Spring 5-30-2018

Interplay of Transcription Factor E and Spt4/5 During Transcription Initiation in *Pyrococcus furiosus*

Kimberly Kay Sheffield
Portland State University

Interplay of Transcription Factor E and Spt4/5 During Transcription Initiation in

Pyrococcus furiosus

by

Kimberly Kay Sheffield

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Biology

Thesis Committee:
Michael Bartlett, Chair
Justin Courcelle
Kenneth Stedman

Portland State University
2018

Abstract

Transcription, the first step in gene expression, is a highly regulated process which relies on a multi-protein complex to occur. Among these proteins are transcription factors, including initiation and elongation factors, which play differing roles in early and late stages of transcription. The mechanisms of transition from transcription initiation to elongation are not well understood in archaea, nor are the structures of the transcription factors involved. For transcription to occur *in vitro*, transcription factors TATA binding protein (TBP) and Transcription Factor B (TFB) are sufficient to allow RNA polymerase (RNAP) to synthesize RNA from template DNA. Another factor, Transcription Factor E (TFE), can also join the initiation complex and is likely to be essential *in vivo*. TFE is known to contribute to initiation by enhancing promoter opening, and while it has been shown to persist in elongation complexes, its role after initiation is unknown. Spt4/5, the archaeal homolog of the only universally conserved RNAP-associated factor, is known to join complexes in elongation steps and enhance processivity of the polymerase. However, if Spt4/5 joins pre-initiated complexes, it has been shown to inhibit transcription activity. The experiments in this thesis show that TFE and Spt4/5 participate in a crucial interchange at the upstream fork of the transcription bubble that helps define the timing of Spt4/5 binding. Using unnatural amino acid crosslinking techniques, the points of proximity between specific regions of these two factors and the template DNA have been mapped to identify possible sites of interaction. Competitive crosslinking assays indicate the exact timing of the shift in affinity between TFE and Spt4/5 for their shared binding site on RNAP. These data, combined with transcription assays, suggest a new role for

TFE in preventing premature Spt4/5 binding, corresponding with a unique localized

mobility within the winged helix of TFE.

Acknowledgements

I would like to acknowledge my advisor, Dr. Bartlett, for his enduring patience and enthusiasm with all aspects of this project, as well as my committee members, Drs. Courcelle and Stedman, for their encouragement and advice. I also must show my deepest appreciation to Paweł Buczkowski, Jennifer Phelan, and Lucille Braun, for all the pineapples.

Table of Contents

List of Tables

List of Figures

Introduction

Gene Expression

All organisms on earth must express the information in their DNA to perform the necessary functions for life. Gene expression is the process by which DNA is transcribed into mRNA, which is then translated into amino acid chains, which fold to form functional proteins. The steps of gene expression are highly regulated and highly conserved throughout evolutionary history.

Gene expression begins with transcription, when promoter DNA is bound by RNA polymerase with assistance from general transcription factors. RNA polymerase then synthesizes an mRNA strand complementary to the template DNA strand. The mRNA strand may then go through post-transcriptional processing steps including splicing, to remove unnecessary nucleic acids, and mRNA capping and polyadenylation, to improve recognition and durability. After transcription and any post-transcriptional processing, the mRNA strand interacts with a ribosome for translation. tRNAs recognize specific triplet codons in the mRNA and incorporate a single amino acid into a growing peptide chain. The polypeptide chain folds, with or without the help of a chaperone, to attain its structure, contributing to cellular function.

Evolution of Transcription Machinery

Transcription machinery has been highly conserved throughout evolutionary history. The core structure of RNA polymerase is conserved in all cellular organisms, but the number of subunits is variable across the evolutionary tree. The most highly

1

conserved regions are those including and surrounding the main channel and clamp of the RNA polymerase, where template DNA is loaded for transcription. Variation between species and domains occurs primarily on the surface of the polymerase (Jun, et al., 2011). Bacterial RNAP is composed of four to six subunits, while archaeal RNAP is composed of up to 13 subunits (Langer, et al., 1995). Eukaryotic RNA polymerase II is more similar to archaeal than bacterial RNA polymerase, with only one subunit not found in archaea and an overall structure that is essentially superimposable on the archaeal structure (Hirata, et al., 2008). Additionally, eukaryotic and archaeal RNA polymerase feature an additional two subunit stalk structure which is not found in bacterial RNA polymerase (Jun, et al., 2011).

The basal general transcription factors are well conserved from archaeal to eukaryotic systems, and contrast with the sigma factors used by bacteria for initiation. Archaeal transcription factor B, which aids in RNA polymerase recruitment to template DNA, is highly conserved with eukaryotic TFIIB, although it also bears weak similarities to regions of the bacterial RNA polymerase subunit sigma (Burton, et al., 2014). Archaeal TATA-binding protein, which helps to orient DNA for RNA polymerase binding, also has a homolog in a subunit of eukaryotic TFIID, while bacteria do not possess any TBP homologs (Brindefalk, et al., 2013). RNA polymerase, TBP, and TFB or TFIIB are necessary and sufficient for *in vitro* transcription in both archaea and eukaryotes (Hethke, et al., 1996; Tyree, et al., 1993). An additional general transcription factor, Transcription Factor E or TFE, is also conserved in archaea and eukaryotes, with archaeal TFE homologous to eukaryotic TFIIEα (Bell, et al., 2001). Eukaryotes feature

an additional TFIIE subunit, TFIIEβ, which has homologs in some archaea, including the crenarchaea (Blombach, et al., 2018). While TFEα is not necessary for *in vitro* transcription, it has not been successfully deleted from an archaeal genome and is therefore believed to be essential (Blombach, et al., 2018). Eukaryotic TFIIE recruits TFIIH to transcription complexes to enhance promoter opening (Ohkuma, et al., 1995). Archaeal TFE has similarly been shown to enhance melting at the promoter region, and to participate in transcription elongation complexes (Hanzelka, et al., 2001; Grunberg, et al., 2007).

One RNA polymerase associated factor has been identified as conserved across all three domains of life (Werner, 2007). This factor is known as NusG in bacteria, Spt5 in archaea and most eukaryotes, and DRB Sensitivity Inducing Factor (DSIF) in humans. This factor is known to participate in elongation steps in all domains. Bacterial NusG is recruited to transcribing complexes late in elongation, and can induce pausing and recruit termination factors (Li, et al., 1992; Yakhnin, et al., 2016). Archaeal Spt4/5 is recruited close to the promoter on most genes and enhances elongation (Smollett, et al., 2017). However, Spt4/5 binding to pre-initiated archaeal complexes results in an inhibition of transcription (Grohmann, et al., 2011). Eukaryotic Spt4/5 is recruited close to the promoter and likely plays a role in the transition from initiation to elongation (Mayer, et al., 2010).

Pyrococcus furiosus and Archaeal Transcription Complexes

Prior to 1977, biological life was divided into two domains: prokaryotes and eukaryotes, defined by the presence or absence of membrane bound organelles. In 1965, Linus Pauling and Emile Zuckerkandl proposed what is now known as phylogenetics to more explicitly define life and evolutionary history based on similarities between organisms at the genetic and molecular level (Zuckerkandl & Pauling, 1965). Phylogenetic comparisons using ribosomal RNA genes in prokaryotes, performed by Carl Woese and others in 1977, identified archaea as a distinct domain of prokaryotic life independent from bacteria (Woese & Fox, 1977).

Archaeal transcription complexes are of particular interest due to their similarity to eukaryotic transcription complexes and also due to the unique characteristics of archaeal life (Woese, et al., 1990). Many archaea are extremophiles, living and transcribing their genes in environments that would be fatal for other organisms. The stability of their proteins, including those involved in transcription complexes, in these extreme conditions has led to research interest in the structures of these proteins and how they differ from similar structures in organisms not adapted to extremes. Despite their uniqueness, archaea contain many genes with homologs in eukaryotic systems, and have thus been studied as a potentially simpler system for understanding the function of the conserved parts of the more complex eukaryotic transcription system (Bell & Jackson, 1998).

Pyrococcus furiosus is an archaeal study system of particular interest due to its eukaryotic similarities, its extreme hyperthermophily, and its potential use in biotechnological applications. *P. furiosus* belongs to the archaeal phylum euryarchaeota,

a diverse phylum including thermophiles, halophiles, and methanogens (Woese, et al., 1990). This species was identified in shallow underwater vents off the coast of Italy (Fiala & Stetter, 1986). This environment requires high temperature resistance as well as resistance to oxidative stress, since *P. furiosus* is a hyperthermophilic anaerobe. *P. furiosus* is an important research model due to its ease of culture in the lab, low risk of culture contamination due to an optimal growth temperature of 95ºC, and the existence of a naturally competent strain known as COM1 (Lipscomb, et al., 2011). Additionally, *P. furiosus* produces hydrogen gas as a metabolic byproduct, and recent research efforts have focused on attempts to optimize this hydrogen production process as a renewable fuel source (Lipscomb, et al., 2014).

The *P. furiosus* transcription complex contains proteins that are highly conserved within the core eukaryotic transcription machinery. *In vitro* assays have shown that the *P. furiosus* proteins necessary to form a functional transcription complex are RNA polymerase (homologous to eukaryotic RNA polymerase II), transcription factor B (homologous to eukaryotic TFIIB), and TATA-binding protein (homologous to the TBP subunit of eukaryotic TFIID) (Hethke, et al., 1999). Other transcription factors, including transcription factor E (homologous to eukaryotic TFIIEα) and Spt4/5 (the only universally conserved transcription factor), are not necessary for *in vitro* transcription but are likely to be essential *in vivo*, as their homologs are essential in other domains of life (Bell, et al., 2001; Herbert, et al., 2011).

Transcription Factor E Function & Structure

5

Archaeal transcription factor E (TFE) is homologous to the alpha subunit of eukaryotic TFIIE (Hanzelka, et al., 2001). A TFIIEβ subunit also exists in eukaryotes, but an archaeal TFIIEβ homolog has only been identified in a select few archaeal species, including crenarchaea and halophilic euryarchaea (Blombach, et al., 2018). A TFIIEβ homolog has not yet been observed in *P. furiosus.* While TFE is not necessary for transcription *in vitro,* it has never been successfully deleted from the *P. furiosus* genome and is likely to be essential (A. Wright & M. Bartlett, personal communication).

TFE has been shown in *P. furiosus* to be a part of transcription complexes prior to initiation as well as throughout initiation and into elongation (Grunberg, et al., 2007). A complete structure for *P. furiosus* TFE does not yet exist, though partial structures are available (Meinhart, et al., 2003). Hypothetical *P. furiosus* TFE structures, as well as TFE structures modeled in other archaea, indicate the presence of winged helix and zinc ribbon domains (Grohmann, et al., 2011). TFE is known to bind to RNA polymerase on the clamp domain of the largest subunit, but has not been shown to have any DNA binding activity (Werner & Weinzerl, 2005). The winged helix has been mapped through previous crosslinking experiments, positioning it proximal to the nontranscribed strand in the opened transcription bubble, near the upstream edge (M. Bartlett, personal communication). No interactions between the transcribed strand and any part of TFE have previously been shown.

In *in vitro* transcription assays, TFE enhances melting of the transcription bubble (Naji, et al., 2007). TBP recruitment is also impacted by TFE, with TATA box recognition enhanced by TFE (Bell, et al., 2001). TFE also efficiently restores function to

complexes negatively impacted by other deficiencies in the transcription mechanism. Among these deficiencies are reduction in available NTPs for RNA synthesis, or the substitution of the essential transcription factor B1 with a weaker variant found in *P. furiosus* known as TFB2 (Micorescu, et al., 2008; A. Bhattarai, R. Eustis, & M. Bartlett, personal communication).

Current Understanding of Spt4/5 Function & Structure

Archaeal Spt4/5 is a heterodimeric protein composed of Spt4 and Spt5 monomers. Spt4/5 is the only universally conserved RNA polymerase-associated factor. Spt5 is composed of the NusG N domain (NGN) and a single Kyprides-Ouzounis-Woese (KOW) domain (Martinez-Rucobo, et al., 2011). The bacterial homolog, NusG, is homologous only to Spt5 (Sevostyanova & Artsimovich, 2010). The eukaryotic homolog consists of Spt4, Spt5, additional KOW domains, and a C-terminal repeat domain (Hartzog & Fu, 2014).

Variations in NusG/Spt4/5 structure correlate with similar variations in function within the transcription complex. Bacterial NusG has been determined to associate primarily with elongating RNA polymerase far from the promoter (Mooney, et al., 2010). In *Escherichia coli*, NusG has been shown to enhance processivity of RNA polymerase by promoting translocation (Herbert, et al., 2011). The bacterial homolog has also been shown to enhance RNA polymerase pausing, in association with a complex of other Nus factors, through sequence-specific interactions with the nontranscribed strand of DNA (Yakhnin & Babitzke, 2002). In addition, bacterial NusG is essential for transcription

termination through recruitment of Rho termination factors (Sevostyanova &

Artsimovich, 2010). Eukaryotic Spt4/5 has been shown to associate with the RNAP II

transcription complex after initiation and enhance transcription elongation (Mayer, et al.,

2010). Eukaryotic Spt4/5 plays additional roles in transcription coupled repair and post-

transcriptional processing, using C-terminal repeats and extra KOW domains not present

in other homologs (Li, et al., 2014). Archaeal Spt4/5 has been shown to have two modes

of recruitment that vary based on the gene being transcribed. The majority of genes

recruit Spt4/5 close to the promoter, while few genes like rRNA and CRISPR genes

demonstrate a much later recruitment, after at least 500 nucleotides have been transcribed

(Smollett, et al., 2017). Despite the difference in timing of joining the transcription

complex, both archaeal and eukaryotic Spt4/5 contribute to processive elongation in a

manner similar to bacterial NusG (Martinez-Rucobo, et al., 2011; Li, et al., 2014).

However, archaeal Spt4/5 has been shown *in vitro* to inhibit transcription when added to

pre-initiated complexes, but not when Spt4/5 binds during elongation (Grohmann, et al.,

2011).

Archaeal Spt4/5 binds to the RNA polymerase clamp and shares its binding site

with TFE. The two factors compete for this binding site, with affinity for each factor

appearing to vary depending on the stage of transcription. TFE efficiently binds RNA

polymerase prior to template loading, and also efficiently displaces Spt4/5 from pre-

initiated complexes (Grohmann, et al., 2011). However, Spt4/5 has been shown to bind

preferentially in complexes assembled on elongation scaffolds with an RNA strand 20

nucleotides in length (Grohmann, et al., 2011). While it is evident that a binding shift

between the two factors occurs at some point between pre-initiation and elongation, there is currently no information pinpointing the precise timing of this binding shift, or how it relates to the shift in archaeal Spt4/5 function between inhibitory and enhancing behavior.

Current Understanding of *Pyrococcus furiosus* Transcription Initiation & Promoter Escape

The transcription cycle progresses through a series of stages: pre-initiation, initiation, elongation, and termination. These stages are loosely defined by the length of the synthesized RNA strand, as well as the proteins participating in the transcription complex.

Transcription begins with a pre-initiation complex (PIC). In archaea, transcription factor B (TFB) and TATA-binding protein (TBP) first bind the DNA and orient it for DNA loading to RNA polymerase. Together, these factors recruit RNA polymerase to the promoter region. Other factors, like transcription factor E (TFE), may also bind at this stage. A complex is called pre-initiated when RNA polymerase and associated factors are bound to DNA, but transcription has not yet begun (Thomm, 2001).

Transcription initiates when RNA polymerase binds to NTP substrates and begins RNA synthesis from the DNA template. The transcription complex may then cycle through abortive initiation, in which the complex transcribes short RNA transcripts without translocating on the DNA template. Little research has been done on abortive initiation in archaea, but bacterial studies indicate that scrunching of the non-template

DNA strand within the RNA polymerase may eventually cause the shift from abortive to productive transcription initiation (Kapanidis, et al., 2006; Winkelman, et al., 2015).

Once the transcription complex begins productive transcription and the RNA strand reaches a specific length, the RNA polymerase will then translocate on the DNA template in a process termed promoter escape. Promoter escape occurs when the RNA polymerase begins translocation and moves away from the promoter, leaving behind TFB and TBP, which are only required at initiation. *In vitro* footprinting assays performed using *Pyrococcus furiosus* RNA polymerase and transcription factors indicate that, on the glutamate dehydrogenase promoter, RNA polymerase begins to translocate away from the promoter once the RNA strand reaches a length of nine nucleotides. RNA polymerase was also shown to undergo small shifts relative to the DNA prior to +9 nucleotides transcribed, possibly due to conformational changes within the transcription complex (Spitalny & Thomm, 2003).

Once the RNA polymerase begins translocation, the complex enters the elongation stage of transcription, and moves freely along the DNA template to transcribe more RNA. The RNA polymerase will then transcribe the remainder of the gene until transcription is terminated. While termination is not well studied in archaea, it is understood that termination can occur in a factor-dependent or independent manner, and recently a termination factor, Euryarchaeal termination activity (Eta), was identified in the archaea *Thermococcus kodakarensis* (Walker, et al., 2017). When transcription is terminated, the RNA polymerase disengages from the template DNA and the synthesized

RNA is released. The RNA polymerase can then be recruited by TBP and TFB to a new DNA template, and the transcription cycle will begin again.

Unnatural Amino Acids & Genetic Code Expansion

Non-canonical or unnatural amino acids are a new tool being developed for many uses in biological research. An unnatural amino acid contains amino and carboxyl groups identical to a canonical amino acid, with variations in the R group that are not found in life's 20 canonical amino acids. Chemistry of these side chains can be altered for any number of desired effects, including crosslinking, fluorophore labeling, insertion of unique chemically reactive groups, and other investigative tools (Xie & Schultz, 2006).

Initially, unnatural amino acids were incorporated into proteins by building proteins one amino acid at a time through solid state synthesis (Kauers, et al., 1986). While this was an effective method, it was inefficient and had low yields, and was also inapplicable to *in vivo* uses. More recently, research in the field of genetic code expansion has allowed construction and expression of proteins with unnatural amino acids efficiently *in vivo* (Chin, et al., 2002).

Genetic code expansion takes advantage of the redundancy of the genetic code. 64 triplet codons encode 20 amino acids, resulting in many amino acids being encoded by multiple codons. Rates of codon usage for each amino acid vary from organism to organism, with some codons significantly more rare than others. Notably, the genetic code includes three codons that code for no amino acid; stop codons serve to indicate the end point of translation, and thus are only featured once per gene. Among these, TAG,

also known as the amber codon, is the least commonly used in *Escherichia coli*, a common organism for use in protein synthesis. The amber codon's infrequent use in *E. coli* genes makes it a good target for manipulation, as alteration of its function will impact fewer genes than the other two stop codons.

To incorporate a specific unnatural amino acid at a TAG codon, tRNAs were mutated to efficiently receive the target amino acid. *Methanococcus janaschii* tRNA/amino acyl tRNA synthetase (aaRS) pairs were manipulated to allow incorporation of the unnatural amino acid *p*-benzoyl-*l*-phenylalanine (Bpa) at the site of a TAG codon *in vivo*, resulting in high yield and high efficiency of unnatural amino acid incorporation into a protein in *E. coli*. Thus, Bpa can be site-specifically incorporated by co-transformation of *E. coli* with two plasmids, one that contains the orthogonal tRNA/aaRS pair, and the other with a mutant gene containing the TAG codon at the site of desired benzoylphenylalanine placement (Chin, et al., 2002).

P-benzoyl-*l*-phenylalanine (Bpa) is an unnatural amino acid commonly used as a crosslinker. Its R group features a photoreactive benzophenone, that when exposed to 365 nm UV, is excited and forms a crosslink to a carbon-hydrogen bond within approximately 3.1 angstroms of the oxygen in the reactive ketone group (Dorman & Prestwich, 1994). Bpa is useful for crosslinking assays that require highly specific crosslinking, due to the relatively short reach of its R group compared to other unnatural amino acid crosslinkers featuring longer sidechains, such as aryl azides. Bpa's crosslinking wavelength is also useful for protein-DNA crosslinking; 365 nanometers is distant on the light spectrum

from ranges which can damage DNA or protein. Bpa is thus ideal for use in highly specific protein-DNA crosslinking applications.

Research Questions

The work in this thesis addresses questions surrounding the role of transcription factors in archaeal transcription complexes, specifically where and when these factors may interact with DNA. First, the positions of TFE and Spt4/5 relative to template DNA in *P. furiosus* transcription complexes were mapped using Bpa crosslinking. Movement of these transcription factors relative to DNA as the complex moves from transcription initiation to elongation were mapped with mutant promoters allowing for site-specific stops in transcription. The precise point of the binding shift between TFE and Spt4/5 was identified using competitive Bpa crosslinking assays to measure the impact of the competitor on transcription factor binding at each step of early transcription. The role of this binding shift was then further quantified using transcription activity assays to determine how TFE and Spt4/5 impact transcription activity both before and after their RNA polymerase binding affinity changes.

This work indicates that TFE is positioned at the upstream fork of DNA and is highly mobile relative to the template DNA throughout early elongation. This mobility appears to occur primarily in the tip of the winged helix of TFE, and likely serves to destabilize the transcription factor after it has contributed to stabilizing the open transcription bubble. TFE is efficiently displaced by Spt4/5, which also associates with DNA near the upstream fork, likely in a sequence-specific manner. The binding shift

13

appears to occur after the polymerase has transcribed eight nucleotides and before it begins translocation. This work also shows that Spt4/5 not only inhibits transcription in pre-initiation complexes, as previously observed, but continues inhibition until eight nucleotides have been transcribed. It is also demonstrated that the presence of TFE is sufficient to prevent Spt4/5 from inhibiting transcription until its functional shift occurs. This work is the first demonstration of high resolution mapping of these transcription factors in dynamic archaeal transcription complexes, and the first data to suggest that the binding shift between Spt4/5 and TFE occurs concurrently with promoter escape and a shift in Spt4/5 function in *P. furiosus.*

Methods

Synthesis of Bpa Mutants

pET24D plasmids containing the gene for *Pyrococcus furiosus* Spt4/5 were replicated with mutagenic primers containing an amber codon mutation in place of the target amino acid codon (Martinez-Rucobo, et al., 2011). PCR amplification generated two DNA fragments extending from the T7 gene to the Spt5 mutagenic region using 1x Phusion HF buffer, 0.2 mM dNTPs, 50 pmols each primer (Table 1), and 5 ng of the Spt4/5 plasmid, and 0.6 units Phusion DNA polymerase. Plasmid fragments were then gel purified. Fragments were reassembled into replicating plasmids through Gibson assembly. 25 fmols of each fragment were combined in preheated tubes with Gibson assembly master mix (1.33x isothermal buffer, 6.66% PEG 3350, 5.33 units Taq ligase, 0.01 units T5 exonuclease, and 0.03 units Phusion DNA polymerase (Thermo Scientific)) and heated to 50°C for 30 minutes. $CaCl_2$ competent BL21 gold cells were then transformed with Gibson assembly reactions and grown overnight for plasmid replication. Plasmids were purified with a GeneJet Plasmid Miniprep Kit (Thermo Scientific). Mutations were confirmed through sequencing.

Table 1. Primers used for amber codon mutagenesis. Primers were used to incorporate TAG in place of a target codon in *Pyrococcus furiosus* Spt5 on a pET24D plasmid.

Site of amber codon incorporation in Spt5	Forward primer	Reverse Primer
Threonine 11	5'-cagttgtctcttcctgtccatgc tacactctcacagcaaagattttc-3'	5'-gaaaatctttgctgtgagagtgt agcatggacaggaagagacaactg-3'
Glutamine 14	5'-gttgtctcttcctatccatgagt cactctcacag-3'	5'-ctgtgagagtgactcatggatag gaagagacaac-3'
Histidine 65	5'-gaacacctctagcctatctaatt ccccttattgcctcatcc-3'	5'-ggatgaggcaataaggggaatta gataggctagaggtgttc-3'
Proline 71	5'-tagacacgctagaggtgttc tttagggagaggttccctttaagga-3'	5'-tccttaaagggaacctctcccta aagaacacctctagcgtgtcta-3'

 Escherichia coli BL21 gold CaCl$_2$ competent cells were sequentially transformed to contain mutant Spt4/5 plasmids and plasmids for Bpa incorporation. 100 μL cells were first combined with Spt4/5 TAG mutant plasmids and incubated on ice for 30 minutes. Cells were then heated to 37°C for five minutes, treated with 1 mL LB growth medium, and then incubated at 37°C for 60 minutes. Cells were then spun down and plated on selective media, and successful mutants were then made competent with CaCl$_2$. Competent Spt4/5 mutants were then combined with 250 ng pSup Bpa-PheRS-6TRN plasmid containing orthogonal tRNAs for Bpa incorporation (Ryu & Schultz, 2006), and transformed as described above.

16

Protein Overexpression and Purification

E. coli BL21 Gold strains containing the desired plasmids were streaked on LB agar plates containing appropriate antibiotic and grown overnight. Cells were scraped from plates and suspended in 1 mL 2x Yeast Tryptone growth medium. 0.5 mL of suspended cells were used to inoculate 60 mL of sterile 2xYT in the presence of the appropriate antibiotic. For Bpa substitutions, 1 mM Bpa dissolved in 1N HCl was added to growth flasks prior to inoculation, followed by 1N NaOH to restore neutral pH. Culture flasks were incubated at 37°C until cultures reached an A600 value of 1.0. Protein synthesis was then induced by the addition of 1 mM IPTG. Cultures were then cooled to room temperature and incubated with shaking overnight.

Centrifuged cells were resuspended in lysis buffer and incubated on ice for 30 minutes. After lysis, cells were sonicated on ice to shear DNA, and then centrifuged to remove cell debris. The supernatant was then mixed with imidazole and $MgCl_2$ and loaded onto a syringe column filled with NiNTA agarose beads for His-tag binding to transcription factors (TBP, TFB, TFE, and Spt4/5). The column was then washed with increasing concentrations of imidazole and elutioned fractions were analyzed on SDS polyacrylamide gels to confirm presence of target proteins. Fractions containing the protein of interest were centrifuged in Amicon Ultra columns to concentrate the proteins and exchange into storage buffer (0.5M NaCl, 1 mM EDTA, 2 mM Tris pH 8.0). Samples of purified protein were then sequentially diluted and run on SDS polyacrylamide gels. Gels were scanned and analyzed in ImageQuant software to quantify pixel intensity of bands and calculate protein concentration relative to a BSA dilution curve.

Synthesis of U-less Cassettes

U-less cassettes were synthesized using *Pyrococcus furiosus* glutamate
dehydrogenase promoter with end points at -95 and +85. Target primers were generated
to eliminate thymine in the sequence until the desired stop site (Table 2). Polymerase
chain reactions were assembled with 1x Phusion HF buffer, 0.2 mM dNTPs, 1 μM each
forward and reverse primer, 5 ng GDH promoter template, and 0.8 units of Phusion DNA
polymerase. Amplicons were gel purified and used as megaprimers for mutagenesis of
GDHP DNA. Full length products were gel purified and sequenced to confirm mutations.

Table 2. Mutagenic primers used for U-less GDHP cassette synthesis. Primers were annealed and
extended to generate a megaprimer amplicon which was used for full length replication. Synthesis of a +5
cassette was not necessary as wild type GDHP is U-less through +5 nt.

Transcript Length (nt)	Mutagenic primer sequence (5'-3')
+6	CATCCCTCCAAATTAGGTATTTGGCGATACATTTTTGGGCAATAGC
+7	CATCCCTCCAAATTAGGAGTTTGGCGATACATTTTTGGGCAATA
+8	TGTTCATCCCTCCAAATTAGATGTTTGGCGATACATTTTTGGGC
+9	ATGTTCATCCCTCCAAATTAAGTGTTTGGCGATACATTTTTGGGC
+10	CCCTCCAAATTAGGTGTTTGGCGATACATTTTTGGGC

Crosslinking Assays

Transcription complexes were assembled *in vitro* by combining 120 nM TBP, 240
nM TFB1, 240 nM of Bpa substituted TFE or Spt4/5, and 40 nM RNAP on 40 nM *P.
furiosus* glutamate dehydrogenase promoter DNA with endpoints of -95 and +85. *P.*

18

furiosus TBP, TFB1, TFE, and Spt4/5 were synthesized through overexpression in *E. coli* as detailed above, while native RNA polymerase was purified from *P. furiosus*. Complexes were assembled in thin-walled PCR tubes to a volume of 25 μL, at 200 mM NaCl, 40 mM NaHEPES, 0.1 mM EDTA, 2.5 M MgCl$_2$, 5% glycerol, 5 mM β-mercaptoethanol, and 0.1 μg/μL bovine serum albumin, and with 30 μL mineral oil layered on top to prevent evaporation. Tubes were then placed open into a 65ºC heat block under a 365 nm UV lamp and incubated for 10 hours (Sato, et al., 2011).

Competitive Crosslinking Assays

To determine crosslinking intensity in the presence of a competitor, complexes were assembled in the buffer described above using 120 nM TBP, 240 nM TBP, 240 nM TFE (wild type or Bpa mutant), and 40 nM RNA polymerase assembled on 40 nM *P. furiosus* glutamate dehydrogenase promoter with endpoints at -95 and +85. 0.2 mM GTP, ATP, and CTP were added to enable elongation to the first U in sequence. Complexes were then heated to 65ºC for five minutes to ensure transcription initiation. 240 nM Spt4/5 (wild type or Bpa mutant) was then added, 25 μL reactions were topped with 30 μL mineral oil, and open tubes were placed in a 65ºC heat block underneath a 365 nm handheld UV lamp and incubated for 10 hours.

Crosslinked DNA Precipitation & Complex Denaturation

Crosslinked complexes were purified from solution using Ni-NTA magnetic Dynabeads, allowing isolation of transcription factors in solution containing 6x histidine tags. Complexes were mixed with 2 μL Dynabeads in 675 μL wash buffer (50 mM

NaPhosphate pH 8.0, 0.3 M NaCl, 0.01% TWEEN 20) and incubated at room

temperature for 10 minutes, then purified on a magnet. Beads were then washed with 70

µL wash buffer before being suspended in wash buffer with 8M urea added. Tubes were

incubated with 8M urea wash buffer at room temperature for 2 minutes before being

transferred to a 95ºC water bath for 45 seconds for further denaturation. Urea wash buffer

was then removed and denatured complexes and bound DNA were rinsed twice with

wash buffer.

Fluorescent Primer Extension

Crosslinked complexes and bound DNA, purified and attached to 6xHis-Tag

Dynabeads, were combined with 20 pmols Cy5-labeled glutamate dehydrogenase

promoter primer (forward primer for transcribed strand analysis or reverse primer for

nontranscribed strand analysis)(Table 3). Reactions were assembled using 4X Thermopol

buffer (New England Biolabs), 1 mM dNTPs, and 0.2 units Taq DNA polymerase (New

England Biolabs). Reactions were thermally cycled through the following program:

1. 95°C, 5 minutes

2. 95°C, 30 seconds

3. 48°C, 30 seconds

4. 72°C, 1 minute

5. 34 times to step 2

6. 72°C, 20 minutes

DNA was then purified by ethanol precipitation and dissolved in 4.5 µL formamide. Sequencing reactions were prepared using Affymetrix Thermosequenase Cycle Kit and cycled through the same program, followed by addition of 4 µL formamide. Dissolved DNA and corresponding sequencing reactions were then run on an 8M urea 10% polyacrylamide gel at 1250 V for approximately 75 minutes. Gels were scanned on a Typhoon Trio fluorescence scanner with a 670 nm laser, and relative pixel intensities of crosslinking regions were quantified in ImageQuant 1D Gel Analysis software.

Table 3. Primer sequences used for fluorescent primer extension.

	Sequence
Forward primer	5'-Cy5-TTGGCCAGCTCTTTCAAGTT-3'
Reverse primer	5'-Cy5-GAATTTTAGATTCTTTGAGCCTAATCA-3'

Transcription Assays

Transcription complexes were assembled *in vitro* using 40 nM DNA, 60 nM TBP, 120 nM TFB1, and 10 nM RNA polymerase assembled in buffer same as that used for crosslinking. Complexes were heated to 65ºC for 10 minutes prior to addition of 500 µM NTPs containing α32-P CTP. Complexes were heated to 65ºC for 20 minutes. Following transcription, an equal volume of urea stop buffer was added to each tube. Complexes were heated to 90°C and then run on urea polyacrylamide gels. Gels were then dried and exposed overnight to storage phosphor plate. Plates were then scanned on a Typhoon Trio phosphor scanner.

U-less Chase Assays

U-less chase assays were utilized to determine how timing of Spt4/5 incorporation impacts transcription output. Transcription complexes were assembled on 40 nM U-less GDH promoter DNA using 60 nM TBP, 120 nM TFB1, and 10 nM RNA polymerase. Complexes were heated to 65°C for five minutes to ensure complex formation. 500 μM GTP, ATP, and radiolabeled CTP were added, and complexes were again heated to 65°C to transcribe to uracil pause sites. 450 nM Spt4/5 was then added, followed by 500 μM UTP. Complexes were incubated at 65°C for 20 minutes before addition of an equal volume of urea stop buffer. Reactions were then heated to 90°C for 30 seconds and run on an 8 M urea 8% polyacrylamide gel at 850 V for approximately 40 minutes. Gels were dried and exposed overnight to storage phosphor plates. Plates were then scanned on a Typhoon Trio phosphor imager.

Results

TFE crosslinks to the nontranscribed strand

Archaeal transcription factor E binds RNA polymerase prior to transcription initiation and assists with promoter melting (Werner & Weinzierl, 2005). However, it is unknown where TFE is positioned relative to the template DNA and how its position might change as RNA is synthesized. To determine sites on TFE proximal to DNA, a crosslinkable unnatural amino acid was inserted at various positions in the protein, followed by protein to DNA crosslinking assays. Individual amino acids were selected from within the conserved patch of aromatic residues in TFE's winged helix, which was suspected to be proximal to DNA, as well as several residues distant from the tip of the winged helix but still on the more highly conserved side of the domain (Figure 1) (Meinhart, et al., 2003). Mutant proteins were synthesized as described in Methods to incorporate benzoylphenylalanine in place of an amino acid of interest. Crosslinking was induced by 10 hours of 365 nm UV exposure, as previous research indicates that 10 hour exposures result in maximum crosslinking efficiency without damaging protein (Sato, et al., 2011). After crosslinking, complexes were purified on Ni-NTA beads, proteins were denatured with urea and heat, and crosslinks were analyzed by fluorescent primer extension and gel electrophoresis.
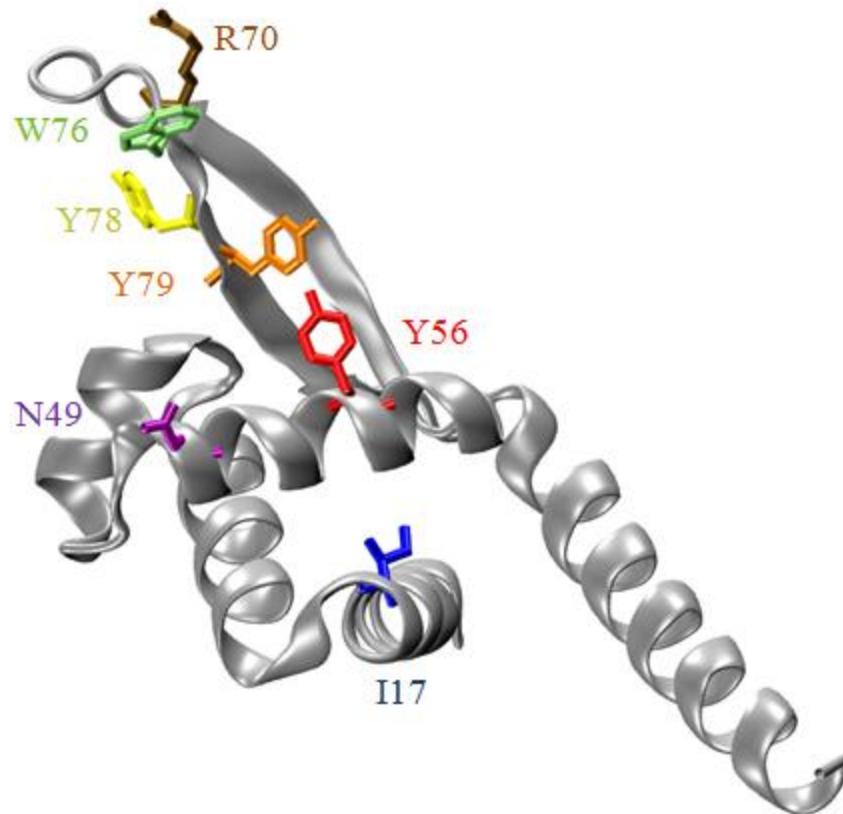
Figure 1. TFE winged helix and target amino acids for Bpa substitution. The TFE winged helix structure, with target amino acids for Bpa substitution emphasized by bond models (K. Murakami, personal communication). R70 is indicated in brown, W76 in green, Y78 in yellow, Y79 in orange, Y56 in red, N49 in purple, and I17 in blue.

Transcription assays were performed to confirm that UV exposure and 10 hour incubation did not negatively impact transcription activity (Figure 2). Complexes containing TBP, TFB, and RNAP on a GDHP promoter were incubated at 65°C for 10 hours in the presence of 365 nm UV prior to transcription (Figure 2, lane 1), resulting in similar levels of RNA synthesis as complexes incubated at 65°C for 10 hours without UV

(Figure 2, lane 2) and complexes which did not endure the long incubation or UV

exposure (Figure 2, lane 3).

| 65° C, 10 hours | + | + | - |
| 365 nm UV | + | - | - |

| | | |
| 1 | 2 | 3 |
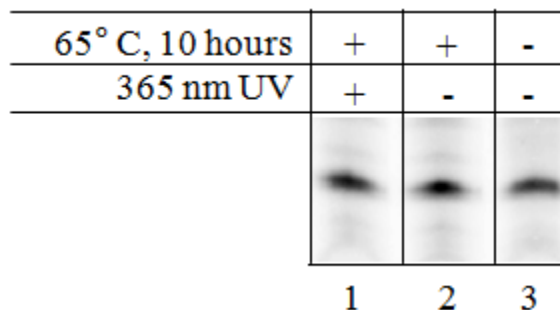
Figure 2. Transcription activity of UV-exposed complexes. Transcription assays were performed using

RNAP, TBP, and TFB on a GDHP template. Prior to addition of 32P-radiolabeled NTPs, complexes were

incubated for 10 hours at 65°C (lanes 1 & 2) with exposure to 365 nm UV (lane 1). Reactions were then

analyzed on an 8M urea 10% polyacrylamide gel. The gel was dried and exposed overnight to a storage

phosphor plate prior to imaging on a Typhoon Trio phosphor imager. The +85 runoff band, pictured, has

relative pixel volumes of 1.00 in lane 1, 1.00 in lane 2, and 0.97 in lane 3.

To gain a sense for the positioning of TFE relative to DNA prior to initiation, TFE

was added to transcription complexes (TBP, TFB, RNAP, and GDH promoter DNA) in

the absence of NTPs (Figure 3). Crosslinks were induced with 10 hour UV exposure and

purified and detected as described in Methods, with primer extension performed on the

nontranscribed strand. As expected, the wild type TFE control, lacking any Bpa

substitutions, did not produce crosslinks to DNA (Figure 3, lane 1). TFE with Bpa

substituted for W76 did produce two crosslinks to DNA at -10/-11 base pairs (Figure 3,

lane 2). Bpa substituted for R70 did not crosslink (Figure 3, lane 3). The TFE Y78

substitution did crosslink in approximately the same region as W76, but produced fainter

25

bands (Figure 3, lane 4). Substitutions in I17 and Y56 did not produce crosslinks (Figure 3, lanes 5 & 6). N49 substitutions crosslinked to DNA at -11/-12 base pairs (Figure 3, lane 7). Y79 substitutions did not crosslink (Figure 3, lane 8). Of the amino acids tested in TFE, Bpa substitutions in W76, Y78, and N49 were the only mutants capable of crosslinking to the nontranscribed strand in pre-initiated complexes. W76, located in the conserved aromatic patch of TFE's winged helix and near the end of β strand 3, and N49, located apart from the winged helix tip in α helix 3, produce the most intense bands. While Y78 is also located in the aromatic patch in the winged helix in β strand 3 (Meinhart, et al., 2003), it produces weaker bands when substituted with Bpa.
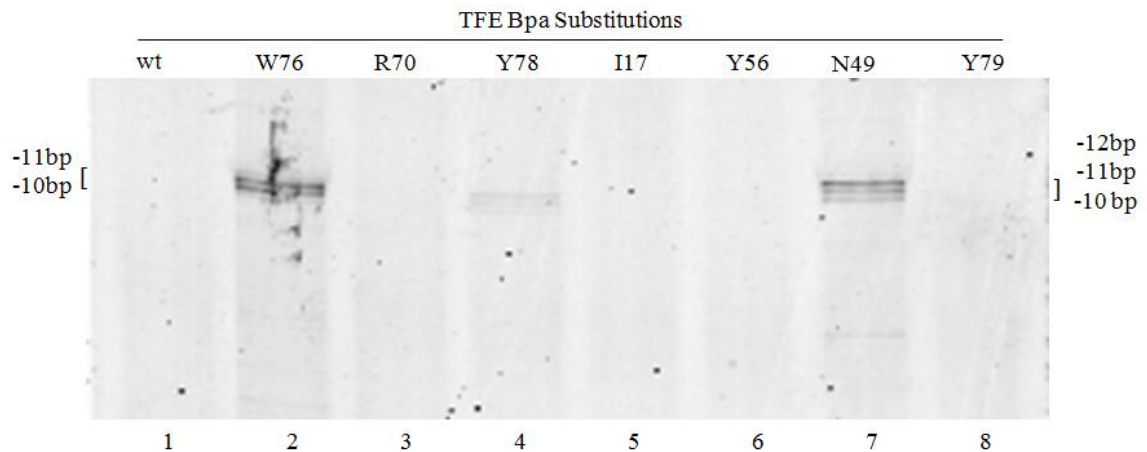
Figure 3. Crosslinking to the nontranscribed DNA strand in pre-initiated complexes by TFE Bpa variants. Transcription complexes were assembled *in vitro* using *P. furiosus* RNA polymerase, TBP, TFB, and TFE variants with Bpa substitutions at the listed positions. The *P. furiosus* glutamate dehydrogenase promoter with endpoints of -95 and +85 relative to the transcription start site was used as the template. Complexes were crosslinked using 365 nm UV exposure for 10 hours. His-tagged transcription factors were then purified with Ni-NTA beads and denatured with urea and heat prior to primer extension with a reverse Cy5-labeled primer for nontranscribed strand replication. Fluorescent DNA was then visualized on

an 8M urea 10% polyacrylamide gel and scanned with a Typhoon Trio imager. Crosslink positions, indicated at left and right, were determined based on position relative to sequencing reactions in other experiments (Figures 4 & 5).

To determine if crosslinks change as transcription proceeds, transcription complexes were assembled with TFE Bpa mutants on the GDH promoter, and transcription was initiated by selective NTP addition (see Methods for details). Complexes were supplied with G, A, and CTP to allow transcription to proceed to the first uracil incorporation, located at the +6 position in wild type GDHP, resulting in complexes halted after transcription of five nucleotides. After heating complexes to allow transcription to the halt site, proteins were crosslinked by UV exposure as described previously, and crosslink positions in DNA were identified by primer extension. In the absence of nucleotides, W76 crosslinks to the nontranscribed strand in pre-initiated (+0 nt) complexes, at -10/-11 base pairs (Figure 4, lane 2). Next, transcription initiation was permitted by the addition of G, A, and CTP, allowing formation of a five nucleotide transcript. W76 crosslinks extend further downstream relative to DNA, approximately five base pairs, in tandem with the complex transcribing five nucleotides (Figure 4, lane 1). These data position the winged helix of TFE near the upstream fork of the nontranscribed strand.
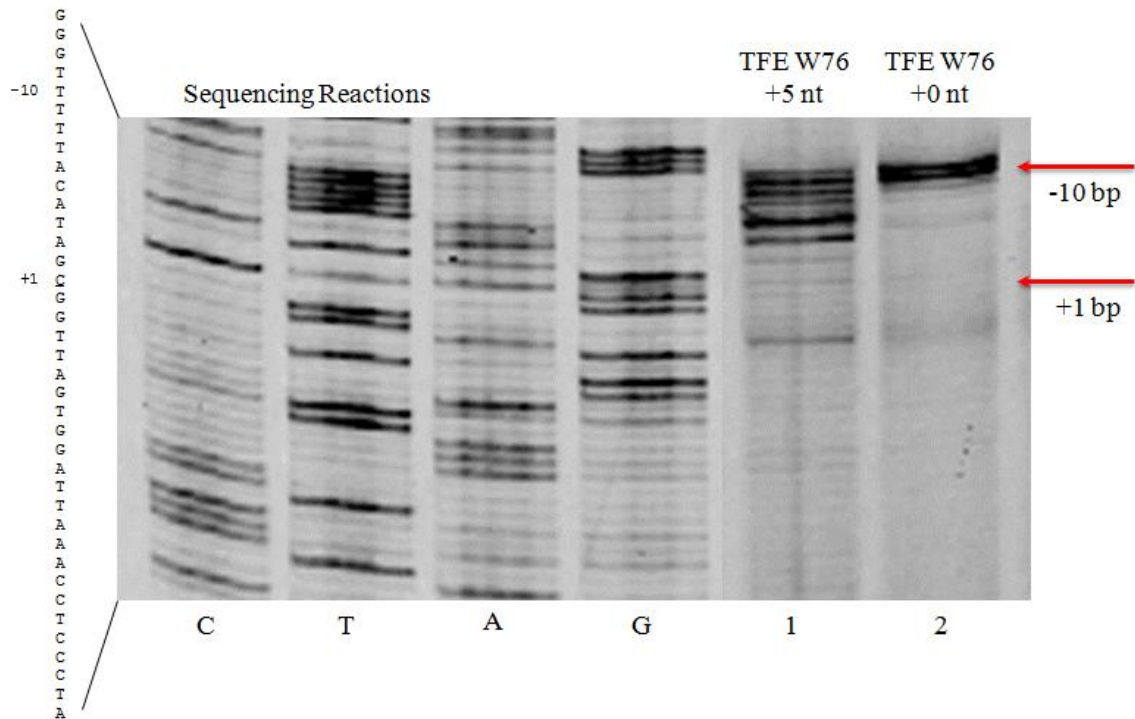
Figure 4. TFE W76 crosslinks to the nontranscribed strand in pre-initiation and initiated complexes. Complexes were assembled *in vitro* as described previously, using TFE mutants with Bpa substituted in place of W76 in the winged helix tip. In lane 1, transcription was initiated using G, A, and CTP to allow transcription to the first uracil incorporation in the GDH promoter sequence at +6, resulting in a halted complex after transcription of five nucleotides. Reactions in lane 2 were not supplied with NTPs. Proteins were then crosslinked, purified, and denatured for primer extension, and visualized on an 8M urea 10% polyacrylamide gel prior to imaging. Template sequence is indicated at left with the transcription start site underlined.

To examine TFE movement relative to DNA more closely, U-less GDHP templates were made to allow formation of initiation complexes halted at +6, +7, +8, +9, and +10 nucleotides (see Methods for details and Figure 9 for partial sequences). Complexes were also halted at +3 using the wild type GDHP sequence with the addition of only G and CTP. W76 mutants crosslink to the nontranscribed strand in pre-initiated

28

complexes as observed previously (Figure 5, lane 1). Synthesis of three nucleotides does not appear to change the location of W76 crosslinks (Figure 5, lane 2). W76 crosslinks do begin to move relative to DNA after synthesis of five nucleotides, extending approximately five nucleotides downstream from the initial crosslinking sites (Figure 5, lane 3). Crosslinks positions appear to remain unchanged as elongation proceeds to six nucleotides (Figure 5, lane 4). W76 crosslinks move again once a 7-mer has been synthesized, now positioned 6 base pairs downstream of PIC crosslinks (Figure 5, lane 5). W76 crosslinks appear to increase in intensity at +8 nucleotides transcribed (Figure 5, lane 6), and then decrease as the unnatural amino acid loses proximity to the nontranscribed strand of DNA after synthesis of nine nucleotides (Figure 5, lane 7). Crosslinks are almost entirely absent by the time the complex reaches +10 nucleotides transcribed (Figure 5, lane 8). W76 crosslinks do not move absolutely, instead creating a spread of crosslinks, which may be indicative of some crosslinking occurring in complexes performing abortive transcription that have not fully reached the intended pause site, or of flexibility in the TFE winged helix relative to the nontranscribed strand in these complexes.

Figure 5. TFE W76 crosslinks to the nontranscribed strand throughout early transcription steps. TFE W76 crosslinks to the NTS throughout early transcription. Complexes were assembled as described previously, using TFE mutants with Bpa in place of W76. U-less variants of the *P. furiosus* glutamate dehydrogenase promoter were used as templates to allow complex pausing at specific sites when supplied with G, A, and CTP (see Methods for details). Template sequence is indicated next to sequencing reactions.

To investigate if other regions of TFE behave similarly to the surface containing W76, similar crosslinking assays were performed with other Bpa substitutions. Y78, located proximal to W76 in β strand 3 of the winged helix, was suspected to display similar crosslinking behavior relative to the nontranscribed strand. Transcription complexes were assembled on U-less cassettes as described previously, in the presence of TFE with Bpa substituted for Y78. Y78 in pre-initiated complexes crosslinked as observed previously, producing faint crosslinks at -11/-12 base pairs (Figure 6, lane 7). After initiation with G, A, and CTP permitting transcription of five nucleotides, Y78 crosslinks appear to extend one more nucleotide downstream relative to PIC crosslinks (Figure 6, lane 6). Transcription of six nucleotides results in Y78 crosslinks extending further, now encompassing a total of four nucleotides from -12 to -9 (Figure 6, lane 5). After this stage, Y78 appears to begin to lose proximity to the nontranscribed strand,

crosslinking to only three nucleotides in +7, +8, and +9 complexes (Figure 6, lanes 4, 3, and 2). Y78 bands appear to decrease in intensity after +6 nucleotides transcribed, until they are nearly undetectable in complexes that have synthesized 10 nucleotides (Figure 6, lane 1).
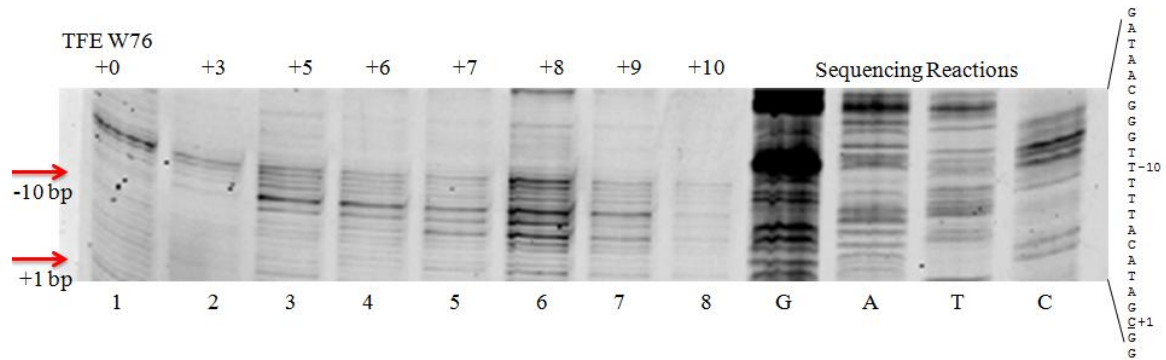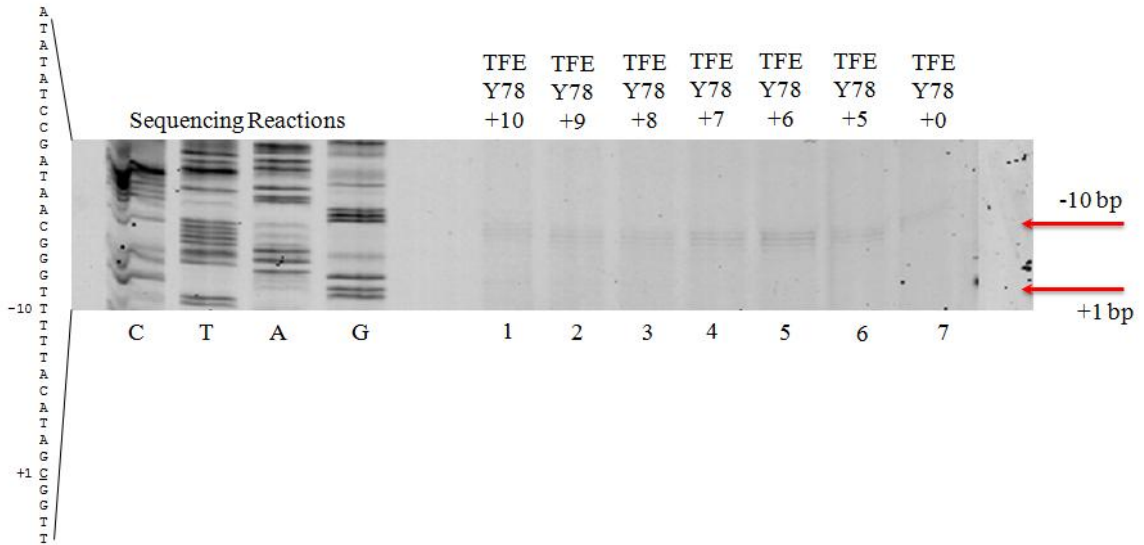

Figure 6. TFE Y78 crosslinks to the nontranscribed strand throughout early transcription steps. TFE Y78 crosslinks to the NTS throughout early transcription. Complexes were assembled using the same templates as in Figure 5, with Y78 Bpa substitutions in TFE. Proteins were crosslinked, purified, and denatured for primer extension prior to visualization as described above. Template sequences are indicated next to sequencing reactions.

While Y78 was observed to move relative to DNA during RNA synthesis in a similar pattern to W76, N49 is located distant from these two amino acids and may behave differently. N49 is part of α helix 3 on the conserved surface of the TFE N-terminal winged helix (Figure 1)(Meinhart, et al., 2003). TFE N49 Bpa substitutions crosslink to the nontranscribed strand at -11/-12 base pairs in pre-initiated complexes

31

(Figure 7, lane 1). Transcription was then initiated through the addition of G, A, and CTP

to allow synthesis of five nucleotides prior to crosslinking. In these initiated complexes,

N49 crosslinks do not move relative to PIC complex crosslinks (Figure 7, lane 2). These

data position N49 at the upstream fork of the transcription bubble, slightly upstream of

W76. While the surfaces containing W76 and Y78 do move throughout transcription

initiation, the surface containing N49 did not move relative to DNA at any point between

+5 and +10 nucleotides (Figure 9 and data not shown).

Figure 7. TFE N49 crosslinks to the nontranscribed strand pre- and post-initiation. Complexes were

assembled using *P. furiosus* RNA polymerase, TBP, TFB, and TFE with Bpa substituted for N49. *P.*

furiosus glutamate dehydrogenase promoter DNA was used as template. In lane 2, transcription was

initiated prior to crosslinking by the addition of G, A, and CTP to allow transcription up to the first uracil

incorporation at +6 bp, resulting in halted complexes after transcription of 5 nucleotides. Complexes were

then crosslinked through 10 hour exposure to 365 nm UV light, purified on Ni-NTA beads and denatured

with urea and heat. Primer extension was then performed using reverse Cy5 labeled primer to visualize

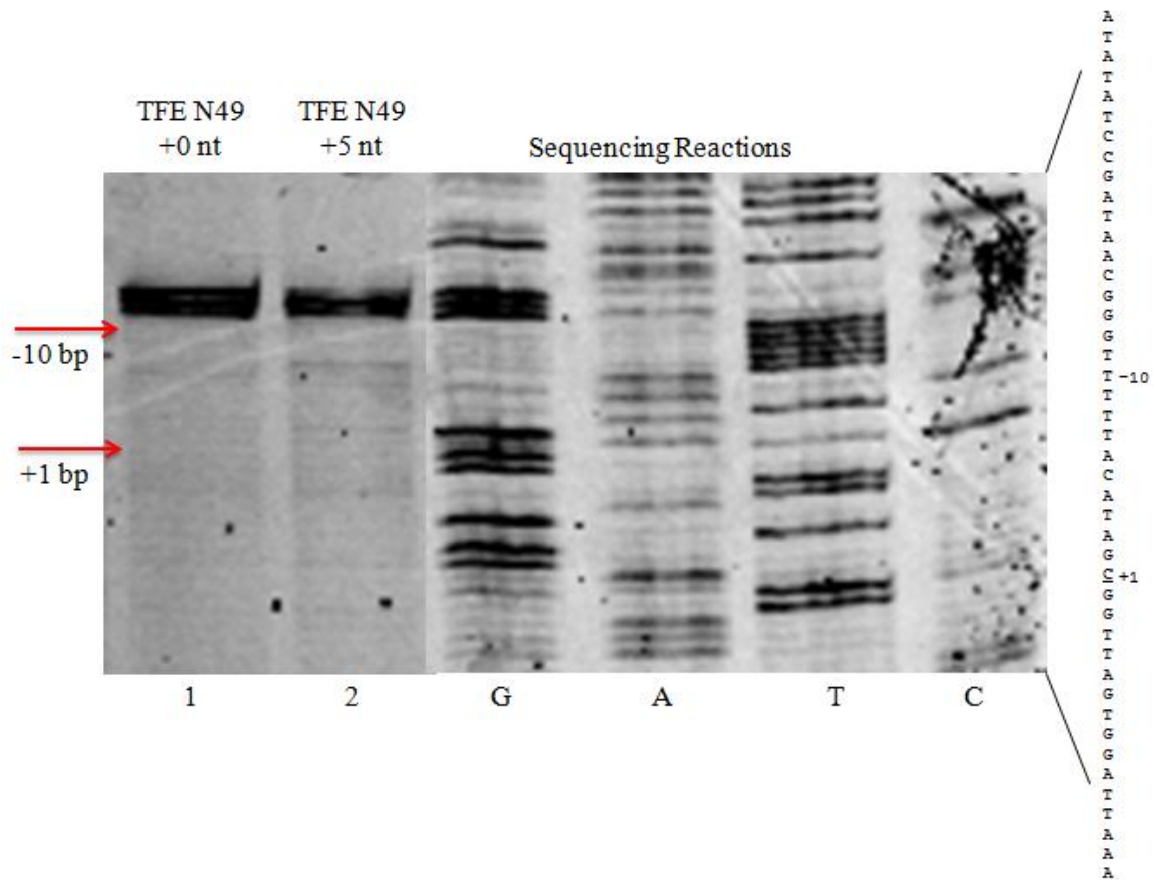crosslinks on the nontranscribed strand. Fluorescent DNA was then purified through ethanol precipitation,

dissolved in formamide, and run on an 8M urea 10% polyacrylamide gel prior to imaging on a Typhoon

Trio scanner. Template sequence is indicated to the right with the transcription start site underlined.

TFE crosslinks to the transcribed strand after transcription initiation.

Previous experiments indicated that no tested Bpa substitutions in TFE (Figure 1)

crosslink to the transcribed (T) strand in preinitiated complexes (Figure 9 & data not

shown). However, regions of the winged helix were observed to have a degree of

mobility after transcription initiation, which may bring crosslinking surfaces closer to the

transcribed strand after initiation. To determine if crosslinks to the transcribed strand

might appear following initiation, complexes were formed with TFE W76 or N49 Bpa

substitutions on U-less cassettes. Transcription was initiated with G, A, and CTP, and

crosslinks were induced with UV exposure. Once the complex has transcribed +5

nucleotides, W76 crosslinks appear faintly at -8/-9/-10 base pairs, consistent with the

NTS crosslinks previously observed near the upstream fork (Figure 8A, lane 1, compare

with figure 5). Crosslinks do not move relative to DNA as transcription proceeds,

however they do appear to increase in intensity as the complex synthesizes six and seven

nucleotides (Figure 8A, lanes 2 & 3). As elongation continues, crosslinks appear to

decrease in proximity in complexes halted at +8, +9, and +10 (Figure 8A, lanes 4-6).

Assays performed with N49 Bpa substitutions produce similar crosslinking patterns. In

complexes that have synthesized five nucleotides, N49 does not appear to produce any

distinct crosslinks (Figure 8B, lane 1). Crosslinks appear once the complex has

transcribed six nucleotides, in the -11/-12 base pair range, similar to nontranscribed

strand crosslinks (Figure 8B, lane 2). These crosslinks do not move as transcription

proceeds but do appear to increase in intensity in complexes halted at +7 (Figure 8B, lane

3). Crosslinking bands then appear to decrease in intensity, without moving relative to

DNA, in complexes halted at +8, +9, and +10 nucleotides (Figure 8B, lanes 4-6). In both

cases, TS crosslinks do not move relative to DNA as transcription proceeds. Both N49

and W76 crosslinks are strongest at +7 nucleotides transcribed.
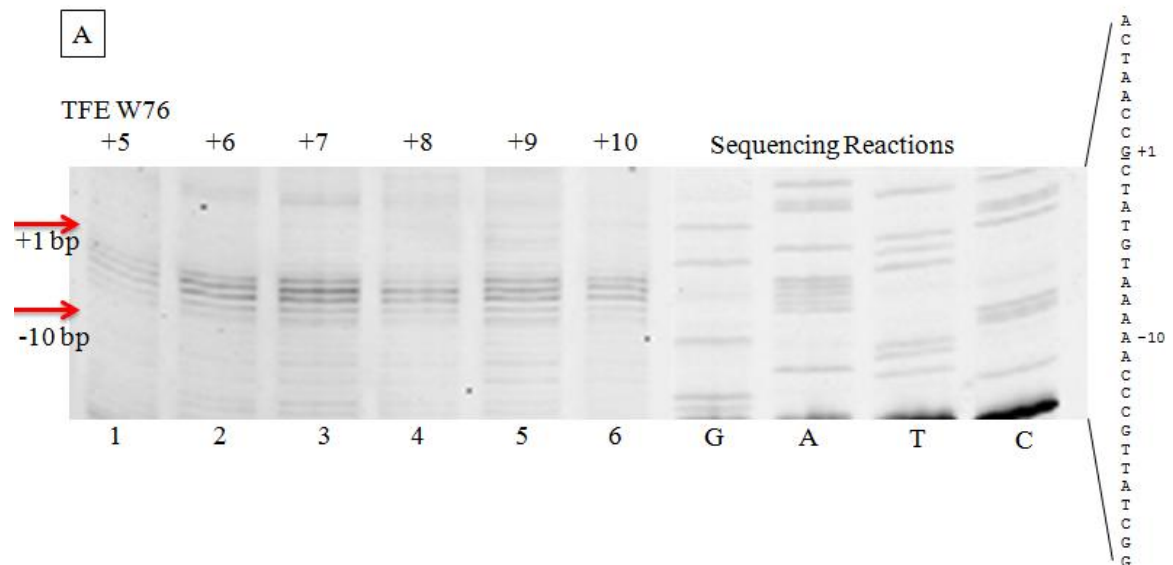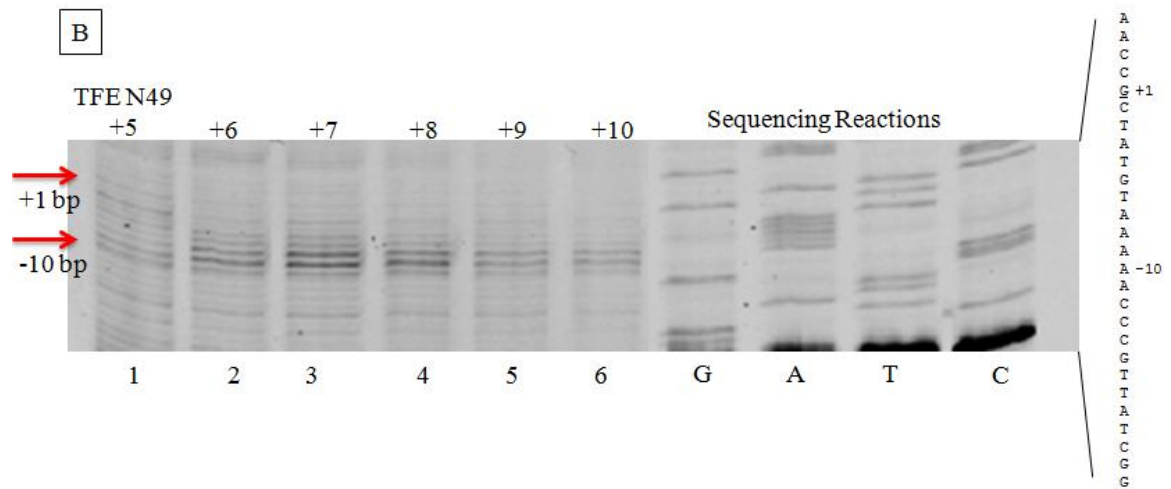
Figure 8. TFE crosslinks to the transcribed strand in initiated complexes. (A) *In vitro* transcription complexes were assembled with *P. furiosus* RNA polymerase, TBP, TFB, and mutant TFE with Bpa substituted for W76. U-less variants of the glutamate dehydrogenase promoter were used as template (see Methods for details). Transcription was initiated with the addition of G, A, and CTP, allowing for transcription up to the first uracil incorporation and resulting in complexes halted at the indicated step. Complexes were then crosslinked using 10 hour exposure to 365 nm UV, purified on Ni-NTA beads, and denatured with urea and heat prior to primer extension using a forward Cy5-labeled primer for visualization of transcribed strand crosslinks. Fluorescent DNA was visualized on an 8M urea 10% polyacrylamide gel using a Typhoon Trio imager. Template sequence is indicated to the right with the transcription start site underlined. (B) Complexes were assembled as in A, with N49 Bpa substitutions in TFE. Crosslinks were purified and detected as described above.

+0
```
                                    aatgtatcGcc
NT strand   ctttatataggctattgcccaaa              aatcacctaatttggag

 T strand   gaaatatatccgataacgggtttt    ● ttagtggattaaacctc
                                    ttacatagCgg
```

+5
```
                                    atgtatcGccaa
NT strand   ctttatataggctattgcccaaaa             tcacctaatttggag

 T strand   gaaatatatccgataacgggtttt    ● agtggattaaacctc
                                    tacatagCggtt
```

+6
```
                                    atgtatcGccaaa
NT strand   ctttatataggctattgcccaaaa             tacctaatttggag

 T strand   gaaatatatccgataacgggtttt    ● atggattaaacctc
                                    tacatagCggttt
```

+7
```
                                    atgtatcGccaaactc
NT strand   ctttatataggctattgcccaaaa             ctaatttggag

 T strand   gaaatatatccgataacgggtttt    ● gattaaacctc
                                    tacatagCggtttgag
```

+8
```
                                    tgtatcGccaaacatc
NT strand   ctttatataggctattgcccaaaaa            taatttggag

 T strand   gaaatatatccgataacgggttttt   ● attaaacctc
                                    acatagCggtttgtag
```

+9
```
                                    tgtatcGccaaacactt
NT strand   ctttatataggctattgcccaaaaa            aatttggag

 T strand   gaaatatatccgataacgggttttt   ● ttaaacctc
                                    acatagCggtttgtgaa
```

+10
```
                                    tatcGccaaacacc
NT strand   ctttatataggctattgcccaaaaatg          taatttggag

 T strand   gaaatatatccgataacgggtttttac ● attaaacctc
                                    atagCggtttgtgg
```
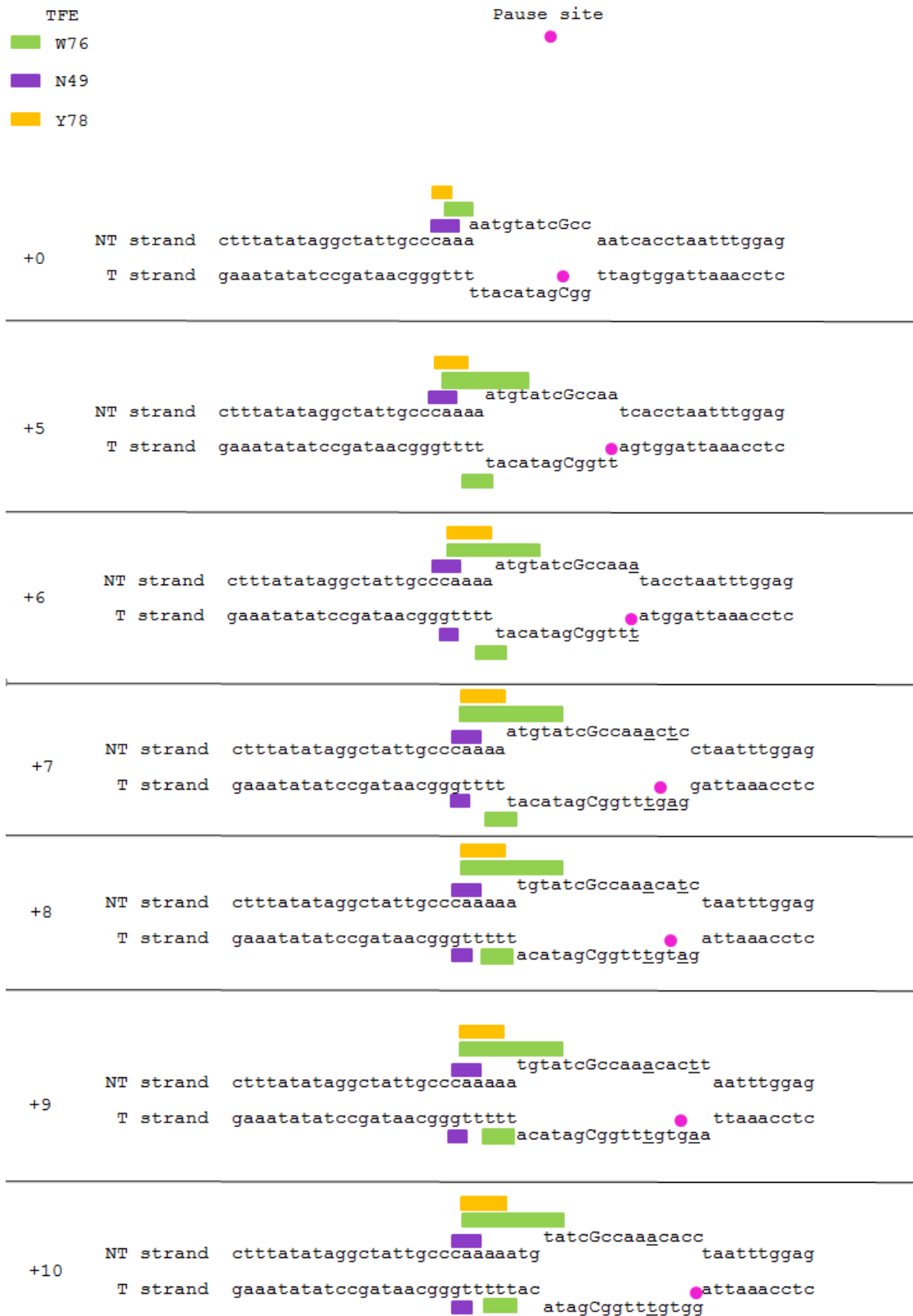
36

Figure 9. Position of TFE crosslinks relative to the open transcription bubble. Dimensions of the transcription bubble, as measured on the *P. furiosus* glutamate dehydrogenase promoter by Spitalny & Thomm (2003), with crosslinking positions for TFE Bpa substitutions indicated by colored bands. W76 crosslinks are in green, N49 in purple, and Y78 in yellow. Transcription start site is indicated by a capital C, and RNA polymerase pause site in U-less reactions is indicated by a pink circle. Mutated bases in U-less cassettes are underlined.

Spt5 crosslinks to both strands of DNA after initiation

Spt4/5 is a heterodimeric transcription elongation factor shown to promote RNA polymerase translocation and processivity and is suspected to be essential in archaea, though it is unknown where Spt5 associates with DNA in *P. furiosus*. Spt5 has been shown to enhance transcription elongation, but inhibits transcription initiation if incorporated into pre-initiation complexes (Grohmann, et al., 2011). This dramatic shift in function raises the question of how Spt5's interactions with DNA change as its function changes, as well as when this shift occurs.

To test this, *P. furiosus* Spt5 mutants were created with substitutions in the surfaces likely to be close to DNA. Bpa substitution positions were selected based on hypothesized proximity to DNA in existing models (Figure 10). Of these, H65, T11, and Q14 are clustered on an Spt5 surface predicted to be near double-stranded DNA at the upstream side of the transcription bubble. P71 is located further away from this cluster but near the nontranscribed strand in the transcription bubble. To examine proximity of these positions of Spt5 to DNA, pre-initiation complexes were made in the presence of the mutant Spt4/5 heterodimer. In pre-initiation complexes, none of these mutants

produced crosslinks to either strand of DNA (Figure 11A lane 1, 11C lane 1, 11 D lane 1, and data not shown). This is consistent with results showing that Spt4/5 added to PIC inhibits transcription, possibly by restriction of DNA binding (Grohmann, et al., 2011). To test if these regions of Spt5 approach DNA following initiation, the mutant Spt4/5 heterodimers were added to complexes allowed to initiate and transcribe short transcripts. In initiated transcription complexes, H65 and Q14 crosslink exclusively to the transcribed strand, while T11 and P71 crosslink exclusively to the nontranscribed strand.
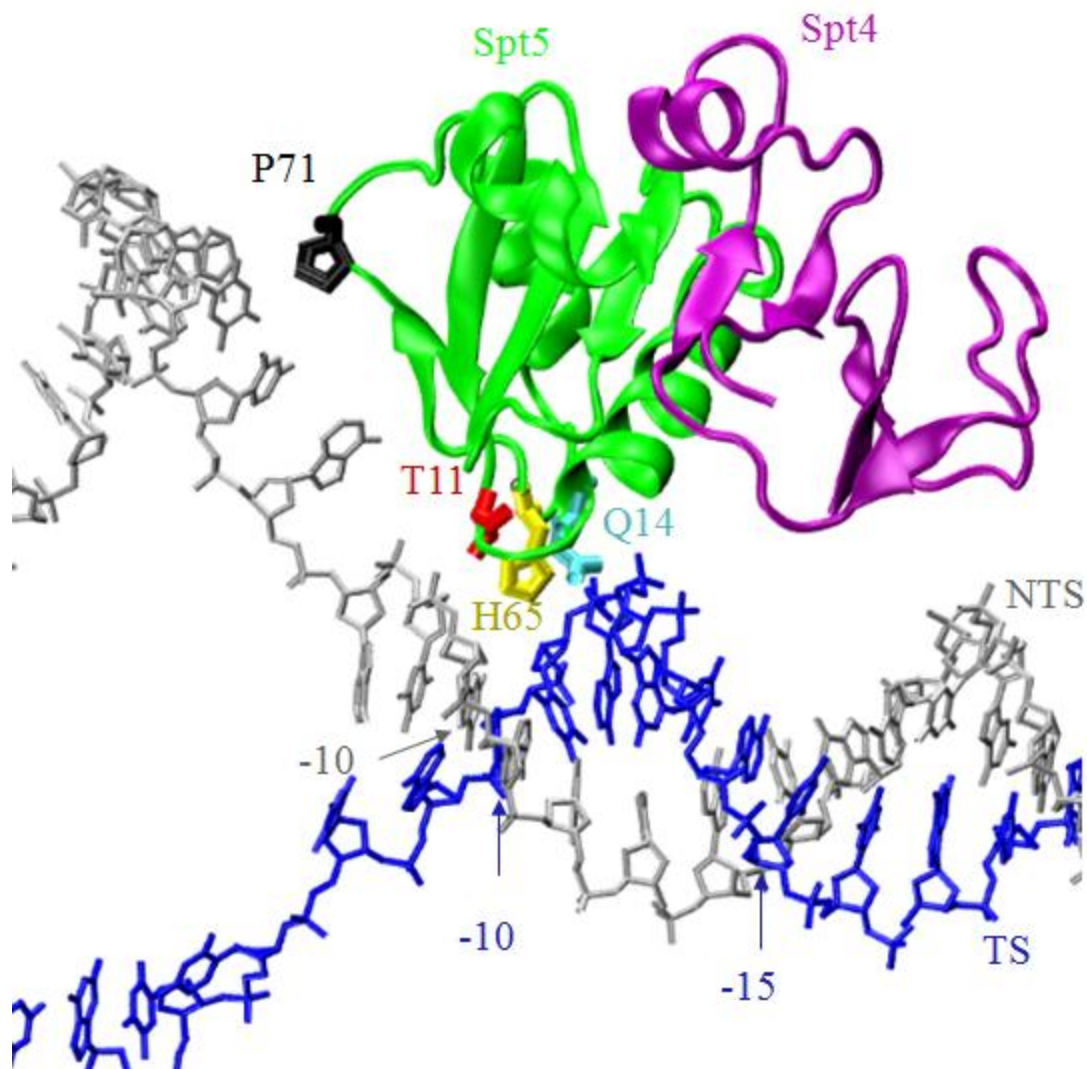
Figure 10. Spt4/5 structure and amino acids selected for Bpa substitution. *P. furiosus* Spt4/5 and DNA model provided by K. Murakami, with Spt4 in purple and Spt5 in green. Amino acids selected for Bpa substitution are indicated with P71 in black, T11 in red, H65 in yellow and Q14 in blue. The nontranscribed strand is modeled in grey, and the transcribed strand in blue. DNA positions relative to the nucleotide being transcribed at the active site (+1, not shown) are indicated.

The Spt5 H65 Bpa mutant crosslinks to the transcribed strand of DNA following initiation (Figure 11A). The crosslink positions extend from +1 base pairs to +7, and they

do not change at any point between the +5 and +10 complex (Figure 11A, lanes 2 & 3, Figure 13, and data not shown). H65 does not crosslink to the nontranscribed strand in these complexes (Figure 13 & data not shown). Q14 also crosslinks to the transcribed strand after initiation, from +3 to -7 base pairs (Figure 11B). These crosslinks do not move at any point between synthesis of five and ten nucleotides (Figure 11B, lanes 1 & 2, Figure 13, and data not shown). Q14 does not crosslink to the nontranscribed strand (Figure 13 and data not shown).

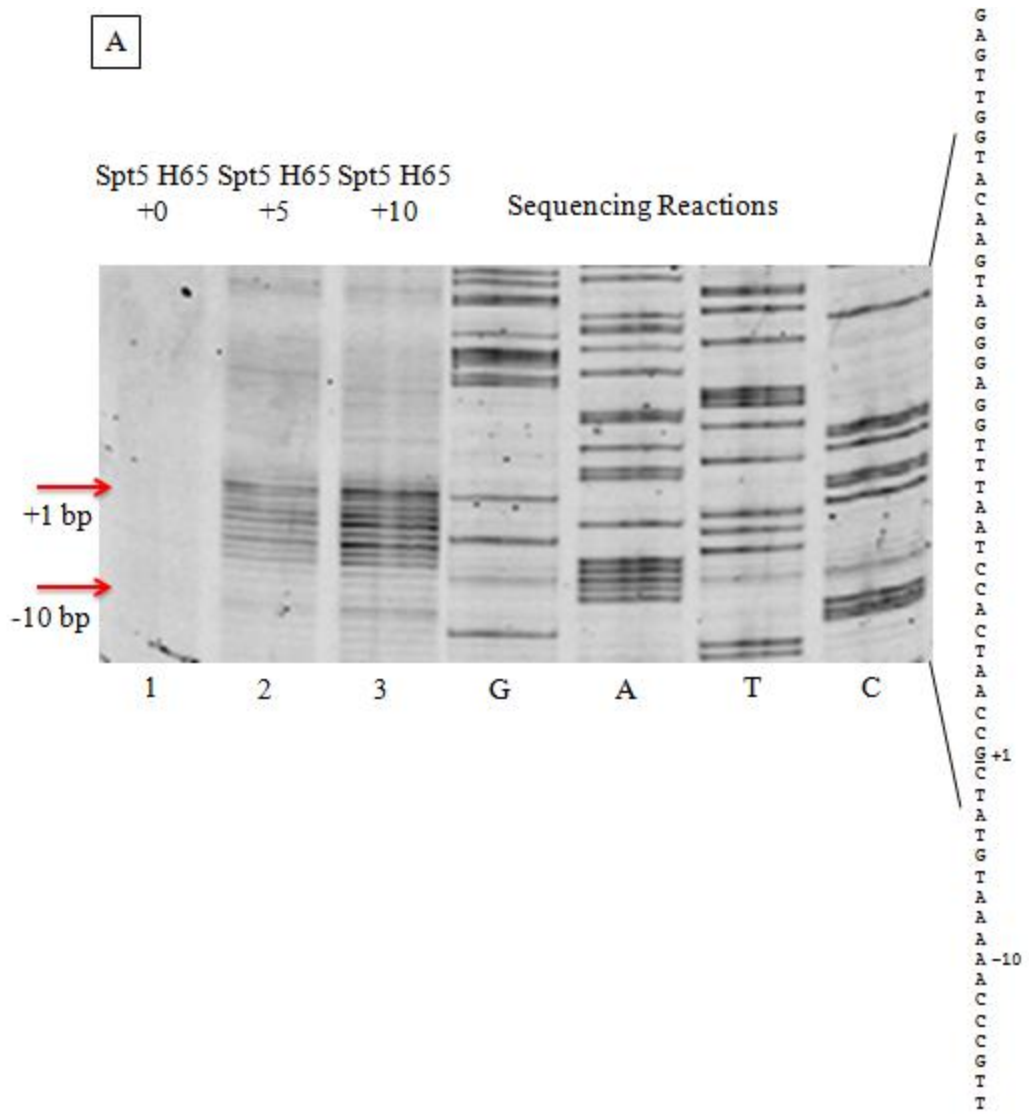T11 crosslinks to the nontranscribed strand in a similar pattern after transcription initiation, ranging from +4 base pairs to -4 base pairs (Figure 11C). These crosslinks do not move relative to DNA throughout initiation (Figure 11C, lanes 2 & 3, Figure 13, and data not shown). T11 does not crosslink to the transcribed strand at any point between pre-initiation and synthesis of 10 nucleotides (Figure 13 and data not shown). P71, located distant from the other three crosslinkers in Spt5, behaves differently. At +5 nucleotides transcribed, P71 crosslinks to the nontranscribed strand solely at +7 bp (Figure 11D, lane 2). As transcription continues, P71's crosslinks change patterns to appear across a wide range in a triplet pattern (Figure 12, lanes 3-5), downstream of the transcription start site. P71 crosslinks disappear once the complex reaches +9 nucleotides transcribed (Figure 12, lanes 1 & 2). P71 does not crosslink to the transcribed strand (Figure 13 and data not shown).

With the exception of P71, all tested Spt5 Bpa substitutions crosslink to a wide range of nucleotides on one strand of DNA (Figure 13). These crosslinks do not move relative to DNA as transcription proceeds, instead crosslinking to the same base pairs at

all observed stages of transcription. H65, Q14, and T11 crosslinks are proximal to each other, as expected from their proximity in the model (Figure 10).

G
A
G
T
T
G
G
T
A
C
A
A
G
T
A
G
G
A
G
G
T
T
T
A
A
T
C
C
A
C
T
A
A
C
C
G +1
C
T
A
T
G
T
A
A
A
A –10
A
C
C
C
G
T
T

A

Spt5 H65   Spt5 H65   Spt5 H65
  +0   +5   +10   Sequencing Reactions

→ +1 bp

→ –10 bp

1   2   3   G   A   T   C

C

Spt5 T11
+0

Spt5 T11
+5

Spt5 T11
+10

Sequencing Reactions

-10 bp

+1 bp

1   2   3   G   A   T   C

A
A
C
G
G
G
T
T
T   -10
T
A
C
A
T
A
G
C   +1
G
G
T
T
A
G
T
G
G
A
T
T
A
A
A
C
C
T
C
C

D

Spt5 P71
+0

Spt5 P71
+5

Spt5 P71
+10

Sequencing Reactions

-10 bp

+1 bp

1   2   3   G   A   T   C

G
T
T   -10
T
T
T
A
C
A
T
A
G
C   +1
G
G
T
T
A
G
T
G
G
A
T
T
A
A
A
C
C
T
C
C
C
T

44

Figure 11. Spt5 crosslinks to DNA throughout early transcription. Complexes were assembled with *P. furiosus* RNA polymerase, TBP, and TFB, with wild type glutamate dehydrogenase promoter as the template for +0 and +5 complexes and a mutant U-less GDH promoter for +10 complexes (see Methods for details). Transcription was then initiated in +5 and +10 lanes using G, A, and CTP to allow transcription to proceed five or ten nucleotides, respectively. After tra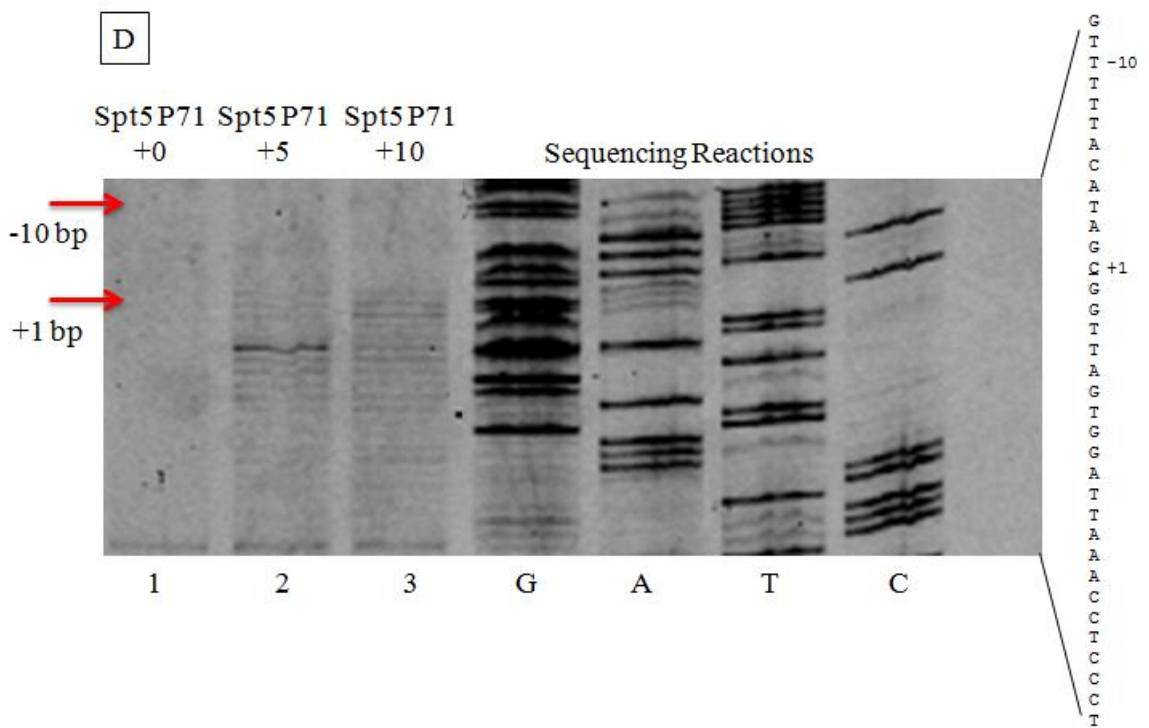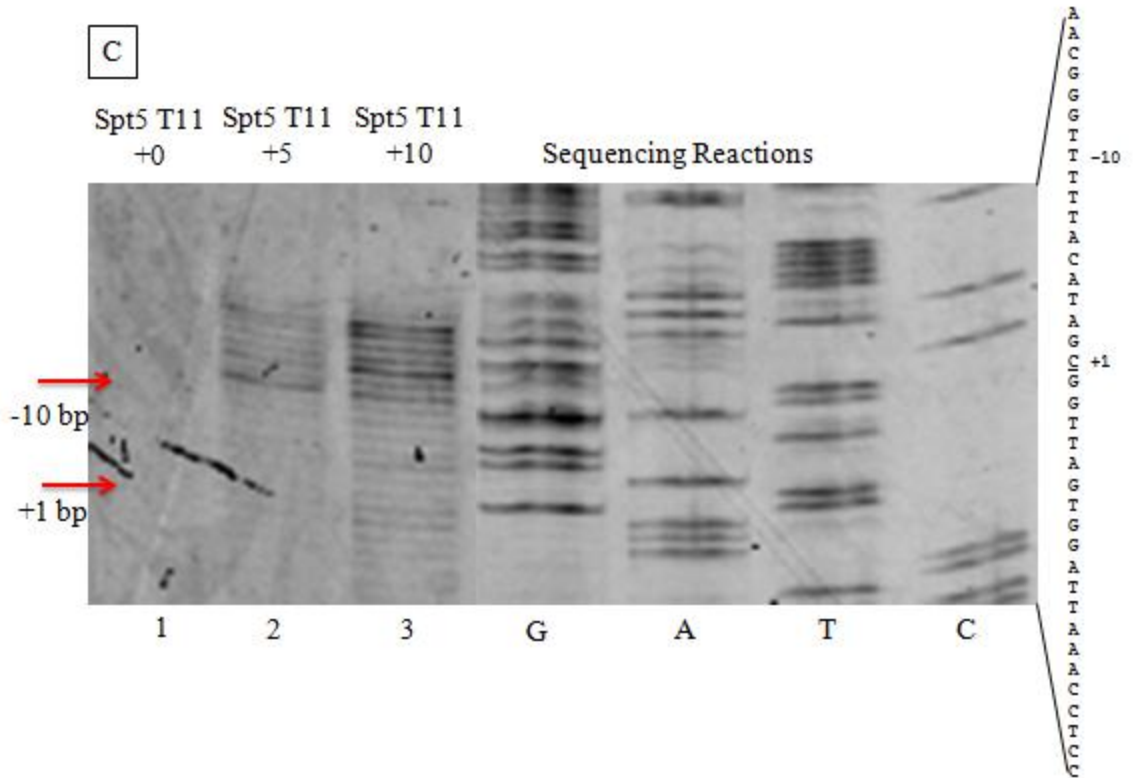nscription was initiated with NTP addition, mutant *P. furiosus* Spt4/5 with Bpa substituted for H65 in panel A, Q14 in panel B, T11 in panel C, and P71 in panel D in Spt5 was added to complexes. Proteins were crosslinked through 10 hour exposure to 365 nm UV. Proteins were then captured using Ni-NTA beads and denatured with urea and heat prior to primer extension. A Cy5-labeled forward primer was used to replicate the transcribed strand in panels A & B, and the nontranscribed strand in panels C & D. After primer extension, fluorescent DNA was ethanol precipitated and dissolved in formamide before visualization in an 8M urea 10% polyacrylamide gel using a Typhoon Trio phosphorimager. Template sequence is indicated at right with transcription start site underlined. (A) H65 crosslinks to the transcribed strand in pre-initation (lane 1) and initiated complexes (lanes 2 & 3). (B) Q14 crosslinks to the transcribed strand in initiated complexes. (C) T11 crosslinks to the nontranscribed strand in pre-initiation (lane 1) and initiated complexes (lanes 2 & 3). (D) P71 crosslinks to the nontranscribed strand in pre-initiation (lane 1) and initiated complexes (lanes 2 & 3).
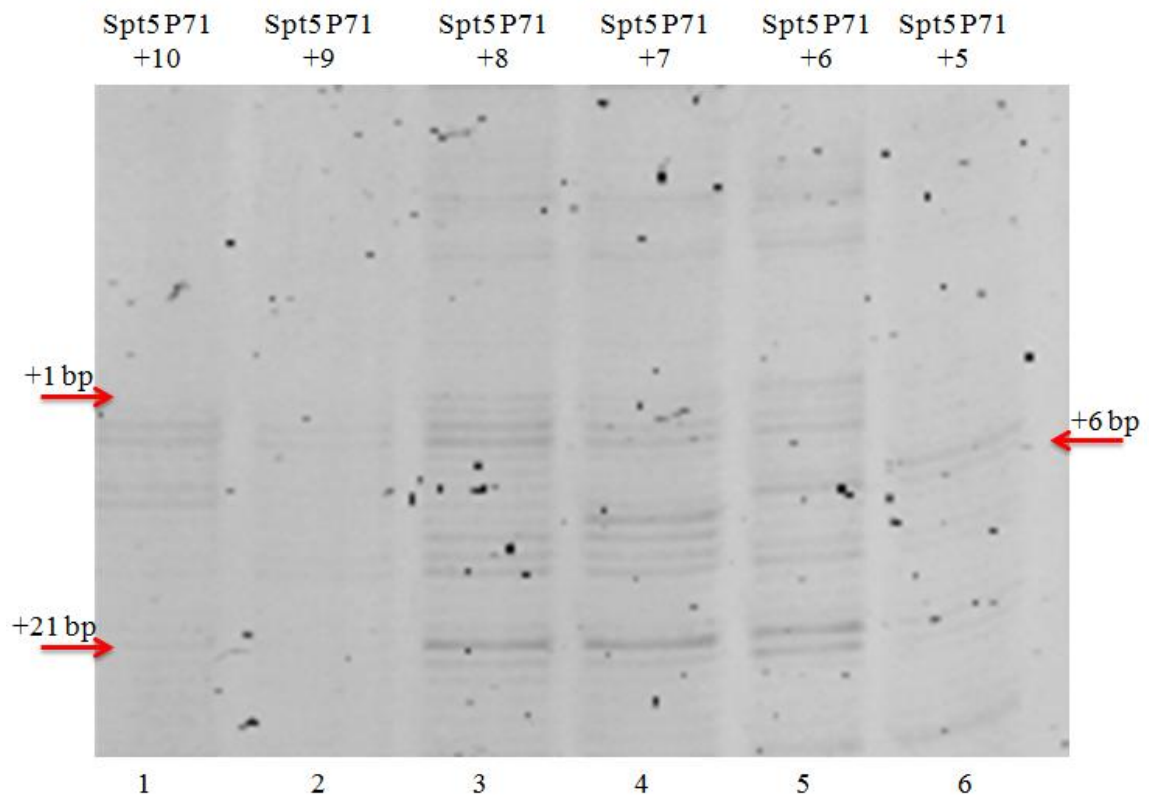
Figure 12. Spt5 P71 crosslinks to the nontranscribed strand throughout early transcription.

Complexes were assembled as described previously, using U-less glutamate dehydrogenase promoter templates (see Methods for details). After initiation with select NTPs, mutant Spt4/5 with Bpa substituted for Spt5 P71 was then added and crosslinked to DNA by 10 hour exposure to 365 nm UV light. Crosslinked proteins were purified and processed as described previously and fluorescent primer extension was performed on the nontranscribed strand. Fluorescent DNA was purified as described previously and analyzed on an 8M urea 10% acrylamide gel, visualized on a Typhoon Trio imager. Crosslink positions were determined by comparison to +5 crosslinks observed in Figure 11D.

Spt5   Pause site

■ T11

■ Q14

■ H65

■ P71

              aatgtatcGcc
+0 NT strand ctttatataggctattgcccaaa   aatcacctaatttggag
   T strand gaaatatatccgataacgggttt ● ttagtggattaaacctc
              ttacatagCgg

              atgtatcGccaa
+5 NT strand ctttatataggctattgcccaaaa   tcacctaatttggag
   T strand gaaatatatccgataacgggtttt ● agtggattaaacctc
             tacatagCggtt

             atgtatcGccaaa
+6 NT strand ctttatataggctattgcccaaaa   tacctaatttggag
   T strand gaaatatatccgataacgggtttt ● atggattaaacctc
             tacatagCggtt<u>t</u>

             atgtatcGccaa<u>a</u>ctc
  NT strand ctttatataggctattgcccaaaa   ctaatttggag
   T strand gaaatatatccgataacgggtttt ● gattaaacctc
+7   tacatagCggttt<u>tg</u>ag

             tgtatcGccaa<u>a</u>ca<u>t</u>c
+8 NT strand ctttatataggctattgcccaaaaa taatttggag
   T strand gaaatatatccgataacgggttttt ● attaaacctc
            acatagCggttt<u>tg</u>t<u>a</u>g

             tgtatcGccaa<u>a</u>cac<u>t</u>t
  NT strand ctttatataggctattgcccaaaaa aatttggag
+9 T strand gaaatatatccgataacgggttttt ● ttaaacctc
            acatagCggttt<u>tg</u>tg<u>a</u>a

             tatcGccaa<u>a</u>cacc
  NT strand ctttatataggctattgcccaaaaatg taatttggag
+10 T strand gaaatatatccgataacgggtttttac ● attaaacctc
            atagCggttt<u>tg</u>tgg

47

Figure 13. Position of Spt5 crosslinks relative to the open transcription bubble. Dimensions of the transcription bubble, as measured on the *P. furiosus* glutamate dehydrogenase promoter by Spitalny & Thomm (2003), with crosslinking positions for Spt5 Bpa substitutions indicated by colored bands. T11 crosslink positions are indicated in red, Q14 in blue, H65 in yellow, and P71 in black. Transcription start site is indicated by a capital C, and RNA polymerase pause site in U-less reactions is indicated by a pink circle. Mutated bases in U-less cassettes are underlined.

Interaction between Spt4/5 and TFE during early transcription

To determine whether an exchange between TFE and Spt4/5 may occur in transcription complexes, competitive crosslinking assays were performed. Transcription complexes were formed with Spt4/5 heterodimers containing Bpa in Spt5, competing with wild type TFE. To determine when TFE has any effect on Spt4/5 crosslinking in transcription complexes, crosslinking was performed with the Spt5 Q14 Bpa mutant in the absence versus presence of wild type TFE at differing transcript lengths in early transcription (Figure 14). Use of a single Bpa mutant in the presence of a wild-type competitor demonstrates decreases in crosslinking band intensity as the crosslinker is efficiently displaced by its competitor, with sharper decreases in intensity indicating a higher affinity for the competitor to bind the complex. When complexes have only transcribed five nucleotides, addition of TFE reduces Q14 crosslinking band intensity (Figure 14A, lane 1 vs. lane 2). Similar changes are observed in complexes halted at +6 nucleotides (Figure 14A, lane 3 vs. lane 4) and those halted at +7 nucleotides (Figure 14A, lane 5 vs. lane 6). When transcription proceeds eight nucleotides, Q14 crosslinking bands in the absence (Figure 14A, lane 7) and presence (Figure 14A, lane 8) of TFE appear to be nearly equal in intensity. In +9 complexes, TFE addition again reduces Q14

48

crosslinking band intensity relative to complexes without TFE (Figure 14A, lane 9 vs. lane 10). Q14 crosslinking band intensities again appear to be mainly unaffected by TFE addition in +10 complexes (Figure 14A, lane 11 vs. lane 12). Results show a minimum decrease in Q14 crosslinking band intensity in the presence of TFE when the complex is paused at +8 nucleotides transcribed (Figure 14B), with a mean percent change of 2.05% in the presence of TFE (+/- 0.04). In contrast, earlier in transcription at +5 nucleotides transcribed, the presence of TFE decreases Q14 crosslinking band intensity by 73.85% (+/- 16.1). TFE's impact increases slightly after +8 nucleotides transcribed, with a mean decrease in crosslinking band intensity of 18.3% (+/- 4.67) at +9 nucleotides transcribed, and a mean decrease of 5.09% (+/- 2.71) at +10 nucleotides transcribed.

A

Spt5 Q14

Transcript Length (nt)	+5		+6		+7		+8		+9		+10	
TFE	−	+	−	+	−	+	−	+	−	+	−	+
	1	2	3	4	5	6	7	8	9	10	11	12

B

Figure showing % Decrease in Crosslinking Band Intensity in the Presence of TFE versus Nucleotides transcribed.
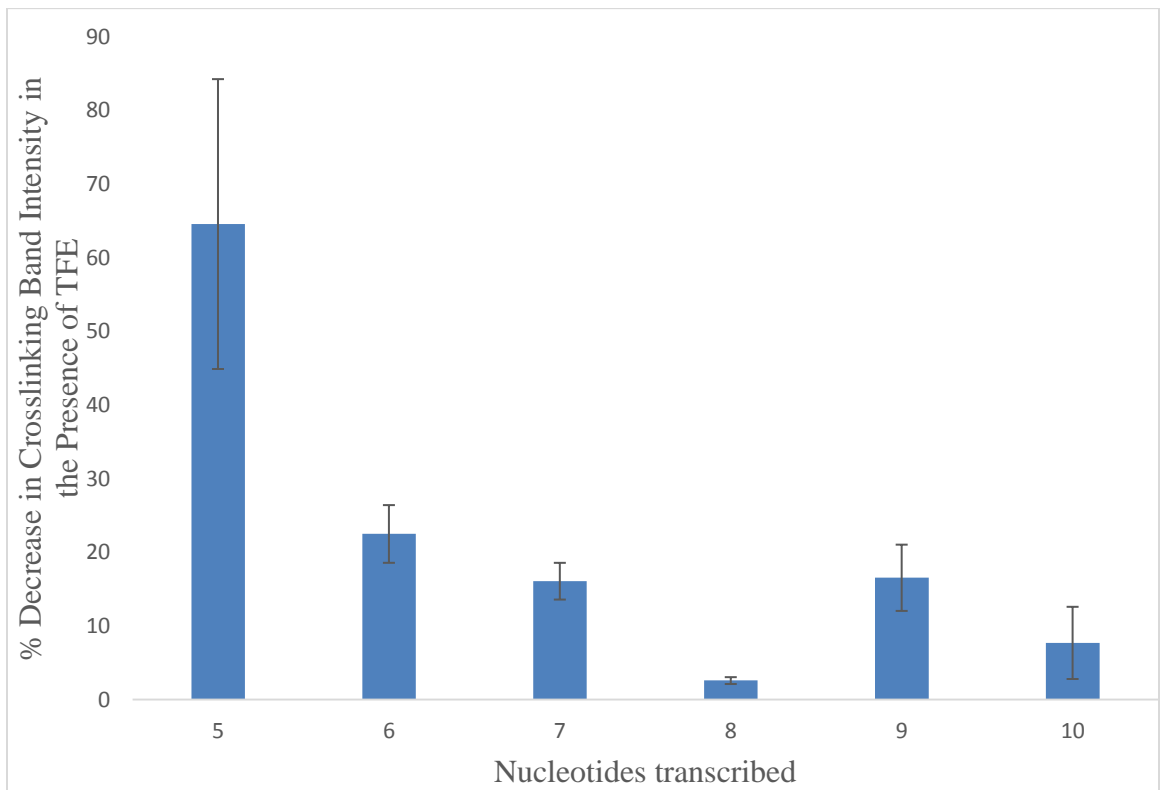
50
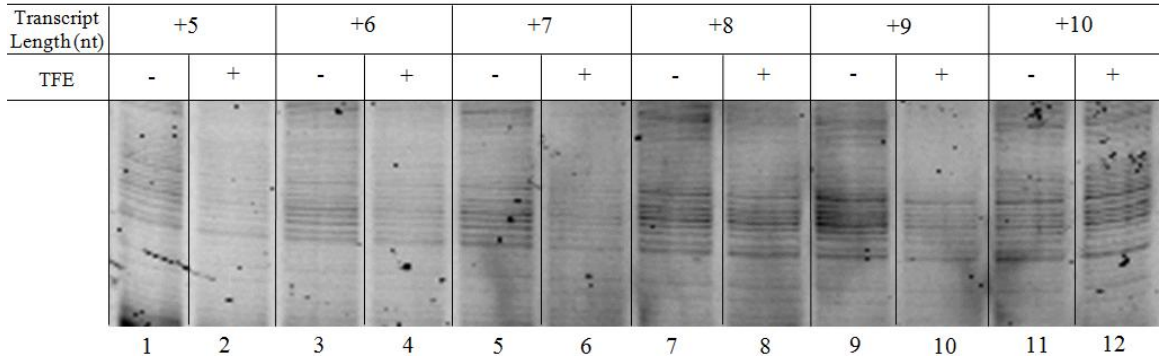
Figure 14. Spt5 Q14 crosslinks to the transcribed strand in the presence and absence of TFE. (A) Complexes were assembled with *P. furiosus* RNAP, TBP, and TFB, as well as wild type TFE (even numbered lanes only), on U-less GDHP templates (see Methods for details). Transcription was initiated with G, A, and CTP, and followed by addition of Spt4/5 with Bpa substituted for Spt5 Q14. Complexes were then crosslinked, purified, and processed as described above. Fluorescent primer extension was performed on the transcribed strand and fluorescent DNA was purified as described above. Crosslinks were analyzed on an 8M urea 10% polyacrylamide gel using a Typhoon Trio scanner. (B) Crosslinking assays were performed as described in A. Spt5 Q14 crosslink bands were quantified using ImageQuant software to determine relative pixel volumes. Percent decrease in crosslinking band intensity with TFE addition was quantified from multiple trials. Error bars represent one standard deviation above and below the mean. (n=3)

Spt4/5 inhibits transcription activity prior to synthesis of a +8mer RNA

The data in figure 14 suggest a specific window of opportunity for Spt5 incorporation, with premature binding reduced by the presence of TFE. These data, along with existing hypotheses that TFE serves to prevent premature Spt4/5 incorporation (Grohmann, et al., 2011), raise the question of what effect Spt4/5 may have on transcription if present early. UTP chase assays were performed on U-less cassettes to analyze the effect of timed Spt4/5 incorporation on transcription activity. As expected, transcription was lowest in complexes lacking TBP, which should be unable to form complexes and thus showed background radiation in samples (Figure 15A, lane 1). For comparison to uninhibited transcription, complexes were assembled and allowed to transcribe without Spt4/5 addition (Figure 15A, lane 2). When Spt4/5 was incorporated, levels of RNA decreased, as expected from inhibition of transcription initiation (Figure 15A, lane 3). Inhibition was also observed when complexes were paused after initiating

51

transcription, combined with Spt4/5, and supplied with UTP to continue elongation

(Figure 15A, lanes 4-5). However, when Spt4/5 was incorporated to complexes paused at

+8 nucleotides, inhibition was no longer observed and transcription activity increased

relative to levels in the complete absence of Spt4/5 (Figure 15A, lane 6). This

enhancement effect was further observed when Spt4/5 was incorporated to complexes

paused later, at +9 nucleotides transcribed (Figure 15A, lane 7). Spt4/5 has been

previously shown to inhibit transcription if incorporated into pre-initiated (+0) complexes

(Grohmann, et al., 2011), and these results were reproduced (Figure 15A, lane 3). A

significant shift in transcription activity between Spt4/5 incorporation after transcription

of seven versus eight nucleotides was confirmed through multiple trials (Figure 15B).
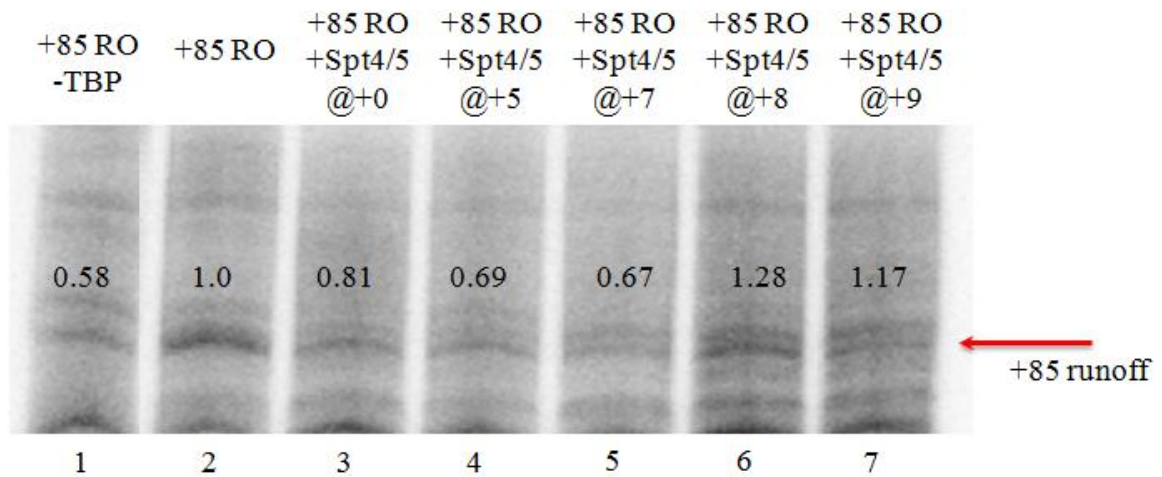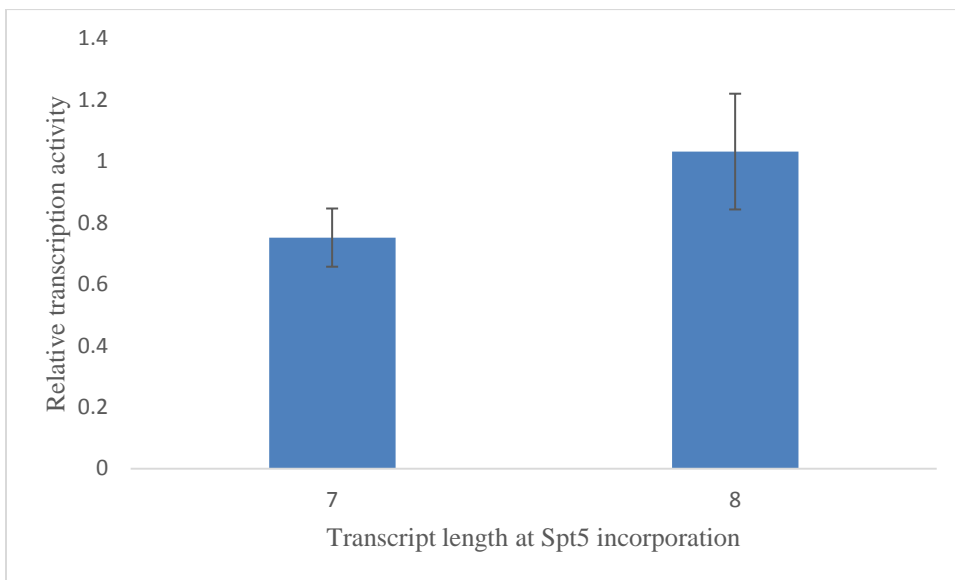
Figure 15. Transcription assays with staggered Spt4/5 incorporation. (A) Transcription assays were performed on U-less cassettes, with transcription complexes halted at U-less sites by limited NTP addition, followed by wild type *P. furiosus* Spt4/5 addition, and chased with UTP for full extension. Transcription reactions were then loaded directly to an 8M urea 8% polyacrylamide gel. The gel was then dried and exposed to a storage phosphor plate overnight, and radiolabeled RNA was visualized using a Typhoon Trio

phosphorimager. Numbers in lanes indicate relative pixel volume of the +85 runoff band. Location of the runoff band was determined by comparison to shorter transcript lengths (not shown). (B) Relative transcription activity with Spt4/5 incorporated after transcription of 7 or 8 nucleotides, determined by measuring relative pixel volumes on multiple transcription assays as in A. Error bars represent one standard deviation above and below the mean. (n=6)

TFE prevents transcription inhibition by Spt4/5 until functional shift at +8 nucleotides transcribed

As shown above, Spt4/5 has a negative effect on transcription if incorporated prematurely. Existing hypotheses and the crosslinking data presented here suggest that TFE may serve a role in preventing Spt4/5 binding and negating its detrimental effects. To determine if the presence of TFE reverses Spt4/5 inhibition of transcription, transcription assays were performed with staggered Spt4/5 incorporation using U-less chase assays as detailed above, this time in the presence or absence of TFE. As before, Spt4/5 showed a shift in transcription function, slightly reducing activity when incorporated at +7 nucleotides transcribed (Figure 16, lane 5) and slightly enhancing activity when incorporated at +8 nucleotides transcribed (Figure 16, lane 4) relative to transcription levels in the absence of TFE and Spt4/5 (Figure 16, lane 6). When similar reactions were performed in the presence of TFE, this inhibitory effect at +7 nucleotides transcribed was not observed and there was instead an increase in relative transcription activity (Figure 16, lane 2). When Spt4/5 was incorporated at +8 nucleotides transcribed in the presence of TFE, transcription levels were similar to those in the same reaction without TFE (Figure 16, lane 1 vs. lane 4).
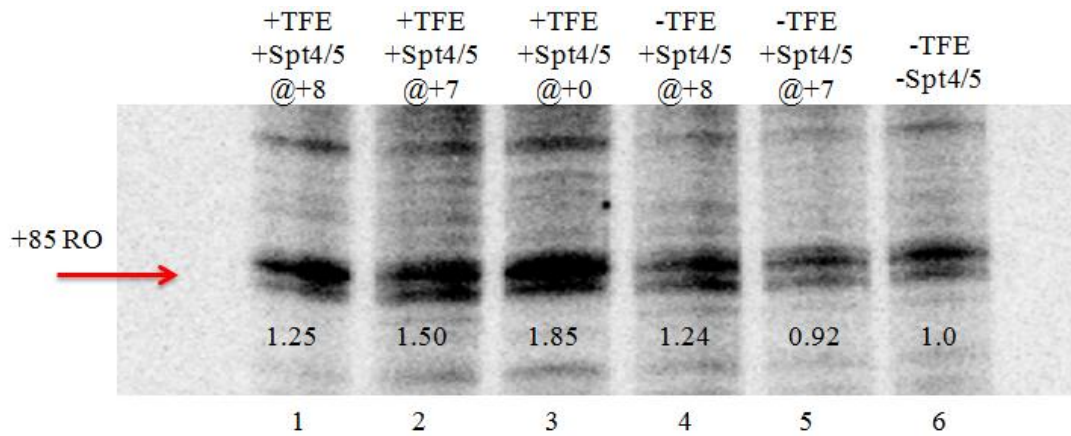
54

Figure 16. Staggered Spt4/5 incorporation in the presence or absence of TFE. Transcription complexes were assembled with *P. furiosus* RNA polymerase, TBP, TFB, and TFE (lanes 1-3 only) on U-less GDHP templates (see Methods for details). Transcription was initiated with G, A, and radiolabeled CTP, followed by addition of Spt4/5 (lanes 1-5 only). After Spt4/5 incorporation, complexes were treated with UTP to enable full-length elongation. After transcribing for 20 minutes, radiolabeled RNA was purified through phenol extraction and ethanol precipitation. RNA was then run on an 8M urea 8% polyacrylamide gel, which was then dried and exposed to a storage phosphor plate overnight prior to scanning on a Typhoon Trio phosphorimager.

Discussion

The research presented in this thesis has focused on the events following transcription initiation, in which the transcription complex shifts from the conformation associated with initiation to that of the elongation complex. This mechanistic transition leads to formation of a stable elongation complex, but the details of the transition are not well understood. A key component of the transition may be the escape from the promoter, coupled with changed interactions with transcription factors.

TFE positioning and role in transcribing complexes

The crosslinking data presented here positions TFE near the upstream fork of the transcription bubble as the transcription complex assembles and transcription initiates. These data agree with previous crosslinking data which position W76 in TFE's winged helix proximal to -9 of the nontranscribed strand (M. Bartlett, personal communication). Notably, a new amino acid was identified as being similarly proximal to the nontranscribed strand of DNA at the upstream fork, N49. This amino acid is located on a surface distant from the winged helix tip and provides more information for orienting the whole of TFE.

While TFE W76 crosslinks move relative to DNA as transcription proceeds, N49 crosslinks do not. This discrepancy may indicate flexibility within TFE that allows the winged helix tip to move independently of N49. Alternatively, the DNA contacting the wing tip may be more mobile than the DNA contacting the TFE surface where N49 is located. The independent movement of TFE's winged helix tip relative to DNA may

indicate further participation by TFE (and specifically by the winged helix) in some function of the transcribing complex. While TFE is known to enhance opening of the transcription bubble prior to transcription initiation, little is known about its roles later in transcription (Naji, et al., 2007). Footprinting assays by Spitalny and Thomm indicate that the transcription bubble increases in size, from 13 to 16 bases, after transcription proceeds +7 nucleotides (Spitalny & Thomm, 2003). It is possible that TFE may assist in these late changes to the transcription bubble as well as previously documented early changes.

TFE has also not been shown previously to interact with the transcribed strand, however the data presented above indicate that TFE does gain proximity to the transcribed strand after transcription initiation. TFE surfaces proximal to the transcribed strand include both W76 and N49, indicating that proximity is not unique to the winged helix tip and may include the full body of TFE. This shift in DNA proximity following initiation may indicate positioning of TFE at the fork in between the two strands at this step in transcription. Additionally, crosslinking bands to the transcribed strand are strongest at +7, coinciding with Spitalny and Thomm's data which demonstrate an increase in transcription bubble size at +7 nucleotides transcribed (Spitalny & Thomm, 2003). Although this transcription bubble widening was demonstrated in the absence of TFE, the remodeling of the bubble may be partially or fully responsible for the change in transcribed strand positioning relative to TFE. The approach to the transcribed strand, and in particular its timing, may further indicate a role for TFE in stabilizing the transcription bubble immediately following initiation (Figure 9).

57

While TFE has been previously shown to be a part of transcription elongation complexes up to +20 nucleotides transcribed (Grunberg, et al., 2007), the crosslinkers tested here began losing proximity to the transcribed strand after +7 nucleotides transcribed and to the nontranscribed strand after +8 nucleotides transcribed. These data indicate that, while TFE may stay bound to RNA polymerase, it appears to lose proximity to DNA shortly after transcription initiation. This loss of proximity suggests that TFE may no longer have a functional role related to DNA after RNA polymerase translocation begins.

Spt4/5 positioning in transcribing complexes

In pre-initiated transcription complexes, none of the four tested Spt4/5 Bpa mutations produce crosslinks to either strand of DNA. Spt4/5 has been shown to inhibit transcription if it joins pre-initiated complexes, and it has been hypothesized that this inhibition is the result of Spt4/5 prematurely closing the RNA polymerase clamp and preventing DNA loading into the active site. The lack of crosslinking indicates a lack of proximity to DNA; therefore, these results would support the hypothesis of Spt4/5 preventing DNA loading. Bpa mutants are still able to bind transcription complexes, as demonstrated by experiments comparing Bpa substituted Spt4/5 to wild type Spt4/5 inhibition activity in pre-initiated transcription complexes (data not shown), so the lack of crosslinks is not likely to be due to a lack of incorporation into the complex.

According to crosslinking data presented here, Spt4/5 surfaces display differential mobility relative to DNA during early transcription steps. The surfaces of Spt5 containing

H65, Q14, and T11 in Spt5 were immobile relative to DNA throughout transcription, while P71 exhibited a wide range of mobility before losing proximity to DNA at +9 nucleotides transcribed. The lack of mobility of surfaces containing H65, Q14, and T11 relative to DNA may suggest interaction with specific nucleotides, as specific amino acids in Spt5 are remaining proximal to specific nucleotides throughout transcription. While sequence-specific interactions between Spt4/5 and DNA have not been observed in archaea, sequence-specific DNA binding by NusG has been observed in bacteria (Yakhnin, et al., 2016). The fact that these crosslinks appear across a broad range rather than single nucleotides likely indicates that these specific amino acids are not involved in sequence-specific interactions, and are instead slightly flexible around other regions responsible for interaction. In contrast, the region containing P71 is much more mobile. This region of Spt5 is homologous to a region in bacterial NusG which has been identified as being highly flexible (Liu & Steitz, 2017). The pattern of P71 crosslinks, a triplet pattern downstream of the transcription bubble (Figures 12 & 13), may arise from crowding of the DNA where it approaches Spt4/5. DNA crowding at the entry channel of RNA polymerase could be related to Spt4/5's inhibitory effect on early transcription. P71 crosslinks lose proximity to DNA after the complex has transcribed +9 nucleotides, the same point at which RNA polymerase begins to translocate away from the promoter (Spitalny & Thomm, 2003). The shift in P71 proximity may indicate a conformational change in Spt4/5 (or another part of the complex) as RNA polymerase begins translocation. It is hypothesized that Spt4/5's enhancement of elongation may be the

result of enhanced translocation and processivity, and it is possible that this conformational change may contribute to this effect.

TFE and Spt4/5 competition

It was previously shown that TFE and Spt4/5 compete for an overlapping binding site on RNA polymerase, but the details of when and if this competition occurs during initiation are unclear. Competitive crosslinking assays performed here indicate that Spt4/5 most efficiently displaces TFE from transcription complexes once the complex has transcribed +8 nucleotides. Efficiency of displacement increases up until +8 nucleotides transcribed, and then decreases slightly after, indicating that +8 nucleotides transcribed may be the best window of opportunity for Spt4/5 incorporation in early transcription.

The shift in binding affinity at +8 nucleotides transcribed coincides with a decrease in intensity in TFE crosslinks to the transcribed strand (Figure 8). Crosslinking of the winged helix of TFE to nontranscribed strand DNA also appears to be strongest at +8 nucleotides transcribed, suggesting mobility between TFE's winged helix and both strands of DNA at this position. At +7 nucleotides transcribed, TFE W76 crosslinks also move further downstream, placing the wing tip at its maximal distance from N49 crosslinks. The mobility of TFE and the extension of the winged helix relative to the rest of TFE may serve to destabilize the protein's interactions with RNA polymerase, making it more vulnerable to displacement from the clamp binding site by Spt4/5.

While crosslinks arising from the Spt4/5 surface containing T11, Q14, and H65 do not move relative to DNA as transcription proceeds from initiation to +10, the DNA positions they crosslink are predicted to be separated into the two single strands of the open transcription bubble as transcription proceeds, and likely do move relative to the transcribing polymerase. Therefore, as transcription proceeds, Spt4/5 crosslinks likely change position relative to the DNA in the transcription bubble; Spt4/5 crosslink positions are expected to approach the upstream fork junction as DNA moves through the polymerase. This DNA movement would position Spt4/5 near the upstream exit of DNA, in accordance with the hypothesis that Spt4/5 increases elongation activity by promoting forward translocation by RNA polymerase (Herbert, et al., 2011). Preliminary archaeal structural data suggest that Spt4/5 could be positioned at the upstream fork (Liu & Steitz, 2017). Eukaryotic structures also indicate that Spt4/5 contributes surfaces to the exit channel of RNA polymerase extensively through Spt4 and eukaryote-specific KOW domains (Crickard, et al., 2016; Ehara, et al., 2017). Thus Spt4/5 may not be preferentially bound to RNA polymerase until DNA specific interactions reach a given location in the transcription bubble relative to single and double strands, and premature association may negatively impact the integrity of the transcription bubble and contribute to Spt4/5's inhibitory affects early in transcription.

While some regions of DNA crosslinked by Spt4/5 become single stranded within the transcription bubble, others become double stranded as transcription proceeds. These newly double stranded nucleotides in the glutamate dehydrogenase promoter include a 5 nucleotide stretch of adenines and thymines. Previous studies of bacterial NusG DNA

binding activity indicate that its DNA interactions are specific to A/T regions (Yakhnin, et al., 2016). Additionally, NusG from the hyperthermophilic bacteria *Thermotoga maritima* has been shown to preferentially bind double stranded DNA over single stranded DNA (Laio, et al., 1996). This preferential binding may indicate that Spt4/5's function in transcription is dependent on sequence-specific interactions with double stranded A/T regions. Spt4/5 may not be able to perform its function while the A/T region is single stranded, and thus may not be preferentially bound to RNA polymerase until DNA interactions are stable.

Accordingly, the data presented here show that Spt4/5 maintains its inhibitory effect on transcription not only in pre-initiated complexes, but also through initiation until the RNA strand reaches a length of +8 nucleotides. This inhibition indicates that association of Spt4/5 with the transcribing complex prior to +8 nucleotides transcribed has a negative impact on transcription activity. These data suggest that the binding shift between TFE and Spt4/5 occurs at +8 nucleotides because earlier shifts would be detrimental. Similar experiments performed in the presence and absence of TFE indicate that TFE prevents transcription inhibition by Spt4/5 not only at initiation, as previously observed by other researchers in similar transcription assays (Grohmann, et al., 2011), but also through +7 nucleotides transcribed. At +8 nucleotides transcribed, when Spt4/5 was shown to shift function and begin enhancement of elongation, transcription activity was virtually unaffected by presence of TFE. This result supports the hypothesis that TFE prevents premature Spt4/5 association, with Spt4/5 incorporation to the transcription complex

prior to +8 nucleotides transcribed having a detrimental effect on transcription. These data suggest a novel function for TFE in initiated transcription complexes.

While bacterial species do have an Spt4/5 homolog (NusG), they do not have a TFE homolog. However, previous research suggests that the sigma subunit of bacterial RNAP may obstruct NusG binding early in transcription. *In vivo*, NusG primarily associates with complexes located distant from the promoter, after the sigma factor has been displaced by NusA (Mooney, et al., 2010). After binding late in transcription, NusG contributes to pausing and termination events (Yakhnin, et al, 2016; Smollett, et al., 2017). Results suggest that in archaea, Spt4/5 binding ability in early transcription seems to require a TFE-like factor to prevent premature Spt4/5 association. It is possible that eukaryotic TFIIE performs a similar function. Additionally, chromatin immunoprecipitation experiments in eukaryotes have indicated Spt4/5 presence close to the promoter, suggesting a role for eukaryotic Spt4/5 in the transition between initiation and elongation (Smollett, et al., 2017). These results are in agreement with the data presented above, which propose Spt4/5 association with the transcription complex prior to the beginning of translocation. Further hypotheses suggest that Spt4/5 enhances elongation through driving the RNA polymerase forward on the DNA template, based on similar observations from *Escherichia coli* NusG (Herbert, et al., 2011). The data presented above indicate that Spt4/5 is positioned near the upstream fork of the transcription bubble. Spt4/5 association is also optimal at +8 nucleotides transcribed, but decreases slightly once the polymerase begins translocation. This timing may be optimal for Spt4/5

63

to bind the polymerase and enhance translocation; a delay in binding may lead to reduced affinity.

In summary, the data in this thesis support a model in which TFE and Spt4/5 both approach DNA near the upstream fork. TFE is optimally bound in pre-initiated complexes. Once transcription is initiated, the winged helix of TFE moves independently of the rest of the protein to approach new nucleotides as well as the transcribed strand of DNA. This motion reaches its maximum at or before +8 nucleotides transcribed, resulting in full extension of the winged helix of TFE relative to the rest of the protein. This combination of motion and extension likely destabilizes TFE and makes it vulnerable to displacement by Spt4/5, with Spt4/5 having maximum displacement efficiency at +8 nucleotides transcribed. This shift in binding affinity corresponds with a shift in function for Spt4/5, from a negative to positive effect on transcription activity, which may indicate a new function for TFE in preventing premature association of Spt4/5.

References

1. Bell, S. D., Brinkman, A. B., Oost, J. Van Der & Jackson, S. P. The archaeal TFIIE α homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Rep.* **2,** 133–138 (2001).

2. Bell, S. & Jackson, S. Transcription and translation in archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol.* **6,** 222–228 (1998).

3. Blombach, F. *et al.* Structural and functional adaptation of *Haloferax volcanii* TFE. *Nucleic Acids Res.* **46,** 2308–2320 (2018).

4. Brindefalk, B., Dessailly, B.H., Yeats, C., Orengo, C., Werner, F., Poole, A.M. Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res.* **41,** 2832–2845 (2013).

5. Burton, S. P. & Burton, Z. F. The sigma enigma: Bacterial sigma factors, archaeal TFB and eukaryotic TFIIB are homologs. *Transcription* **5,** 1–8 (2014).

6. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550,** 53–60 (2017).

7. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. *Proc. Natl. Acad. Sci.* **99,** 11020–11024 (2002).

8. Crickard, J. B., Fu, J. & Reese, J. C. Biochemical analysis of yeast suppressor of ty 4 / 5 (Spt4 / 5) reveals the importance of nucleic acid interactions in the prevention of RNA Polymerase II arrest. *J. Biol. Chem.* **291,** 9853–9870 (2016).

9.   Dorman, G. & Prestwich, G. D. Benzophenone photophores in biochemistry. *Biochemistry* **33,** 5661–5673 (1994).

10.   Ehara, H., Yokoyama, T., Shigematsu, H., Yokoyama, S., Shirouzu, M., Sekine, S.I. Structure of the complete elongation complex of RNA polymerase II with basal factors. *Science* **357,** 921–924 (2017).

11.   Fiala, G. & Stetter, K. O. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Arch. Microbiol.* **145,** 56–61 (1986).

12.   Grohmann, D., Nagy, J., Chakraborty, A., Klose, D., Fielden, D., Ebright, R.H., Michaelis, J., Werner, F. The initiation factor TFE and the elongation factor Spt4/5 compete for the RNAP clamp during transcription initiation and elongation. *Mol. Cell* **43,** 263–274 (2011).

13.   Grünberg, S., Bartlett, M. S., Naji, S. & Thomm, M. Transcription factor E is a part of transcription elongation complexes. *J. Biol. Chem.* **282,** 35482–35490 (2007).

14.   Hanzelka, B. L., Darcy, T. J. & Reeve, J. N. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEα. *J. Bacteriol.* **183,** 1813–1818 (2001).

15.   Hartzog, G. A. & Fu, J. The Spt4-Spt5 complex: a multi-faceted regulator of transcription elongation. *Biochim Biophys Acta* **1829,** 105–115 (2014).

16.   Herbert, K. M., Zhou, J., Mooney, R.A., Porta, A.L., Landick, R., Block, S.M. *E. coli* NusG inhibits backtracking and accelerates pause-free transcription by

promoting forward translocation of RNA polymerase. *J. Mol. Biol.* **399,** 17–30 (2011).

17. Hethke, C., Bergerat, A., Hausner, W., Forterre, P. & Thomm, M. Cell-free transcription at 95°: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. *Genetics* **152,** 1325–1333 (1999).

18. Hethke, C., Geerling, A.C., Hausner, W., de Vos, W.M., Thomm, M. A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* **24,** 2369–2376 (1996).

19. Hirata, A., Klein, B. J. & Murakami, K. S. The x-ray crystal structure of RNA polymerase from archaea. *Nature* **451,** 851–854 (2008).

20. Jun, S., Reichlen, M. J., Tajiri, M. & Murakami, K. S. Archaeal RNA polymerase and transcription regulation. *Crit Rev Biochem Mol Biol* **46,** 27–40 (2011).

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

22. Kauers, J. C., Erickson-Viitanen, S., Wolfe, H. R. & Degrado, W. F. *p*-Benzoyl-*L*-phenylalanine, a new photoreactive amino acid. *J. Biol. Chem.* **261,** 10695–10700 (1986).

23. Langer, D., Hain, J., Thuriauxt, P. & Zillig, W. Transcription in archaea: similarity to that in eucarya. *Proc. Natl. Acad. Sci.* **92,** 5768–5772 (1995).

24. Li, J., Horwitzjll, R., Mccrackens, S. & Greenblattso, J. NusG , a new *Escherichia coli* elongation factor involved in transcriptional antitermination by the N protein of phage X. *J. Biol. Chem.* **267,** 6012–6019 (1992).

25. Li, W., Giles, C. & Li, S. Insights into how Spt5 functions in transcription elongation and repressing transcription coupled DNA repair. *Nucleic Acids Res.* **42,** 7069–7083 (2014).

26. Liao, D., Lurz, R., Dobrinski, B. & Dennis, P. P. A NusG-like protein from *Thermotoga maritima* binds to DNA and RNA. *J. Bacteriol.* **178,** 4089–4098 (1996).

27. Lipscomb, G. L., Schut, G.J., Thorgersen, M.P., Nixon, W.J., Kelly, R.M., Adams, M.W. Engineering hydrogen gas production from formate in a hyperthermophile by heterologous production of an 18-subunit membrane-bound complex. *J. Biol. Chem.* **289,** 2873–2879 (2014).

28. Lipscomb, G. L., Stirrett, K., Schut, G.J., Yang, F., Jenney, F.E., Scott, R.A., Adams, M.W., Westpheling, J. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Appl. Environ. Microbiol.* **77,** 2232–2238 (2011).

29. Liu, B., & Steitz, T.A. Structural insights into NusG regulating transcription elongation. *Nucleic Acids Research* 45.2 (2017): 968–974.

30. Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. M. & Cramer, P. Architecture of the RNA polymerase – Spt4 / 5 complex and basis of universal transcription processivity. *EMBO J.* **30,** 1302–1310 (2011).

31. Mayer, A., Lidschrieber, M., Siebert, M., Leike, K., Soding, J., Cramer, P. Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17,** 1272–1278 (2010).

32. Meinhart, A., Blobel, J. & Cramer, P. An extended winged helix domain in general transcription factor E / IIEα. *J. Biol. Chem.* **278,** 48267–48274 (2003).

33. Micorescu, M., Grunberg, S., Franke, A., Cramer, P., Thomm, M., Bartlett, M. Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus*. *J. Bacteriol.* **190,** 157–67 (2008).

34. Naji, S., Gru, S. & Thomm, M. The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation. *J. Biol. Chem.* **282,** 11047–11057 (2007).

35. Ohkuma, Y., Hashimoto, S., Wang, C. K., Horikoshi, M. & Roeder, R. G. Analysis of the role of TFIIE in basal transcription and TFIIH-mediated carboxy-terminal domain phosphorylation through structure-function studies of TFIIE-α. *Mol. Cell. Biol.* **15,** 4856–4866 (1995).

36. Petushkov, I., Esyunina, D., Mekler, V., Severinov, K., Pupov, D., Kulbachinskiy, A. Interplay between σ region 3.2 and secondary channel factors during promoter escape by bacterial RNA polymerase. *Biochem. J.* **474,** 4053–4064 (2017).

37.   Sato, S., Mimasu, S., Sato, A., Hino, N., Sakamoto, K., Umehara, T., Yokoyama, S. Crystallographic study of a site-specifically cross-linked protein complex with a genetically incorporated photoreactive amino acid. *Biochemistry* **50,** 250–257 (2011).

38.   Sevostyanova, A. & Artsimovitch, I. Functional analysis of *Thermus thermophilus* transcription factor NusG. *Nucleic Acids Res.* **38,** 7432–7445 (2010).

39.   Smollett, K., Blombach, F., Reichelt, R., Thomm, M. & Werner, F. A global analysis of transcription reveals two modes of Spt4/5 recruitment to archaeal RNA polymerase. *Nat. Microbiol.* **17021,** 1–11 (2017).

40.   Spitalny, P. & Thomm, M. Analysis of the open region and of DNA-protein contacts of archaeal RNA polymerase transcription complexes during transition from initiation to elongation. *J. Biol. Chem.* **278,** 30497–30505 (2003).

41.   Thomm, M. Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions. *J. Bacteriol.* **183,** 3025–3031 (2001).

42.   Tyree, C. M., George, C.P., Lira-DeVito, L.M., Wampler, S.L., Dahmus, M.E., Zawel, L., Kadonaga, J.T. Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes Dev.* **7,** 1254–1265 (1993).

43.   Walker, J. E., Luyties, O. & Santangelo, T. J. Factor-dependent archaeal transcription termination. *Proc. Natl. Acad. Sci.* **114,** 6767–6773 (2017).

44.   Werner, F. Structure and function of archaeal RNA polymerases. *Mol. Microbiol.* **65,** 1395–1404 (2007).

45.   Werner, F. & Weinzierl, R. O. J. Direct modulation of RNA polymerase core functions by basal transcription factors. *Mol. Cell. Biol.* **25,** 8344–8355 (2005).

46.   Winkelman, J. T., Winkelman, B.T., Boyce, J., Maloney, M.F., Chen, A.Y., Ross, W., Gourse, R.L. Crosslink mapping at amino acid base resolution reveals the path of scrunched DNA in initial transcribing complexes. *Mol. Cell* **59,** 1–13 (2015).

47.   Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci.* **74,** 5088–5090 (1977).

48.   Woese, C. R., Kandlert, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains. *Proc. Natl. Acad. Sci.* **87,** 4576–4579 (1990).

49.   Xie, J. & Schultz, P. A chemical toolkit for proteins--an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7,** 775–782 (2006).

50.   Yakhnin, A. V & Babitzke, P. NusA-stimulated RNA polymerase pausing and termination participates in the *Bacillus subtilis* trp operon attenuation mechanism in vitro. *Proc. Natl. Acad. Sci.* **99,** 11067–11072 (2002).

51.   Yakhnin, A. V, Murakami, K. S. & Babitzke, P. NusG Is a sequence-specific RNA polymerase pause factor that binds to the non-template DNA within the paused transcription bubble. *J. Biol. Chem.* **291,** 5299–5308 (2016).

52.   Zuckerkandl, E. & Pauling, L. Molecules as documents of evolutionary history. *J. Theor. Biol.* **8,** 357–366 (1965).