, & Steitz, 1999). Promoter-specific recognition is mediated by interaction between a small insertion between the polymerase fingers and the palm sub-domains (739-770), and template bases -8 to -11 in the major groove. The residues that make specific contacts with the bases are N748, Q758, R746, and N748 (Cheetham, Jeruzalmi, & Steitz, 1999c; Rong, He, McAllister, & Durbin, 1998). The N748 residue is involved in discriminating bases -10 and -11, and substituting N748 with a corresponding Asp residue results in a switch in promoter recognition specificity especially between bases -10 and -11 (Raskin, Diaz, Joho, & McAllister, 1992). Q758 makes specific contact with the base at the -8 position which is reasonable for its location (Figure 1.3). ΔQ758K base substitution led to the altered promoter specificities for Sp6 or K11 polymerases. Consistent with this, mutant RNAPs with Q758K and Q758R base substitutions with altered specificities also exhibited decreased affinity to synthetic promoters (Rong et al., 1998).

Moreover, residues 93-101 in the NTD, called the AT-rich recognition loop, recognizes the DNA sequence -17 to -13 in the minor groove (Cheetham et al., 1999). In addition, the open promoter complex is stabilized by a stacking interaction mediated by V237 in the β-intercalating hairpin (amino acid residues (232-242) which also plays a role in melting the promoter DNA to expose single-stranded template DNA for transcription to start (Brieba & Sousa, 2001). Consistent with this observation, mutation of V237 or deletion of amino acids in the hairpin region showed smaller amounts of pre-initiation open complex at equilibrium with no effect on affinity of

mutants for promoter DNA, indicating an alteration in the rate of change of promoter opening and closing steps (Stano & Patel, 2002). Effect of deletion of promoter regions involved in strand recognition and opening were also kinetically studied in detail using synthetic promoters, and partially deleted non-template strands (Martin & Coleman, 1987; Milligan, Groebe, Witherell, & Uhlenbeck, 1987). Using partially single stranded promoter template which was base-paired only from -17 to +1, it was shown that this has no effect in transcription. It was also shown that the top strand of the promoter (5'-3' strand) can be shorter by 3 nucleotides from both the ends with no effect on transcription.

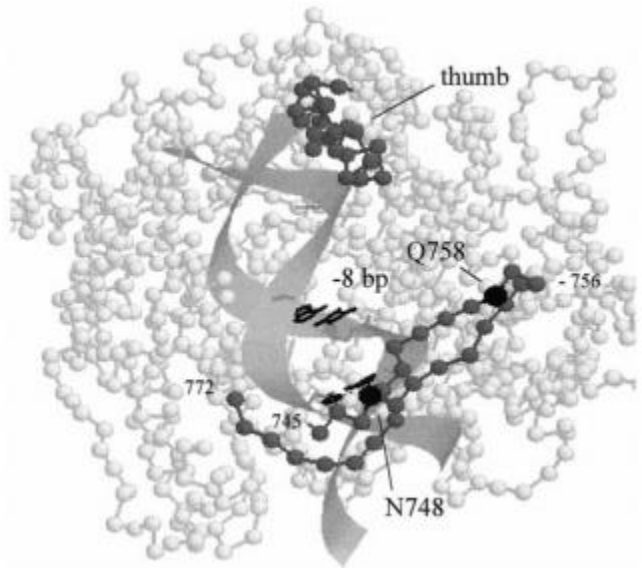


Figure 1.3. Specificity loop residues N748 and Q758 play a role in specific contacts with T7 RNA polymerase. (Adapted from: Rong, M., He, B., McAllister, W. T. and Durbin, R. K. (1998) Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci. U. S. A.,* **95,** 515-519. Copyright © (1998) National Academy of Sciences, USA.)

Moreover, promoter truncation from the non-template strand from bases 17 to -14 and -3 to -6 does not have any effect on the transcription reaction, showing that T7 RNAP does not require

DNA from the non-template strand in this region for activity (Milligan et al., 1987). Kinetic analysis using synthetic oligonucleotide-based promoters indicated that removal of non-template strand from +1 to +5 positions has little effect on binding while there was a 2-fold increase in kinetics. Further deletion through base -4 has no effect on binding, but deletion through base -6 weakens binding while specific activity is unaffected (Maslak & Martin, 1993).

After successful identification and binding to the promoter sequence, the T7 RNAP melts the promoter to initiate transcription by generating single stranded template DNA. As mentioned above, this melting process is mediated by specific interaction of V237 in the β-intercalating hairpin loop. Once the template DNA is melted at the transcription initiation site, the DNA undergoes bending (Tang & Patel, 2006). A similar type of bending followed by promoter melting has also been observed in *E. coli*, where the DNA is bent by 45^{o} near the transcription initiation site (Craig, Suh, & Record, 1995; Meyer-Almes, Heumann, & Porschke, 1994; Rees, Keller, Vesenka, Yang, & Bustamante, 1993). It is observed that the yeast *Saccharomyces cerevisiae* mtRNA (mitochondrial) polymerase also induces bending after successful binding to the promoter DNA, which is correlated to the efficiency of transcription, and bending is an integral process during transcription initiation and promoter utilization (Schinkel, Groot Koerkamp, Teunissen, & Tabak, 1988). Modeling studies with T7 RNA polymerase using long DNA sequences incorporated into the ternary initiation complex observed an 80^{o} bend in the promoter DNA centered around the transcription start site. Steady state and time resolved fluorescence resonance energy transfer (FRET) was used to measure FRET efficiency using a duplex promoter and a pre-melted bubble promoter (Tang & Patel, 2006). The FRET efficiency,

and therefore shortening of end-to-end distance, is correlated to the amount of promoter bending. FRET efficiency changes were relatively larger when pre-melted bubble promoter bound T7 RNAP or when the double promoter was supplied along with initiating nucleotides. FRET efficiency was smaller when only double promoter was used. In the absence of the initiating nucleotide, there is an equilibrium existing between closed and open complex, and the DNA is bent $<40^{o}$ in the absence if the initiating nucleotide in the closed complex, but the DNA is bent at 86^{o} in the open complex. This observation supports the proposal that the promoter opens spontaneously when the DNA is bent by gaining the free energy obtained through the interactions with the melted promoter (Tang & Patel, 2006).

1.4.2 *De novo* initiation and abortive synthesis of short RNAs

As mentioned earlier, T7 RNAP can initiate RNA synthesis *de novo* without the presence of a primer or co-factors, unlike large multi-subunit RNA polymerases as bacterial or eukaryotic RNA polymerases (Ling, Risman, Klement, McGraw, & McAllister, 1989; Kennedy, Momand, & Yin, 2007; T. A. Steitz, 2004). During this phase, the enzyme catalyzes the formation of a phosphodiester bond between the first two nucleotides, followed by elongation from this dinucleotide. The mechanism of *de novo* initiation differs considerably from the subsequent steps of transcription. Kinetically, the rate of addition of initiating nucleotides is very slow as indicated by the rate of formation of first phosphodiester bond at 7.8 s^{-1}, although the rate of formation of phosphodiester bond formation during the elongation phase is ~220 s^{-1}. The efficiencies of the initiation phase and the elongation phase, therefore, determine the overall yield of the full-length transcript (Anand & Patel, 2006). Affinity of the incoming nucleotides also differs substantially between the initiation and elongation phases. The affinity of the polymerase to the initiating

nucleotide triphosphate (NTP), indicated by the dissociation constant, k_d, was found to be 2 mM for the initiating NTP (Bandwar, Jia, Stano, & Patel, 2002), whereas the k_d value during elongation was found to be ~5 µM (Guajardo, Lopez, Dreyfus, & Sousa, 1998). Thus, a higher concentration of NTP is required during the initiation process compared with the transcription extension during the elongation stage. T7 RNAP has a strong preference for GTP as the initiating nucleotide (Kuzmine, Gottlieb, & Martin, 2003). This choice of GTP as the initiating nucleotide is also found in other members of this T7 family like phages T3, SP6 (Basu & Maitra, 1986), as well as in many pathogenic viruses like the dengue hepatitis C, and West-Nile viruses (Selisko et al., 2006). Sequence alignment of 17 T7 RNAP promoters showed that 15 promoters had GTP at the +1 position, whereas two had ATP at the same position (Dunn & Studier, 1983). Nucleotide-specific pausing experiments with SP6 RNAP demonstrated a marked decrease in the level of RNA production on promoters encoded with an A at the +1 site (Nam & Kang, 1988). The basis for this strong bias for GTP as the initiating nucleotide, and the differences in the mechanism for NTP selection during initiation and elongation were delineated by studying the structures of the ternary complexes formed by T7 RNAP with partial duplexes (Kennedy, Momand, & Yin, 2007). NTP-binding sites and mechanism for selection of the initiating NTP are different during the initiation and the elongation stages. While binding of the initiating nucleotide does not induce a conformational change, there is a conformational change induced in the fingers domain when an elongating nucleotide binds. The initiating nucleotide, in this case, GTP makes specific contacts with R425 and Y427 of the polymerase. T7 RNAP takes advantage of these specific interactions to stabilize the heteroduplex formation during *de novo* initiation and shifts the reaction equilibrium towards open promoter complex.

During *de novo* initiation, the polymerase enters a phase of abortive initiation during which short transcripts of ~2-6 nucleotides (nt) are produced before the initiation complex (IC) transitions into the elongation complex (EC) (Tahirov et al., 2002; Yin & Steitz, 2002). After the synthesis of 10-12 nucleotides, the reaction transitions from the initiation phase to the elongation phase, completing the transcription in a processive manner until termination when the complete RNA transcript is released (Steitz, 2004). Abortive initiation has been observed in both *E. coli* and eukaryotic RNA polymerases and therefore it is a fundamental process during early transcription in all RNA polymerases (Martin, Muller, & Coleman, 1988). Several models have been proposed to explain the process of abortive initiation (Cheetham & Steitz, 1999). In the "inchworming" model (Figure 1.4), during each cycle of abortive initiation, one module in the RNAP detaches from the second module using a flexible element. Once the abortive transcript is released, this module retracts back to its initial position. In the "DNA scrunching" model (Figure 1.4), the transcribed template strand collects within the active site pocket of the polymerase 1 base pair per phosphodiester bond, with the polymerase remaining unchanged (Kapanidis et al., 2006). Fluorescence resonance energy transfer (FRET), and confocal optical microscopy using alternating laser excitation studies on *E. coli* RNA polymerase and a consensus *E. coli* promoter lacCONS concluded that initial transcription involves DNA scrunching and not transient excursion or inchworming (Kapanidis et al., 2006).

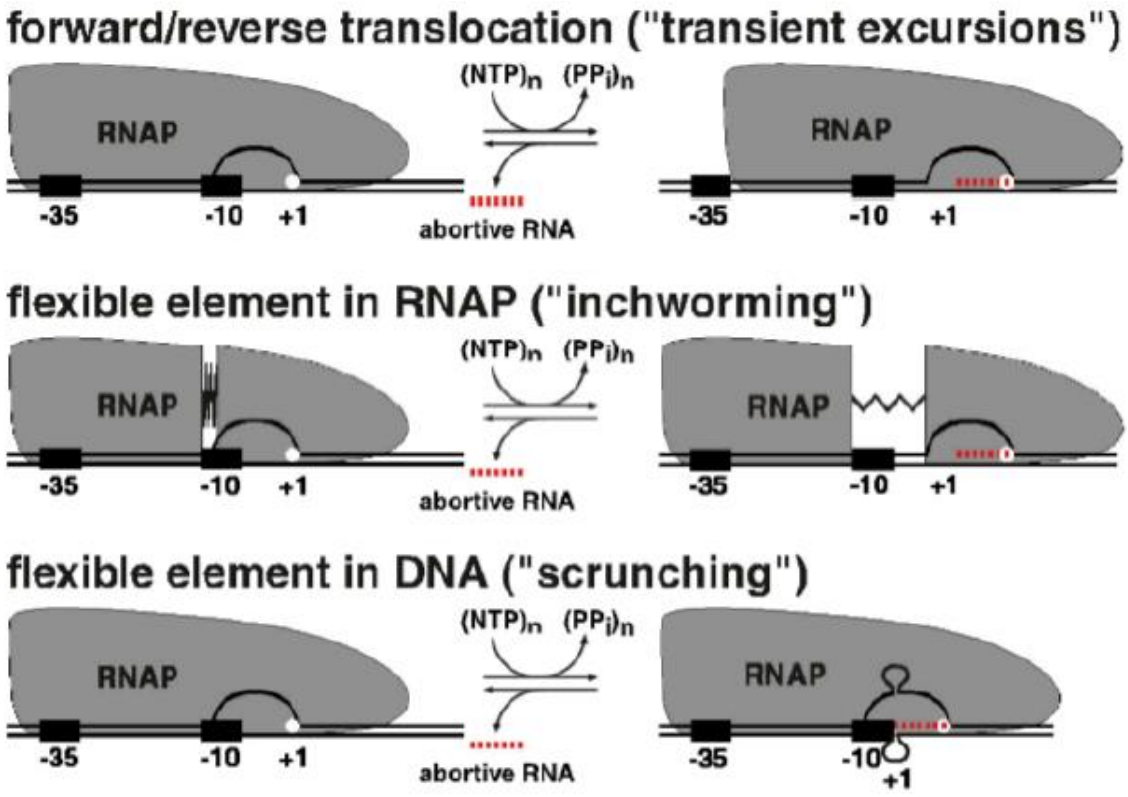


Figure 1.4. The three models initially proposed for abortive initiation of *E.coli* RNAP. As per the transient excursion model, the RNAP translocated forward as a unit. In the inchworming model, the active center of the RNAP translocated downstream with the rest of the enzyme intact with the help of a flexible region. In the Scrunching model, the polymerase accumulates the DNA template as a bulging forms and at the end of the cycle, the enzyme extrudes the accumulated DNA (Adapted from: 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).

Two other models proposed for abortive initiation account for the stress induced by the RNAP deformations and DNA scrunching in transition to the elongation state (Tang, Roy, Ha, & Patel, 2008). In RNAP translation or shifting model, the NTD shifts gradually from the fixed CTD during initiation (Theis, Gong, & Martin, 2004; Turingan, Theis, & Martin, 2007). In the rotation or twist model, the NTD rotates to provide space for the expanding RNA: DNA heteroduplex

which begins when ~4-5 nucleotides are added, and continues until a stable EC is formed (Tahirov et al., 2002).

Ensemble and FRET experiments were conducted to investigate the DNA scrunching model and rotation of NTD during transcription initiation by T7 RNAP (Kapanidis et al., 2006). In this experiment, they, tagged fluorescent donor and acceptor dye pairs at specific sites in the T7 RNAP promoter consensus sequence. The average donor-acceptor distances calculated from observed ensemble FRET using free DNA, RNAP-DNA complex, RNAP-DNA initiation complex making

2-7 nucleotide RNA, and RNAP-DNA elongation complex making 10 nt RNA. The mean donor-acceptor distance decreased from 80 Å in the free DNA to 4 Å in the 2-4 nucleotide state and further decreased in complexes making 7 nucleotide RNA. These results indicate that the template DNA is scrunched in the initiation complex. The FRET measurements also concluded a $18^{o}\pm4^{o}$ rotation of the promoter during 4 to 7 nucleotide synthesis. This degree of rotation is small compared to other studies from the crystal structures of IC and EC (Tahirov et al., 2002; Yin & Steitz, 2002). Other studies showed the inherent instability of the initiation complexes due to DNA scrunching during abortive initiation. Yield and stability measured on stalled complexes concluded that the instability of initiation complexes is due to downstream bubble collapse (Gong & Martin, 2006). Mutational studies have also been carried out to investigate the instability of abortive initiation. A polymerase mutation that reduces this instability during initiation has been isolated (Ramirez-Tapia & Martin, 2012). In this study, a mutant P266L was

isolated where this mutation was distant from the promoter-binding interface. Fluorescence anisotropy was used to compare the binding affinity of this mutant enzyme and the wild-type enzyme to a partially single-stranded promoter. It was found that both versions had similar binding affinities indicated by K_d values of 4.1μM (P266L) and 4.5μM (wild-type), in contrast to a previous study with the same mutation which reported a low binding affinity of the promoter to the P266L mutant enzyme promoter (Guillerez, Lopez, Proux, Launay, & Dreyfus, 2005). Limited proteolysis studies using the same mutant also showed that the transition to elongation takes place at much longer RNA lengths, prompting them to propose a protein "push-back" model for the inherent instability during abortive cycling (Ramirez-Tapia & Martin, 2012). This model proposed that there is an equal and concomitant pushing occurring on the DNA-RNA hybrid and the N-terminal promoter-binding domain of the polymerase. This push back promotes a rotation of the N-terminal promoter binding platform that leads to promoter release and transition to elongation. The P266L mutant showed a characteristic delay in transition to elongation due to a positionally delayed barrier to rotation. Another study explored the effect of structure and sequential changes on promoter function (Jiang, Rong, Martin, & McAllister, 2001). In this study modified promoter sequences were used by introducing nicks between the promoter-binding region and the initiation region. It was found that the nicked promoters not only showed a greater flexibility in the choice of the initiation site, but also, a reduced synthesis of poly (G) products and eliminated the synthesis of abortive products. These studies clearly indicate the importance of transitioning from an unstable abortive initiation to a stable elongation, and during this transition the polymerase undergoes extensive structural transition.

1.4.3 Transition to elongation: Entry into a stable, processive phase

For the transition from the unstable initiation phase to a more stable and processive elongation phase, the polymerase must get promoter clearance and break up all promoter contact. This is considered as the most important step in this transition (Martin, Muller, & Coleman, 1988). The average length of transition is about 10 base pairs suggesting that this is a basic feature of transcription in all RNA polymerases (Liu & Martin, 2002). In the case of T7 RNA polymerase transcription, this transition occurs when the RNA-DNA duplex reaches a length of 8-9 base pairs resulting in the collapse of the initiation bubble and promoter release (Tahirov et al., 2002). Comparing the structure of the initiation complex and elongation complex reveals extensive structural alterations in the N-terminal domain (NTD) of the enzyme, and formation of a channel that accommodates a 7-bp heteroduplex and a tunnel through which the transcript passes after detachment from the heteroduplex, all of which account for the increased stability and processivity of the EC (Durniak, Bailey, & Steitz, 2008). Structural studies of the EC have revealed a gradual transition from the IC to the EC with majority of the changes happening in the NTD which leads to abandoning of sequence-specific contacts with promoter DNA, with relatively less changes happening in the CTD (Yin & Steitz, 2002). Structure of a co-crystal containing T7 RNAP, transcription bubble, and mRNA was studied at 2.1 Å, revealing a gradual transition from initiation to the elongation phase. The NTD underwent notable conformational changes leading to altered shape and function (Yin & Steitz, 2002). Most notable of the changes include- a rigid body rotation of helices D, E, F, G, I, J and the intercalating β-hairpin loop by 140^{o} which abrogates the initial contacts of the T7 RNAP with the promoter, leading to promoter clearance. The second conformational change occurred in the C1 helix which extended from 22-

30A^{o} and this extended C1 helix now protrudes into the space where the six-helix bundle previously occupied during initiation. The third conformational change involves a refolding of residues 160-190 forming a new subdomain H. This domain moves 70 Å forming an RNA exit tunnel. Similar conformational changes in the NTD were also observed at 2.9 Å resolution using T7 RNAP elongation complex (Tahirov et al., 2002). These conformational changes exhibited multifunctional interactions during elongation. One major role of the NTD conformation during the elongation complex is to make sufficient contacts with the template and the non-template DNA strands.

At the CTD, the thumb subdomain rotates by 15^{o} from its initial conformation during the initiation and along with the H subdomain, it contacts the non-template DNA strand. Another change in the CTD involves the specificity loop. The specificity loop makes sequence-specific contacts during initiation by interacting with the major groove. This interaction blocks the tunnel through which RNA exits. During elongation, the specificity loop makes a turn in such a way that it opens the exit tunnel interacting with the 5' end of the mRNA in the exit tunnel during elongation. This conformational change in the specificity loop not only helps in promoter clearance, but this stretch of hydrophobic residues is also found to be important in responding to termination signals. The specificity loop, therefore, functions differently during the initiation and elongation phase.

1.4.4 Formation and translocation of a mature elongation complex

Formation of a complete elongation complex occurs when the RNA transcript reaches 8-12 nucleotides in length, becoming resistant to salt denaturation, and changing the footprint of the

enzyme on the template (Mentesana, Chin-Bow, Sousa, & McAllister, 2000). This final step in transition has been found to be stabilized by interactions between the displaced transcript with the exit tunnel (Tahirov et al., 2002). Whereas the exit tunnel interacts with the displaced transcript, the entry pore interacts with the upstream substrate elements which become more accessible in the EC due to the 15º movement of the thumb subdomain. Similar entry pore and exit tunnel occur in the ECs of multi-subunit RNA polymerases including *E. coli* RNAP. One major contrasting feature between these RNAPs is in the rate of polymerization. Whereas the lengths of these pores are longer in multi-subunit RNAPs, they are much shorter in T7 RNAP which accounts for a higher rate of polymerization of 200 nucleotides/second for T7 RNAP compared with 30 nucleotides/second for *E. coli* RNAP (Transcription is catalyzed by RNA polymerase - biochemistry - NCBI bookshelf.; Tahirov et al., 2002).

At least two distinct steps are involved in the transition to a stable and mature elongation complex. First, the nascent transcript associates with the product-binding region in the NTD of the polymerase and second, the thumb sub-domain folds to close the DNA-binding site (Bonner, Lafer, & Sousa, 1994; Muller, Martin, & Coleman, 1988). By halting T7 RNAP at defined intervals downstream from the promoter, it was shown that the transition to a stable elongation complex occurs in multiple steps, and the non-template DNA strand plays a major role in stabilizing the EC (Mentesana et al., 2000). Wild-type (WT) polymerase, mutant polymerase defective in the product binding site (Δ172-173), and polymerase containing mutant thumb domain behaved differently with duplex template, partially single stranded template and super-coiled template. WT enzyme is most stable with duplex DNA, favoring displacement of nascent

transcript whereas it is very unstable with super-coiled and PSS templates. Δ172-173 showed destabilization with all templates as it could not bind the nascent transcript. The polymerase that contained mutant thumb sub-domain could bind the RNA, but was very unstable with the duplex DNA, as it lost the residues that contributed to complex stabilization. These observations corroborate the importance of the two distinct steps mentioned above in stabilizing the mature elongation complex. This finding of the role of the non-template strand in stabilizing the EC was in complete contrast to the destabilizing effect of the non-template strand observed on binary and initiation complexes (Villemain, Guajardo, & Sousa, 1997). Analysis of halted ECs at different positions on a single template by digestion with exonuclease III, λ exonuclease, RNase T1 and $KMnO_4$ digestion revealed that the transcription bubble spans 9 bases in length, and the RNA-DNA hybrid spans 7-8 base pairs (Huang & Sousa, 2000). Probing the ECs with and without NTP revealed that the NTP stabilizes the polymerase in the post-translocated position with decreased lateral mobility, whereas in the absence of NTP, the polymerase was free to slide along the template (Huang & Sousa, 2000).

Two models have been proposed for the movement of EC during translocation. The power stroke mechanism proposes that the EC is propelled along the template by coordinated conformational alterations, and the energy for translocation derives from the breakdown of the incoming NTP which is then converted into mechanical energy by protein conformational changes which is then used for propelling the polymerase down the template (Yin & Steitz, 2004). This model assumes that the polymerase has completed all conformational changes and is ready to accept an NTP. The second model is called the Brownian ratchet or more commonly known as the passive

sliding model, or translocational or positional equilibrium model (Bar-Nahum et al., 2005; Guo & Sousa, 2006; Huang & Sousa, 2000; Komissarova & Kashlev, 1997). In this model, the polymerase slides back and forth based on its interactions with the RNA and DNA, as well as on the relative stabilities of competing conformations between the polymerase and the RNA:DNA hybrid or the transcription bubble. The passive sliding model allows for a temporal equilibrium that exists between the pre- and the post-translocated conformations. Evidence in support of the passive sliding mechanism came from studies involving the use of translocation protein roadblocks immediately downstream of the promoter and evaluation of the effects of varying NTP concentration (Guajardo, Lopez, Dreyfus, & Sousa, 1998). The roadblock stalls translocation by creating a physical barrier for the movement of the ternary complex rather than destabilizing it. Supporting evidence for the passive sliding mechanism has also been obtained by studying protein roadblocks in *E. coli* (Abbondanzieri, Greenleaf, Shaevitz, Landick, & Block, 2005; Mosrin-Huaman, Turnbough, & Rahmouni, 2004). Such studies also demonstrate the important role played by the fluctuations of intracellular NTP concentrations on transcription elongation. Exertion of an opposing mechanical force stalls the translocation of the EC (Thomen, Lopez, & Heslot, 2005). In this single molecule study, a mechanical force exerted on the EC at steady-state transcription elongation condition acted as a competitive inhibitor of NTP binding which ultimately affected the translocation rate.

1.4.5 Transcription termination: dissociation of the transcription apparatus and the release of the nascent RNA

Termination is the final step in transcription where the transcription complex dissociates to release the nascent RNA. Termination is characterized by a reversal of steps from the formation

of an unstable IC to a stably transcribing EC, the mechanism of which has undergone less investigation (Lyakhov et al., 1998; Sousa, Patra, & Lafer, 1992; Zhang & Studier, 1997). Termination is a highly specific process because only characteristic configurations of RNA are recognized as termination signals, and recognition of these signals alone triggers the reversal of the isomerization and dissociation of transcription complex (Macdonald, Durbin, Dunn, & McAllister, 1994). T7 RNAP, like *E. coli* RNAP, is prone to terminate at any GC rich hairpin followed a poly-U string (Dunn & Studier, 1983). T7 RNAP will also terminate at its own terminator encoded in T7 DNA and at *E. coli* threonine attenuator (Jeng, Gardner, & Gumport, 1990).

Two classes of T7 RNAP terminators have been identified- class I and class II termination signals. Class I termination signals are typified by T7-Tφ present in the late region of the T7 DNA (Dunn & Studier, 1983; Sousa et al., 1992). T φ termination sites are characterized by the ability to form stable stem-loop structures followed by a stretch of U residues. Such sites are reminiscent of Rho-independent termination signals recognized by *E. coli* RNAP, and many similar sites cause termination by both T7 and *E. coli* RNAPs (Garcia & Molineux, 1995; Zavriev & Shemyakin, 1982). The second type of termination site does not involve stem-loop formation, and was first identified in a cloned human gene for preproparathyroid (PTH) hormone gene (Macdonald et al., 1994; Mead, Szczesna-Skorupa, & Kemper, 1986). The class II site contains a conserved 8 bp sequence CATCTGTT and termination occurs 7-8 nucleotides downstream of this sequence (He et al., 1998). Another class II termination site is found in the concatemer junction (CJ) of the replicating T7 DNA which is required for the normal growth of

T7 phage. Consistent with this observation, a mutant polymerase fails to recognize class II signals affecting the subsequent packaging and maturation of the phage particles (Lyakhov et al., 1997).

The reversal of isomerization model for transcription termination was based on observations that proteolytically modified T7 RNAP dissociates more readily at natural termination sites (Ikeda & Richardson, 1987; Macdonald et al., 1994; Muller, Martin, & Coleman, 1988). A trypsinized T7 RNAP containing 80 kDa and 20 kDa fragments generates a less stable elongation complex, and is blocked at the transition from IC to EC. This instability of forming a ternary complex was attributed to the failure of the nicked polymerase to bind single-stranded RNA. However, the process of termination was demonstrated to be a highly specific event because the nicked enzyme was less efficient in terminating at T7-Tφ and PTH sites (Macdonald, Zhou, & McAllister, 1993). Furthermore, insertion of single glycine residues or deleting a stretch of amino acids between 177-188 resulted in production of a polymerase that had similar termination properties as that of the nicked enzyme (Sousa, Patra, & Lafer, 1992b). The efficiency of termination at class I and class II sites depends on the relative rates of dissociation of the transcription complex to resuming the elongation conformation (Lyakhov et al., 1998b). The mechanism by which the T7 lysozyme inhibits T7 RNAP is by interfering with the conformational transition from IC to EC, and by forcing it to pause at termination sites (Kumar & Patel, 1997).

1.5 References

Processivity in early stages of transcription by T7 RNA polymerase - American Chemical Society. doi:- 10.1021/bi00411a012

Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R., & Block, S. M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature, 438*(7067), 460-465. doi:nature04268 [pii]

Anand, V. S., & Patel, S. S. (2006). Transient state kinetics of transcription elongation by T7 RNA polymerase. *The Journal of Biological Chemistry, 281*(47), 35677-35685. doi:M608180200 [pii]

Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B., & Mallet, F. (1998). Transcription of RNA templates by T7 RNA polymerase. *Nucleic Acids Research, 26*(15), 3550-3554. doi:gkb588 [pii]

Bandwar, R. P., Jia, Y., Stano, N. M., & Patel, S. S. (2002). Kinetic and thermodynamic basis of promoter strength: Multiple steps of transcription initiation by T7 RNA polymerase are modulated by the promoter sequence. *Biochemistry, 41*(11), 3586-3595. doi:bi0158472 [pii]

Bar-Nahum, G., Epshtein, V., Ruckenstein, A. E., Rafikov, R., Mustaev, A., & Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. *Cell, 120*(2), 183-193. doi:S0092867404011493 [pii]

Basu, S., & Maitra, U. (1986). Specific binding of monomeric bacteriophage T3 and T7 RNA polymerases to their respective cognate promoters requires the initiating ribonucleoside triphosphate (GTP). *Journal of Molecular Biology, 190*(3), 425-437. doi:0022-2836(86)90013-6 [pii]

Beese, L. S., Derbyshire, V., & Steitz, T. A. (1993). Structure of DNA polymerase I klenow fragment bound to duplex DNA. *Science (New York, N.Y.), 260*(5106), 352-355.

Bonner, G., Lafer, E. M., & Sousa, R. (1994). Characterization of a set of T7 RNA polymerase active site mutants. *The Journal of Biological Chemistry, 269*(40), 25120-25128.

Bonner, G., Lafer, E. M., & Sousa, R. (1994). The thumb subdomain of T7 RNA polymerase functions to stabilize the ternary complex during processive transcription. *The Journal of Biological Chemistry, 269*(40), 25129-25136.

Bonner, G., Patra, D., Lafer, E. M., & Sousa, R. (1992). Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *The EMBO Journal, 11*(10), 3767-3775.

Borukhov, S., & Nudler, E. (2003). *RNA polymerase holoenzyme: Structure, function and biological implications* doi:http://dx.doi.org/10.1016/S1369-5274(03)00036-5

Borukhov, S., & Severinov, K. (2002). *Role of the RNA polymerase sigma subunit in transcription initiation* doi:http://dx.doi.org/10.1016/S0923-2508(02)01368-2

Brieba, L. G., Padilla, R., & Sousa, R. (2002). Role of T7 RNA polymerase His784 in start site selection and initial transcription. *Biochemistry, 41*(16), 5144-5149. doi:bi016057v [pii]

Brieba, L. G., & Sousa, R. (2001). The T7 RNA polymerase intercalating hairpin is important for promoter opening during initiation but not for RNA displacement or transcription bubble stability during elongation. *Biochemistry, 40*(13), 3882-3890. doi:bi002716c [pii]

Cermakian, N., Ikeda, T. M., Cedergren, R., & Gray, M. W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Research, 24*(4), 648-654. doi:5t0662 [pii]

Cheetham, G. M., Jeruzalmi, D., & Steitz, T. A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature, 399*(6731), 80-83. doi:10.1038/19999 [doi]

Cheetham, G. M., & Steitz, T. A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science (New York, N.Y.), 286*(5448), 2305-2309. doi:8116 [pii]

Chen, B., Kubelik, A. R., Mohr, S., & Breitenberger, C. A. (1996). Cloning and characterization of the neurospora crassa cyt-5 gene. A nuclear-coded mitochondrial RNA polymerase with a polyglutamine repeat. *The Journal of Biological Chemistry, 271*(11), 6537-6544.

Craig, M. L., Suh, W. C., & Record, M. T.,Jr. (1995). HO. and DNase I probing of E sigma 70 RNA polymerase--lambda PR promoter open complexes: Mg2+ binding and its structural consequences at the transcription start site. *Biochemistry, 34*(48), 15624-15632.

Cramer, P. (2002). Multisubunit RNA polymerases. *Current Opinion in Structural Biology, 12*(1), 89-97. doi:https://doi.org/10.1016/S0959-440X(02)00294-4

Davanloo, P., Rosenberg, A. H., Dunn, J. J., & Studier, F. W. (1984). Cloning and expression of the gene for bacteriophage T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America, 81*(7), 2035-2039.

Delarue, M., Poch, O., Tordo, N., Moras, D., & Argos, P. (1990). An attempt to unify the structure of polymerases. *Protein Engineering, 3*(6), 461-467.

Diaz, G. A., Raskin, C. A., & McAllister, W. T. (1993). Hierarchy of base-pair preference in the binding domain of the bacteriophage T7 promoter. *Journal of Molecular Biology, 229*(4), 805-811. doi:S0022283683710867 [pii]

Dunn, J. J., & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *Journal of Molecular Biology, 166*(4), 477-535.

Durniak, K. J., Bailey, S., & Steitz, T. A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science (New York, N.Y.), 322*(5901), 553-557. doi:10.1126/science.1163433 [doi]

Garcia, L. R., & Molineux, I. J. (1995). Rate of translocation of bacteriophage T7 DNA across the membranes of escherichia coli. *Journal of Bacteriology, 177*(14), 4066-4076.

Gong, P., & Martin, C. T. (2006). Mechanism of instability in abortive cycling by T7 RNA polymerase. *The Journal of Biological Chemistry, 281*(33), 23533-23544. doi:M604023200 [pii]

Guajardo, R., Lopez, P., Dreyfus, M., & Sousa, R. (1998). NTP concentration effects on initial transcription by T7 RNAP indicate that translocation occurs through passive sliding and reveal that divergent promoters have distinct NTP concentration requirements for productive initiation. *Journal of Molecular Biology, 281*(5), 777-792. doi:S0022-2836(98)91988-X [pii]

Guillerez, J., Lopez, P. J., Proux, F., Launay, H., & Dreyfus, M. (2005). A mutation in T7 RNA polymerase that facilitates promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America, 102*(17), 5958-5963. doi:0407141102 [pii]

Gunderson, S. I., Chapman, K. A., & Burgess, R. R. (1987). Interactions of T7 RNA polymerase with T7 late promoters measured by footprinting with methidiumpropyl-EDTA-iron(II). *Biochemistry, 26*(6), 1539-1546.

Guo, Q., & Sousa, R. (2006). Translocation by T7 RNA polymerase: A sensitively poised brownian ratchet. *Journal of Molecular Biology, 358*(1), 241-254. doi:S0022-2836(06)00166-5 [pii]

He, B., Kukarin, A., Temiakov, D., Chin-Bow, S. T., Lyakhov, D. L., Rong, M., . . . McAllister, W. T. (1998). Characterization of an unusual, sequence-specific termination signal for T7 RNA polymerase. *The Journal of Biological Chemistry, 273*(30), 18802-18811.

Hertzberg, R. P., & Dervan, P. B. (1984). Cleavage of DNA with methidiumpropyl-EDTA-iron(II): Reaction conditions and product analyses. *Biochemistry, 23*(17), 3934-3945.

Huang, J., & Sousa, R. (2000). T7 RNA polymerase elongation complex structure and movement. *Journal of Molecular Biology, 303*(3), 347-358. doi:10.1006/jmbi.2000.4150 [doi]

Ikeda, R. A. (1992). The efficiency of promoter clearance distinguishes T7 class II and class III promoters. *The Journal of Biological Chemistry, 267*(16), 11322-11328. .

Ikeda, R. A., Lin, A. C., & Clarke, J. (1992). Initiation of transcription by T7 RNA polymerase as its natural promoters. *The Journal of Biological Chemistry, 267*(4), 2640-2649.

Ikeda, R. A., & Richardson, C. C. (1986). Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proceedings of the National Academy of Sciences of the United States of America, 83*(11), 3614-3618. .

Ikeda, R. A., & Richardson, C. C. (1987). Interactions of a proteolytically nicked RNA polymerase of bacteriophage T7 with its promoter. *The Journal of Biological Chemistry, 262*(8), 3800-3808.

Jeng, S. T., Gardner, J. F., & Gumport, R. I. (1990). Transcription termination by bacteriophage T7 RNA polymerase at rho-independent terminators. *The Journal of Biological Chemistry, 265*(7), 3823-3830.

Jeruzalmi, D., & Steitz, T. A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *The EMBO Journal, 17*(14), 4101-4113. doi:10.1093/emboj/17.14.4101 [doi]

Jiang, M., Rong, M., Martin, C., & McAllister, W. T. (2001). Interrupting the template strand of the T7 promoter facilitates translocation of the DNA during initiation, reducing transcript slippage and the release of abortive products. *Journal of Molecular Biology, 310*(3), 509-522. doi:10.1006/jmbi.2001.4793 [doi]

Jorgensen, E. D., Durbin, R. K., Risman, S. S., & McAllister, W. T. (1991). Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *The Journal of Biological Chemistry, 266*(1), 645-651.

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., & Ebright, R. H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science (New York, N.Y.), 314*(5802), 1144-1147. doi:314/5802/1144 [pii]

Kennedy, W. P., Momand, J. R., & Yin, Y. W. (2007). Mechanism for *de novo* RNA synthesis and initiating nucleotide specificity by t7 RNA polymerase. *Journal of Molecular Biology, 370*(2), 256-268. doi:S0022-2836(07)00396-8 [pii]

Kim, Y., Eom, S. H., Wang, J., Lee, D. S., Suh, S. W., & Steitz, T. A. (1995). Crystal structure of thermus aquaticus DNA polymerase. *Nature, 376*(6541), 612-616. doi:10.1038/376612a0 [doi]

Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., & Steitz, T. A. (1992). Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science (New York, N.Y.), 256*(5065), 1783-1790.

Komissarova, N., & Kashlev, M. (1997). RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *The Journal of Biological Chemistry, 272*(24), 15329-15338.

Kumar, A., & Patel, S. S. (1997). Inhibition of T7 RNA polymerase: Transcription initiation and transition from initiation to elongation are inhibited by T7 lysozyme via a ternary complex with RNA polymerase and promoter DNA. *Biochemistry, 36*(45), 13954-13962. doi:10.1021/bi971432y [doi]

Kuzmine, I., Gottlieb, P. A., & Martin, C. T. (2003). Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *The Journal of Biological Chemistry, 278*(5), 2819-2823. doi:10.1074/jbc.M208405200 [doi]

Ling, M. L., Risman, S. S., Klement, J. F., McGraw, N., & McAllister, W. T. (1989). Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. *Nucleic Acids Res, 17*(4), 1605-1618.

Liu, C., & Martin, C. T. (2002). Promoter clearance by T7 RNA polymerase. initial bubble collapse and transcript dissociation monitored by base analog fluorescence. *The Journal of Biological Chemistry, 277*(4), 2725-2731. doi:10.1074/jbc.M108856200 [doi]

Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J., & McAllister, W. T. (1997). Mutant bacteriophage T7 RNA polymerases with altered termination properties. *Journal of Molecular Biology, 269*(1), 28-40. doi:S0022-2836(97)91015-9 [pii]

Lyakhov, D. L., He, B., Zhang, X., Studier, F. W., Dunn, J. J., & McAllister, W. T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *Journal of Molecular Biology, 280*(2), 201-213. doi:S0022-2836(98)91854-X [pii]

Macdonald, L. E., Durbin, R. K., Dunn, J. J., & McAllister, W. T. (1994). Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *Journal of Molecular Biology, 238*(2), 145-158. doi:S0022-2836(84)71277-0 [pii]

Macdonald, L. E., Zhou, Y., & McAllister, W. T. (1993). Termination and slippage by bacteriophage T7 RNA polymerase. *Journal of Molecular Biology, 232*(4), 1030-1047. doi:S0022-2836(83)71458-0 [pii]

Martin, C. T., & Coleman, J. E. (1987). Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry, 26*(10), 2690-2696.

Martin, C. T., Muller, D. K., & Coleman, J. E. (1988). Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry, 27*(11), 3966-3974.

Maslak, M., & Martin, C. T. (1993). Kinetic analysis of T7 RNA polymerase transcription initiation from promoters containing single-stranded regions. *Biochemistry, 32*(16), 4281-4285.

McAllister, W. T. (1993). Structure and function of the bacteriophage T7 RNA polymerase (or, the virtues of simplicity). *Cellular & Molecular Biology Research, 39*(4), 385-391.

McAllister, W. T., & Carter, A. D. (1980). Regulation of promoter selection by the bacteriophage T7 RNA polymerase in vitro. *Nucleic Acids Research, 8*(20), 4821-4837.

Mead, D. A., Szczesna-Skorupa, E., & Kemper, B. (1986). Single-stranded DNA 'blue' T7 promoter plasmids: A versatile tandem promoter system for cloning and protein engineering. *Protein Engineering, 1*(1), 67-74.

Mentesana, P. E., Chin-Bow, S. T., Sousa, R., & McAllister, W. T. (2000). Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *Journal of Molecular Biology, 302*(5), 1049-1062. doi:S0022-2836(00)94114-7 [pii]

Meyer-Almes, F. J., Heumann, H., & Porschke, D. (1994). The structure of the RNA polymerase-promoter complex. DNA-bending-angle by quantitative electrooptics. *Journal of Molecular Biology, 236*(1), 1-6. doi:S0022-2836(84)71112-0 [pii]

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research, 15*(21), 8783-8798.

Mosrin-Huaman, C., Turnbough, C. L.,Jr, & Rahmouni, A. R. (2004). Translocation of escherichia coli RNA polymerase against a protein roadblock in vivo highlights a passive sliding mechanism for transcript elongation. *Molecular Microbiology, 51*(5), 1471-1481. doi:10.1111/j.1365-2958.2003.03926.x [doi]

Muller, D. K., Martin, C. T., & Coleman, J. E. (1988). Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry, 27*(15), 5763-5771.

Murakami, K. S., & Darst, S. A. (2003). *Bacterial RNA polymerases: The wholo story* doi:http://dx.doi.org/10.1016/S0959-440X(02)00005-2

Murakami, K. S., Trakselis, M. A., & SpringerLink. (2014). *Nucleic acid polymerases*. Berlin, Heidelberg: Springer Berlin Heidelberg : Imprint: Springer.

Nam, S. C., & Kang, C. W. (1988). Transcription initiation site selection and abortive initiation cycling of phage SP6 RNA polymerase. *The Journal of Biological Chemistry, 263*(34), 18123-18127.

Ramirez-Tapia, L. E., & Martin, C. T. (2012). New insights into the mechanism of initial transcription: The T7 RNA polymerase mutant P266L transitions to elongation at longer RNA lengths than wild type. *The Journal of Biological Chemistry, 287*(44), 37352-37361. doi:10.1074/jbc.M112.370643 [doi]

Raskin, C. A., Diaz, G., Joho, K., & McAllister, W. T. (1992). Substitution of a single bacteriophage T3 residue in bacteriophage T7 RNA polymerase at position 748 results in a switch in promoter specificity. *Journal of Molecular Biology, 228*(2), 506-515. doi:0022-2836(92)90838-B [pii]

Rees, W. A., Keller, R. W., Vesenka, J. P., Yang, G., & Bustamante, C. (1993). Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science (New York, N.Y.), 260*(5114), 1646-1649.

Rong, M., He, B., McAllister, W. T., & Durbin, R. K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America, 95*(2), 515-519.

Schinkel, A. H., Groot Koerkamp, M. J., Teunissen, A. W., & Tabak, H. F. (1988). RNA polymerase induces DNA bending at yeast mitochondrial promoters. *Nucleic Acids Research, 16*(19), 9147-9163.

Selisko, B., Dutartre, H., Guillemot, J. C., Debarnot, C., Benarroch, D., Khromykh, A., . . . Canard, B. (2006). Comparative mechanistic studies of de novo RNA synthesis by flavivirus RNA-dependent RNA polymerases. *Virology, 351*(1), 145-158. doi:S0042-6822(06)00197-8 [pii]

Smeekens, S. P., & Romano, L. J. (1986). Promoter and nonspecific DNA binding by the T7 RNA polymerase. *Nucleic Acids Research, 14*(6), 2811-2827.

Sousa, R. (2013). T7 RNA polymerase. In W. J. Lennarz, & M. D. Lane (Eds.), *Encyclopedia of biological chemistry* (pp. 355-359). Waltham: Academic Press. doi:https://doi.org/10.1016/B978-0-12-378630-2.00267-X

Sousa, R., Chung, Y. J., Rose, J. P., & Wang, B. C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 A resolution. *Nature, 364*(6438), 593-599. doi:10.1038/364593a0 [doi]

Sousa, R., Patra, D., & Lafer, E. M. (1992). Model for the mechanism of bacteriophage T7 RNAP transcription initiation and termination. *Journal of Molecular Biology, 224*(2), 319-334. doi:0022-2836(92)90997-X [pii]

Sousa, R., Rose, J. P., Chung, Y. J., Lafer, E. M., & Wang, B. C. (1989). Single crystals of bacteriophage T7 RNA polymerase. *Proteins, 5*(4), 266-270. doi:10.1002/prot.340050403 [doi]

Sousa, R. (2004). T7 RNA polymerase. In W. J. Lennarz, & M. D. Lane (Eds.), *Encyclopedia of biological chemistry* (pp. 147-151). New York: Elsevier. doi:https://doi.org/10.1016/B0-12-443710-9/00240-4

Sousa, R., Rose, J. P., Je Chung, Y., Lafer, E. M., & Wang, B. (1989). Single crystals of bacteriophage T7 RNA polymerase. *Proteins: Structure, Function, and Bioinformatics, 5*(4), 266-270. doi:10.1002/prot.340050403

Stahl, S. J., & Zinn, K. (1981). Nucleotide sequence of the cloned gene for bacteriophage T7 RNA polymerase. *Journal of Molecular Biology, 148*(4), 481-485. doi:0022-2836(81)90187-X [pii]

Stano, N. M., & Patel, S. S. (2002). The intercalating beta-hairpin of T7 RNA polymerase plays a role in promoter DNA melting and in stabilizing the melted DNA for efficient RNA synthesis. *Journal of Molecular Biology, 315*(5), 1009-1025. doi:10.1006/jmbi.2001.5313 [doi]

Steitz, T. A. (1997). DNA and RNA polymerases: Structural diversity and common mechanisms. *Harvey Lectures, 93*, 75-93.

Steitz, T. A. (1999). DNA polymerases: Structural diversity and common mechanisms. *The Journal of Biological Chemistry, 274*(25), 17395-17398.

Steitz, T. A. (2004). The structural basis of the transition from initiation to elongation phases of transcription, as well as translocation and strand separation, by T7 RNA polymerase. *Current Opinion in Structural Biology, 14*(1), 4-9. doi:10.1016/j.sbi.2004.01.006 [doi]]

Steitz, T., Smerdon, S., Jager, J., & Joyce, C. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science, 266*(5193), 2022-2025. doi:10.1126/science.7528445

Tabor, S., & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proceedings of the National Academy of Sciences of the United States of America, 82*(4), 1074-1078.

Tahirov, T. H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W. T., Vassylyev, D. G., & Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature, 420*(6911), 43-50. doi:10.1038/nature01129 [doi]

Tang, G. Q., Bandwar, R. P., & Patel, S. S. (2005). Extended upstream A-T sequence increases T7 promoter strength. *The Journal of Biological Chemistry, 280*(49), 40707-40713. doi:M508013200 [pii]

Tang, G. Q., & Patel, S. S. (2006). Rapid binding of T7 RNA polymerase is followed by simultaneous bending and opening of the promoter DNA. *Biochemistry, 45*(15), 4947-4956. doi:10.1021/bi052292s [doi]

Tang, G. Q., Roy, R., Ha, T., & Patel, S. S. (2008). Transcription initiation in a single-subunit RNA polymerase proceeds through DNA scrunching and rotation of the N-terminal subdomains. *Molecular Cell, 30*(5), 567-577. doi:10.1016/j.molcel.2008.04.003 [doi]

Theis, K., Gong, P., & Martin, C. T. (2004). Topological and conformational analysis of the initiation and elongation complex of t7 RNA polymerase suggests a new twist. *Biochemistry, 43*(40), 12709-12715. doi:10.1021/bi0486987 [doi]

Thomen, P., Lopez, P. J., & Heslot, F. (2005). Unravelling the mechanism of RNA-polymerase forward motion by using mechanical force. *Physical Review Letters, 94*(12), 128102. doi:10.1103/PhysRevLett.94.128102 [doi]

Transcription is catalyzed by RNA polymerase - biochemistry - NCBI bookshelf. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK22546/

Tunitskaya, V. L., & Kochetkov, S. N. (2002). Structural-functional analysis of bacteriophage T7 RNA polymerase. *Biochemistry.Biokhimiia, 67*(10), 1124-1135. doi:BCM67101360 [pii]

Turingan, R. S., Theis, K., & Martin, C. T. (2007). Twisted or shifted? fluorescence measurements of late intermediates in transcription initiation by T7 RNA polymerase. *Biochemistry, 46*(21), 6165-6168. doi:10.1021/bi700058b [doi]

Villemain, J., Guajardo, R., & Sousa, R. (1997). Role of open complex instability in kinetic promoter selection by bacteriophage T7 RNA polymerase. *Journal of Molecular Biology, 273*(5), 958-977. doi:S0022-2836(97)91358-9 [pii]

Woody, A. Y., Eaton, S. S., Osumi-Davis, P. A., & Woody, R. W. (1996). Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription. *Biochemistry, 35*(1), 144-152. doi:10.1021/bi952037f [doi]

Yin, Y. W., & Steitz, T. A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science (New York, N.Y.), 298*(5597), 1387-1395. doi:10.1126/science.1077464 [doi]

Yin, Y. W., & Steitz, T. A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell, 116*(3), 393-404. doi:S0092867404001205 [pii]

Zavriev, S. K., & Shemyakin, M. F. (1982). RNA polymerase-dependent mechanism for the stepwise T7 phage DNA transport from the virion into E. coli. *Nucleic Acids Research, 10*(5), 1635-1652.

Zhang, X., & Studier, F. W. (1997). Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *Journal of Molecular Biology, 269*(1), 10-27. doi:S0022-2836(97)91016-0 [pii]

CHAPTER 2

THE LENGTH OF PROMOTER SEQUENCE AFFECTS THE *DE NOVO* INITIATION OF T7 RNA POLYMERASE IN VITRO: NEW INSIGHTS INTO THE EVOLUTION OF PROMOTERS FOR SINGLE SUBUNIT RNA POLYMERASES.

Authors: Ramesh Padmanabhan and Dennis Miller

The Department of Biological Sciences, BSB12

The University of Texas at Dallas

800 West Campbell Road

Richardson, TX USA 75080

2.1 Abstract

RNA polymerases (RNAPs) differ from other polymerases in that they can bind promoter sequences and initiate transcription *de novo*. Promoter recognition requires the presence of specific DNA binding domains in the polymerase. The structure and mechanistic aspects of transcription by the bacteriophage T7 RNA polymerase (T7 RNAP) are well characterized. This single subunit RNAP belongs to the family of RNAPs which also includes the T3, SP6 and mitochondrial RNAPs. High specificity for its promoter, the requirement of no additional transcription factors, and high fidelity of initiation from a specific site in the promoter makes it the polymerase of choice to study the mechanistic aspects of transcription. The structure and function of the catalytic domains of this family of polymerases are highly conserved suggesting a common mechanism underlying transcription. Although the two groups of single subunit RNAPs, mitochondrial and bacteriophage, have remarkable structural conservation, they recognize quite dissimilar promoters. Specifically, the bacteriophage promoters recognize a 23 nucleotide promoter extending from -17 to + 6 nucleotides relative to the site of transcription initiation, while the well characterized promoter recognized by the yeast mitochondrial RNAP is nine nucleotides in length extending from -8 to +1 relative to the site of transcription initiation. Promoters recognized by the bacteriophage RNAPs are also well characterized with distinct functional domains involved in promoter recognition and transcription initiation. Thorough mutational studies have been conducted by altering individual base-pairs within these domains. Here we describe experiments to determine whether the prototype bacteriophage RNAP is able to recognize and initiate at truncated promoters similar to mitochondrial promoters. Using an *in vitro* oligonucleotide transcriptional system, we have assayed transcription initiation activity by

T7 RNAP. When a complete or almost complete (20 to 16 nucleotide) double stranded T7 RNAP promoter sequence is present, small RNAs are produced through template-independent and promoter-dependent stuttering corresponding to abortive initiation. When partial (10 to 12 nucleotide) double stranded promoter sequences are supplied, template dependent *de novo* initiation of RNA occurs, but at a site different from the canonical +1-initiation site. The site of transcription initiation is determined by a recessed 3' end based paired to the template strand of DNA rather than relative to the partial promoter sequence. Oligonucleotides with less than about 10 nucleotides of promoter do not initiate de novo transcription. However, if the oligonucleotides have the potential to form transient duplexes with a recessed 3' end, template-specific addition of a nucleotide at the 3' end of the oligonucleotides occurs. Understanding the mechanisms underlying these observations helps us to understand the role of the elements in the T7 promoter, and provides insights into the promoter evolution of the single-subunit RNAPs.

2.2 Introduction

The evolution of genetic information depends in part on the coevolution of polymerases that can synthesize the informational molecule and at the same time transfer the genetic information in the template through the formation of Watson-Crick base pairing. The earliest polymerases in an RNA World are thought to be ribozymes with RNA-directed RNA synthesis. During the transition from the RNA world to an RNA-protein world the ribozymes were replaced by single subunit protein polymerases with RNA-directed RNA synthesis [RNA replicase]. The transition from the RNA-Protein world through the RNA-DNA Retro world to the modern DNA world required a co-evolution of polymerase properties to give RNA-directed DNA synthesis [reverse transcriptases], and DNA-directed DNA polymerases [DNA polymerases, DNA Replicases],

respectively. One argument that genetic information has evolved through these different "Worlds" is the modern existence of these coevolved single subunit polymerases. All of these single subunit polymerases presumably initiated synthesis through a primer extension mechanism, since almost all modern single subunit polymerases retain this property.

Polymerases can be classified into four categories based on the type of nucleic acids synthesized and templates used: 1] DNA-directed DNA polymerases (DNAPs; DdDp), 2] DNA-directed RNA polymerases (RNAPs; DdRp), 3] RNA-directed DNA polymerases (reverse transcriptases or RTs), and 4] RNA-directed RNA polymerases (RNA replicase; RdRp) (Ahlquist, 2002; Castro et al., 2009; Chen & Romesberg, 2014). These polymerases can be further divided into single subunit RNA polymerases and multi-subunit RNA polymerases (Werner & Grohmann, 2011). The single subunit RNAPs have motifs in common and are thought to derive from a common ancestor but the evolutionary divergence of these polymerases is still obscure (Cermakian et al., 1997).

A major evolutionary step was the evolution of the promoter which defined genes with specific functions from the genetic information, along with the coevolution of domains in RNA polymerases which would recognize promoter sequences. This innovation results in *de novo* initiation of transcription and provides a way to specifically regulate gene expression. RNAPs are distinguished by their ability to recognize promoter sequences and initiate transcription *de novo*, rather than extend from a primer (Cheetham & Steitz, 2000; Kennedy, Momand, & Yin, 2007; Steitz, Smerdon, Jager, & Joyce, 1994). This initiation phase is transient and generally

occurs while the RNAP is bound to the promoter (Gong & Martin, 2006; Liu & Martin, 2002). The transition to the elongation phase requires a change in RNAP structure to allow promoter release and processive movement on the template (promoter clearance) (Gong, Esposito, & Martin, 2004; Martin, Muller, & Coleman, 1988). This initiation phase is unique to RNAPs and this ability to recognize promoters and initiate *de novo* is a key step in the evolution of organisms' ability to transfer specific genetic information from DNA.

The bacteriophage T7 RNA Polymerase is the prototype of the single subunit RNA polymerases. It is an ideal model system for studying polymerase and promoter evolution. It is related to other bacteriophage RNA polymerases and the mitochondrial RNA polymerases. This group of RNAPs has very conserved structure and sequence. However, the mitochondrial RNA polymerases generally have different promoter sequences from the bacteriophage polymerases (Figure 2.1). Bacteriophage polymerases generally have a 23 nucleotide promoter that overlaps the site of transcription initiation by six nucleotides (-17 to + 6). Mitochondrial RNAPs recognize a diverse variety pf promoter sequences which are typically about nine nucleotides in length and run from -8 to +1. In figure 2.1 the well characterized mitochondrial promoter from the yeast mtRNAP is used as a reference for comparison with the bacteriophage RNAPs. The yeast promoter sequence has similarity with the -8 to +1 portion of the T7 promoter.

The 23 nucleotide promoter of the T7 RNA polymerase can be divided into several functional domains (Figure 2.2). The initiation region covers the ten nucleotides from - 4 to +6. Nucleotides +1 to +6 are called the transcription start site. The +1 site is conserved as a G in all

the bacteriophage promoters. Sites +1 to + 6 are conserved as purines in the bacteriophage promoters. Polymerase contacts in this region are primarily with the pyrimidine in the template strand (Weston, Kuzmine, & Martin, 1997). Substitution of these nucleotides decreases promoter strength and initiation efficiency only slightly.

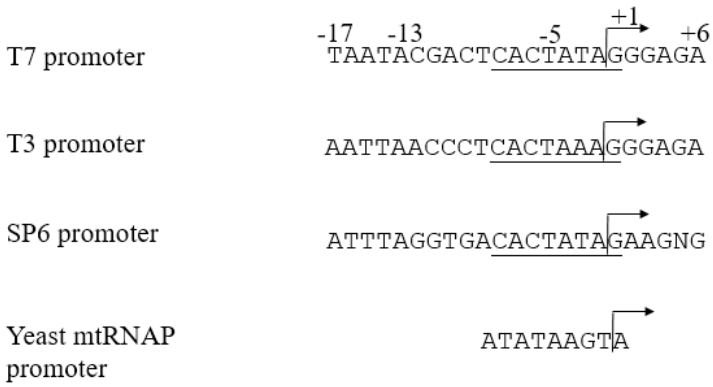


Figure 2.1. Comparison of three bacteriophage RNAP promoters and yeast mtRNAP (YmtRNAP) promoter. All sequences shown in the 5'-3' orientation. The -8 to +1 portion of the bacteriophage promoter is underlined. The nine-nucleotide yeast mtRNAP promoter sequence is aligned with the bacteriophage -8 to +1 sequence to highlight their analogous positions relative to transcription initiation.

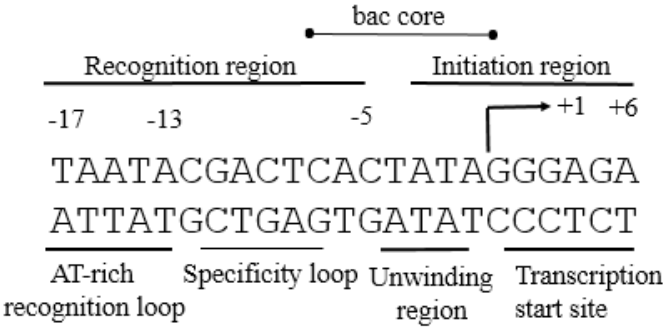


Figure 2.2. The T7 RNAP promoter showing functional domains

The region -4 to -1 is called the unwinding region. Positions -1 [A], -3 [A], -4 [T] are conserved among the bacteriophage promoters. This region is invariably AT-rich presumably to aid in melting of the DNA strands at the initiation site, although the conservation of positions -1, -3 and -4 may indicate polymerase contact sites.

The promoter recognition region (13 nucleotides) extends from -5 to -17 and consists of polymerase interaction sites important for polymerase binding and positioning. These promoter contact sites interact with two regions of the polymerase. Promoter positions -5 to -12 interact with a T7 RNAP structural domain called the specificity loop, located near the carboxyl terminus (Temiakov et al., 2000), while positions -13 to -17 interact with a T7 RNAP structural domain called the AT-rich recognition loop located near the amino terminus of the T7 RNA polymerase (Imburgio, Rong, Ma, & McAllister, 2000).

The phylogenetic tree depicting the evolutionary relationships among the single subunit RNAPs has not been rooted and so the evolutionary sequence of the two classes of promoter is unknown (Cermakian et al., 1997). The T7 RNAP can recognize a range of promoter sequences which are closely related to each other by this consensus sequence (Dunn & Studier, 1983; Tang, Bandwar, & Patel, 2005). Details concerning the T7 RNAP promoter structure and function were deduced from studies on mutant promoter sequences (Ikeda, Ligman, & Warshamana, 1992; Klement et al., 1990). Single base changes either in the recognition region or the initiation domain affects the efficiency of promoter recognition and initiation of transcription respectively, but a mutation in the recognition region does not affect initiation and vice versa (Ikeda, Ligman, et al., 1992).

The promoter region from -17 to -13 is dispensable, and can be deleted with only small effects on initiation activity, however, optimal promoter recognition requires contacts within this region (Chapman & Wells, 1982; Martin & Coleman, 1987; Osterman & Coleman, 1981). Comparison of the promoter sequences from the three phage promoters T3, T7 and SP6 revealed that all the three promoters share a similar core sequence from -7 to +1 pointing to a common role of this region in the promoter function (Brown, Klement, & McAllister, 1986; Jorgensen, Durbin, Risman, & McAllister, 1991), Figure 2.1). There is considerable sequence divergence from -8 to -12 corresponding to the specificity loop-recognition region in the T7 RNAP promoter, suggesting that this region plays a key role in sequence-specific contacts. Although there is an 82% homology in the amino acid sequence between T3 and T7 RNAPs, neither of the enzymes can efficiently initiate transcription at promoters used by the other (Klement et al., 1990; Rong, He, McAllister, & Durbin, 1998). Promoter specificity studies using base-pair substitutions at -10, -11, and -12 positions in the T7 RNAP promoter by residues of T3 RNAP promoter revealed that base pairs -10 to -12 have an important role with -11 base having a significant role in promoter binding (Klement et al., 1990). Base substitutions at positions -7 (A or G for C), -8 (A for T), -9 (A or T for C, -11 (T for G) completely inactivated the T7 promoter (Chapman, Gunderson, Anello, Wells, & Burgess, 1988). Methylation interference studies show that binding of T7 RNAP to the promoter was interfered by methylation of the G-residues at -7 and -9 in the template strand and -11 in the non-template strand suggesting that T7 RNAP makes important contacts in the major groove between -7 and -11 (Jorgensen et al., 1991).

The yeast *Saccharomyces cerevisiae* mitochondrial RNA polymerase (YmtRNAP) is homologous to the single subunit bacteriophage T3 and T7 RNAPs (Cermakian, Ikeda, Cedergren, & Gray, 1996; Masters, Stohl, & Clayton, 1987; Matsunaga & Jaehning, 2004). YmtRNAP recognizes the simple nine-nucleotide-long promoter consensus sequence ATATAAGTA for transcription initiation, which differs in sequence and length from the phage RNAPs (Nayak, Guo, & Sousa, 2009). However, this sequence can be aligned with the -8 to +1 core region of phage RNAPs (Figure 2.1)**.** The YmtRNAP has a region analogous to the specificity loop of T7 RNAP which by analogy would interact with -8 to -5 [ATAT]. However, there is no analogous region in yeast mtRNAP to the A-T rich binding region in T7RNAP and so, it is not surprising that the yeast promoter consensus does not extend beyond -8.

To study T7 RNAP's ability to use truncated promoters similar to mitochondrial promoters, we have developed an *in vitro* transcription system based on the *in vitro* transcription system developed by Milligan, Groebe, Witherell, & Uhlenbeck (1987). Classically, two complementary oligonucleotides of equal length containing the T7 consensus sequence (Figure 2.3A and 2.3B) could be used to produce run off transcripts *in vitro*. Milligan *et al*. (1987) have shown that the 18 nucleotides, -17 to +1, when double stranded and attached to a 5' extended template is sufficient to act as a promoter *in vitro* (Figure 2.3C). We have modified their procedure by extending the double stranded promoter region to 20 nucleotides (-17 to +3) in order to increase initiation frequency. We further modified this model by creating an oligonucleotide which formed a intramolecular double stranded region of the type shown in Figure 2.3D. Under the same conditions it produced an RNA of the same length and composition.

We have used this oligonucleotide system to determine whether T7 RNAP in the presence of such an oligonucleotide and a single ribonucleotide triphosphate can correctly and efficiently initiate transcription on truncated promoter sequences resembling mitochondrial promoters. Here we show that T7 RNAP can initiate on truncated promoter sequences similar to mitochondrial promoters. However, the site of transcription initiation is not at the canonical initiation site, but initiates instead on the first unpaired base in the template. In addition, we show that with complete promoters, or mostly complete promoters, and a recessed 3' end, nontemplated, *de novo* initiated transcripts are produced. In the absence of a promoter and with transiently stable hairpin duplexes, the oligonucleotide itself can be labeled at the recessed 3' end. These results are analyzed in the light of T7 promoter flexibility, promoter evolution, and the use of this technique to produce defined RNAs without 5'- sequence constraints and 3'- variability.

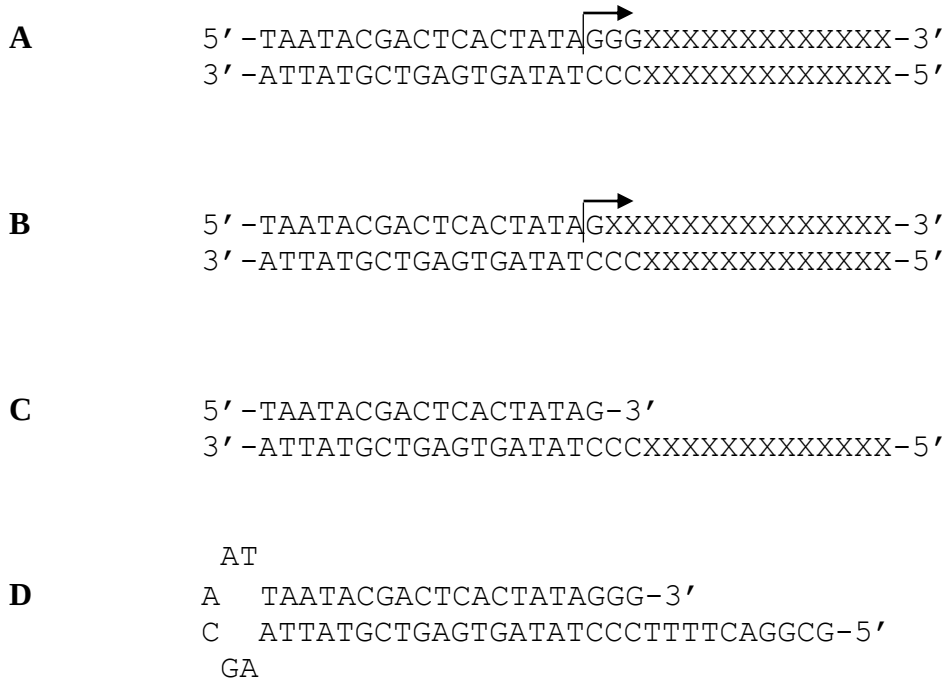


Figure 2.3. Oligonucleotides used for run-off transcription with T7RNAP *in vitro*.

Panel A shows two complementary oligonucleotides of equal length annealed to create a 20 bp double stranded promoter producing a run off transcript with three G nucleotides at the 5' end. Panel B shows two complementary oligonucleotides of equal length annealed to create an 18 bp double stranded promoter producing a run off transcript with one G nucleotide at the 5' end, but with slightly decreased initiation frequency. Panel C shows two complementary oligonucleotides of different length annealed to produce a double stranded promoter with a recessed 3' end of the type used by Milligan, Groebe, Witherell, & Uhlenbeck(1987).Panel D shows an oligonucleotide of the type used in the experiments reported in this paper i.e. a single oligonucleotide internally annealed to produce a hairpin loop with a 20 nucleotide double stranded promoter and a recessed 3' end.

2.3 Materials and methods

2.3.1 Oligonucleotides, Radiolabeled nucleotide triphosphate and T7 RNAP

Oligonucleotides designed for this study were procured from Eurofins MWG operon, USA. The desiccated oligonucleotides were resuspended in water to a concentration of 100µM. The radiolabeled α^{32}P ATP was procured from Molecular Bioproducts, USA. T7 RNA polymerase enzyme (50,000 U/mL) was procured from New England Biolabs, USA.

2.3.2 Reaction conditions

50µL reaction mixtures containing 5µl of the 100 µM oligonucleotides, 2µL of radiolabeled α^{32}P- ATP (3000Ci/mmol), 5µL of 10X RNA polymerase reaction buffer (1X concentrations of 40mM Tris-HCl, 6mM $MgCl_2$, 2mM spermidine, 1mM dithiothreitol), supplied along with T7 RNA polymerase, 1µL of RNAase inhibitor-RNasin (Promega, USA) and 1µL (50U) of T7 RNA polymerase were incubated at 37^{o}C for 60 minutes. Reactions were stopped and run on 15% Polyacrylamide gel at 75V. The gels were then stained with ethidium bromide to visualize the nucleic acids, and subsequently exposed to a phosphor screen (Amersham GE Healthcare) for 5 minutes and scanned using Storm 840 (Amersham GE Healthcare, USA).

2.3.3 Hairpin Oligonucleotide Nomenclature

Hairpin oligonucleotides are designated by a three number designation. The first number is the total length of the oligonucleotide in nucleotides, the second number is the length of the duplex portion of the hairpin in base pairs, and the third number is the length of the retained promoter

consensus sequence. For some oligonucleotides an additional designation indicating the number and the type of complementary sequences in the 5' extended template is added.

2.4 Results

The 5' ends of the 20 nucleotide (-17 to +3) double stranded promoter region on hairpin oligonucleotides were systematically deleted to determine if they would (1) label at the 3' end of the oligonucleotide; (2) initiate promoter-dependent, template-independent *de novo* RNA synthesis; (3) initiate promoter-dependent, template-dependent *de novo* transcription; or (4) fail to incorporate label under the conditions of having the T7 RNAP and a single radiolabeled ribonucleotide triphosphate. The starting oligonucleotide for the deletion mapping was the oligonucleotide shown in Figure 2.3D. It is able to form intra- and intermolecular base pairing to produce a recessed 3'end on an extended 5' template. Next to the 3' end on the unpaired template are four T nucleotides that will base pair with the radiolabeled α-^{32}P-rATP used as the only nucleotide triphosphate in the experiment. As the double stranded promoter region is removed, the loop length is, in some cases, increased to maintain the overall length. Table 2.1 shows the results of these experiments and Figure 2.4 shows examples of the four different results.

The oligonucleotide shown in Figure 2.3C (56-21-20) has a full 20 base pair promoter sequence (-17 to +3). With all four ribonucleotide triphosphates present T7 RNAP polymerase would be expected to initiate at +1 and make a 13 nucleotide RNA run-off transcript starting with GGG. With only rATP present T7 RNAP reproducibly makes RNAs that separate upon gel electrophoresis as two high mobility, labeled spots (Figure 2.4, lane 5) designated DS for

"double spot" in Table 2.1, line 1. This double spot product is RNAase sensitive (not shown) and will be called the "double spot" RNA product below.

Table 2.1. Results of the sequential deletion of the 5' end of the double stranded promoter region of hairpin loop oligonucleotides

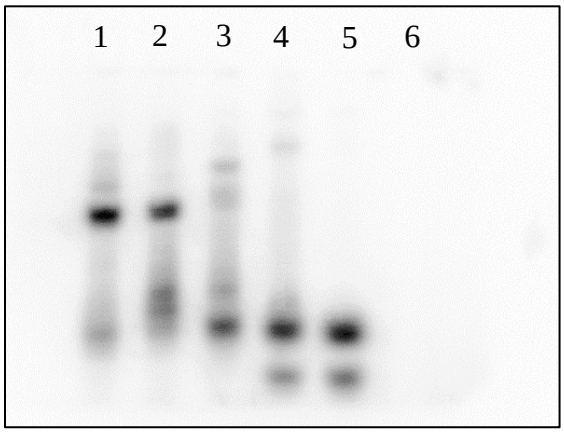
T7 promoter sequence	Promoter length	Duplex length	Type of labeling	Representative oligonucleotides
TAATACGACTCACTATAGGG	20	21	DS	56-21-20
……...ACGACTCACTATAGGG	16	16/20	DS	56-16-16, 56-20-16
……..CGACTCACTATAGGG	15	20	DS	56-20-15
……GACTCACTATAGGG	14	20	DS	56-20-14
……..CGACTCACTATAGGG	15	15	SS	56-15-15
……GACTCACTATAGGG	14	14	SS	56-14-14
……ACTCACTATAGGG	13	13	SS	56-13-13
…… CTCACTATAGGG	12	12	SS	56-12-12
……. TCACTATAGGG	11	12	SS	56-12-11
…… CACTATAGGG	10	12	SS	56-12-10
……...CTATAGGG	8	10	SS	50-10-8
……...CTATAGGG	8	8/9	OEL	48-9-8
……ATAGGG	6	6/7/8	OEL	46-8-6,44-7-6,42-6-6
…… TAGGG	5	5	OEL	40-5-5
……. AGGG	4	4	OEL	38-4-4
……GGG	3	3	OEL	36-3-3
……GG	2	2	NEG	34-2-2
……. G	1	1	NEG	36-3-1
……...	0	0	NEG	30-0-0

DS = double spot RNA, SS = single spot RNA, OEL = oligonucleotide labeling, NEG = no RNA

Oligonucleotide 66-31-20 is identical to 56-21-20 except that the recessed 3' end is extended on the template strand to produce a 31 bp double stranded region with a full promoter present and a blunt end terminus. This oligonucleotide fails to produce a "double spot" RNA product under

the same conditions as with 56-21-20 (Figure 2.4, lane 6) indicating, that under the conditions used in our assay, a recessed 3' end is required to produce the "double spot" RNA.

A



B

1. GCGGAC**TTTT**CCCTATCAGACAACGCAAGACCAGCAGATAGGG **44-7-6**

2. GCGGAC**TTTT**CCCTAGACAACGCAAACCAGCAGAAGGG **38-4-4**

3. GCGGAC**TTTT**CCCTATAGGCAGACAACGCAAACCAGCCTATAGGG **50-10-8**

4. GCGGAC**TTTT**CCCTATAGTGAGTCGTAACCAGCAATACCAACGACTCACTATAGGG **56-16-16**

5. GCGGAC**TTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20**

6. GCGGAC**TTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGGAAAAGTCCGC **66-31-20**

Figure 2.4. Panel A: Autoradiograph of 15% polyacrylamide gel used to separate the RNA labeled products produced from incubation of the various oligonucleotides with T7 RNA polymerase and radiolabeled rATP. The oligonucleotides used, listed from left to right (1 to 6) are 44-7-6, 38-4-4, 50-10-8, 56-16-16, 56-21-20, 66-31-20. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate promoter or partial promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

Oligonucleotide 56-16-16 is identical to 56-21-20 except that four nucleotides are removed from the 5' end of the double stranded promoter region to remove the AT-rich region. Eight

nucleotides are added to the single stranded loop region to maintain the size of the oligonucleotide at 56 nucleotides. The -13 to +3 region of the promoter is retained (16 nucleotide promoter). Oligonucleotide 56-16-16 (Table 1, line 2) also produces a "double spot" RNA (Figure 2.4, lane 4).

Oligonucleotides with 7, 8, 9, 10, 11, or 12 base pairs of double stranded promoter region deleted, to give 13 to 8 bp truncated promoters (from -12 to +3 to -7 to +3), reproducibly produce a high mobility, labeled spot designated SC (single spot) in Table 1 and fails to produce the "double spot" RNA product, indicating that a 14 bp promoter or larger is necessary for the "double spot" RNA and that the failure to produce the "double spot" RNA may allow the production of the "single spot" RNA product. These oligonucleotides lack all or part of the specificity loop region, and therefore, most of the recognition region of the promoter. Oligonucleotide 50-10-8 (-7 to +3) produces a single spot RNA product (Figure 2.4, lane 3). The "single spot" product is also sensitive to RNAase (not shown) and will be designated the "single spot" RNA product below.

Oligonucleotides with 11, 12, 13, 14, 15, 16 and 17 base pairs of 5' double stranded promoter region deleted, to give 3 to 9 bp truncated promoters (from -6 to +3 to +1 to +3), reproducibly add a labeled nucleotide to the 3' end of the oligonucleotide, designated OEL (oligonucleotide end labeling) in Table 2.1, and fail to produce "single spot" RNA product indicating that end labeling may prevent, or compete with, the production of "single spot" RNA. These oligonucleotides have a double stranded region with a recessed 3' end equal to the truncated

promoter length. Oligonucleotides 44-7-6 (-3 to +3) and 38-4-4 (+1 to +3) label the 3' end of the oligonucleotide (Figure 2.4, lanes 1 and 2).

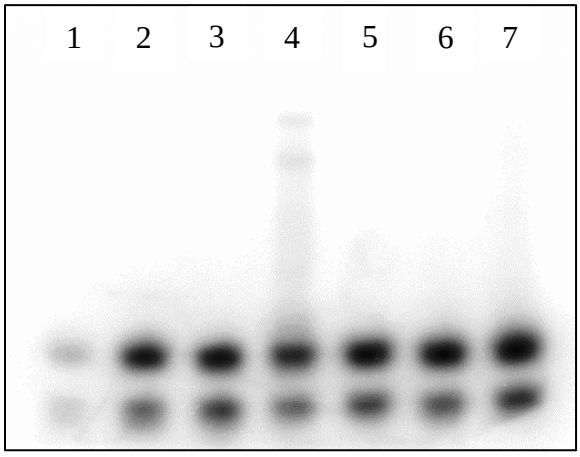
Oligonucleotides with 18, 19, and 20 base pairs of 5' double stranded promoter region deleted, to give 0 to 2 bp truncated promoters (+2 and +3, just +3, none), do not produce an RNA product or label the 3' end of the oligonucleotide (data not shown).

2.4.1 Characterization of the "double spot" RNA product.

An RNA product remarkably similar to the "double spot" has been observed by Ling et al. (1989) using T7 RNAP with a full double strand promoter with four ribonucleotide triphosphates present (Ling, Risman, Klement, McGraw, & McAllister, 1989). They ascribed this RNA product to abortive initiation. To determine if the "double spot" is abortive initiation, we examined the dependence of "double spot" on both template and promoter.

In order to test the effect of the template portion of the nucleotide on "double spot" RNA production, the number of T nucleotides in the template portion of 56-21-20 were varied from 0 to 8 or substituted with A nucleotides. All of these modified oligonucleotides were able to produce the double spot RNA product (Figure 2.5), indicating that double spot RNA synthesis is template independent.

A



B

1. GCGGAC**AAAA**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 T0A4**

2. GCGG**TTTTAA**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 A2T4**

3. GCGGACAA**TT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 T2A2**

4. GCGGAC**TTTT**CCCTATAGTGAGTCGTAACCAGCAATAACCACGACTCACTATAGGG **56-16-16 T4**

5. GCGGAC**TTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 T4**

6. GCGG**TTTTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 T6**

7. GC**TTTTTTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20 T8**

Figure 2.5. Panel A: Autoradiograph of 15% polyacrylamide gel showing double spot RNA production after incubation with T7 RNA polymerase and radiolabeled rATP of the 56-21-20 and 56-16-16 oligonucleotides modified in their template region. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

To determine if promoter sequences are necessary for the production of "double spot" RNA product, scrambled promoter sequences were substituted for the promoter sequence. While the classic "double spot" nucleotides 56-21-20, and 56-16-16 produced "double spot" RNA product,

the oligonucleotides with scrambled promoter sequences did not produce “double spot” RNA product (Figure 2.6), indicating that double spot production is dependent on promoter sequence. These results indicating that “double spot” RNA production is template independent and promoter dependent are consistent with the report that “double spot” RNA is abortive initiation.

A

B

1. GCGGAC**TTTT**CCCTATAGTGAGTCGTAACCAGCAATAACCACGACTCACTATAGGG **56-16-16**

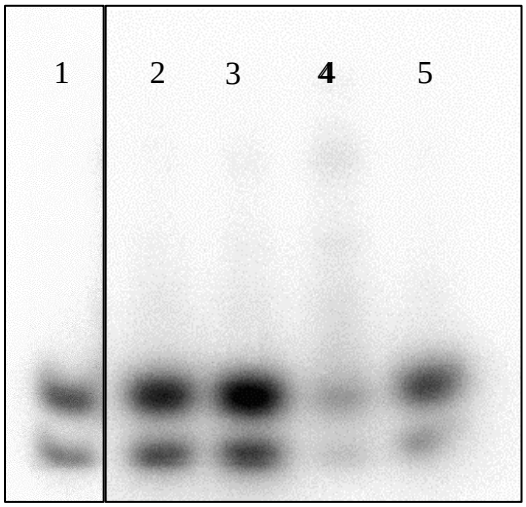
2. GCGGAC**TTTT**CGTACTAGATCTGACGTACGCAAGAACCTACGTCAGATCTAGTACG **56-18 Scr**

3. GCGGAC**TTTT**CGTACTAGATCTGACGTATAAGCAATTATACGTCAGATCTAGTACG **56-21 Scr**

4. GCGGAC**TTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20**

Figure 2.6. Panel A: Autoradiograph of 15% polyacrylamide gel showing double spot RNA production after incubation with T7 RNA polymerase and radiolabeled rATP with the 56-21-20 and 56-16-16 oligonucleotides modified to have scrambled promoter sequences. Panel B: The DNA sequences are written 5’ to 3’ for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

A



B

1. GCGGAC**TTTT**CCCTATAGTGAGTCGTATTAAGCAATTAATACGACTCACTATAGGG **56-21-20**

2. GCGGAC**TTTT**CCCTATAGTGAGTCTATATATGCAATTATATAGACTCACTATAGGG **56-20-14**

3. GCGGAC**TTTT**CCCTATAGTGAGTCGATATATGCAATTATATCGACTCACTATAGGG **56-20-15**

4. GCGGAC**TTTT**CCCTATAGTGAGTCGCCGCGTGCAATCGCGGCGACTCACTATAGGG **56-20-15**

5. GCGGAC**TTTT**CCCTATAGTGAGTCGTTACCTGCAATGGTAACGACTCACTATAGGG **56-20-16**

Figure 2.7. Panel A: Autoradiograph of 15% polyacrylamide gel showing double spot RNA production after incubation with T7 RNA polymerase and radiolabeled rATP with the 56-21-20, 56-20-16, 56-20-15 oligonucleotides. The 56-20-16 and 56-20-15 oligonucleotides have substituted bases in the AT-rich binding region. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

5' end deletions of 4, 5, and 6 nucleotides (oligonucleotides 56-20-16, 56-20-15, and 56-20-14) can produce "double spot" RNA (Figure 2.7). Oligonucleotide 56-20-15 has the AT-rich recognition sequence 5'-TAATA-3' substituted with 5'-TATAT-3', a similar sequence. This substitution has little effect on "double spot" RNA production (Figure 2.7, lane 3). To determine

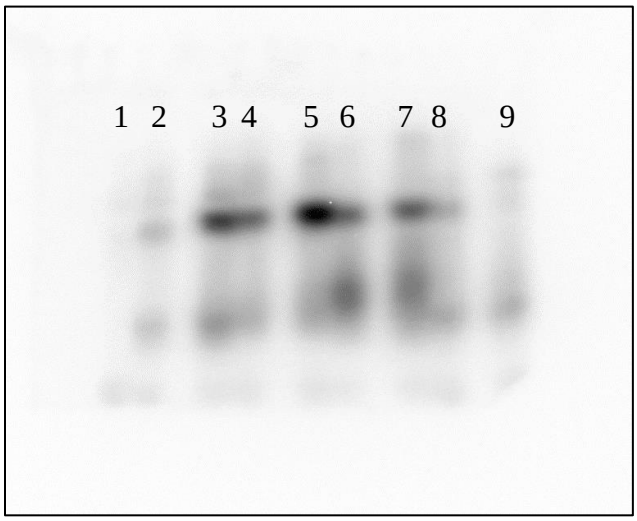
if a GC–rich sequence can be tolerated as a substitute for the AT-rich recognition sequence, the sequence CGCGG was substituted in 56-20-15. This oligonucleotide produces only a very small amount of double spot" DNA (Figure 2.7, lane 4), indicating that while the primary sequence of the AT-rich region is not critical, the base composition of the region should be AT-rich to give optimum double spot RNA production.

Our interpretation of these results is that "double spot" RNA production is equivalent to abortive *de novo* initiation at the recessed 3' end. A complete 20 nucleotide promoter (56-21-20; -17 to +13) or nearly complete 5' truncated promoters (56-16-16, 56-20-16, 56-20-15, and 56-20-14; -13, -12, -11 to +13) bind strongly enough so that it cannot leave the promoter during promoter clearance during the transition from initiation to elongation. The presence of only a single ribonucleotide triphosphate also prevents elongation. The transcripts produced prior to promoter clearance produce the abortive initiation products that we designate "double spot" RNA.

2.4.2 Oligonucleotide 3' end labeling

The labeling of oligonucleotides with recessed 3' ends by T7 RNA polymerase has been reported and characterized by Sarcar & Miller (manuscript submitted) as a type of DNA editing. They report a promoter-independent, template dependent, recessed 3' end-dependent addition of rNTPs to the 3' end of the oligonucleotide hairpins. A ribonucleotide is added to the 3' end when the 3' end is positioned next to a single stranded region where the first unpaired nucleotide is complementary to the labeling ribonucleotide triphosphate. While this process absolutely requires a double stranded region with a recessed 3' end, it is limited by the length of the base paired region.

A



B

1. GCGGAC**TTTT**CCAGACAACGCAAACCAGCAGAGG **34-2-2**

2. GCGGAC**TTTT**CCCAGACAACGCAAACCAGCAGAGGG **36-3-3**

3. GCGGAC**TTTT**CCCTAGACAACGCAAACCAGCAGAAGGG **38-4-4**

4. GCGGAC**TTTT**CCCTAAGACAACGCAAACCAGCAGATAGGG **40-5-5**

5. GCGGAC**TTTT**CCCTATAGACAACGCAAACCAGCAGAATAGGG **42-6-6**

6. GCGGAC**TTTT**CCCTATCACAACGCAAACCAGCAGAGATAGGG **44-7-6**

7. GCGGAC**TTTT**CCCTATAGAGACAACGCAAACCAGCAGACTATAGGG **46-8-8**

8. GCGGAC**TTTT**CCCTATAGGAGACAACGCAAACCAGCAGACCTATAGGG **48-9-8**

9. GCGGAC**TTTT**CCCTATAGGCAGACAACGCAAACCAGCAGAGCCTATAGGG **50-10-8**

Figure 2.8. Panel A: Autoradiograph of 15% polyacrylamide gel showing oligonucleotide 3' end labeling after incubation with T7 RNA polymerase and radiolabeled rATP. Each hairpin oligonucleotide has a duplex length one bp longer than the previous oligonucleotide. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

Oligonucleotides with duplex regions shorter that 3 base pair or longer than about 8 base pairs, depending on their base composition, do not label. In our initial detection of oligonucleotide 3' end labeling, we observed labeling of oligonucleotide with duplex potentials of 3 - 9 base pairs (Table 2.1).

However, in this initial screen the partial promoter length was equal to, or approximately equal to the duplex length of the hairpin oligonucleotide (Table 2.1). To uncouple these two parameters, we made oligonucleotides with a constant 20 bp duplex length, but with variable promoter lengths (see for example, 56-20-16, 56-20-15, 56-20-14, Figure 2.7). These showed "double spot" RNA production but did not show end labeling since the 20 nucleotide duplex length exceeded the 3-9 nucleotide length necessary for end labeling. Conversely, oligonucleotides without promoter sequences, but with hairpin oligonucleotides with duplex lengths between 3 and 9 bp and recessed 3'ends with complementary sequences, labeled efficiently.

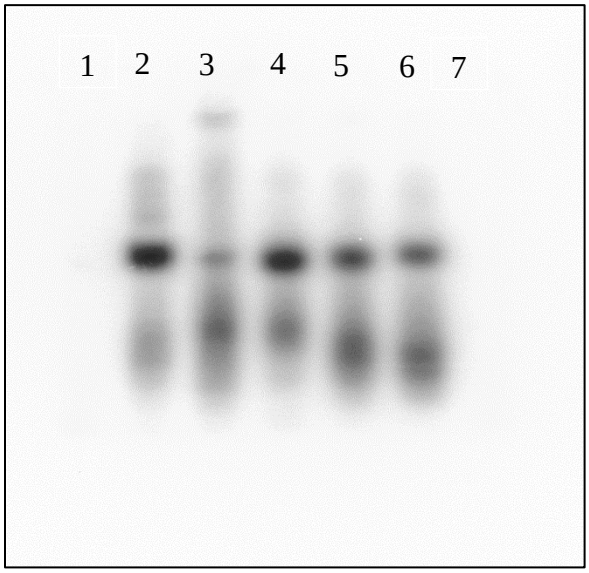
Figure 2.8 shows oligonucleotides with constant loop length (15 nucleotides) and template length (10 nucleotides) and increasing duplex length from 2 bp to 10 bp. Oligonucleotides with duplexes 4 to 8 bp label efficiently; oligonucleotides with duplexes of 3 and 9 bp label slightly. Oligonucleotides with smaller or larger duplex length (2 bp and 10 bp) did not 3' end label (Figure 2.8).

To determine template dependence in oligonucleotide 3' end labeling, the 5' extension/template of the 44-7-6 oligonucleotide was systematically altered. Substitution of A nucleotides for T nucleotides in the labeling site resulted in no labeling (Figure 2.9, lanes 1 and 7), indicating that the labeling site next to the 3' end must base pair with the radionucleotide triphosphate used in the assay (rATP in this case). Lanes 2, 3, 4, 5, and 6 of Figure 2.9 have oligonucleotides with a variable number of T nucleotides in the labeling site giving the potential for addition of multiple A nucleotides to the 3' end of the oligonucleotide. Although at least one A nucleotide is added in each case since labeling occurs, it is unlikely that multiple A nucleotides are added. This conclusion is based on the fact that the intensity of labeling is essentially equal, except for lane 3 where labeling intensity is decreased not increased for an eight T nucleotide template, and the migration of the labeled oligonucleotide is not changed. Addition of multiple nucleotides to the 3' end of the oligonucleotide would be expected to change the migration of the oligonucleotide in gel electrophoresis, but this is not observed. Interestingly, there seems to be concurrent end labeling and *de novo* initiation in the autoradiograph of Figure 2.9, and the *de novo* initiation does seem to elongate up to 8 nucleotides (see the similar experiment in Figure 2.13 below).

2.4.3 Charaterization of the "single spot" RNA product.

Under the conditions used in these experiments, it was not expected that the oligonucleotides would be capable of *de novo* intiaition, since promotor dependent *de novo* intiation requires three G nucleotides for intiation and rGTP is not provided in these experiments. Nonetheless, a small RNA, presumably 5'-pApApApA-3', was produced when four T nucleotides are positioned next to the recessed 3' end in hairpin oligonucleotides with 5' truncated promoter sequences between 10 and 15 nucleotides.

A



B

1. GCGG**TTTTAA**CCCTATCAGACAACGCAAACCAGCAGAGATAGGG **44-7-6A2T4**

2. GCGGAC**TTTT**CCCTATAAGACAACGCAAACCAGCAGATATAGGG **44-7-6 A**

3. GC**TTTTTTTT**CCCTATCAGACAACGCAAACCAGCAGAGATAGGG **44-7-6 T8**

4. GCGG**TTTTTT**CCCTATCAGACAACGCAAACCAGCAGAGATAGGG **44-7-6 T6**

5. GCGGAC**TTTT**CCCATACAGACAACGCAAACCAGCAGAGATAGGG **44-7-6 T4.**

6. GCGGACAA**TT**CCCTATCAGACAACGCAAACCAGCAGAGATAGGG **44-7-6 T2**

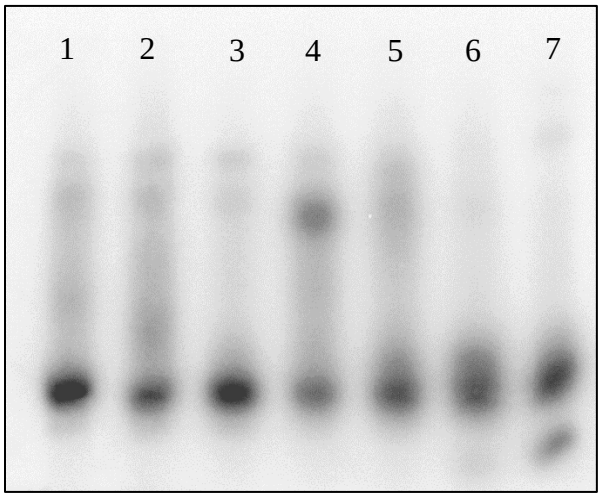
7. GCGGAC**AAAA**CCCTATCAGACAACGCAAACCAGCAGAGATAGGG **44-7-6 T0A4**

Figure 2.9. Panel A: Autoradiograph of 15% polyacrylamide gel showing oligonucleotide 3' end labeling after incubation with T7 RNA polymerase and radiolabeled rATP. Each hairpin oligonucleotide has a constant length (44 nucleotides), constant duplex length (7 bp), and constant loop length (15 nucleotides). The template 5' extension is a constant length (10 nucleotides) but has a variable number of T nucleotides at the labeling site or A nucleotide substitutions at the site. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

In Figure 2.10 six oligonucleotides with promoter lengths between 10 and 15 are able to produce "single spot" RNAs, but scrambling of promoter sequences completely eliminates "single spot" RNA production (Figure 2.11), indicating that this activity is dependent on a partial promoter sequence. While the duplex region of the hairpin oligonucleotides requires a partial promoter sequence, variable loop lengths do not affect the single spot production (Figure 2.12).

However, the single spot production is dependent on the 5' extension/template region of the oligonucleotide (Figure 2.13). Oligonucleotide 56-12-10 A4 (Figure 2.13, lane 1) has A substitutions for the template T nucleotides and does not produce RNA. This indicates "single spot" RNA production is template dependent and requires a complementary base in the site next to the 3' end. The remaining oligonucleotides in Figure 2.13 have 2, 4, 5, 6, and 8 T nucleotides in the template region next to the 3' end. The RNAs made from these oligonucleotides correspond in length to the template, indicating that they can initiate template-dependent *de novo* RNA synthesis from a recessed 3' end when a partial promoter is present and elongate the RNA on the template. This *de novo* initiation site is located at +4 in contrast to the classical site of transcription initation at +1.

A



B

1. GCGGAC**TTTT**CCCTATAGTGACAGACAACGCAAGAACCAGCAGAGTCACTATAGGG **56-12-11**

2. GCGGAC**TTTT**CCCTATAGTGAGAGACAACGCAAGAACCAGCAGACTCACTATAGGG **56-12-12**

3. GCGGAC**TTTT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10**

4. GCGGAC**TTTT**CCCTATAGTGAGTACGAACGCAAGAACCAGTACACTCACTATAGGG **56-13-13**

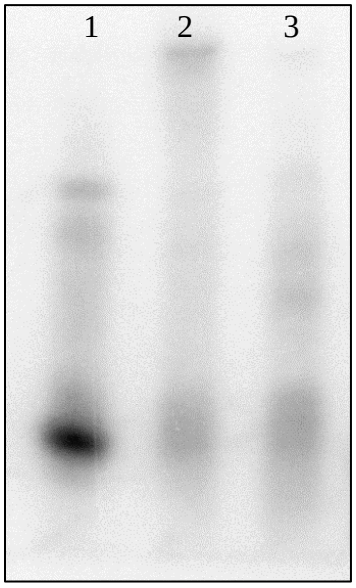
5. GCGGAC**TTTT**CCCTATAGTGAGTCCGAACGCAAGAACCAGTAGACTCACTATAGGG **56-14-14**

6. GCGGAC**TTTT**CCCTATAGTGAGTCGGAACGCAAGAACCAGTCGACTCACTATAGGG **56-15-15**

7. GCGGAC**TTTT**CCCTATAGTGAGTCGTAACCAGCAATACCAACGACTCACTATAGGG **56-16-16**

Figure 2.10. Panel A: Autoradiograph of 15% polyacrylamide gel showing oligonucleotide "single spot" RNA production with variable length partial promoter sequences (10 – 15 bp) labeling after incubation with T7 RNA polymerase and radiolabeled rATP. Oligonucleotides with partial promoter sequences longer than 15 bp (56-16-16) produce "double spot" RNA. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

A



B

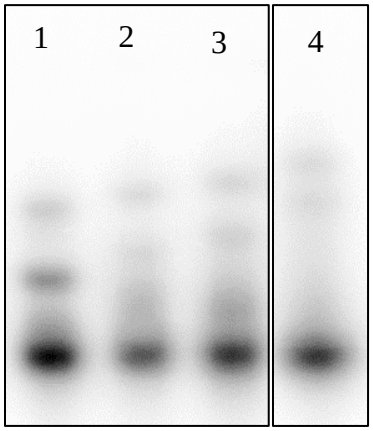
1. GCGGAC**TTTT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10**

2. GCGGAC**TTTT**CGTACTAGATCTGACGAACGCAAGAACCAGTATCAGATCTAGTACG **56-14 Scr**

3. GCGGAC**TTTT**CGTACTAGATCTGACGAACGCAAGAACCAGCGTCAGATCTAGTACG **56-16 Scr**

Figure 2.11. Panel A: Autoradiograph of 15% polyacrylamide gel showing single spot RNA production after incubation with T7 RNA polymerase and radiolabeled rATP. Oligonucleotide 56-12-10 (lane 1) has a 10 bp partial promoter sequence and produces "single spot" RNA. Oligonucleotides 56-14 Scr and 56-16 Scr contain scrambled promoter sequences and do not produce single spot RNA. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

A



B

1. GCGGAC**TTTT**CCCTATAGGCAGACAACGCAAACCAGCAGAGCCTATAGGG **50-10-8** (20 bp-loop)

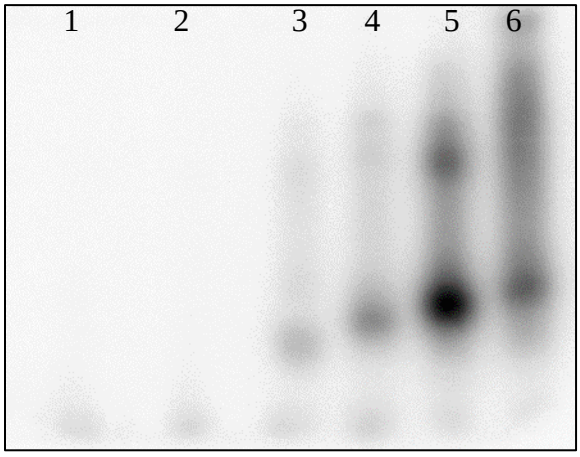
2. GCGGAC**TTTT**CCCTATAGGCAGACGACCAGCAGAGCCTATAGGG **44-10-8** (14 bp-loop)

3. GCGGAC**TTTT**CCCTATAGGCAGACCAGAGCCTATAGGG **38-10-8** (8 bp-loop)

4. GCGGAC**TTTT**CCCTATAGGCTTGCCTATAGGG **32-10-8** (2 bp-loop)

Figure 2.12. Panel A: Autoradiograph of 15% polyacrylamide gel showing "single spot" RNA after incubation with T7 RNA polymerase and radiolabeled rATP. Each hairpin oligonucleotide has a duplex length of 10 bp, a 5' truncated promoter to 8 bp, and a 10 nucleotide 5'extension. Each hairpin oligonucleotide has a different loop length. Panel B: The DNA sequences are written 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

A



B

1. GCGGAC**AAAA**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 A4**

2. GCGG**TTTTAA**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 A2T4**

3. GCGGACAA**TT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 T2**

4. GCGGAC**TTTT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 T4**

5. GCGG**TTTTTT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 T6**

6. GC**TTTTTTTT**CCCTATAGTGGCAGACAACGCAAGAACCAGCAGAGCCACTATAGGG **56-12-10 T8**

Figure 2.13. Panel A: Autoradiograph of 15% polyacrylamide gel showing *de novo* initiation with template directed elongation from oligonucleotides with recessed 3' ends and partial promoter sequences after incubation with T7 RNA polymerase and radiolabeled rATP. Panel B: The DNA sequences written are 5' to 3' for the oligonucleotides used in this experiment. Underlines indicate double stranded promoter regions or partial double stranded promoter sequences, red nucleotides indicate the potential for duplex beyond the promoter region, bold letters indicate potential templates for ATP radionucleotides.

2.5 Discussion

When presented with an oligonucleotide which can form a hairpin loop with a recessed 3' end and a single ribonucleotide triphosphate, T7RNAP displays several novel activities depending on the amount of the promoter sequence retained in the duplex region of the hairpin. When a full promoter (20 nucleotides, -17 to +3) or up to a four base pair 5' deletion of the full double stranded promoter (16 nucleotide, -13 to +3) is used in these circumstances a double spot, abortive transcript is produced. Although this *de novo* initiation can only occur if a recessed 3' end is present, it is template independent and produces the same double spot pattern independent of the sequence of the 5' extension. This is presumably the result of a strong promoter binding due to the retention of the -12 to -5 region of the promoter which is bound by the specificity loop.

Five to ten base pair deletions of the 5' end of the double stranded promoter result in hairpin oligonucleotides with 15 bp (-12 to +3) to 10 bp (-7 to +3) partial promoters that can initiate *de novo* template-dependent RNA synthesis. These RNAs initiate at the first unpaired base after the recessed 3' end at a +4 relative to classical full promoter driven transcription. These partial promoters resemble mitochondrial promoters in length, 9 nucleotides (-8 to +1), however, mitochondrial promoters initiate at +1. The minimal T7 RNAP promoter sequence able to support *de novo* initiation at a recessed 3' end *in vivo* is 5'-CACTATAGGG-3'. Deletion of promoters beyond 10 base pairs results in the loss of *de novo* initiation.

However, hairpin oligonucleotides with recessed 3’ ends and a duplex shorter than about 9 base pairs of AT–rich sequence or longer than about 3 base pairs of GC-rich sequence can add a Watson-Crick specified ribonucleotide to the 3’ end of DNA oligonucleotides (Sarcar & Miller, manuscript submitted), producing a DNA-RNA phosphodiester bond. The addition of a specific ribonucleotide to a DNA oligonucleotide constitutes DNA editing. This activity may be similar to the co-transcriptional, non-DNA templated, insertional RNA editing activity observed for the mtRNAP in mitochondria of myxomycetes (Miller, Padmanabhan, & Sarcar, 2017; Miller & Miller, 2008; Visomirski-Robic & Gott, 1997).

The one common feature of these three activities is a hairpin oligonucleotide with a recessed 3’ end. A recessed 3’ end is sufficient for DNA polymerase binding and primer extension and is also sufficient for T7 RNAP to bind oligonucleotides in the absence of a promoter. In the absence of a promoter hairpin with duplexes between 3 and 9 depending on base composition can bind a NTP at a complementary base at the 3’ end of the oligonucleotide and form a DNA-RNA phosphodiester bond. Oligonucleotides with stable duplexes greater than about 9 bp do not label oligonucleotide 3’ ends. However, duplexes greater than 9 nucleotides containing a partial promoter sequence are able to bind T7 RNAP and bind an NTP at a complementary base next to the base paired 3’ end. Under these conditions a phosphodiester bond is not formed between DNA and RNA. Rather, templated *de novo* transcription occurs albeit at an ectopic initiation site (+4 in these experiments rather than +1). With 15 nucleotides to full promoters, de novo initiation can also occur. Here, a NTP can be found by the polymerase without a complementary

sequence in the template. Here, "double spot" RNA is repetitively synthesized by a template-independent mechanism.

The evolutionary relationship of the single subunit polymerases is obscure even though they are clearly related through highly conserved motifs. DNA-dependent DNA polymerases and RNA-dependent DNA polymerases utilize RNA or DNA primers to initiate DNA synthesis at specific sites determined by the complementarity of the primer. Although primer extension is also used during elongation by RNA polymerases, they have uniquely evolved the ability to initiate *de novo* transcription through promoter recognition. Here we show that in the correct sequence context (partial promoters), primers can be used to specify *de novo* initiation of transcription *in vitro.* This activity may represent a transitional activity from primer-initiated nucleic acid synthesis to promoter-directed initiation in the evolution of RNA polymerases.

In any case, the discovery of primer-specified *de novo* initiation *in vitro* suggests a method of synthesizing from oligonucleotide hairpin templates, RNAs without G nucleotides at the 5' end and without non-DNA templated nucleotides at the 3' end. Experiments to determine whether the use of partial promoter sequences similar to mitochondrial promoters will work with T7 polymerase *in vivo* are in progress.

2.6 References

Ahlquist, P. (2002). RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science, 296*(5571), 1270-1273. doi:10.1126/science.1069132

Brown, J. E., Klement, J. F., & McAllister, W. T. (1986). Sequences of three promoters for the bacteriophage SP6 RNA polymerase. *Nucleic Acids Res, 14*(8), 3521-3526.

Castro, C., Smidansky, E. D., Arnold, J. J., Maksimchuk, K. R., Moustafa, I., Uchida, A., . . . Cameron, C. E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol, 16*(2), 212-218. doi:10.1038/nsmb.1540

Cermakian, N., Ikeda, T. M., Cedergren, R., & Gray, M. W. (1996). Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. *Nucleic Acids Res, 24*(4), 648-654.

Cermakian, N., Ikeda, T. M., Miramontes, P., Lang, B. F., Gray, M. W., & Cedergren, R. (1997). On the evolution of the single-subunit RNA polymerases. *J Mol Evol, 45*(6), 671-681.

Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D., & Burgess, R. R. (1988). Bacteriophage T7 late promoters with point mutations: quantitative footprinting and in vivo expression. *Nucleic Acids Res, 16*(10), 4511-4524.

Chapman, K. A., & Wells, R. D. (1982). Bacteriophage T7 late promoters: construction and in vitro transcription properties of deletion mutants. *Nucleic Acids Res, 10*(20), 6331-6340.

Cheetham, G. M. T., & Steitz, T. A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Current Opinion in Structural Biology, 10*(1), 117-123. doi:https://doi.org/10.1016/S0959-440X(99)00058-5

Chen, T., & Romesberg, F. E. (2014). Directed polymerase evolution. *FEBS Lett, 588*(2), 219-229. doi:10.1016/j.febslet.2013.10.040

Dunn, J. J., & Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol, 166*(4), 477-535.

Gong, P., Esposito, E. A., & Martin, C. T. (2004). Initial bubble collapse plays a key role in the transition to elongation in T7 RNA polymerase. *J Biol Chem, 279*(43), 44277-44285. doi:10.1074/jbc.M409118200

Gong, P., & Martin, C. T. (2006). Mechanism of instability in abortive cycling by T7 RNA polymerase. *J Biol Chem, 281*(33), 23533-23544. doi:10.1074/jbc.M604023200

Ikeda, R. A., Ligman, C. M., & Warshamana, S. (1992). T7 promoter contacts essential for promoter activity in vivo. *Nucleic Acids Res, 20*(10), 2517-2524.

Imburgio, D., Rong, M., Ma, K., & McAllister, W. T. (2000). Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry, 39*(34), 10419-10430.

Jorgensen, E. D., Durbin, R. K., Risman, S. S., & McAllister, W. T. (1991). Specific contacts between the bacteriophage T3, T7, and SP6 RNA polymerases and their promoters. *Journal of Biological Chemistry, 266*(1), 645-651.

Kennedy, W. P., Momand, J. R., & Yin, Y. W. (2007). Mechanism for de novo RNA synthesis and initiating nucleotide specificity by t7 RNA polymerase. *J Mol Biol, 370*(2), 256-268. doi:10.1016/j.jmb.2007.03.041

Klement, J. F., Moorefiedl, M. B., Jorgensen, E., Brown, J. E., Risman, S., & McAllister, W. T. (1990). Discrimination between bacteriophage T3 and T7 promoters by the T3 and T7 RNA polymerases depends primarily upon a three base-pair region located 10 to 12 base-pairs upstream from the start site. *Journal of Molecular Biology, 215*(1), 21-29. doi:https://doi.org/10.1016/S0022-2836(05)80091-9

Ling, M. L., Risman, S. S., Klement, J. F., McGraw, N., & McAllister, W. T. (1989). Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. *Nucleic Acids Res, 17*(4), 1605-1618.

Liu, C., & Martin, C. T. (2002). Promoter clearance by T7 RNA polymerase. Initial bubble collapse and transcript dissociation monitored by base analog fluorescence. *J Biol Chem, 277*(4), 2725-2731. doi:10.1074/jbc.M108856200

Martin, C. T., & Coleman, J. E. (1987). Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry, 26*(10), 2690-2696.

Martin, C. T., Muller, D. K., & Coleman, J. E. (1988). Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry, 27*(11), 3966-3974.

Masters, B. S., Stohl, L. L., & Clayton, D. A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell, 51*(1), 89-99.

Matsunaga, M., & Jaehning, J. A. (2004). Intrinsic promoter recognition by a "core" RNA polymerase. *J Biol Chem, 279*(43), 44239-44242. doi:10.1074/jbc.C400384200

Miller, D., Padmanabhan, R., & Sarcar, S. N. (2017). Chapter 4 - Genomics and Gene Expression in Myxomycetes A2 - Stephenson, Steven L. In C. Rojas (Ed.), *Myxomycetes* (pp. 107-143): Academic Press.

Miller, M. L., & Miller, D. L. (2008). Non-DNA-templated addition of nucleotides to the 3' end of RNAs by the mitochondrial RNA polymerase of Physarum polycephalum. *Mol Cell Biol, 28*(18), 5795-5802. doi:10.1128/MCB.00356-08

Milligan, J. F., Groebe, D. R., Witherell, G. W., & Uhlenbeck, O. C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Research, 15*(21), 8783-8798. doi:10.1093/nar/15.21.8783

Nayak, D., Guo, Q., & Sousa, R. (2009). A promoter recognition mechanism common to yeast mitochondrial and phage t7 RNA polymerases. *J Biol Chem, 284*(20), 13641-13647. doi:10.1074/jbc.M900718200

Osterman, H. L., & Coleman, J. E. (1981). T7 ribonucleic acid polymerase-promotor interactions. *Biochemistry, 20*(17), 4884-4892.

Rong, M., He, B., McAllister, W. T., & Durbin, R. K. (1998). Promoter specificity determinants of T7 RNA polymerase. *Proceedings of the National Academy of Sciences, 95*(2), 515-519.

Steitz, T. A., Smerdon, S. J., Jager, J., & Joyce, C. M. (1994). A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. *Science, 266*(5193), 2022-2025.

Tang, G. Q., Bandwar, R. P., & Patel, S. S. (2005). Extended upstream A-T sequence increases T7 promoter strength. *J Biol Chem, 280*(49), 40707-40713. doi:10.1074/jbc.M508013200

Temiakov, D., Mentesana, P. E., Ma, K., Mustaev, A., Borukhov, S., & McAllister, W. T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proc Natl Acad Sci U S A, 97*(26), 14109-14114. doi:10.1073/pnas.250473197

Visomirski-Robic, L. M., & Gott, J. M. (1997). Insertional editing in isolated Physarum mitochondria is linked to RNA synthesis. *RNA, 3*(8), 821-837.

Werner, F., & Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol, 9*(2), 85-98. doi:10.1038/nrmicro2507

Weston, B. F., Kuzmine, I., & Martin, C. T. (1997). Positioning of the start site in the initiation of transcription by bacteriophage T7 RNA polymerase. *J Mol Biol, 272*(1), 21-30. doi:10.1006/jmbi.1997.1199

CHAPTER 3

DISCUSSION AND FUTURE DIRECTION

In this study, experiments were performed to evaluate T7 RNAP's ability to initiate transcription *de novo* from truncated promoter sequences similar to mitochondrial promoters in an *in vitro* transcription system in which T7 RNAP was supplied with oligonucleotides containing T7 RNAP promoter which can form inter- and intra-molecular double stranded regions and a single radiolabeled ribonucleotide triphosphate. Using this *in vitro* system, we show that T7 RNAP can correctly and efficiently initiate transcription on truncated promoter sequences resembling mitochondrial promoters. In order to determine the ability of T7 RNAP to initiate at a truncated promoter, the promoter sequence was systematically deleted from the 5' end of the 20 nucleotide double stranded region on the oligonucleotides.

When a complete or almost complete [20 (-17 to +3) to 16 (-13 to +3) nucleotides] double stranded T7 RNAP promoter with a 3' recessed end is present, small double spot RNA products are produced through template-independent and promoter-dependent stuttering corresponding to abortive initiation. This effect was lost when a scrambled promoter was used. Moreover, oligonucleotides containing complete or almost complete promoter with blunt end terminus (66-30-20) failed to produce a double spot RNA product, showing that a 3' recessed end is required to produce the double spot RNA product.

Oligonucleotides with 15 (-12 to +3) to 10 (-7 to +3) bp truncated promoter failed to produce double spot RNA products, but instead, reproducibly produced a high mobility band designated as single spot RNA product, indicating that a 16-long promoter or higher is required for the production of double spot RNA product. The single spot RNA product was sensitive to RNAase. When the partial promoter region was scrambled, these oligonucleotides failed to produce the single spot RNA product indicating that this activity is dependent on the presence of partial promoter sequence and also dependent on the 5' extension of the template region of the oligonucleotide. Consistent with this observation, the single spot RNAs made from oligonucleotides containing 2 to 8 T nucleotides in the 5' extension of the template produced single spot RNAs corresponding to the length of the template indicating that the polymerase can initiate template-dependent *de novo* RNA synthesis from a 3' recessed end. Further, the site of the *de novo* initiation is located at +4 in contrast with the canonical site of transcription initiation at +1. These partial promoters resembled mitochondrial promoters in length.

Deletion of promoters beyond 10 base pairs with the duplex length from 2 to 10 base pairs with constant loop length results in the loss of *de novo* initiation, but. In some cases the 3' end of the hairpin oligonucleotide is labeled. This labeling disappeared when the duplex length was below 2 or above 10. A promoter-independent and template-dependent addition of rNTPs to oligonucleotides with recessed 3' end has been reported and characterized by Sarcar & Miller (manuscript submitted). This labeling absolutely requires a double stranded region with a recessed 3' end, but it is limited by the length of the base paired region. The labeling activity was also template dependent consistent to what has been reported previously. Systematic alteration of

the 5’ extension/template of the 44-7-5 oligonucleotide by substitution of A nucleotides for T nucleotides resulted in no labeling showing that the labeling site next to the 3’ end must base pair with the radiolabeled ATP. It is unlikely though, that multiple A nucleotides are added as labeling intensity was essentially equal with multiple T nucleotides at the 5’ extension/template except for 8 T nucleotides where labeling intensity decreased.

The addition of a specific ribonucleotide to a DNA oligonucleotide constitutes DNA editing, similar to the co-transcriptional, non-DNA-templated, insertional RNA editing activity observed for the mtRNAP in mitochondria of myxomycetes (Miller, Padmanabhan, & Sarcar, 2017; Miller & Miller, 2008; Visomirski-Robic & Gott, 1997). The evolutionary relationship of the single subunit polymerases is obscure even though they are clearly related through highly conserved motifs. DNA-dependent DNA polymerases and RNA-dependent DNA polymerases utilize RNA or DNA primers to initiate DNA synthesis at specific sites determined by the complementarity of the primer. Although primer extension is also used during elongation by RNA polymerases, they have uniquely evolved the ability to initiate *de novo* transcription through promoter recognition. we have shown that in the correct sequence context (partial promoters) primers can be used to specify *de novo* initiation of transcription *in vitro.* This activity may represent a transitional activity from primer-initiated nucleic acid synthesis to promoter-directed initiation in the evolution of RNA polymerases.

In any case, the discovery of primer-specified *de novo* initiation *in vitro* suggests a method of synthesizing from oligonucleotide templates, RNAs without G nucleotides at the 5' end and without non-DNA templated nucleotides at the 3' end. Experiments to determine whether the use of partial promoter sequences similar to mitochondrial promoters will work with T7 polymerase *in vivo* are in progress.

3.1 References

Miller, D., Padmanabhan, R., & Sarcar, S. N. (2017). Chapter 4 - Genomics and Gene Expression in Myxomycetes A2 - Stephenson, Steven L. In C. Rojas (Ed.), *Myxomycetes* (pp. 107-143): Academic Press.

Miller, M. L., & Miller, D. L. (2008). Non-DNA-templated addition of nucleotides to the 3' end of RNAs by the mitochondrial RNA polymerase of Physarum polycephalum. *Mol Cell Biol*, *28*(18), 5795-5802. doi:10.1128/MCB.00356-08

Visomirski-Robic, L. M., & Gott, J. M. (1997). Insertional editing in isolated Physarum mitochondria is linked to RNA synthesis. *RNA*, *3*(8), 821-837.

BIOGRAPHICAL SKETCH

Ramesh Padmanabhan was born in Calcutta (Kolkata), India. After finishing his High School from Don Bosco High School, Irinjalakkuda, Kerala, India in 1995, Ramesh finished his Pre-Degree from University of Calicut, India in 1997. Ramesh earned his Baccalaureate degree in Biotechnology from University of Kerala, India in 2002, and finished his MS in Industrial Biotechnology from Bharathiar University, Coimbatore, Tamil Nadu, India. Ramesh then earned his MS degree in Biological Sciences in Summer 2009 from Wichita State University, Wichita, KS, USA and later joined the PhD program in Molecular and Cell Biology in the Department of Biological Sciences, at The University of Texas at Dallas in 2010. Ramesh earned his MS degree in Molecular and Cell Biology in Fall 2013 from The University of Texas at Dallas.

CURRICULUM VITAE

Ramesh Padmanabhan

Address: FO 3.218 Molecular and Cell Biology, The University of Texas at Dallas

800W Campbell Rd Richardson TX 75080.

E-mail: rxp104120@utdallas.edu.

Education:

PhD The University of Texas at Dallas 2018

MS Molecular and Cell Biology The University of Texas at Dallas 2013

MS Biological Sciences Wichita State University2009

M.Sc. Industrial Biotechnology Bharathiar University 2004

B.Sc. Biotechnology University of Kerala 2002

Professional experience:

Research Assistant Department of Biological Sciences Protein Core Wichita State University Jan 2010-June 2010.

Research Intern: Reliance Life Sciences Private Limited Mumbai, India Dec 2003-Apr 2004.

Research Fellow: Reliance Life Sciences Private Limited Mumbai, India Jun 2004-Dec 2004.

Research Intern Sugar cane Breeding Institute Indian Council for Agricultural Research Coimbatore India March 2003- May 2003.

Teaching/Research Assistant Experience:

Graduate research assistant Training and Technology Teams Wichita State University Jan 2007-May 2007.

Graduate Teaching Assistant Department of Biological Sciences Wichita State University May 2007-July 2009.

Graduate Teaching Assistant Department of Biological Sciences The University of Texas at Dallas Aug 2010- May 2017.

Research Interests:

Gene expression during embryonic development and cancer, Developmental Biology of cancer

Cancer drug discovery and drug targeting

Cancer drug resistance, drug transporters

Cancer pathology

Cancer cell metabolism

Cell death pathways

Cancer epigenetics

Publications:

1. Padmanabhan, R., I. R. Hendry, J. R. Knapp, B. Shuai and W. J. Hendry (2017). "Altered microRNA expression patterns during the initiation and promotion stages of neonatal diethylstilbestrol-induced dysplasia/neoplasia in the hamster (*Mesocricetus auratus*) uterus." Cell Biology and Toxicology **33**(5): 483-500.

2. Genomics and Gene Expression in Myxomycetes. Dennis Miller, Ramesh Padmanabhan, and Subha N. Sarcar. In Myxomycetes Biology, Systematics, Biogeography and Ecology. Academic Press 2017.

3. The length of promoter sequence affects the *de novo* initiation of T7 RNA polymerase *in vitro*: New insights into the evolution of promoters for single subunit RNA polymerases (Manuscript submitted).

Conference Presentations:

- Gilbert S. Greenwald symposium on reproduction (2007) University of Kansas Medical Center, Kansas City, KS, USA
- Annual SSR (society for the study of reproduction) conference, Pittsburgh, PA (July 2009) platform presentation
- "Altered micro-RNA expression patterns during the initiation and promotion stages of neonatal diethylstilbestrol-induced dysplasia/neoplasia in the hamster uterus". Ramesh Padmanabhan, Isabel R. Hendry, William J. Hendry. Abstract published in *Biol Reprod June 15, 2009 81 (1 Supplement)* 101.

- Poster presentation**,** Gilbert S. Greenwald Symposium on Reproduction (October 2009)**.** University of Kansas Medical Center, Kansas City, Kansas. “Altered micro-RNA expression patterns during the initiation and promotion stages of neonatal diethylstilbestrol-induced dysplasia/neoplasia in the hamster uterus”.