CRYSTAL STRUCTURE AND BIOCHEMICAL ANALYSIS OF ARCHAEL
RNA POLYMERASE TRANSCRIPTION ELONGATION FACTORS

A Dissertation in
Biochemistry, Microbiology and Molecular Biology

by

Brianna Joy Klein

© 2010 Brianna Joy Klein

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2010
The dissertation of Brianna Joy Klein was reviewed and approved* by the following:

Katsuhiko Murakami  
Associate Professor of Biochemistry and Molecular Biology  
Dissertation Adviser  
Chair of Committee  

Paul Babitzke  
Professor of Biochemistry and Molecular Biology  

David S. Gilmour  
Professor of Molecular and Cell Biology  

B. Tracy Nixon  
Professor of Biochemistry and Molecular Biology  

Biao He  
Associate Professor and GRA Distinguished Investigator  
Department of Infectious Diseases  
University of Georgia School of Veterinary Medicine  

Scott B. Selleck  
Professor and Head,  
Department of Biochemistry and Molecular Biology  

*Signatures are on file in the Graduate School
ABSTRACT

Transcription factors regulate each of the stages of transcription: initiation, elongation, and termination. Throughout my dissertation I focused on three directions of elongation. The first, Spt4/5 in archaea, eukaryotes, and bacteria (NusG), is the only conserved elongation factor in all three domains of life. Here, I report the X-ray crystal structure of the Pyrococcus furiosus Spt4/5 complex solved at 1.8 Å resolution. This is the first complete Spt4/5 X-ray crystal structure containing Spt4 (zinc and N-terminal NusG (NGN) binding domains) and the Spt5 N-terminal, linker and C-terminal Kyprides-Ouzounis-Woese (KOW) domains. Spt4 contacts Spt5 through the linker, KOW, and NGN. Thus, Spt4 may restrict the location of the KOW with respect to the Spt5NGN domain, suggesting the Spt5 linker between NGN and KOW is rigid. With our collaborators we solved a cryo-EM reconstruction of Spt4/5-archaeal RNAP (aRNAP) and engineered a model of the Spt4/5-aRNAP elongation complex providing new features about Spt4/5.

The second direction is RNA synthesis, occurs during transcription elongation, is not foolproof and incorrect bases can be introduced. Removal of mismatches occurs by a proofreading mechanism. I demonstrate that aRNAP is capable of intrinsic and TFS-enhanced RNA cleavage of a transcription elongation complex (TEC) assembled on a DNA/RNA scaffold. Altering reaction conditions, it is possible to inhibit RNA cleavage and capture steps of the cleavage mechanism for crystallization, to determine the X-ray crystal structures of the proofreading TEC in the presence and absence of TFS.

The final direction focuses on hepatitis delta virus (HDV), an RNA satellite virus of the hepatitis B virus that infects about 20 million people worldwide. HDV has a single-stranded circular RNA genome without its own RNA-dependent RNAP. Instead, HDV replication and mRNA synthesis is carried out by host RNAP without DNA intermediates. Using aRNAP and a short model RNA template containing a hairpin loop and dinucleotide bulge, I determined the aRNAP forms a stable complex with the model RNA. In the presence of Mg²⁺ aRNAP cleaves the RNA around the dinucleotide bulge...
and carries out RNA extension. This shows how a DNA-dependent aRNAP cleaves and synthesizes RNA, extending the known abilities of aRNAP.
# TABLE OF CONTENTS

LIST OF FIGURES........................................................................................................................ix
LIST OF TABLES........................................................................................................................xii
ACKNOWLEDGEMENTS.............................................................................................................xiii
ABBREVIATIONS......................................................................................................................xv

Chapter 1 Introduction.................................................................................................................1
  1.1 Archaea.................................................................................................................................1
  1.2 Archaeal Transcription...........................................................................................................3
    1.2.1 Stages of transcription.....................................................................................................3
    1.2.2 Transcription pausing......................................................................................................5
  1.3 Co-transcriptional regulation in Archaea.............................................................................7
    1.3.1 Archaeal Spt4/5.............................................................................................................8
    1.3.2 Eukaryotic Spt4/5.........................................................................................................9
    1.3.3 Bacterial Spt4/5..........................................................................................................13
  1.4 Factors involved in transcription.........................................................................................17
    1.4.1 RNA polymerase..........................................................................................................17
    1.4.2 General transcription factors.......................................................................................20
    1.4.3 Transcription factor S....................................................................................................20
  1.5 Dissertation format..............................................................................................................21
  1.6 Acknowledgment..................................................................................................................23
  1.7 References............................................................................................................................23

Chapter 2 Crystal Structure of an Archaeal Homolog of the RNA Polymerase II
  Transcription Elongation Factor Spt4/5..................................................................................31
  2.1 Abstract...............................................................................................................................31
  2.2 Introduction...........................................................................................................................32
    2.2.1 Archaeal Spt4/5.............................................................................................................32
    2.2.2 Prior structural studies.................................................................................................33
    2.2.3 Experimental summary...............................................................................................35
  2.3 Materials and Methods........................................................................................................35
    2.3.1 Cloning *P. furiosus* Spt4 and Spt5.................................................................35
2.3.2 Spt4/5 expression and purification

2.3.3 Crystallization and determination of the structure of

\[ P. furiosus \] Spt4/5

2.4 Results

2.4.1 Purification of \( P. furiosus \) Spt4/5

2.4.2 Crystallization of \( P. furiosus \) Spt4/5

2.4.3 Evaluation of cryo conditions for freezing the \( P. furiosus \)

Spt4/5 crystals

2.4.4 Home source evaluation of crystals

2.4.5 Data collection at MacCHESS and structure determination of \( P. furiosus \)

Spt4/5

2.4.6 Structure of \( P. furiosus \) Spt4/5

2.4.7 The interface between \( P. furiosus \) Spt4 and Spt5 are residues

on \( \beta \)-strands

2.5 Discussion

2.5.1 Structural overview

2.5.2 Location of the Spt5 KOW with respect to the Spt5 NGN

2.5.3 Structural components at the interface between \( P. furiosus \)

Spt4 and Spt5

2.5.4 Comparison with structural homologues

2.5.5 Spt4/5 interaction with RNAP

2.6 Acknowledgments

2.7 References

Chapter 3 Biochemical Characterization of RNA-Dependent RNA Transcription by Archaeal RNA Polymerase

3.1 Abstract

3.2 Introduction

3.2.1 Hepatitis delta virus

3.2.2 6S RNA

3.2.3 Archaeal RNAP
Chapter 3. Inhibition of An Archaeal RNA Polymerase Capable of Intrinsic and Transcription Factor S Enhanced RNA Cleavage

3.1 Abstract........................................................................................................................................102
3.2 Introduction....................................................................................................................................103
3.2.1 Proofreading................................................................................................................................103
3.2.2 Prior structural summary..............................................................................................................106
3.2.3 Experimental summary...............................................................................................................107
3.3 Methods..........................................................................................................................................107
3.3.1 Purification of RNAP..................................................................................................................107
3.3.2 Purification of recombinant TFS...............................................................................................107
3.3.3 Oligonucleotides.......................................................................................................................108

Chapter 4. Inhibition of An Archaeal RNA Polymerase Capable of Intrinsic and Transcription Factor S Enhanced RNA Cleavage

4.1 Abstract...........................................................................................................................................102
4.2 Introduction.....................................................................................................................................103
4.2.1 Proofreading.............................................................................................................................103
4.2.2 Prior structural summary............................................................................................................106
4.2.3 Experimental summary..............................................................................................................107
4.3 Methods.........................................................................................................................................107
4.3.1 Purification of RNAP..................................................................................................................107
4.3.2 Purification of recombinant TFS...............................................................................................107
4.3.3 Oligonucleotides.......................................................................................................................108
4.3.4 Electrophoretic mobility shift assay (EMSA) ......................... 109
4.3.5 Nucleolytic Assay ......................................................... 110
4.3.6 Extension of RNA cleaved products .................................. 110
4.4 Results ................................................................................ 111
  4.4.1 Purification of TFS and TFS double mutant ......................... 111
  4.4.2 Archaeal RNAP can bind a DNA/RNA scaffold to form
      the TEC ........................................................................ 113
  4.4.3 RNAP can induce intrinsic RNA cleavage in backtracked TEC... 116
  4.4.4 Intrinsic RNAP RNA cleavage was inhibited at lower
      temperatures ................................................................. 119
  4.4.5 Different conditions observed can prevent intrinsic RNA
      cleavage by RNAP .......................................................... 120
  4.4.6 TFS was able to enhance RNAP RNA cleavage .................... 123
  4.4.7 TFS dipeptide mutation prevents stimulation of RNAP
      RNA cleavage .............................................................. 124
4.5 Discussion ........................................................................... 125
4.6 References ........................................................................... 132
Chapter 5 Summary and Future Directions .................................. 137
  5.1 Introduction ......................................................................... 137
  5.2 Summary and future directions for an archaeal homolog of the RNA
      polymerase II transcription elongation factor Spt4/5 ............... 137
    5.2.1 Further examine the structural interaction between
      Spt4/5 and aRNAP .......................................................... 139
    5.2.2 Follow-up to preliminary experiments ............................... 140
  5.3 References ........................................................................... 141

Appendix A Growth and Handling of Pyrococcus furiosus Spt4/5 Crystals ............ 143
Appendix B Structure-based Biochemical Characterization of Pyrococcus furiosus
     Spt4/5 .............................................................................. 152
# LIST OF FIGURES

| Figure 1.1 | Phylogenetic tree | .......................................................... | 2 |
| Figure 1.2 | Assembly of the preinitiation complex (PIC) | .......................................................... | 4 |
| Figure 1.3 | Formation of the transcription “bubble” | .......................................................... | 4 |
| Figure 1.4 | Cartoon of RNAP with DNA and RNA in transcription | elon| gation complex bubble | .......................................................... | 6 |
| Figure 1.5 | Transcription elongation as a cartoon | .......................................................... | 7 |
| Figure 1.6 | Linear Maps of Spt4 and Spt5 in the Three Domains of Life | .......................................................... | 8 |
| Figure 1.7 | Details about promoter proximal pausing and phosphorylation | .......................................................... | 12 |
| Figure 1.8 | Species specificity for binding between RNAP and Spt4/5 | .......................................................... | 13 |
| Figure 1.9 | Pathway marking bacterial NusG involvement in translocation | .......................................................... | 14 |
| Figure 1.10 | Schematic of composition of the domains of NusA | .......................................................... | 15 |
| Figure 1.11 | Model of *B. subtilis* *trp* operon regulation | .......................................................... | 17 |
| Figure 1.12 | Cellular RNA polymerase structures from three domains of life | .......................................................... | 18 |
| Figure 1.13 | Two-metal ion catalysis | .......................................................... | 19 |
| Figure 2.1 | Linear Maps of Spt4 and Spt5 in the three domains of life | .......................................................... | 33 |
| Figure 2.2 | Column chromatogram and coomassie blue stained gel of purification of *P. furiosus* Spt4/5 | .......................................................... | 40 |
| Figure 2.3 | Crystals of *P. furiosus* Spt4/5 | .......................................................... | 41 |
| Figure 2.4 | The *P. furiosus* Spt4/5 crystal structure | .......................................................... | 44 |
| Figure 2.5 | Two *P. furiosus* Spt4/5 structures per asymmetric unit | .......................................................... | 47 |
| Figure 2.6 | Interfaces of Spt4/5 structures | .......................................................... | 48 |
| Figure 2.7 | The interface structure of the *P. furiosus* Spt4 and Spt5 heterodimer formation | .......................................................... | 49 |
| Figure 2.8 | Sequence alignment of Spt4 | .......................................................... | 50 |
| Figure 2.9 | The acid-dipole interaction between Spt4-Ser42 and Spt5-Glu49 | .......................................................... | 51 |
| Figure 2.10 | Interaction between Spt4 and the Spt5 KOW domain | .......................................................... | 53 |
| Figure 2.11 | Interaction between Spt4 His10 and the Spt5 linker Glu85 | .......................................................... | 54 |
| Figure 2.12 | Two crystal structures of *Aquifex aeolicus* NusG from one | .......................................................... |
asymmetric unit..................................................................................................................55
Figure 2.13 A. aeolicus NusG structure contains the bacterial specific insertion within the NGN domain.................................................................56
Figure 2.14 Structural comparisons of P. furiosus Spt4/5, A. aeolicus NusG and human KOW of the surface charge residues on the KOW domains..............58
Figure 2.15 Comparison between the P. furiosus Spt4/5 and human Spt4/5NGN structures...........................................................................................................60
Figure 2.16 Comparison between P. furiosus Spt4/5 and human KOW structures.......................................................................................................................61
Figure 2.17 Comparison between P. furiosus Spt4/5 and bacterial NusG structures..............................................................................................................62
Figure 2.18 The P. furiosus RNAP-Spt4/5 complex reconstruction...............................64
Figure 2.19 Model of cryo-EM reconstruction of Spt4/5-aRNAP-DNA/RNA..............66
Figure 2.20 Sliding clamp.................................................................................................67
Figure 3.1 HDV life cycle within the host.......................................................................75
Figure 3.2 Genomic (G), anti-genomic (AG) and an 800-nucleotide mRNA..............75
Figure 3.3 Secondary structure of the template RNA....................................................77
Figure 3.4 6S RNA sequence compared to HDV RNA...................................................78
Figure 3.5 RNAP purification chromatograms and SDS-PAGE coomassie blue stained gels...........................................................................................................88
Figure 3.6 HDV RNA-RNAP complexes.......................................................................90
Figure 3.7 Identifying HDV RNA cleavage sites............................................................91
Figure 3.8 Detailed in vitro RNA extension of HDV RNA with P. furiosus RNAP .......94
Figure 3.9 In vitro RNA extension of HDV RNA with P. furiosus RNAP.....................95
Figure 3.10 Model for HDV RNA cleavage.................................................................96
Figure 4.1 TFIIS domain orientation and TFIIS-RNAP structure...............................104
Figure 4.2 Backtracked Pol II structure.......................................................................105
Figure 4.3 Scaffolds involved in this study.................................................................109
Figure 4.4 Purification of TFS.....................................................................................112
Figure 4.5 Purification of double mutant TFS[D90A/E91A].......................................113
Figure 4.6 Evaluation of binding affinity between aRNAP and scaffolds……………115
Figure 4.7 Intrinsic RNA cleavage by RNAP..........................................................117
Figure 4.8 Intrinsic RNA backtrack cleavage by RNAP..............................................118
Figure 4.9 Increasing the concentration of magnesium alters the RNA cleavage products................................................................................................................119
Figure 4.10 Evaluation of temperature gradient for intrinsic RNA cleavage by RNAP.................................................................................................................120
Figure 4.11 Calcium cleavage reactions.................................................................120
Figure 4.12 Modified internal RNA alters intrinsic RNA RNAP cleavage phosphorothioate bonds to inhibit intrinsic tec_bt2*........................................123
Figure 4.13 TFS dependent RNA cleavage...........................................................125
Figure 4.14 Calcium as a catalytic metal ion........................................................127
Figure 4.15 Model for the modified RNA in the DNA/RNA scaffold in the TEC-backtracked form.........................................................................................129
Figure A.1 Hanging drop crystallization technique..............................................144
Figure A.2 Sitting drop vapor diffusion crystallization technique......................144
Figure A.3 Data collection of P. furiosus Spt4/5 crystals........................................149
Figure B.1 Comparison between the structures of RfaH and NusG......................153
Figure B.2 Sequence alignment of Spt5 from several organisms.........................155
Figure B.3 X-ray crystal structure of the Spt4/5 complex at 1.8 Angstrom resolution with mutation sites.................................................................156
Figure B.4 DNA/RNA scaffold..............................................................................159
Figure B.5 Pull-down assay binding predictions................................................162
Figure B.6 Binding interaction between Spt4/5 variants and RNAP....................163
Figure B.7 Binding interaction between Spt4/5 variants and RNAP....................164
Figure B.8 EMSA...............................................................................................166
LIST OF TABLES

Table 2.1 Several of the Spt4, NusG, and Spt5 solved structures...........................34
Table 2.2 Data collection and refinement statistics of the P. furiosus Spt5/Spt4 ........45
Table 3.1 Sensitivity to α-amanitin.................................................................76
Table 3.2 HDV RNA.......................................................................................81
Table 3.3 Primers used for identification of HDV RNA cleavage site...............84
ACKNOWLEDGEMENTS

My thesis advisor Dr. Katsuhiko Murakami has provided me support and guidance in my doctoral research, for which I am thankful for; however, I am even more thankful to him for the opportunities that he has provided to me. Dr. Murakami supported my research and took me to two FASEB conferences and provided the opportunity to collect data at two different synchrotron radiation facilities. I want to thank Dr. Murakami for the confidence in me by continually providing me challenging projects and having the patience with me as we worked through the hurdles together.

A special thank you to all of the past members of the laboratory. They were very supportive at the start of my experiments. Dr. Akira Hirata was instrumental in getting me starting with purified RNAP and an initial course on how transcription works. Drs. Mike Gleghorn and Rieko Yajima are both very knowledgeable about structural details whom provided suggestions about my work.

The current members of laboratory have also provided constructive suggestions about my research. The current researcher in archaeal details is Dr. Mathew Reichlen whom critically reviews my research and has help with a plethora of other items throughout the past years. Dr. Raka Ghosh, Ritwika Basu have both become great mentors in their own ways and have aided in presentation editing. Brittany Kirk provided editing assistance of my thesis. Kevin Baker has aided in providing much of the supplies around the laboratory as well as much of the purified proteins as mentioned throughout my thesis. Zahirah Md Yusoff, helped with the purification of some of the Spt5 mutant variants as well as setting up many of the initial crystallization conditions. She also gave me the opportunity to teach many of my favorite laboratory techniques.

I have been very fortunate that my committee members have been willing to help with both my writing and experimental details. I am very thankful for these collaborations with Dr. Anamika Missra in Dr. Dave Gilmour’s laboratory and with Dr. Alex Yakhnin in Dr. Paul Babitzke’s laboratory. In addition to these collaborations, my committee members have been available to meet with me to discuss any topic, and for this I am very thankful.
Lastly I am thankful for my family and friends. To my grandparents for their support and love. I thank my Aunt and Uncle for taking me in and giving me a place to live and the extra support. My parents have always supported my dreams and have helped me to make them a reality. To my parents and my brother, who have always been there to provide me with assistance. I also thank my younger brother, who often surprises me and acts more like an older brother. Last but not least, I thank Cheyenne for all her patience while I have been working and writing.
ABBREVIATIONS

aRNAP  aRNA polymerase
RNA    Ribonucleic acid
RNAP   RNA polymerase
RdRP   RNA-dependent RNAP
EC     Elongation Complex
TEC    Transcription Elongation Complex
TFs    Transcription factors
TFS    Transcription factor S
TFB    Transcription factor B
TBP    TATA-binding protein
N-terminal amino-terminal
NGN    N-terminal NusG
C-terminal carboxyl-terminal
DTT    dithiothreitol; (2S,3S)-1,4-Bis-sulfanylbutane-2,3-diol
EDTA   ethylenediaminetetraacetic acid; 2-[2-(Bis(carboxymethyl)amino)ethyl-(carboxymethyl)amino]
Hepes  4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
SDS    sodium dodecyl sulfate
Tris   tris(hydroxymethyl)aminomethane; 2-Amino-2-hydroxymethyl-propane-1,3-diol
PAGE   polyacrylamide gel electrophoresis
PCR    polymerase chain reaction
RT-PCR reverse transcriptase polymerase chain reaction
bp     basepair
BPB    bromophenol blue
BSA    bovine serum albumin
BLAST  Basic Local Alignment Search Tool
CCP4   Collaborative Computational Project No. 4
CNS    Crystallography and NMR System
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cryo-EM</td>
<td>cryo-electron microscopy</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine 5'-triphosphate</td>
</tr>
<tr>
<td>DNAP</td>
<td>DNA polymerase</td>
</tr>
<tr>
<td>dNTP</td>
<td>2'-deoxy nucleotide triphosphate</td>
</tr>
<tr>
<td>dsDNA</td>
<td>double-stranded DNA</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>FT</td>
<td>flow-through</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-D-1-thiogalactopyranoside</td>
</tr>
<tr>
<td>kDa</td>
<td>kilo Daltons</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani media</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut-off</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotides</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleotide triphosphate</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulphonyl fluoride</td>
</tr>
<tr>
<td>Pol II</td>
<td>DNA polymerase II</td>
</tr>
<tr>
<td>rNTP</td>
<td>ribonucleotide tri-phosphate</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris, borate, EDTA</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane;2-Amino-2-(hydroxymethyl)-1,3-propanediol</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>wt</td>
<td>wild-type</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Archaea

There are three domains of life, Eukaryota, Eubacteria, and Archaea, based on the DNA sequence of the ribosomal RNA. Prior to the late 1970s, there was only a distinction between the prokaryotic (Bacterial) and Eukaryotic kingdoms, the latter including protists, fungi, plants, and animals. Dr. Carl Woese discovered that there were distinct differences in the bacterial ribosome sequences that he was studying. He grouped them into two sets, Bacteria and what he named Archaebacteria, because of the strong sequence differences (Woese, C.R. et al., 1990).

Archaebacteria, which Dr. Woese shortened to Archaea, became a separate domain from Bacteria; however, they are still classified as prokaryotes. Within the Archaeal domain there are four tentative kingdoms, Crenarchaeota, Euryarchaeota, Korarchaeota, and Nanoarchaeota (Figure 1.1). The Crenarchaeota are mainly hyperthermophiles that grow optimally between 80-100°C, mainly sulfur-dependent, and often live in extremely acidic environments. Euryarchaeota are methanogens (produce methane during metabolism) and halophiles (live in high salt). Korarchaeota are characterized by the 16S rRNA sequences from terrestrial hot springs and very little physiology is known about this kingdom. The nanoarchaeota cells are about 400 nm in diameter and can be spherical in shape and have genomes as small as 0.5 megabases in size (Huber, H. et al., 2002).

Species from the Archaeal domain can live in extreme environments including temperatures above 100°C or in extremely acidic or alkaline waters. Archaeal species have been found in salt marshes, the deep ocean, oil reserves, and in the guts of animals, including humans. Interestingly, archaea have similarities to both bacteria and eukaryotes. Several of the archaea cellular mechanisms are similar to bacteria, such as the metabolic pathway, genome organization, and regulation of transcription. However,
archaea are more similar to eukaryotes in the factors involved in and the mechanism of transcription.

**Figure 1.1 Phylogenetic tree**

Phylogenetic tree from the 16S rRNA genes to detail the four archaeal kingdoms Crenarchaeota, Euryarchaeota, Korarchaeota, and Nanoarchaeota. Highlighted in yellow are species used throughout my study. *Pyrococcus furiosus* is a sulfur-dependent archaea which was isolated from sulfurous volcanoes and shallow marine areas, and has growth temperatures of 70-103°C in anaerobic conditions with a pH environment of 5-9. *Sulfolobus solfataricus* optimally grows at 80°C in pH environment of 2-4 in sulfur hot springs. *S. solfataricus* was discovered by Wolfram Zillig and Karl Stetter near Naples in Italy, and live in volcanic hot springs (Image from Alquéres, S.M.C. *et al.*, 2007).

Archaea are potentially involved in periodontal disease, colon cancer, diverticulosis and obesity (Conway, E. and Macario, A.J.L., 2009). Consequently, a deeper understanding of archaeal systems, such as gene regulation, is important for
responding to these conditions. A process essential to gene regulation is transcription, the synthesis of RNA from DNA by RNA polymerase (RNAP) and additional factors. Besides furthering our knowledge of Archaea, the structural similarity to eukaryotic RNAP (Hirata, A. et al., 2008) and eukaryotic-like basal transcription machinery (Guo, M. et al., 2008), renders it an ideal model system to better investigate eukaryotes.

1.2 Archaeal transcription

1.2.1 Stages of transcription

Transcription by RNAP is one of the key events in gene expression. RNA is synthesized antiparallel and complementary to DNA template by RNAP with the aid of additional factors. The multi-subunit cellular DNA-dependent RNAP including archaeal RNAP (aRNAP) synthesizes RNA in three stages: initiation, elongation and termination. During initiation, RNAP along with initiation factors locates and binds to the promoter sequence (Figure 1.2). Prior to initiating RNA synthesis, the DNA strands are separated to form a “bubble” with single-stranded DNA template positioned in the active site (open complex) (Figure 1.3). The addition of ribonucleotides occurs in the 5’ to 3’ direction with template DNA sequence specificity. The RNA continues to grow and may trigger a structural change in the RNAP that leads to a stable transcription elongation complex (TEC), thus displacing transcription initiation factors.
Figure 1.2 Assembly of the preinitiation complex (PIC)
Formation of the PIC compared across the domains of life displayed in schematic representation (Modified from Bell, S.D. and Jackson, S.P., 1998).

Figure 1.3 Formation of the transcription “bubble”
This figure shows the open complex formation through DNA strand separation. Both upstream and downstream DNAs are shown. RNAP moves along the DNA following the direction of the arrow. The template DNA strand is shown in red and the nontemplate DNA strand is shown in blue.
During elongation, the RNAP may tighten around the template and increase the rate of ribonucleotide addition. Several functions are performed by the transcribing RNAP including pausing and proofreading. Pausing is the interruption of normal transcription elongation within the TEC caused by structural changes initiated by the nucleic acid strands, a regulator that contacts the RNAP, or a combination of the two (Landick, R., 2006). Proofreading occurs when RNAP detects an erroneously incorporated nucleotide and transcription pauses.

1.2.2 Transcription pausing

Pausing in a TEC is characterized by the fraction of RNAP that enter into a pause state and for how long before the RNAP resumes transcription. There are several reasons proposed for why RNAP may enter into a pause (Landick, R., 2006). Archaeal RNAP may pause as a proofreading mechanism, allowing transcription factors to reach the polymerase to cleave misincorporated nucleotides and conserve transcriptional fidelity (Lange, U. and Hausner, W., 2004). In many eukaryotes promoter proximal pausing occurs, effecting transcription attenuation and developmental regulation. Bacterial pausing can occur to allow factors (NusG, Gre, and ribosome) to catch up in order to keep transcription and translation coupled or to recruit regulators (Landick, R., 2006).

The key structural features in the TEC that cause pausing are the downstream DNA, the DNA/RNA hybrid, the RNA in the RNA exit channel, the active site, and the secondary channel that connects the active site to the outside of the RNAP. This last feature is thought to function as an NTP entry or PPi exit route (Figure 1.4; Landick, R., 2006). There are two types of pausing that are possible, elemental and long-lived pauses. The elemental pause is one that occurs so that the RNAP does not translocate along the DNA template. A long-lived pause happens when rearrangement occurs to keep the RNAP in a pause state longer than a typical elemental pause and slow down the rate of pause escape.

One well known mechanism to accomplish long-lived pausing is through backtracking (Artsimovitch, I. and Landick, R., 2000; Nudler, E. et al., 2000; Reeder, T.
and Hawley, D., 1996). Backtracking occurs when the TEC pauses and the 3’ OH of the RNA in the active site in the i, or product site (P), is displaced from the active site and walks backwards to produce an altered DNA/RNA hybrid. Backtracking is favored when a less stable DNA/RNA hybrid is present (Landick, R., 2006).

In all three domains of life, intrinsic RNA cleavage by RNAP is possible when RNA is backtracked by a couple of bases. However, when RNAP backtracks a greater distance, the presence of an additional protein factor in the TEC such as TFS, Gre, or TFIIS is necessary for RNA cleavage, and these three factors are homologues in archaea, bacteria, and eukaryotes, respectively.

Figure 1.4 Cartoon of RNAP with DNA and RNA in transcription elongation complex bubble
Bacterial RNAP subunits shown here are β and β’. Key pausing features: downstream DNA (black), the DNA/RNA hybrid, the RNA in the RNA (grey) exit channel, the active site (circle in the middle), and the secondary channel. Also shown are the active site locations P, A and E, the main channel and the upstream DNA. The direction of transcription is shown by the black arrow. These are shown on a cartoon of bacterial DNA (Image from Landick, R., 2006).
**Figure 1.5 Transcription elongation as a cartoon**

This figure shows the transcription elongation from the PIC formation that is presented in figure 1.2. The RNAPs from archaea and bacteria, and eukaryotic Pol II are in orange. NusA is in red, as is NELF, NusG and DSIF in dark green, P-TEFb in light green and TFS, TFIIS or GRE in light blue. The dashed arrow displays movement along the cartoon has occurred. The double hash mark on the DNA transcript suggests transition along the transcript. TRAP is in multi-colored circles and Spt4 and Spt5 is in maroon in green respectively. Each factor is described in the text.

### 1.3 Co-transcriptional regulation in Archaea

A way to understand the increasing transcription elongation processivity through pausing is by the archaeal complex Spt4/5 (Figures 1.5 and 1.6). Archaeal Spt5 (aSpt5), which is also called NusG, is about 14 kDa and consists of a C-terminal domain, a linker region, and a conserved N-terminal NusG (NGN) domain. Spt5 forms a heterodimer with Spt4. Spt4 in archaea (aSpt4), which is also called RpoE”, is approximately 7 kDa in size and contains a zinc finger domain and an NGN binding domain.

Spt4/5 has been shown to increase transcription elongation by suppressing pausing and backtracking of RNAP (Herbert, K.M. et al., 2010). Previous studies in both eukaryotes and bacteria have shown that Spt4/5 or NusG regulate the processivity of...
RNAP transcription elongation (Hartzog, G.A., 2003; Hartzog G.A. et al., 1998; Herbert, K.M. et al., 2010). One specific study has demonstrated in bacterial NusG to prevent backtracking and entrance into posttranslocation (Herbert, K.M. et al., 2010). Regulation by NusG in transcription termination has also been implicated (Ciampi, M.S., 2006; Knowlton, J.R. et al., 2003; Cardinale, C.J. et al., 2008).

1.3.1 Archaeal Spt4/5

The archaeal Spt4/5 complex is a transcription elongation complex that can stimulate transcription elongation and is known to form a stable complex with RNAP (Hirteiter, A. et al., 2010). Spt4/5 is composed of two subunits Spt4 and Spt5 that are both stable at 50°C and these two subunits interact to form the Spt4/5 heterodimer (Appendix B). Spt4 enhances a thermostability of Spt5 at or around 75°C (Hirteiter, A. et al., 2010). Spt4 is composed of both a conserved zinc binding motif and an N-terminal binding domain (Figure 1.6). Spt5 is composed of a C-terminal domain, also called Kyprides-Ouzounis-Woese (KOW) domain, a linker region, and a conserved N-terminal NusG (NGN) domain (Kyprides, N.C. and Ouzounis, C.A., 1999; Kyprides, N.C., Woese, C.R. and Ouzounis, C.A., 1996). KOW motifs are known to interact with both nucleic acids and proteins (Kyprides, N.C., Woese, C.R. and Ouzounis, C.A., 1996).

Figure 1.6 Linear Maps of Spt4 and Spt5 in the Three Domains of Life
Boxed regions depict the conserved and/or functional domains and motifs. In eukaryotic Spt5, numbering is based on Saccharomyces cerevisiae. Spt4 contains a zinc binding motif and N-terminal binding NusG (NGN) domain. Spt5 is composed of KOW and NGN domains. Additional features are the acidic region and C-terminal repeats in Eukaryotic Spt5, C-terminal region in Eukaryotic Spt4, and Bacterial insert in NusG (Image modified from Guo, M. et al., 2008).
Patrick Cramer and Finn Werner’s groups studied the archaeon, *Methanocaldococcus jannaschii*, Spt4/5 complex and determined it capable of stimulating transcription processivity *in vitro* (Hirtreiter, A. *et al.*, 2010). The Spt5 NGN was identified as the domain necessary for elongation stimulation based on deletion assays of variants of the Spt5. Mutations within the hydrophobic pocket of Spt5 NGN (Y42A and L44A) reduced interaction with RNAP and transcription elongation stimulation assays displayed parallel results (Hirtreiter, A. *et al.*, 2010). Archaeal Spt4/5 interacts with a tip of the coiled-coil motif in the RNAP clamp. The 10 amino acids at the tip of the coiled-coil motif of RNAP were replaced by a tetra-glycine linker (RNAP CC-Gly$_4$). This replacement resulted in the elimination of Spt4/5 binding and the stimulation of transcription elongation by Spt4/5 (Hirtreiter, A. *et al.*, 2010).

I determined the X-ray crystal structure of archaeal Spt4/5 from *P. furiosus* and also characterized its function *in vitro*. This work is described in Chapter 2 and Appendices A and B.

### 1.3.2 Eukaryotic Spt4/5

The eukaryotic homologue of Spt4/5 is much larger than its archaeal counterpart. Eukaryotic Spt4 has the conserved zinc binding motif and N-terminal binding domain; however, there is also an extra C-terminal region that is absent in archaeal Spt4 (Figure 1.6). Spt4 is not essential in yeast (Malone, E.A. *et al.*, 1993), but is necessary for correct transcription of DNA sequences that are long and GC-rich (Rondón, A.G. *et al.*, 2003). Eukaryotic Spt5 is composed of multiple KOW domains (approximately five), a linker region, a NGN domain, an acidic region, and C-terminal repeats. The C-terminal repeats are targets of regulatory kinases (reviewed by Guo, M. *et al.*, 2008).

The crystal structure of *Saccharomyces cerevisiae* Spt4/5 has been determined (Guo, M. *et al.*, 2008) and it revealed an acid-dipole interaction between the Spt4 and Spt5 interface. Mutations that disrupt the acid-dipole (Ser58 in Spt4 and Glu338 plus Ser324 in Spt5) decrease the affinity between Spt4 and Spt5 (Guo, M. *et al.*, 2008). Based on the crystal structure, authors suggested that Spt4 provides a possible spatial
restriction of the Spt5 NGN and KOW domains upon binding to Spt5 (Guo, M. et al., 2008).

The studies completed on gene regulation primarily have focused on transcription initiation and not after polymerase starts transcription. In eukaryotic RNA polymerase II (Pol II) transcription systems, specifically in *Drosophila melanogaster*, the DRB (5,6-dichloro-1-β-D-ribobenzimidazole; transcription inhibitor) sensitivity-inducing factor (DSIF) and the negative elongation factor (NELF) work together to regulate transcription elongation, by mediating promoter proximal pausing (Wu, C.H. et al., 2003). DSIF is composed of Spt4 and Spt5 subunits and has properties of both negative and positive elongation factors. There are four subunits in NELF, namely A, B, C/D, and E.

Pol II stalling by NELF in gene regulation is part of a complicated interaction with chromatin (Gilchrist, D.A. et al., 2008). During promoter proximal pausing, NELF enhanced a chromatin free environment that facilitated transcription (Gilchrist, D.A. et al., 2008). During promoter proximal pausing, Pol II pauses 20 or 30 nucleotides (nt) downstream from the transcription start site of the Drosophila heat-shock gene *hsp70* (Rasmussen, E.B. and Lis, J.T., 1993). Permanganate genomic footprinting on more than 60 Drosophila genes shows pausing around 20-50 nucleotides from the start site of transcription (Lee, C. et al., 2008).

A promoter proximal pausing model indicates that Spt5 of DSIF requires an 18 nt nascent RNA transcript from the elongation complex to make a stable complex with Pol II (Missra, A. and Gilmour D.S., 2010). DSIF stabilizes the NELF association with Pol II (Yamaguchi, Y. et al., 2002) and induces pausing (Figure 1.7). RNA-crosslinking studies suggest a model that DSIF interacts first with the nascent RNA. NELF then binds to form a stable complex that slows the transcription elongation complex (TEC) (Missra, A. and Gilmour, D.S., 2010). Unidentified additional factors are necessary to halt the TEC (Renner, D.B. et al., 2001; Missra, A. and Gilmour, D.S., 2010). In vivo chromatin are present and may interact with DSIF, NELF and Pol II during pausing to halt transcription (Brown, S.A. et al., 1996; Mavrich, T.N. et al., 2008). At this point the Ser-5, but not the Ser-2, of the Pol II CTD is already phosphorylated (Reviewed by Chiba, K. et al., 2010). Kinase P-TEFb, composed of cyclin-dependent kinase 9 (Cdk9) and Cyclin and sensitive
to DRB, phosphorylates the repeating heptads of Pol II-CTD (serine 2) and NELF-E (Chiba, K. et al., 2010; Peterlin, B.M. and Price, D.H., 2006; Renner, D.B. et al., 2001; Yamaguchi, Y. et al., 2002; reviewed by Gilmour, D.S., 2009). When NELF dissociates from RNAP, DSIF remains bound to Pol II and may stimulate elongation (Wu, C.H. et al., 2003; Yamada, T. et al., 2006). The C-terminal region (CTR) of Spt5 has a structure similar to the repetitive sequence in the Pol II CTD that acts to regulate proteins during transcription elongation. The CTR of DSIF (Spt5) is phosphorylated by P-TEFb (Figure 1.7), converting DSIF to an activator (Yamada, T. et al., 2006). Following figure 1.7, once phosphorylated, DSIF recruits additional factors, Paf1C, and Tat-SF1 to aid in productive transcription elongation (Chen, Y. et al., 2009). Additional unknown factors may also be involved as displayed in figure 1.7. Interestingly, in the case of Caenorhabditis elegans, promoter proximal pausing occurs due to starvation in the presence of DSIF, however lacking NELF (Baugh, L.R. et al., 2009).
Figure 1.7 Details about promoter proximal pausing and phosphorylation

An enlarged portion of figure 1.5 to analyze in detail promoter proximal pausing and the phosphorylation by P-TEFb (light green) of Pol II (orange), NELF (red) and DSIF (maroon and green). Additional factors shown here are Tat-SF1 (light blue), Paf1C (light red) and unknown (grey) factors that are involved (Modeled after several details in Chiba, K. et al., 2010).

While I was examining the structure and function of archaeal Spt4/5 (Chapters 2 and Appendices A and B), my laboratory provided aRNAP from Pyrococcus furiosus to Dr. Missra in Dr. Gilmour’s laboratory here at The Pennsylvania State University. Dr. Missra ascertained that the interaction between RNAP and Spt4/5 is species specific by comparing the results of P. furiosus to that of Drosophila. As shown in figure 1.8, the native gel results displayed that archaeal Spt4/5 only interacts with aRNAP, while Drosophila DSIF and NELF only interacts with Drosophila Pol II (Missra, A., 2010). This assay was completed using yeast Pol II as a control, because it had shown species specific interactions (communication with Dr. Missra).
Species specificity for binding between RNAP and Spt4/5

A. Cartoon model for experimental setup. Different polymerases combined with different Spt4/5 complexes were evaluated for interaction with each other. The DNA templates used were a G-less cassette to the length of RNA desired. The example shown here in A is 27 ntds. (nucleotides) in the nascent RNA transcript with DNA that stops at repeating Gs on the tailed template. B. Using elongation complexes of 70 nucleotides (EC70) run on a native gel Drosophila NELF and DSIF were analyzed for compatibility with yeast and Drosophila Pol II. C. On EC27 a similar assay was repeated with *P. furiosus* Spt4/5 or Spt4 or Spt5 compared to Drosophila Pol II or archaea RNAP. However increasing amounts were used as shown above the gel. D. On EC70 NELF, DSIF, and Spt4/5 were evaluated for interactions with archaea RNAP (Modified from Missra, A., 2010).

1.3.3 Bacterial NusG

In Bacteria, the transcription elongation factor NusG is a Spt5 homolog, however, there is no Spt4 counterpart (Figure 1.6). NusG is an approximately 21 kDa protein that interacts with RNAP and regulates transcription elongation and termination. NusG is composed of a C-terminal KOW domain, a linker region, and NGN.

Most of the bacterial biochemical studies of NusG have been completed on *Escherichia coli*. The single molecule study of bacterial RNAP transcription in the presence of NusG, has found that NusG is able to enhance transcription as shown in
**Figure 1.9.** This effect is involved in both backtracking and entrance into a postranslocated state and modulates the RNAP to accelerate synthesis of transcripts in the presence and absence of pauses during elongation (Herbert, K.M. *et al.*, 2010). Bacterial NusG is also involved in increasing Rho-dependent termination, anti-termination complex formation for ribosomal and λ phage gene expression as well as in the regulation of phage HK022 Nun termination (Ciampi, M.S., 2006; Knowlton, J.R. *et al*., 2003; Cardinale, C.J. *et al*., 2008).

Biochemical assays have also evaluated NusG functions in *Bacillus subtilis*. Instead of decreased pausing, *B. subtilis* NusG stimulates pausing at two hairpin-stimulated pause sites (U107 and U144 of the *trp* leader) (Yakhnin, A.V. *et al*., 2008; Yakhnin, A.V. and Babitzke, P., 2010). *B. subtilis* NusG shows a greater effect on pausing at the U144 site (Yakhnin, A.V. *et al*., 2008). In addition, *B. subtilis* NusG works...
with another transcription elongation factor NusA (Figure 1.5 and 1.10) that also stimulates pausing at U107 and U144 (Yakhnin, A.V. and Babitzke, P., 2002).

NusA

NTD S1 KH1 KH2 AR1 AR2

Figure 1.10 Schematic of composition of the domains of NusA
NusA can be organized into the multi-domains of the N-terminal domain (NTD), three RNA-binding domains (S1, KH1, and KH2 domains), and two C-terminal domain (CTD) acidic repeat domains, or auto-inhibitory domains (AR1 and AR2) (Shin, D. H. et al., 2003; Worbs, M. et al., 2001; Gopal, B. et al., 2001; Borukhov, S. Lee, J. and Laptenko, O., 2005).

NusA is a multi-domain factor as shown in figure 1.10. The *B. subtilis* NusA is composed of NTD, S1, KH1 and KH2. In *B. subtilis* the NusA-NTD (residues 1–123) and NusA-CTD (residues 149–343) interact with RNAP (Yang, X. *et al*., 2009). NusA-NTD interacts with the β-flap region of RNAP while NusA-CTD stabilizes the interaction by binding the β-flap tip of RNAP. *E. coli* NusA has two additional domains that are not present in *B. subtilis* or many other organisms; however, these domains can interact with the RNAP α-C-terminal domain (Mah, T.F. *et al*., 1999). More detailed studies have shown that the *E. coli* NusA two AR domains act to prevent the S1 and KH domains from binding RNA. This is reversed when the NusA-AR domains interact with the RNAP α-C-terminal domain (Mah, T.F. *et al*., 2000; Mah, T.F. *et al*., 1999; Traviglia, S.L. *et al*., 1999; Eisenmann, A. *et al*., 2005). There is a homologue of NusA in archaea that contain only the two KH domains and its function is unknown (Shibata, R. *et al*., 2007).

Following PIC formation (Figure 1.5), initiation of RNA synthesis occurs. The release of the sigma factor during initiation allows NusA to bind RNAP. The *B. subtilis* NusA interacts with RNAP by the nascent RNA at the β-flap, similar to where σ binds to the RNAP. This region of RNAP is involved in promoter complex formation and is also involved in later stages of transcription including elongation and termination (Murakami, K.S. and Darst, S.A., 2003; Reviewed by Yang, X. *et al*., 2009). The structure of the *B. subtilis* NusA–RNAP complex was solved by cryo-electron microscopy (cryo-EM)
reconstruction to help clarify the interaction in a structural context (Yang, X. et al., 2009).

During transcription of the trp operon NusA of B. subtilis is binding to RNAP, NusG is also binding RNAP to cooperatively stimulate pausing at two sites on the non-template DNA (Figure 1.5). NusA and NusG stimulated pausing provides an attenuation limit for TRAP interaction. Simultaneously B. subtilis trp RNA-binding attenuation protein (TRAP) acts to negatively regulate the trp operon (reviewed by Gollnick, P. et al., 2005). The promoter for trp is approximately 200 bp upstream of trpE. The initiation of transcription is not known to be regulated by tryptophan concentration (Kuroda, M.I., Henner, D. and Yanofsky, C., 1988); however, is directed by formation of the RNA secondary structures preceding trpE (Figure 1.11). Binding of TRAP determines the formation of the structure as a consequence of the concentration of tryptophan, which leads to TRAP-dependent transcription attenuation.
Figure 1.11 Model of *B. subtilis* trp operon regulation
A and B Transcriptional attenuation. A. During limiting tryptophan conditions TRAP will not bind to the RNA. The antiterminator (AT) forms allowing transcription read-through. B. In excess tryptophan conditions TRAP can bind to the 5' Stem-loop and the (G/U) AG repeats. TRAP promotes release of RNAP, by preventing formation of AT, and allowing formation of the terminator (T). Formation of the T causes termination at 140 or 141; however termination is not always 100 percent efficient, thus a portion will fail to terminate (D). C and D. Translation regulation and RNAP pausing at U144. C. During limiting tryptophan conditions TRAP will not bind to the RNA. While RNAP continues transcription the structure of the RNA folds to make available for the ribosome the trpE Shine-Dalgarno (SD) sequence. D. In excess tryptophan conditions TRAP binds RNA. TRAP promotes formation a hairpin that sequesters the SD and prevents translation of trpE (Figure adapted from Yakhnin, A.V. *et al.*, 2008).

1.4 Factors involved in transcription

1.4.1 RNA polymerase

Archaea have a single type of RNAP, consisting of 11-13 subunits depending on the species, five subunits which have homologues in both bacteria and eukaryote (Figure 1.12). We reported the first crystal structure of the aRNAP at 3.4 Å resolution (Hirata, A. *et al.*, 2008). Striking structural similarities were observed between aRNAP and
eukaryotic Pol II. Additional studies have shown functional similarities and differences of aRNAP with both eukaryotic and bacterial RNAPs.

![Figure 1.12 Cellular RNA polymerase structures from three domains of life](image)

Surface representation of multi-subunit cellular RNA polymerase structures from Bacteria (left, *Thermus aquaticus* core enzyme; Campbell, E.A. et al., 2001), Archaea (center, *S. solfataricus*; Hirata, A. et al., 2008) and Eukarya (right, *S. cerevisiae* Pol II; Armache, K.J. et al., 2005). The overall structure of the archaean RNA polymerase resembles a “crab claw”, with a protruding stalk (E’/F subcomplex). Each subunit is denoted by a unique color. Orthologous subunits are depicted by the same color (image from Hirata, A. and Murakami, K.S., 2009).

The structure of RNAP is commonly referred to as a “crab claw” with the two largest subunits as the pinchers and the active site at the base of the pinchers. The active site has a tunable mechanism which allows for RNA synthesis and nucleolytic cleavage of RNA for proofreading purposes. The RNA synthesis mechanism occurs in the active site with predicted two-metal ion catalysis (Steitz, T.A. and Steitz, J.A., 1993). This mechanism suggests that the binding of two metals in the active site of RNAP promotes RNA synthesis. Three Asp residues (Asp triad) in the active site of the RNAP coordinates
the two Mg ions (Figure 1.13). The Asp triad is flexible on a loop connected to β-strands in a barrel-like structure. The flexibility of the multi-subunit RNAP active center could be responsible for substrate selectivity and the ability to perform regulated RNA synthesis and degradation (Sosunov, V. et al., 2005).

![Diagram of two metal ion catalysis](image)

**Figure 1.13 Two-metal ion catalysis**

Active center aspartates coordinating metal ions. Proposed mechanism of multi-subunit RNAP with Mg-II coordinated by the incoming NTP in blue. The two metal ions with green circles and for polymerization reaction the arrows designate direction of the electron density to be transferred (Sosunov, V. et al., 2003; Image is adapted from Sosunov, V. et al., 2005).

Recent X-ray crystallographic studies of multi-subunit RNAPs have provided numerous insights into RNAP structure-function relationships. All structures have the similar “crab-claw” exterior with the Asp triad in the active center. In some cases, the structure of the TEC displays the RNA in the exit channel and the downstream and upstream DNA (Vassylyev, D.G. et al., 2007; Vassylyev, D.G. et al., 2007b; Gnatt, A.L. et al., 2001; Armache, K.J. et al., 2005). These structures of both the apo and elongation complex allow for analysis of the active site during transcription at various states.
1.4.2 General transcription factors

The transcription machinery of Archaea is very similar, albeit, a simplified version of the eukaryotic machinery (Thomm, M., 1996; Reeve, J.N. et al., 1997), including a Pol II-like RNAP and general transcription factors. Archaeal transcription factor B (TFB), TATA-binding protein (TBP), transcription factor E (TFE), and transcription factors S (TFS) have eukaryotic homologues TFIIB, TBP, TFIIEα, and TFIIS respectively (Bell, S.D. and Jackson, S.P., 2001; Hausner, W. et al., 1996; Lange, U. and Hausner, W., 2004). TFB and TBP are both essential for transcription initiation with RNAP. The TBP binds to the TATA-box, and introduces a kink in the DNA, while the TFB recruits RNAP to the promoter allowing initiation of transcription (Hausner, W. et al., 1996).

1.4.3 Transcription factor S

Other factors that are not necessary for in vitro transcription, however affect transcription efficiency. One such factor is TFS, which is capable of stimulating the RNAP mediated RNA cleavage activity to proofread RNA synthesis (light blue angled towards the “active site” in Figure 1.5) (Lange, U. and Hausner, W., 2004). The TFS homologs contain evolutionarily conserved Asp-Glu dipeptides at the end of the zinc finger domain, which in both bacteria and eukaryotes have been shown to be critical for RNA cleavage (Jeon, C, Yoon, H. and Agarwal, K., 1994; Sosunova, E. et al., 2003). This finding suggests that the catalytic residues are involved in coordinating metals at the RNAP active site, effecting RNA cleavage. Prior to the start of my study, there was no biochemical data available evaluating the role of the possible catalytic Asp-Glu dipeptide in archaeal TFS. In addition, TFS reactivates transcription of an arrested elongation complex, caused by misincorporated nucleotides, in RNA that is moderately backtracked or not at all (Lange, U. and Hausner, W., 2004). Thus, I have characterized the archaeal TFS function in vitro as described in Chapter 4.
1.5 Dissertation format

The objective of my research was to initially gain insight into the mechanism of the archaeal elongation complex. As my project progressed and altered through the years I started to work more with factors as well as elongation complex and RNAP. Thus, in Chapter 2 I will describe the structure of *P. furiosus* Spt4/5 at 1.8 Å. This is the first complete structure containing Spt4 and the Spt5 N-terminal, linker and C-terminal domains. Spt4 contacts Spt5 through all domains, including the linker, KOW, and NGN, suggesting a rigid Spt5 linker region. Along with collaborators we solved the cryo-electron microscopy (cryo-EM) structure of Spt4/5-aRNAP complex and were able to make a model cryo-EM reconstruction of Spt4/5-aRNAP-DNA/RNA. From this model, conclusions can be made to explain how Spt4/5 influences RNAP transcription.

In Chapter 3 I characterize biochemical interactions of aRNAP along RNA, which is similar to what happens with hepatitis delta virus (HDV) in eukaryotic systems. HDV is a RNA satellite virus of the hepatitis B virus infecting about 20 million people worldwide. HDV has a single-stranded circular RNA genome (1.7 kb) that does not encode for a RNA-dependent-RNA polymerase. Instead, HDV replication and mRNA synthesis is carried out by RNAP from their host without DNA intermediates. At the beginning of this investigation, there was no *in vitro* system using a purified Pol II available to investigate HDV RNA replication, making the task of understanding the molecular basis of HDV RNA synthesis very difficult.

*Archaear* is known to have a single RNAP, which is responsible for synthesizing all RNAs in the cell. Our crystal structure of the aRNAP with its striking structural similarity to the eukaryal Pol II, suggests a functional similarity (Hirata, A. *et al.*, 2008; Hirata, A. and Murakami, K.S., 2009). Therefore, I use the aRNAP from *P. furiosus* as a model enzyme to investigate HDV RNA-directed RNA synthesis *in vitro*.

Thus, in Chapter 3 I will explain the model HDV RNA template including a potential RNAP binding site that is comprised of a single 49 nucleotide RNA strand with a 5 nucleotide hairpin loop and a dinucleotide bulge. I have established a stable complex between the aRNAP and the model HDV RNA. In addition, in the presence of Mg$^{2+}$, the
enzyme cleaves the RNA around the dinucleotide bulge to form a free 3’-RNA end, and carries out RNA-directed RNA extension.

In Chapter 4, I detail the abilities of the factor TFS as well as the ability of aRNAP to intrinsically cleave RNA. This is shown by an aRNAP that is capable of intrinsic and TFS enhanced RNA cleavage of a transcription elongation complex (TEC) assembled on a DNA/RNA scaffold. In addition, I show that it is possible to adjust the conditions to inhibit RNA cleavage in order to capture the cleavage mechanism for crystallization, with the goal of determining the X-ray crystal structure of the proofreading TEC, with or without TFS. In one of the modifications a phosphorothioate bond derivative of the RNA in the DNA/RNA scaffolds enables inhibition of RNA cleavage by the introduction of sulfur atoms to replace nonbridging oxygen atoms. The modified RNA in the DNA/RNA scaffold also keeps the TEC in the backtracked form and may help form crystals with homogenous complexes. Additional modifications include altering the metals at the active site and lowering the temperature of the reaction. The final adjustment made was the preparation of the evolutionarily conserved TFS double mutant, TFS[D90A/E91A], using site-directed mutagenesis of Asp90 and Glu91 each to Ala. TFS was able to cleave the backtracked RNA in a stepwise manner; however, TFS[D90A/E91A] did not result in any products for cleavage of the backtracked RNA. Thus, the TFS double mutant would be a good candidate for determining the X-ray crystal structure of aRNAP-TFS-DNA/RNA, as the mutation will inhibit TFS enhanced RNA cleavage and still conserve the structural features of the elongation complex and TFS.

In the final chapter I provide a summary of discoveries presented in this dissertation along with preliminary data and future directions for whoever continues this project in the hope of giving the next person(s) continuing this project a place to start. In the first portion of Chapter 5 I discuss how the structures of Spt4/5 (X-ray crystallography) and Spt4/5-aRNAP (cryo-EM) provide the framework for the model cryo-EM reconstruction Spt4/5-aRNAP-DNA/RNA. From this model some questions arise. Thus, I suggest possible future methods including X-ray crystal structures of both Spt4/5-aRNAP and Spt4/5-aRNAP-DNA/RNA complexes.
In the second portion of Chapter 5 I will address future plans for understanding the results of the preliminary results in Appendix B. The preliminary results showed association between \textit{P. furiosus} Spt4/5 and aRNAP, in the absence of a DNA/RNA scaffold, or with the elongation complex formed on a fluorescently 5’-end labeled-RNA with DNA scaffold. Lastly, in the final portion of this chapter I provide suggestions for future directions.

1.6 Acknowledgment

I thank Dr. Anamika Missra in Dr. David Gilmour’s laboratory at The Pennsylvania State University for use of Figure 1.8.

1.7 References


Lee, C., Li, X., Hechmer, A., Eisen, M., Biggin, M.D., Venters, B.J., Jiang, C., Li, J.,
Pugh, B.F. and Gilmour, D.S. 2008. NELF and GAGA factor are linked to promoter-
proximal pausing at many genes in Drosophila. Mol. Cell Biol. 28(10), 3290-3300.
alpha subunit of E. coli RNA polymerase activates RNA binding by NusA. Genes Dev.
14, 2664–2675.
Mah, T.F., Li, J., Davidson, A.R. and Greenblatt, J. 1999. Functional importance of
regions in Escherichia coli elongation factor NusA that interact with RNA polymerase,
the bacteriophage lambda N protein and RNA. Mol. Microbiol. 34, 523–537.
of SPT4, a gene important for transcription initiation in Saccharomyces cerevisiae,
Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P.,
Missra, A. 2010. Biochemical analysis of interactions of DSIF and NELF with the
Drosophila RNA polymerase II elongation complex. The Pennsylvania State
University, 87-131.
Missra, A. and Gilmour, D.S. 2010. Interactions between DSIF (DRB sensitivity
inducing factor), NELF (negative elongation factor), and the Drosophila RNA
11301-6.
Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. and Imanaka, T. 1994. Purification and
characterization of a thermostable thiol protease from a newly isolated
hyperthermophilic Pyrococcus sp. Applied and Environmental Microbio. 60(12) 4559-
4566.


structures of transcription factor NusG in light of its nucleic acid- and protein-binding

Steitz, T.A. and Steitz, J.A. 1993. A general two-metal-ion mechanism for catalytic RNA.


protein footprinting: where different transcription factors bind to RNA polymerase.
*Biochem.* 38, 15774–15778.


Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I. and

Woese, C.R., Kandler, O. and Wheelis, M.L. 1990. Towards a natural system of
Sci. USA*, 87(12) 4576–4579.

extended RNA binding surface through arrayed S1 and KH domains in transcription

Wu, C.H., Yamaguchi, Y., Benjamin, L.R., Horvat-Gordon, M., Washinsky, J., Enerly,
E., Larsson, J., Lambertsson, A., Handa, H. and Gilmour, D. 2003. NELF and DSIF
cause promoter proximal pausing on the hsp70 promoter in Drosophila. *Genes Dev.* 17,
1402-1414.

Yakhnin, A.V. and Babitzke, P. 2002. NusA-stimulated RNA polymerase pausing and
termination participates in the *Bacillus subtilis trp* operon attenuation mechanism in


Chapter 2

Crystal Structure of an Archaeal Homolog of the RNA Polymerase II Transcription Elongation Factor Spt4/5

2.1 Abstract

Transcription occurs in three stages, initiation, elongation and termination, throughout which occurs regulation by transcription factors. A complex in archaea, Spt4/5, with homologs in eukaryotes (Spt4/5) and bacteria (NusG) is involved in the regulation of transcription elongation (Hirtreiter, A. et al., 2010; Guo, M. et al., 2008; Herbert, K.M. et al., 2010). Eukaryotic Spt4 and Spt5 have been studied both structurally and biochemically, suggesting that the binding region of the N-terminal NusG (NGN) domain of Spt5 with Spt4 is conserved in Archaea (Guo, M. et al., 2008). Although studies have confirmed binding and copurification of archaeal Spt4 and Spt5, the X-ray crystal structure of the complete complex has remained unsolved. Here, I report the X-ray crystal structure of the complete *Pyrococcus furiosus* Spt4/5 complex solved at 1.8 Angstrom resolution.

This is the first complete Spt4/5 X-ray crystal structure containing Spt4 and the Spt5 N-terminal, linker and C-terminal domains. This structure contains Spt4 zinc and NGN binding domains (β1-β4 and α1) and Spt5 Kyprides-Ouzounis-Woese (KOW) domain with the β barrel-like fold (β4-β9). Spt4 contacts Spt5 through all domains, including the linker, KOW, and NGN. Thus Spt4 may restrict the location of the KOW with respect to the NGN domain of Spt5, suggesting that the Spt5 linker between NGN and KOW is rigid and presents the KOW in a fixed position relative to the NGN domain. A sequence alignment of homologues revealed a bacterial specific insert in *Aquifex aeolicus*. Further analysis revealed with a protein Basic Local Alignment Search Tool (BLAST) search of *A. aeolicus* NusG sequence that the insert is not common, as it is found in 10 of the examined 114 sequences (Altschul, S.F. et al., 1990; Altschul, S.F. et
al., 1997; Camacho, C. et al., 2008; Gish, W. et al., 1993; Madden, T.L. et al., 1996; Morgulis, A. et al., 2008; Zhang, J. and Madden, T.L. 1997; Zhang, Z. et al., 2000). Analysis of the A. aeolicus NusG structure demonstrated minimal contacts between the bacterial insert and the rest of the A. aeolicus NusG structure (PDB entry 1NPR Knowlton, J.R, et al., 2003, Surface model comparison) and comparison with the P. furiosus Spt4/5 structure suggests that the bacterial specific insertion is not a counterpart of the Spt4 and cannot fix the orientation of the KOW to the NGN domain. With our collaborators we solved a cryo-EM reconstruction of Spt4/5-archaeal RNAP (aRNAP) and engineered a model of the Spt4/5-aRNAP elongation complex providing new features about Spt4/5.

2.2 Introduction

Transcription is the synthesis of RNA from a DNA template carried out by a conserved enzyme, RNA polymerase (RNAP), across all domains of life. This is completed along with additional transcription factors. Transcription occurs in three stages including initiation, elongation and termination, throughout which occurs proofreading and regulation by transcription factors.

2.2.1 Archaeal Spt4/5

A transcription elongation complex in archaea, Spt4/5, is known to interact stably with RNAP and can stimulate transcription elongation (Hirteiter, A. et al., 2010). Spt4/5 is composed of two subunits Spt4 and Spt5 that are both stable at 50°C and they interact to form the Spt4/5 heterodimer (Appendix 2). Spt4 stabilizes Spt5 at high temperatures, at or around 75°C (Hirteiter, A. et al., 2010). Spt4 is composed of both a conserved zinc binding motif and an N-terminal binding domain (Figure 2.1). Spt5 is composed of a C-terminal domain, also called Kyprides-Ouzounis-Woese (KOW) domain, a linker region, and conserved N-terminal NusG (NGN) domain (Kyprides, N.C. and Ouzounis, C.A., 1999; Kyprides, N.C., Woese, C.R., Ouzounis C.A., 1996). KOW motifs are known to

**Figure 2.1 Linear Maps of Spt4 and Spt5 in the three domains of life**

Boxed regions depict the conserved and/or functional domains and motifs. In the eukaryotic Spt5, numbering is based on *Saccharomyces cerevisiae*. Spt4 contains a zinc binding motif and N-terminal binding NusG (NGN) domain. The Spt5 is composed of KOWs, and NGN domain. Additional features are the acidic region and C-terminal repeats on the Eukaryotic Spt5, C-terminal region on Eukaryotic Spt4, and Bacterial specific insert on NusG (Image modified from Guo, M. *et al.*, 2008).

### 2.2.2 Prior structural studies

Biochemical experiments on archaeal *Methanocaldococcus jannaschii*, including copurification by gel filtration and affinity column chromatography, have shown a tight complex forms between Spt5 and Spt4 (Guo, M. *et al.*, 2008). Spt4 may be responsible for regulating the highly flexible linker region between the Spt5 C- and N-terminal domains. Thus dictating, where the KOW motif is aligned in reference to the NGN domain, and controlling interactions with other transcription factors. Thus, one way to study the linker region between Spt4 and Spt5 is to solve the complete X-ray crystal structure of Spt4/5. In addition, this will aid in a deeper understanding of archaeal systems including gene regulation. Furthermore, structures solved over the last several years have included structures with parts of Spt4 and/or Spt5, however, not the entire Spt4/5 complex.
Table 2.1 Several of the Spt4, NusG, and Spt5 solved structures

<table>
<thead>
<tr>
<th>Name</th>
<th>PDB</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Aae</em> NusG</td>
<td>1NPP</td>
<td><em>Aquifex aeolicus</em> NusG IN P2(1) 2.00</td>
<td>Knowlton, J.R. <em>et al.</em>, 2003.</td>
</tr>
<tr>
<td><em>Aae</em> NusG</td>
<td>1NPR</td>
<td><em>A. aeolicus</em> NusG IN C222(1) 2.21</td>
<td>Knowlton, J.R. <em>et al.</em>, 2003.</td>
</tr>
<tr>
<td><em>E. coli</em> NusE:NusG KOW NMR</td>
<td>2KVQ</td>
<td>NMR <em>Escherichia coli</em> NusE:NusG-KOW</td>
<td>Burmann, B.M. <em>et al.</em>, 2010.</td>
</tr>
<tr>
<td><em>E. coli</em> NGN NMR</td>
<td>2K06</td>
<td>NMR NusG NGN <em>E. coli</em></td>
<td>Schweimer, K. <em>et al.</em>, 2009.</td>
</tr>
<tr>
<td><em>E. coli</em> KOW NMR</td>
<td>2JVV</td>
<td>NMR NusG KOW <em>E. coli</em></td>
<td>Schweimer, K. <em>et al.</em>, 2008.</td>
</tr>
<tr>
<td><em>T. thermophilus</em> NMR NusG NGN</td>
<td>1NZ8</td>
<td>NMR NusG NGN <em>Thermus thermophilus</em></td>
<td>Reay, P. <em>et al.</em>, 2004</td>
</tr>
<tr>
<td><em>T. thermophilus</em> NMR NusG KOW</td>
<td>1NZ9</td>
<td>NMR NusG C-terminal (NGC) <em>T. thermophilus</em></td>
<td>Reay, P. <em>et al.</em>, 2004</td>
</tr>
</tbody>
</table>
2.2.3 Experimental summary

To understand the structural details of archaeal Spt4/5, cloning of *P. furiosus* Spt4 and Spt5 was completed followed by over expression in *E. coli*. I purified and crystallized Spt4/5. Following data collection, the structure of *P. furiosus* Spt4/5 was solved at 1.8 Å resolution.

2.3 Materials and Methods

2.3.1 Cloning *P. furiosus* Spt4 and Spt5

The DNA coding sequences of *spt5* and *spt4* from *P. furiosus* genomic DNA were PCR amplified using oligonucleotides (Genomics Core Facility, University Park):

- **spt4f**: 5’ GGAATTCCATATGATGAGTGAAAAAGCCTGC 3’
- **spt4r**: 5’ GAGGGATCCTCAGCGCACCCTTATGC 3’

and subcloned between NdeI and BamHI into pET21a and pET15b, respectively (Novagen). Thus, the *spt4* gene will have an N-terminal six His-tag. Cloned genes in the expression vectors were confirmed by DNA sequencing.

2.3.2 Spt4/5 expression and purification

2.3.2.1 Spt4 and Spt5 expression

*P. furiosus* Spt5 and N-terminal His6-tagged Spt4 were individually over-expressed in *E. coli* BL21 (DE3) – CodonPlus cells (Stratagene). The proteins were each expressed in separate 4L flasks containing 1.5 L Luria-Bertani media (LB) with 100 µg/ml ampicillin at 37°C until OD$_{600}$ reached 0.4-0.6. The flasks were then moved to room temperature without shaking for 10 min followed by adding 0.5 mM IPTG to start induction. Expression of Spt4 required the addition of 10 µM ZnSO$_4$ to the growth media.
at the time of induction. The flasks were transferred to a shaker at 22°C for ~18 hrs and cells were harvested by centrifugation.

2.3.2.2 Spt4/5 purification

This section discusses the processes where Spt4/5 was purified and prepared for further use. Immediately following centrifugation purification began, the pellets were never stored at -80°C. The pellets containing Spt4 and Spt5 induced cells were each resuspended on ice in 25 ml lysis buffer (10 mM Tris-HCl (pH 8.0 at room temperature), 15 % glycerol, 10 mM β-mercapto ethanol (BME), 10 µM ZnSO₄, and 1 mM phenylmethylsulfonyl fluoride) and were mixed together. The cells were lysed by sonication followed by heat treatment at 50°C for 20 min. The heat treated cell lysates were centrifuged (13,000 rpm, 15 min) at 6°C to remove cell debris and denatured proteins by the heat treatment, and the supernatant was used for further Spt4/5 purification.

After centrifugation, the supernatants containing Spt4/5 were loaded onto a 5 ml Ni-NTA (Qiagen) column equilibrated with loading buffer that contains TGMZ (10 mM Tris-HCl (pH 8.0 at room temperature), 5 % glycerol, 10 mM BME, 10 µM ZnSO₄), 0.3 M NaCl, and 0.01 M Imidazole. The column then was washed two times with 5 ml of wash buffer (TGMZ, 0.3 M NaCl, and 0.05 M Imidazole). The Spt4/5 proteins were eluted in approximately 1 ml fractions by TGMZ containing 0.3 M NaCl and 0.5 M Imidazole and stored at 4°C.

The fraction containing Spt4/5 was diluted with TGMZ to reduce the NaCl concentration to 0.05 M and then applied to a 1 ml cation exchange SP column (GE Healthcare) equilibrated with TGMZ and 0.05 M NaCl. The column was washed with TGMZ and 0.05 M NaCl followed by stepwise elution of Spt4/5 with TGMZ and 1 M NaCl. Fractions containing the Spt4/5 were pooled together and applied to a size-exclusion column chromatography (Superdex75 (16/60), GE Healthcare). Spt4/5 was eluted with a flow rate of 1 ml/min with TGMZ and 0.3 M NaCl (Figure 2.2a). The peak fractions were analyzed by SDS-polyacrylamide gel electrophoresis and then visualized.
by Coomassie Brilliant Blue (CBB) straining (Figure 2.2b). Protein concentrations were quantified by the Bradford method using BSA as a standard (Bradford, M.M., 1976). About 12 mg of Spt4/5 per liter of culture was obtained after the gel filtration column.

2.3.3 Crystallization and determination of the structure of *P. furiosus* Spt4/5

*P. furiosus* Spt4/5 was crystallized by vapor diffusion in sitting drops at 22°C against a reservoir containing 0.1 M Bis-Tris (pH 5.5), 0.2 M NaCl, and 25% (w/v) polyethylene glycol (PEG) 3,350. Within each drop, there are several crystals of varying sizes. The crystals used for data collection reached their full size within several days (~3-5 days); however, the crystals were harvested approximately 1.5 months from date of set up.

For cryocrystallography, the crystals were transferred stepwise over a period of approximately five min through mother liquor with the addition of 25 then 30% (w/v) PEG 3,350, followed by 0-20% glycerol in rapid 5% increments. The crystals were then immediately flash-frozen by immersion in liquid nitrogen. The complete 1.8 Å resolution diffraction data (Table 2.2) were collected by synchrotron radiation (F1 line) at the MacCHESS (Cornell University, Ithaca, NY) and processed with HKL2000 (Otwinowski, Z. and Minor, W., 1997). Primitive hexagonal space group P22_1_2_1 crystal (a=40.53, b=87.21, c=133.11 Å) contained two 21 kDa Spt4/5 heterodimers per asymmetric unit. The structure was determined by a molecular replacement (resolution range from 30 to 2.5 Å [Phaser (McCoy, A.J. *et al.*, 2007)])). The search model, which contains the Spt5 N-terminal NGN domain and the Spt4, was derived from the homology model of *P. furiosus* Spt5, which was constructed based on the X-ray crystal structure of *Saccharomyces cerevisiae* Spt5 (PDB entry 2EXU) and the *P. furiosus* Spt4 (PDB entry 1RYQ). The molecular replacement solution includes two Spt4/5 heterodimers in an asymmetric unit.

The electron density map was calculated using phases from the molecular replacement, and it was further improved using the density modification program Resolve (Terwilliger, T.C., 2000) which included 2-fold noncrystallographic symmetry (NCS).
restraints. The resulting electron density map had several deviations from the molecular replacement solution that included the Spt5 C-terminal KOW domain, which indicated that model bias was effectively removed by density modification (Appendix A).

Following building of the electron density map, a complete model was built by Coot (Emsley, P. and Cowtan, K., 2004). The positional refinement with NCS restraints was performed with Refmac5 (Murshudov, N.G., Vagin, A.A. and Dodson, E.J., 1997). Finally, positional refinement without NCS restraint was carried out, and water molecules were added to the model. The final model contained two Spt4/5 heterodimers (Spt5, residues 2-148; Spt4, residues 2-61), two Zn$^{2+}$ ions (one on each Spt4), five glycerol, six BME, and 111 water molecules ($R_{\text{work}}$=23.6 %; $R_{\text{free}}$=29.2 %). Structures of the six N-terminus His-tag of Spt4 and the C-terminal residues from 149 to 153 of Spt5 were disordered.

2.4 Results

2.4.1 Purification of P. furiosus Spt4/5

Currently, there are no published methods to express recombinant Spt4 and full-length Spt5 from any eukaryotic or archaeal species in E. coli cells. Thus, a new method needed to be established. Instead of co-expression of Spt4 and Spt5 together from a single plasmid, I decided to express each of the proteins individually. Initially, this path was chosen to individually purify Spt4 and Spt5 for crystallization experimentations. This expression method facilitated site-directed mutagenesis and additional biochemical studies to be conducted. Also, ZnSO$_4$ was added to the growth media during the expression of Spt4. Initially, there were concerns that the zinc in Spt4 would not be incorporated by the E. coli expression host due to a lack of available zinc ions present while folding the Spt4 proteins. E. coli cells may use up all available zinc, leaving few metal ions for Spt4. Thus, by adding external ZnSO$_4$ I wanted to enhance the chance of E. coli having available zinc present for Spt4/5.
As with expression, there also was no purification method for Spt4/5. I chose a combination of column chromatographies wherein the first resin would only bind one of the proteins while simultaneously removing the excess of the other protein in order to purify the stoichiometric Spt4/5 complex. This was followed with a series of columns, involving first a Ni-NTA and then a cation exchange (SP) column chromatography. Spt4 binds the Ni-NTA by virtue of its N-terminal His₆-tag; however, it will not bind the SP column, and Spt5 binds the SP column only. The final column chosen for the methodology was a size-exclusion column. The chromatograph and SDS-PAGE of the final step of purification was shown in **figure 2.2**. Although the Ni-NTA and SP columns are able to purify the Spt4/5 complex to near homogeneity, the final gel-filtration column chromatography step was recently added to obtain optimal purity for crystallization.

Samples from several fractions as seen in **figure 2.2** were used to set up the crystallization trials, including the fraction D2. Each of these fractions was concentrated to at least 5 mg/ml using a 5,000 Da molecular weight cutoff concentrator (VivaScience) at room temperature. The Spt4/5 fractions precipitated when the NaCl level dropped below 0.3 M, and if the protein concentration was raised above approximately 8 mg/ml. Thus, there was some difficulty obtaining a sample for crystallization. Fraction D2 was concentrated to 6.5 mg/ml and used for crystallization trials.
Figure 2.2 Column chromatogram and coomassie blue stained gel of purification of *P. furiosus* Spt4/5

A. Elution profile of the size-exclusion column chromatography. The violet line indicates the UV absorbance at 280 nm and the red columns are the fractions that were collected. The green line indicates the salt concentration that was constant at 0.3 M NaCl. The blue arrows at the bottom of the chromatogram are to show the two peaks and to point out the two sets of fractions collected. B. Alternating numbers of the fractions were analyzed by 8-25 % SDS-PAGE to show the purity of the factor in each fraction. On the left side the arrows point to the Spt4 and Spt5 subunits.
2.4.2. Crystallization of *P. furiosus* Spt4/5

Initial screening for crystals using *P. furiosus* Spt4/5 protein at 6.5 mg/ml was completed using the following crystallization screening kits: Index (Hampton Research), SaltRx (Hampton Research), PEGRx (Hampton Research), JCSG+ (Qiagen) and JCSG core I- IV (Qiagen). From these trials, several conditions were found with crystals that were optimized, including evaluation with the Additive Screen Kit (Hampton Research) and microseeding. Following optimization of the crystallization conditions, I found most of the conditions to be very similar and the crystals all had similar sizes and shapes in both sitting and hanging drops (Figure 2.3).

![Figure 2.3 Crystals of *P. furiosus* Spt4/5](image)

*Figure 2.3 Crystals of *P. furiosus* Spt4/5*
Image of *P. furiosus* Spt4/5 crystals taken under microscope in a drop.

2.4.3 Evaluation of cryo conditions for freezing the *P. furiosus* Spt4/5 crystals

I used cryocystallography to enhance the success of the data collected and the solved structure. After the crystals are grown, cryocystallography uses cryoprotectants (glycerol, PEG) to soak the crystals prior to freezing in cryogenic temperatures in liquid nitrogen (**Appendix A**). There are several advantages of using cryocystallography, including a reduction in radiation damage that allows the completion of X-ray crystallography data collection from a single crystal. An additional benefit of
cryocrystallography is that the frozen crystals can be protected and stored either for long range storage or shipping.

Crystals of *P. furiosus* Spt4/5 formed, in drops, over a reservoir of 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25 % (w/v) polyethylene glycol (PEG) 3,350 were used for evaluation of different cryo conditions. For cryocrystallography, the crystals were transferred stepwise over a period of approximately five minutes through mother liquor with the addition of 30 % (w/v) PEG 3,350, followed by 0-20 % glycerol in rapid 5 % increments ([Appendix A](#)).

The first cryo condition examined was mother liquor (0.1 M Bis-Tris (pH 5.5), 0.2 M NaCl, and 25 % (w/v) PEG 3,350) with an increasing amount of PEG 3,350, in 5 % increments, from 25 % to 40 %. Initially, the 5 % increases were completed with several minute incubations in between. The crystals showed cracking and melting the longer the time the crystals were left in mother liquor at higher concentrations of PEG 3,350. Thus, the time between solutions was decreased and the crystals were rapidly transferred from mother liquor through each of the cryo conditions. The final cryoprotectant was mother liquor with 30 % PEG 3,350, in which the crystal was fully immersed. In a cryo-loop, slightly larger than the crystal, I immediately captured the crystal, and submerged the loop in liquid nitrogen. Once in liquid nitrogen, the cryo-loop was placed inside a cryo-tube, which was already labeled and pre-chilled in the liquid nitrogen. The cryo-tube was sealed for storage.

The second cryoprotectant to be examined was mother liquor with an increasing amount of PEG 3,350 from 25 % to 30 % with glycerol (0-20 %, in rapid 5 % increments). The initial plan was to increase to 25 % glycerol; however some of the crystals showed damage at 25 %. Thus, I stopped with 20 % glycerol and continued with the above protocol to reduce damage to the crystals and freeze the cryoprotected crystals in liquid nitrogen.
2.4.4 Home source evaluation of crystals

Several crystals were evaluated for diffraction on the home source at The Pennsylvania State University (Appendix A). These crystals were not the same crystals as those taken to MacCHESS, however they were from the same conditions. I used these crystals to evaluate if the cryoprotectants would provide diffraction quality images (Appendix A).

2.4.5 Data collection at MacCHESS and structure determination of P. furiosus Spt4/5

The X-ray data collections of the P. furiosus Spt4/5 crystals were extensive. Out of the many crystals grown, 18 different diffracting crystals were evaluated at the MacCHESS F1 synchrotron beam line. From these 18, data was collected for six crystals. The crystal used for data collection was from a sitting drop, set up using 1 µl protein mixed with 2 µl reservoir. I used the second cryoprotectant (mother liquor with an increasing amount of PEG 3,350 from 25 % to 30 % with glycerol (0-20 %, in rapid 5 % increments)) to freeze the crystal. The exposure time for each image was 10 seconds with synchrotron radiation. All 366 data images collected for the structure determination used an oscillation rotation of 0.5°. The crystal structure of the P. furiosus Spt4/5 was determined by molecular replacement. Refinement statistics are detailed in the Table 2.2. There were two Spt4/5 complexes in the asymmetric unit. The final crystallographic model contains Spt4 residues 2-61 and Spt5 residues 2-148 with an R-factor of 23.6 % (R_{free} 29.2 %). Structures of the N-terminus His_{6}-tag of Spt4 and the C-terminus residue (from 149 to 153) of Spt5 were disordered.
Figure 2.4 The *P. furiosus* Spt4/5 crystal structure

A. Spt4 (amino acid residues 2-61) in magenta at top of image with zinc, in grey, coordinated by four cysteines (C6, C9, C18, and C21) in yellow, four β-strands (β1-β4) and one α helix (α1). Spt5 (amino acid residues 2-148) shown in green. In darker green the NGN domain, with the linker (83-87) in light green in the middle, and the KOW in bright green on the right. Spt5 contains eight β-strands (β1-β8) and four α helices (α1-α4). Image prepared by using PyMOL. **B.** A topology of the secondary structure with the same color and numbering as in **A.** **C.** A rotation of the topology of the secondary structure with as in **B,** with interactions between Spt4 and Spt5 domains. Interactions with the Spt5 linker are in black, and interactions with Spt5 KOW are in red.
Table 2.2 Data collection and refinement statistics of the *P. furiosus* Spt5/Spt4

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Refinement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wavelength (Å)</td>
<td>0.918</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50-1.8</td>
</tr>
<tr>
<td>No. of Reflections</td>
<td>956,628/44,007</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>98.1 (84.5)*</td>
</tr>
<tr>
<td>R/s(I)</td>
<td>19.3 (1.2)*</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>8.4 (75.1)*</td>
</tr>
</tbody>
</table>

- Data set was collected by synchrotron radiation (F1 line) at the MacCHESS (Cornell University, Ithaca, NY);
- \( R_{sym} = \frac{\sum |I - <I>|}{\sum I} \), where \( I \) is observed intensity and \( <I> \) is average intensity obtained from multiple observations of symmetry related reflections; *highest resolution shell (1.83-1.80 Å) is shown in parenthesis.

2.4.6 Structure of *P. furiosus* Spt4/5

The X-ray crystal structure of *P. furiosus* Spt4 (amino acid residues 2-61) contains four β-strands and one α helix and is composed of the NGN binding domain (residues 29-61) and characteristic zinc binding domain (residues 2-28) with the zinc surrounded by four highly conserved cysteine residues (C6, C9, C18, and C21) (*Figure 2.4*). The six His-tag residues at the N-terminus of Spt4 were not visualized in the electron density map. The third β-strand is the main binding interface with Spt5.

Spt5 is composed of residues 2-148, which includes the N-terminal, linker and C-terminal domains. This is the first complete archaeal Spt4/5 structure, containing both Spt4 and Spt5 with NGN, linker and KOW. The N-terminal domain portion, NGN, of the structure contains three β-strands (β1- β3) and three α helices (α1-α3) in which the β-strands are antiparallel to each other and sandwiched between the α helices (*Figure 2.4*). The fourth β-strand in the NGN which is shown in Spt5 and NusG homologs is not
visible in this P. furiosus Spt4/5 crystal structure. Possible reasons for this are that there are two proline residues that are on both sides of the region for β4 in a sequence alignment with the yeast structure of Spt4/5NGN (Guo, M. et al., 2008) and in another alignment for the Spt4 NGN domain from M. jannaschii (shifting the region for β4 to residues 66 in P. furiosus Spt5 to A, R, G, and V) which also does not form a β-strand. Without this previously mentioned β-strand, there are still a total of eight β-strands in the P. furiosus Spt5 structure overall. The KOW is composed of five antiparallel β-strands which form a barrel-like fold (Figure 2.4).

There are two P. furiosus Spt4/5 complexes per asymmetric unit in the solved X-ray crystal with almost identical structures (Figures 2.5 and 2.6). There were only slight differences observed in the NGN domain particularly in loop regions. Residues in the regions of Spt5 (10-14, 61-68, and 69-72) of the NGN domain showed slight variation. In the KOW loop between β6 and β7 (Spt5 residues 128-132) are also slight differences between the two complexes in the asymmetric unit. The Spt4 region surrounding the zinc atom also contains a small shift in residues 8-25. However, the dimerization interfaces between Spt4 and Spt5 are similar in both complexes. Figure 2.6 details the interface similarity between the two structures in the asymmetric unit. There is a slight shift in the side chains of the Spt4 Leu33 and Ile 35, along with the Spt5 Leu37 and Ser40. However the backbone at the interface between Spt4 and Spt5 from each P. furiosus Spt4/5 structure in the asymmetric unit aligned very well. Since two Spt4/5 structures in the asymmetric unit are almost identical, I used only one Spt4/5 structure for describing structure details.
The two *P. furiosus* Spt4/5 structures found in one asymmetric unit from the structure determination were overlaid to compare differences in structures. The first structure is in green while the second Spt4/5 structure is in cyan and shown are 180° rotation images. Residues 61-68 and 69-72 are pointed out by a black arrow. Images prepared using PyMOL.
2.4.7 The interface between *P. furiosus* Spt4 and Spt5 are residues on β-strands

The interface between *P. furiosus* Spt4 and Spt5 are residues on β-strands Spt4β3 and Spt5β2. These antiparallel β-strands align to form a continuous antiparallel β-sheet as shown in Figure 2.4b. Spt4β3 and β2 of Spt5 contain residues with hydrophobic side chains including alanine, isoleucine, leucine, and valine. Manipulation of the structure of *P. furiosus* Spt4/5 in PyMOL, looking at different angles, gives a clear view of the interface of the hydrogen bonding between the backbones of the antiparallel β-strands (Figure 2.7). Two residues that can be observed at this interface for example are Spt4 Ile35 and Spt5 Ile36.
I evaluated whether Spt4 interacts with Spt5 via any additional locations. To answer this question, I aligned amino acid sequences of Spt4 from several archaeal and eukaryotic species and identified conserved amino acid residues. From the amino acid sequence alignment, I found the Ser42, as show in figures 2.8 and 2.9, is absolutely conserved and is located at the interface between Spt4 and Spt5. In Spt4, Ser42 is located near the α helix and the closest interacting partners are Spt5 Ala2 and Tyr34 (Figure 2.9). In addition, the α helix of Spt4 may interact with residue Phe79 of the α helix three of Spt5.
Figure 2.8 Sequence alignment of Spt4
Alignment of representative Spt4 homologues including: Pfu, Pyrococcus furiosus; Sso Sulfolobus solfataricus; Mja Methanocaldococcus jannaschii; Sc, Saccharomyces cerevisiae; Dm, Drosophila melanogaster; and Hs, Homo sapiens. The symbols under the alignment stand for: a star means the residues in the column are identical, a colon stands for the conserved substitutions have been observed, and a period means that semi-conserved substitutions are observed. Colors represent: red for the small hydrophobic residues (AVFPMLW), blue is the acidic residues (DE), magenta is the basic residues (RK), green is for the hydroxyl and amine and basic (STYHCNGQS), and all others are in gray. The arrow in red and labeling is to highlight the conserved Ser42 residue. Prepared using ClustalW (Larkin, M.A. et al., 2007).
Figure 2.9 The acid-dipole interaction between Spt4-Ser42 and Spt5-Glu49

A. The *P. furiosus* Spt4/5 structure with the focus of this figure boxed in black. Spt4 in magenta with Zn$^{2+}$ in grey, four β-strands (β1-β4) and one α helix (α1). Spt5 shown in green. In darker green the NGN domain, with the linker in light green in the middle, and the KOW in bright green. Spt5 contains eight β-strands (β1-β8) and four α helices (α1-α4). B. Close up view around Spt4-Ser42 with colors the same as in A, with Spt4-Ser42 and Spt5-Glu49 shown as sticks and colored with oxygen atoms as red and nitrogen atoms in blue. The distances between atoms are shown by dashed lines. C. Electron density map superimposed on the final *P. furiosus* Spt4/5 structure shown in sticks (blue net σ=2.4). Images prepared in Coot (Emsley, P. and Cowtan, K., 2004). D. Spt5β2 shown as sticks. Coloring is as shown in B. Images for A, B and D prepared by PyMOL.
2.5 Discussion

2.5.1 Structural overview

In this study, I have prepared recombinant *P. furiosus* Spt4/5, crystallized and determined its structure at 1.8 Å by X-ray crystallography. This is the first complete Spt4/5 X-ray crystal structure containing Spt4 and full length Spt5 that includes the N-terminal, linker and C-terminal domains. As expected by the linear map in figure 2.1, Spt4 contains the zinc binding motif and NGN binding domain (Figure 2.4). *P. furiosus* Spt4/5 reveals the interface between the subunits. The heterodimer interface is formed by Spt4β3 and Spt5β2 and aligns the seven β-strands into an antiparallel β-sheet which is strengthened by hydrophobic bonds (Figure 2.7). Spt5 in the NGN domain contains the characteristic three α helices (α1-α3), however there are only three β-strands (β1-β3). A comparison with bacterial NusG and eukaryotic Spt5 show three α helices (α1-α3) and four β-strands (β1-β4) (Knowlton, J.R. *et al.*, 2003; Schweimer, K. *et al.*, 2009; Guo, M. *et al.*, 2008). In *P. furiosus* Spt4/5 the fourth β-strand is not in the full form, possibly because there is a proline at the end, yet the structure backbone is near that of a β-strand. The *P. furiosus* Spt4/5 KOW structure is that of the characteristic five β-strands that forms the barrel-like fold.

2.5.2 Location of the Spt5 KOW domain with respect to the Spt5 NGN domain

The X-ray crystal structure contains two *P. furiosus* Spt4/5 heterodimers in an asymmetric unit. Although there are some small variations between these two heterodimers, the overall structures of these heterodimers, especially the orientation of the KOW domain toward the NGN domain, are almost identical (Figure 2.5). Spt4 contacts the linker and KOW domain of Spt5 (Figure 2.4), and further analysis of the *P. furiosus* Spt4/5 structure shown in figure 2.10 and 2.11 details the interactions between Spt4 and Spt5. From figure 2.10 the hydrogen bond between Arg61 of Spt4 and the Asp120 of the Spt5 KOW domain and shown in figure 2.11 both hydrogen bonds (or salt
bridge) of Spt4 His10 and the Spt5 linker residue Glu85. Hydrogen bonds are also shown between Spt4 and Spt5 NGN domain in figure 2.7b in the interface between Spt4 and Spt5. In addition to contacting the Spt5 domains, Spt4 may also restrict the location of the KOW domain with respect to the NGN domain. Therefore, I conclude that the linker between the NGN and KOW domains of P. furiosus Spt5 is rigid, establishing the KOW domain in a fixed orientation from the NGN domain.

**Figure 2.10 Interaction between Spt4 and the Spt5 KOW domain**
The P. furiosus Spt4/5 structure with the focus of this figure boxed in black. Spt4 in magenta with Zn$^{2+}$ in grey, four β-strands (β1-β4) and one α helix (α1). Spt5 shown in green. In darker green the NGN domain, with the linker in light green in the middle, and the KOW in bright green. Spt5 contains eight β-strands (β1-β8) and four α helices (α1-α4). Interactions between the Spt5 KOW domain and Spt4 outlined in figure 2.4 by red dashed lines as also shown in this figure. The nitrogen atoms are shown in blue and the oxygen atoms are shown in red. Images prepared by PyMOL.
Figure 2.11 Interaction between Spt4 His10 and the Spt5 linker Glu85

*Top:* The *P. furiosus* Spt4/5 structure with the focus of this figure boxed in black similar to figure 2.10. Spt4 is shown in magenta with Zn$^{2+}$ in grey while Spt5 is shown in green. Shown in darker green is the NGN domain, with the linker in light green in the middle, and the KOW in bright green. *Bottom:* Interactions between the Spt5 linker and Spt4 outlined in figure 2.4 by black dashed lines as also shown in this figure. The interactions between the two nitrogen atoms (in blue) of the Spt4 His10 (histidine) and the two oxygen atoms (in red) of the Spt5 linker Glu85 (glutamic acid), each shown as sticks. The hydrogen bond between His10 and Glu85 could also be salt bridge if the surrounding factors influenced the pK$_a$ by protonating the histidine and deprotonating the glutamic acid. Images prepared by PyMOL.

Spt4 positions the Spt5 KOW relative to the Spt5 NGN domain in the *P. furiosus* Spt4/5 structure. The rigid *P. furiosus* Spt4/5 structure is in stark comparison to the flexible structure that was revealed from the bacterial NusG structural study. For example, X-ray crystal structure of the *A. aeolicus* NusG has been reported that contains four NusG per asymmetric unit. One of the *A. aeolicus* NusG structures, the orientation of KOW domain from NGN domain is completely different than others (Figure 2.12). In
the *A. aeolicus* NusG, there is a bacterial specific insertion in the middle of NGN domain that is suggested to act as a Spt4 structural counterpart (Guo, M. *et al*., 2008). Further analysis revealed with a protein Basic Local Alignment Search Tool (BLAST) search of *A. aeolicus* NusG sequence that the insert is not very common, as it is found in only 10 of the examined 114 sequences (Altschul, S.F. *et al*., 1990; Altschul, S.F. *et al*., 1997; Camacho C. *et al*., 2008; Gish, W. *et al*., 1993; Madden, T.L. *et al*., 1996; Morgulis, A. *et al*., 2008; Zhang, J. and Madden, T.L. 1997; Zhang, Z. *et al*., 2000). However, the surface model in **figure 2.13b** clearly demonstrates that the position of insertion in the *A. aeolicus* NusG is similar to the Spt4 in the Spt4/5 but there is no contact between the insertion and the linker/KOW domain. In addition, there is only a marginal contact between NGN and the insertion. Thus, I conclude that the bacterial specific insertion is not a counterpart of the Spt4 and cannot fix the orientation of KOW to the NGN domains as found possible in the *P. furiosus* Spt4/5 structure.

![Figure 2.12 Two crystal structures of *Aquifex aeolicus* NusG from one asymmetric unit](image)

**Figure 2.12 Two crystal structures of *Aquifex aeolicus* NusG from one asymmetric unit**

Two of the four crystal structures of *A. aeolicus* NusG found in one asymmetric unit from the structure determination were overlaid by using NGN domains to compare differences of KOW domain orientations. The first structure is in green while the second is in cyan. The *A. aeolicus* contains the NGN, linker and KOW domains, and has an additional bacterial insert that may have a similar role as Spt4 (Knowlton, J.R, *et al*., 2003). Image prepared using PyMOL (PDB entry 1NPP, Knowlton, J.R, *et al*., 2003).
A. The *A. aeolicus* is shown as a cartoon model (NGN domain in dark green, the linker in light green in the middle, and the KOW in bright green). There is a bacterial specific amino acid sequence insertion in the middle of NGN domain that is show in blue (Knowlton, J.R, *et al.*, 2003). B. *A. aeolicus* NusG shown as a surface model. The same coloring as in A. C. *P. furiosus* Spt4/5 structure shown as a surface model (Spt4 in magenta, Spt5 NGN domain in dark green, the linker in light green in the middle, and the KOW in bright green). Images prepared using PyMOL (PDB entry 1NPR Knowlton, J.R, *et al.*, 2003).

A purpose of a linker connecting domains in a protein provides flexibility to switch between roles involved in transcription. A flexible bacterial NusG is capable of binding several different factors involved in pausing, translation, and termination. The *E. coli* transcription elongation complex with NusG is able to inhibit backtracking and this function may be associated with the NGN domain (Herbert, K.M. *et al.*, 2010). While the NGN domain associates with RNAP, the KOW domain is able to communicate with
several protein factors that includes NusE, which is also called S10 in the ribosome, completing a connection between transcription and translation machineries (Burrmann, B.M. et al., 2010). The flexibility of the linker may enable NusG to switch between roles as an elongation factor to a termination factor. The NGN domain may then remain bound to the RNAP while the KOW interacts with phage HK022 Nun protein to block termination (Cardinale, C.J. et al., 2008), or bind to Rho to stimulate termination (Ciampi, M.S., 2006). In the case of *B. subtilis* NusG, it enhances the transcription pausing for establishing the U144 pausing complex in the trpEDCFBA operon, which is regulated by the TRAP-dependent transcription attenuation mechanism (Yakhnin, A.V. et al., 2008; Yakhnin, A.V. and Babitzke, P., 2010).

The NusG homologs showed affinity for different types of nucleic acids. This can be demonstrated by the case of *E. coli* NusG interacting with double stranded DNA (dsDNA) and ribosomal RNA (rRNA), and also where *A. aeolicus* NusG binds to dsDNA, single stranded DNA (ssDNA), and rRNA (Steiner, T. et al., 2002). One assay also previously demonstrated *A. aeolicus* NusG KOW interacting with rRNA based on the highly negative charge of the domain (Steiner, T. et al., 2002).

In order to obtain structural insight into whether *P. furiosus* KOW domain in Spt5 is able to interact with nucleic acids, I compared the surface charge distributions of KOW domains of *P. furiosus* Spt5 and *A. aeolicus* NusG (Figure 2.14). It is clear that these structures have neutral residues on one side (Figures 2.14b and c). The reverse sides of KOWs are slightly different. *P. furiosus* Spt4/5 is still neutral with a few scattered positive and negative charged residues while *A. aeolicus* NusG is primarily charged in a region with a mix of negative and positive residues. There is a charged region present in *P. furiosus* Spt4/5; however, it is buried within the linker and Spt4 (Figure 2.14b). Some of the positively charged residues within the *P. furiosus* Spt4/5 structure belong to the KOW (Lys86 and Lys119), but without Spt4 they may be free to interact with nucleic acids. This suggests a possible alternative explanation for why Spt4 restricts the motion between the Spt5 domains. Spt4 could act as both a positive and negative regulator towards Spt5 KOW, in binding NusE and not binding nucleic acids, respectively.
Figure 2.14 Structural comparisons of *P. furiosus* Spt4/5, *A. aeolicus* NusG and human KOW of the surface charge residues on the KOW domains

All figures shown as surface models with charge residues shown as positive in blue and negative in red. A and B. *P. furiosus* Spt4/5 structure shows Spt4 in magenta with Spt5 NGN domain in dark green, with the linker in light green in the middle, and the KOW in bright green. C and D. *A. aeolicus* NusG contains the NGN domain in dark green, with the linker in light green in the middle, and the KOW in bright green. There is an additional bacterial insert in blue (PDB entry 1NPR Knowlton, J.R., *et al.*, 2003). E and F, G and H. The solution structures of the second (yellow) and fifth (brown) KOW motifs of human Spt5, 2E6Z and 2E70 respectively (Tanabe, W., *et al.*, 2007). Images prepared by using PyMOL.
2.5.3 Structural components at the interface between *P. furiosus* Spt4 and Spt5

From the *P. furiosus* Spt4/5 X-ray crystal structure, I investigated the interface between Spt4 and Spt5. The interaction between *P. furiosus* Spt4 and Spt5 is very similar to other archaeal and eukaryotic Spt4/5 structures that were reported previously. In addition to the main interface, which is formed by the hydrogen bonding between the backbones of the antiparallel β-strands Spt4β2 and Spt5β2 (Figure 2.7), there is a characteristic acid dipole between Ser and Glu residues (Figure 2.9). This acid dipole interaction was also observed in the *S. cerevisiae* and *Methanocaldococcus jannaschii* Spt4/5 structures (Guo, M. *et al.*, 2009; Hirtreiter, A. *et al.*, 2010.). Therefore, this acid dipole may have a pivotal role in the specificity between Spt4 and Spt5 in addition to the hydrophobic interface.

2.5.4 Comparison with structural homologues

2.5.4.1 *P. furiosus* Spt4/5 and human Spt4/5

Initial structural comparisons between *P. furiosus* and human Spt4/5 reside around the zinc binding motif. On human Spt4 (hSpt4), there are extra three small α helices and one more considerable sized additional α helix around the zinc binding motif. These additional α helixes may act as a platform for interacting with other protein factors, or to protect the zinc in the zinc binding motif from factors in human cells but not present in archaea (Figure 2.15). Mutation study to the four cysteines of yeast Spt4 resulted in loss-of-function (Basrai, M.A *et al.*, 1996; Malone, E.A. *et al.*, 1993), suggesting that this zinc may be a structural component (Guo, M. *et al.*, 2008).
Human Spt5 has 5 KOW domains and the NMR (PDB entries 2E6Z and 2E70) structures of second and fifth KOW domains have been determined (Tanabe, W. et al., 2007). An overlay of these human KOW structures with the *P. furiosus* Spt5 KOW structure shows a strong overall structural similarity (Figure 2.16). The greatest difference between the KOWs in these structures can be seen in the charge distribution of the residues. Also as mentioned above in section 2.5.2, there are significant differences in the charges of the surface residues between the structures of *P. furiosus* Spt4/5 and *A. aeolicus* NusG. However, a comparison of *P. furiosus* and human KOW structures shows the charge distributions to be extremely different. While one surface of the *P. furiosus* KOW remains neutral, the human KOWs have charged residues on all surfaces (Figure 2.14). In figures 2.14f, g and h a large portion of the charged residues are positively charged suggesting a possibility of binding with negatively charged nucleic acids.
Figure 2.1 Comparison between *P. furiosus* Spt4/5 and human KOW structures

*P. furiosus* Spt4/5 structure shown as a cartoon with Spt4 in magenta with Spt5 NGN domain in dark green, with the linker in light green in the middle, and the KOW in bright green. The solution structures of the second (yellow) and fifth (magenta) KOW motifs of the human Spt5, 2E6Z and 2E70 respectively, are overlaid on the *P. furiosus* Spt4/5 structure. Images prepared using PyMOL (PDB entries 2E6Z and 2E70 Tanabe, W. *et al.*, 2007).

2.5.4.2 *P. furiosus* Spt4/5 and bacterial NusG

The structural comparisons between the *P. furiosus* Spt5 protein and several bacterial NusG structures show very small differences between the KOW domains. However, large variations are found in the NGN domain. The bacterial NusG contains a larger insertion within the NGN domain that was suggested that it is a structural counterpart of the archaeal and eukaryotic Spt4 (Guo, M. *et al.*, 2008). However, the *P. furiosus* Spt4/5 crystal structure from my study and other archaeal and eukaryotic Spt4/5 crystal structures have shown that the bacterial insertion is not a structural counterpart of Spt4 (Figure 2.13). There are some additional differences in a comparison between *P. furiosus* Spt4/5 and *E. coli* NusG structures in the region prior to and surrounding the Spt5α3, also known as the NusGa3. There are extra loops and an α-helix prior to where the KOW will be connected and these additional structural changes could be to position
the KOW for *E. coli* NusG. However, structurally this same region is also predicted for RNAP binding (Figure 2.17). The binding sites for RNAP and the Spt5 NGN domain are supported by data from mutation studies followed by Ni-NTA pull down assays (Appendix B), and supported by past studies by additional authors (Hirtreiter, A. *et al*., 2010; Mooney, R.A. *et al*., 2009; Belogurov, G.A. *et al*., 2007).

![Figure 2.17 Comparison between *P. furiosus* Spt4/5 and bacterial NusG structures](image)

*P. furiosus* Spt4/5 structure shown as a cartoon (Spt4 in magenta, Spt5 NGN domain in dark green, the linker in light green in the middle, and KOW in bright green). *E. coli* NusG domain structures (NGN in pink, and KOW in white) also shown as cartoons. The regions predicted for RNAP bindings are indicated by circles. The red in the bacterial NGN structure is an extension on the α-helix in comparison to the *P. furiosus* Spt4/5 structure. Images prepared using PyMOL (PDB entries 2JVV and 2K06; Schweimer, K. *et al*., 2009).

### 2.5.5 Spt4/5 interaction with RNAP

The structural details of Spt4/5 are important in understanding the function of Spt4/5 and how it interacts with RNAP. The biochemical studies presented in Appendix B indicate a binding interaction between Spt4/5 and archaeal RNAP (aRNAP). Supported by studies by Drs. Cramer’s and Werner’s groups who discovered Spt4/5 stimulates
transcription elongation and that binding occurs between the Spt5 N-terminal hydrophobic domain and the αRNAP clamp coiled-coil motif (Hirtreiter, A. et al., 2010). With this strong interaction, I was interested in how Spt4/5 interacts with RNAP both biochemically and structurally. Thus, we sent the Spt4/5-αRNAP complex for cryo-electron microscopy (cryo-EM) to our collaborators (Appendix C). The results showed αRNAP and extra density, which our collaborators were able to fit in Spt4/5 (Figure 2.18). I was surprised by the need to rotate Spt5KOW approximately 90° towards the upstream DNA away from the crystal structure to fit it into the Spt4/5-αRNAP cryo-EM structure. This suggests that the theory of a rigid linker is possibly a crystallographic artifact. There are several contacts that occur within the asymmetric unit between the two Spt4/5 complexes. This may in turn force the structure to be more rigid by protein-protein interactions. Some ideas as to how to use structural methods to determine if the linker is rigid or not are discussed in Chapter 5.

A DNA/RNA scaffold was added into the cryo-EM reconstruction to model the Spt4/5-αRNAP elongation complex (Figure 2.19). Spt4/5 in the cryo-EM structure is located in the RNAP claw, which closes off the DNA binding channel (Figure 2.18 and 2.19); thus preventing DNA from backwards translocation during RNAP transcription, explaining why Spt4/5 can enhance transcription. The DNA sliding clamp in the DNA replication fork has the similar idea with complete closure of the DNA channel (Figure 2.20).
Figure 2.18 The *P. furiosus* RNAP-Spt4/5 complex reconstruction

A. The crystal structure from *S. sulfataricus* RNAP fitted into the cryo-EM maps from top and front. RpoA’ and RpoA” in dark gray, RpoB in light brown and the other RNAP subunits in light gray. The extra density positions for the Spt4 and Spt5NGN (a) in green and Spt5 KOW (b) in orange. B. A magnified view of the boxed interface between *S. sulfataricus* RNAP and Spt4/5 with cryo-EM map shown as mesh. Spt4 is shown as purple, Spt5NGN in green, the zinc in cyan, the Spt5 linker in blue, and the Spt5KOW in orange. C. Model of the *P. furiosus* RNAP-Spt4/5 complex with *P. furiosus* RNAP is shown as a surface model from top and front. Coloring as in A. Spt4 and Spt5NGN are shown as cartoon models.
This model of the Spt4/5-aRNAP elongation complex (Figure 2.19b) complements the results found from *B. subtilis* NusG, which interacts with the non-template DNA to stabilize pausing at U144 on the *trp* leader while RNAP enters the elemental pause state. Based on mutation studies of non-template DNA at positions 137-139 on the *trp* leader, *B. subtilis* NusG-stimulated pausing, was strongly altered by these positions, and was determined to be sequence specific (Yakhnin, A.V. and Babitzke, P., 2010). Using these non-template positions, I have approximately mapped them on the model of the Spt4/5-aRNAP elongation complex to show possible interactions with *B. subtilis* NusG (Figure 2.19c). The *B. subtilis* NusG structure may be different, thus contacting the DNA at slightly different regions, as also suggested by the basic regions near the interacting surface. This will allow NusG to contact U135 and 137-139 as expected by the mutation studies (Yakhnin, A.V. and Babitzke, P., 2010).
Figure 2.19 Model of cryo-EM reconstruction of Spt4/5-aRNAP-DNA/RNA

A. DNA/RNA used: Template DNA is cyan, non-template DNA is black, and RNA is red. Positions of RNAP domains (D/L and E/F subunits), Spt4/5 domains (NGN and KOW) and DNA (downstream (d-DNA) and upstream (u-DNA) double-stranded DNAs) are indicated. Viewed from the top and side (show active site).

B. Magnified view of the active site same as in A. With the coiled-coil (cc) clamp. The surface of Spt4/5 is colored according to electrostatic surface potential (negative, red; neutral, white; positive, blue).

C. Same as in B, with the addition of labeled RNA U144 and non-template positions 135-139 of the trp leader from B. subtilis NusG-stimulated pausing mapped on the model of the Spt4/5-aRNAP elongation complex to show possible interactions with B. subtilis NusG (Yakhnin, A.V. and Babitzke, P., 2010).
Figure 2.20 Sliding clamp
A. The sliding clamp is shown in dark blue with DNAP in red and DNA template in green with new DNA in light blue. The direction of mechanism follows the black arrow. B and C. X-ray crystal structure of a eukaryotic (S. cerevisiae) sliding clamp in green with DNA in orange and blue. The angle in B is rotated in C. Sliding clamp completely surrounds the DNA (PDB entry 3K4X; McNally, R. and Kuriyan, J. 2010) Images prepared using PyMOL.

2.6 Acknowledgments

I thank Drs. Anamika Missra, David Gilmour and Paul Babitzke for discussion. I also thank the staff at the Cornell High Energy Synchrotron Source and Hemant Yennawar at The Pennsylvania State University for supporting the X-ray crystallography data collections. I thank the collaboration with Daniel Bose and Dr. Xiaodong Zhang at the Imperial College London on cryo-EM. This work was supported by The Pew Scholars Program in the Biomedical Science, the National Institutes of Health.
2.7 References


*Microbiology*, 152(9), 2515-2528.


Cycle Is Suggested by X-ray Crystallography and Supported by Site-Directed Mutants. *Biochem.* 42(8), 2275-2281.


Schweimer, K., Scheckenhofer, U., and Roesch, P. 2008. 2JVV.

Schweimer, K., Scheckenhofer, U., and Roesch, P. 2009. 2K06.

Tanabe, W. *et al.* 2007. 10.2210/pdb2e6z/pdb, 10.2210/pdb2e70/pdb.


Chapter 3

Biochemical Characterization of RNA-Dependent RNA Transcription by Archaeal RNA Polymerase

3.1 Abstract

Hepatitis delta virus (HDV) is a RNA satellite virus of the hepatitis B virus that infects about 20 million people worldwide. HDV has a single-stranded circular RNA genome (1.7 kb) that does not encode for a RNA-dependent-RNA polymerase. Instead, HDV replication and mRNA synthesis is carried out by RNA polymerase (RNAP) from their host without any DNA intermediates. The HDV replication is sensitive to α-amanitin suggesting that eukaryal RNAP II (Pol II) may be involved in HDV RNA replication (Macnaughton, T.B. et al., 2002). At the beginning of this investigation, there was no in vitro system using a purified Pol II available to investigate HDV RNA replication, making the task of understanding the molecular basis of HDV RNA synthesis very difficult.

Archaea is known to have a single RNAP, which is responsible for synthesizing all RNAs in the cell. The transcription apparatus in Archaea can be described as a simplified version of its eucaryal Pol II counterpart, comprising a Pol II-like RNAP as well as two general transcription factors including TATA-binding protein (TBP) and the eucaryal TFIIB ortholog TFB. Our crystal structure of the archaeal RNAP (aRNAP) has shown striking structural similarity with eucaryal Pol II suggesting a functional similarity (Hirata, A. et al., 2008; Hirata, A. and Murakami, K.S., 2009). Therefore, I used the aRNAP from Pyrococcus furiosus as a model enzyme to investigate HDV RNA-directed RNA synthesis in vitro. A model HDV RNA template including a potential RNAP binding site is comprised of a single 49 nucleotide RNA strand with a 5 nucleotide hairpin loop and a dinucleotide bulge. The aRNAP forms a stable complex with the model HDV RNA. In addition, in the presence of Mg$^{2+}$, the enzyme cleaves the RNA
around the dinucleotide bulge to form a free 3’-RNA end, and carries out RNA-directed RNA extension. These results show how a DNA-dependent aRNAP is also able to cleave and synthesize RNA, extending the known abilities of aRNAP.

3.2 Introduction

3.2.1 Hepatitis delta virus

Hepatitis is characterized by the inflammation of the liver. There are several varieties of hepatitis such as A, B, C, D and E. Many patients are diagnosed with hepatitis by testing serum for antibodies or antigens specific to hepatitis. The first reported case of hepatitis delta virus (HDV) was in 1977 by Rizzetto and colleagues (Rizzetto, M. et al., 1977). HDV requires hepatitis B virus (HBV) to supply important components for the formation of new HDV particles. Thus, HDV usually associates with HBV, causing severe liver disease (Macnaughton, T.B. et al., 2002a), which can lead to liver failure, cirrhosis, or hepatocellular carcinoma (Hadziyannis, S.J., 1997). HDV is a RNA satellite virus of hepatitis B virus that infects about 20 million people worldwide. It is transmitted through injection under the skin or sexual contact with infected blood. Those at risk have HBV or are able to contract HBV. HDV is seen throughout the world, usually parallel to HBV infections.

HDV is a single-stranded circular RNA (1.7 kb) that encodes for a protein hepatitis delta antigen (HDAg). There are a minimum of three HDV RNAs involved in the virus life cycle (Figure 3.1), including the genomic (G), anti-genomic (AG) and an 800-nucleotide mRNA (Figure 3.2). The G RNA forms a semi-double-stranded structure that may stabilize the RNA, and the AG RNA encodes for the HDAg protein (Macnaughton, T.B. and Lai, M.M.C., 2006). Replication of both G and AG RNAs is carried out by double rolling-circle-RNA-directed RNA synthesis without any DNA intermediates (Macnaughton, T.B. and Lai, M.M.C., 2006; Chen, P.J. et al., 1986).
Figure 3.1 HDV life cycle within the host
Potential model for HDV RNA replication in infected cells. This model displays the use of Pol II and Pol I or Pol I-like RNA polymerase for rolling circle replication (Image from Macnaughton, T.B. et al., 2002).

Figure 3.2 Genomic (G), anti-genomic (AG) and an 800-nucleotide mRNA
The minimum three HDV RNAs involved in the virus life cycle, consisting of genomic (G), anti-genomic (AG), and an 800-nucleotide mRNA. AG contains an open reading frame for the mRNA, but is not translated into HDAg (Image from Taylor, J., 2006).
HDV has a single-stranded circular RNA genome that does not encode for an RNA-dependent-RNA polymerase (RdRP). Instead, HDV replication and mRNA synthesis is carried out by RNA polymerase (RNAP) from their host without any DNA intermediates. To date, HDV RNA is observed as the only example of RNA species that mammalian enzymes in cells can copy (Macnaughton, T.B., 2002). The HDV replication is sensitive to α-amanitin suggesting that RNAP II (Pol II) is involved in HDV RNA replication. The different sensitivity of each polymerase to α-amanitin during transcription, suggests that mRNA transcription and G RNA synthesis are mediated by Pol II, while AG RNA synthesis is carried out by another RNA polymerase, most likely RNAPI (Pol I) (Table 3.1) (Macnaughton, T.B. and Lai, M.M., 2002; Macnaughton, T.B. et al., 2002; Modahl, L. et al., 2000).

**Table 3.1 Sensitivity to α-amanitin** (Modified from Macnaughton, T.B. and Lai, M.M.C., 2006)

<table>
<thead>
<tr>
<th></th>
<th>HDAg mRNA</th>
<th>G HDV RNA</th>
<th>AG HDV RNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative ratio</td>
<td>~1</td>
<td>~50</td>
<td>~1000</td>
</tr>
<tr>
<td>α-Amanitin</td>
<td>Sensitive</td>
<td>Sensitive</td>
<td>Resistant</td>
</tr>
<tr>
<td>Polymerase</td>
<td>Pol II</td>
<td>Pol II</td>
<td>Pol I question</td>
</tr>
</tbody>
</table>

The start site of RNA synthesis is determined by the secondary structure of the template RNA (Figure 3.3) (Filipovska, J. and Konarska, M.M., 2000). Presently, there is no in vitro system to investigate HDV RNA syntheses using human RNAPs (Gudima, S.O., Chang, J. and Taylor, J.M., 2004; Beard, M.R., Macnaughton, T.B. and Gowans, E.J., 1996; Yamaguchi, Y. et al., 2001), which makes it difficult to understand the molecular basis of HDV RNA synthesis. In addition to understanding DNA and RNA-dependent RNA transcription, there is a potential for evaluating clinical applications, some of which relate to RNA-dependent RNA synthesis.
Figure 3.3 Secondary structure of the template RNA
A. Model for HDV RNA-templated transcription (Image prepared from Filipovska, J. and Konarska, M., 2000). Pol II recognizes and binds to the RNA secondary structure. Cleavage of the RNA occurs and the RNA is used as a template for extension. B. Secondary structures of the HDV RNA used in this study.

The secondary structure of the RNA is important for binding of RNAP. Selection of the transcription start site is determined by the RNA secondary structure, not the sequence (Figure 3.3; Filipovska, J. and Konarska, M., 2000). Using an in vitro transcription system with HeLa cell nuclear extract (NE), the dinucleotide bulge acts as a cleavage site near the terminal loop. The cleaved RNA is now available as the site for synthesis of nascent RNA with AG RNA as a template (Figure 3.3; Filipovska, J. and Konarska, M., 2000).

3.2.2 6S RNA

An example of an additional RNA that is capable of binding with RNAP is 6S RNA. A small RNA (sRNA), 6S is an abundant 184 nucleotide RNA isolated from Escherichia coli initially in the late 1960s (Hindley, J., 1967). Both co-immunoprecipitation and purification experiments showed 6S RNA to interact with RNAP (Wassarman, K.M. and Storz, G., 2000). The folding of the secondary structure of 6S RNA is both similar to and different than that of the HDV RNA. It is similar in that it acts as double stranded, and imitates DNA; however it is also different because the
binding site for the RNAP for 6S RNA imitates a full open complex with a large central bulge rather than a cleavage site like HDV (Figure 3.4). The 6S RNA also is different from HDV in that it regulates transcription by binding RNAP with σ70, which competes for DNA in the active site. This all occurs during stationary cell growth, or when nutrient sources are low. RNAP can use 6S RNA as a template to synthesize product RNA (20 nucleotide long), thus when entering a new nutrient source product RNA is synthesized and RNAP is released along with 6S RNA and product RNA (Wassarman, K.M. and Saecker, R.M., 2006; Wassarman, K.M., 2007).

Figure 3.4 6S RNA sequence compared to HDV RNA
A. Secondary structure of E. coli 6S RNA diagram (Barrick, J.E. et al., 2005; Trotochaud, A.E. and Wassarman, K.M., 2005), displaying a red arrow to mark RNA synthesis where 6S RNA template initiates and yellow shading to indicate the complementary region to the product RNA (20 nucleotide long) (Image from Wassarman, K.M., 2007). B. HDV AG103 containing a minimal segment of RNA to act as a template in the presence of NE and Pol II. Progressive deletions of HDV sequences were used to find a template that did not affect transcription. This template was used to aid in finding the cleavage site (purple arrow) for HDV RNA (Filipovska, J. and Konarska, M., 2000).

3.2.3 Archaeal RNAP

There is a single RNAP in archaea, which is responsible for synthesizing all RNAs in the cell. The transcription apparatus in archaea can be described as a simplified version of its eukaryotic Pol II counterpart, comprising a Pol II-like RNAP as well as two general transcription factors including TATA-binding protein (TBP) and the eucaryal TFIIB ortholog TFB. Our recent crystal structure study of the archaeal RNAP (aRNAP)
has shown striking structural similarity between eucaryal Pol II and aRNAP, suggesting their functional similarity (Hirata, A., et al., 2008). Thus, I am using aRNAP as a model enzyme to investigate HDV RNA-directed RNA synthesis in vitro and as a strategy to provide high resolution X-ray crystal structures of RNAP-HDV RNA complexes. This study will provide structures and a biochemical understanding of a DNA-dependent-RNAP that is able to use both DNA and RNA in the same active site for polymerization and RNA cleavage.

3.2.4 HDAg

The HDAg protein encoded by HDV is in two isoforms containing different C-termini, HDAg-S (195 amino acids) or HDAg-L (214 amino acids). HDAg-S is involved in HDV replication and transcription (Chao, M. et al., 1990) and is a RNA-binding protein composed of basic amino acid residues which interacts with partially double-stranded HDV RNA. HDAg-S interacts with Pol II to stimulate the in vitro step of DNA and RNA transcription elongation (Yamaguchi, Y. et al., 2001; Yamaguchi, Y. et al., 2002).

A weak sequence similarity exists between HDAg-S and the largest subunit (A) of negative elongation factor (NELF) (also described in detail in Chapter 1). Authors suggest that binding locations on Pol II will be similar since both factors have an affinity for Pol II (Narita, T. et al., 2003). An additional involved transcription factor, DRB sensitivity-inducing factor (DSIF) (described in detail in Chapter 1), is also known as Spt4/5. Studies suggest that HDAg-S stimulates transcription elongation through two separate methods. In the first, HDAg-S counteracts the negativity of DSIF and NELF by removing NELF from Pol II. In the second, the HDAg-S interaction with Pol II improves elongation (Yamaguchi, Y. et al., 2001), by increasing forward Pol II translocation (Nedialkov, Y.A. et al., 2003). Based on photocrosslinking HDAg-S is seen to interact with the Pol II clamp. Additional transcription assays help suggest a model where HDAg-S loosens the clamp enabling forward acceleration of translocation (Yamaguchi, Y. et al., 2007).
3.2.5 Experimental summary

In this study, I aimed to understand the structural details of HDV-aRNAP complex in three stages: 1) pre-cleavage, 2) post-cleavage and 3) RNA extension. To achieve this goal, biochemical characterization is necessary to determine how to prepare stable HDV-aRNAP complexes for crystallization. Initial assays were used to purify aRNAP, design and order RNA oligonucleotides, and evaluate binding affinity between aRNAP and the different HDV RNAs. Once binding between aRNAP and RNA was established, nucleolytic assays and identification of the cleavage site of the HDV RNA was required. The last experiments conducted were in vitro RNA extension assays to test the ability of aRNAP to extend the cleaved RNA. Crystallization trials were also examined initially with aRNAP-HDV1 in a pre-cleavage state.

3.3 Materials and Methods

3.3.1 Purification of *P. furiosus* RNAP

*P. furiosus* wet cell mass was purchased from the University of Georgia Fermentation Facility from 500 L fermentations. The wet cells (200 g) were lysed in 400 ml TGEDM (10 mM Tris-HCl (pH 8.0 at room temperature), 15 % glycerol, 0.1 mM EDTA, 1 mM DTT, and 10 mM β-mercapto ethanol (BME)) with 0.2 M KCl using a Microfluidics M-110EH-30 Microfluidizer Processor (Microfluidics) at 20,000 psi. Following centrifugation, the supernatant was submitted to 0.5% Polymin-P and ammonium sulfate precipitations overnight. The pellet from centrifugation was resuspended in 50 ml TGEMD and loaded on an equilibrated (TGEMD 0.2 M KCl) Heparin-Sepharose column (GE Healthcare) and eluted stepwise by TGEDM and 0.5 M KCl. The elutions were dialyzed overnight into TGEDM and 0.05 M KCl followed by loading on a SourceQ column. The protein was eluted with a linear gradient of TGEDM and 0.01 to 0.5 M KCl. Peak fractions were concentrated using a 1 ml Q column (GE Healthcare) and protein fractions were passed over a size-exclusion column (GE.
Healthcare Superdex200 16/60 120 ml). *P. furiosus* RNAP was eluted in three fractions, frozen by liquid nitrogen, and stored at -80°C. Fractions were evaluated by SDS-PAGE and Bradford assay for quality and concentration. From 200 g of wet cells we were able to obtain ~10 mg of pure RNAP.

### 3.3.2 Oligonucleotides

The RNA oligonucleotides were purchased from IDT and are shown in **figure 3.3** (Table 3.2). HDV1 RNA was comprised of a single 49 nucleotide RNA strand with hairpin loop and dinucleotide bulge, HDV2 contained the pre-cleaved segment of RNA only, and HDV3 was the fragment of cleaved RNA from the full RNA. These oligonucleotides were annealed at 90°C for 5 min and the temperature was slowly decreased (0.02°C/second) back to 21°C in a thermocycler (UNO II or Tgradient, Biometra). Annealing was necessary to verify that the RNA oligonucleotides have a uniform correct secondary structure (**figure 3.3**). When needed, RNA oligonucleotides were 5’-end labeled with [γ-32P] ATP and polynucleotide kinase (PNK, NEB) and were purified from unincorporated [γ-32P] ATP using G-25 Sephadex quick spin columns (GE Healthcare) prior to annealing.

A second set of oligonucleotides was needed for several of the assays. This set was labeled with a b following the name: HDV1b, HDV2b, and HDV3b. these oligonucleotides also were annealed as above and 5’ –end labeled as needed (Table 3.2).

<table>
<thead>
<tr>
<th>HDV</th>
<th>Sequence 5’--&gt; 3’</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDV1</td>
<td>AGAGCCACUUUUCUCGAAUUCUAAUCGGAUUCGAAUUGUGG</td>
</tr>
<tr>
<td>HDV1b</td>
<td>ACACCCACUUUUCUCGAAUUCUAAUCGGAUUCGAAUUGUGG</td>
</tr>
<tr>
<td>HDV2</td>
<td>AGAGCCACUUUUCUCGAAUUCUAAUCGGAUUCGAAUUC</td>
</tr>
<tr>
<td>HDV2b</td>
<td>ACACCCACUUUUCUCGAAUUCUAAUCGGAUUCGAAUUC</td>
</tr>
<tr>
<td>HDV3</td>
<td>UAGAGAGAUUUGUGG</td>
</tr>
<tr>
<td>HDV3b</td>
<td>UAGAGAGAAAGUGG</td>
</tr>
</tbody>
</table>
3.3.3 Binding assays of *P. furiosus* RNAP to HDV RNA

To test whether HDV RNA is able to form a stable complex with aRNAP, the HDV AG RNA oligonucleotide was used for an electrophoretic mobility shift assay (EMSA). HDV1 RNA was comprised of a single 49 nucleotide RNA strand with hairpin loop and dinucleotide bulge (Figure 3.3). HDV RNA was 5’-end $^{32}$P radio labeled and was purified by removing unincorporated [$\gamma$-$^{32}$P] ATP using G-25 Sephadex quick spin columns. Purified labeled HDV RNA was annealed at 90°C for 5 min and the temperature was slowly decreased (0.02°C/second) back to 21°C. Annealing was necessary to verify that the RNA oligonucleotides have a uniform correct secondary structure. The annealed RNA was combined with *P. furiosus* RNAP at room temperature for 15 min, loaded on Native-PAGE, and visualized by phosphorimaging with a Typhoon 8600 (GE Healthcare).

3.3.4 Nucleolytic assay

Nucleolytic assays were assembled by combining 25 nM 5’ end-labeled annealed RNA with 50 nM RNAP in buffer (12 mM Hepes pH 8.0 rt, 12 mM KCl, 0.6 mM DTT, 0.12 mM EDTA). The reactions were incubated in a thermocycler (UNO II or Tgradient, Biometra) for times, temperatures, and with additional factors as indicated. Each reaction was stopped with an equal amount of 2x stop buffer (10 M Urea, 200 mM EDTA, and a trace amount of bromophenol blue) and loaded on a 15 % polyacrylamide gel, (30 % 19:1 acrylamide: bisacrylamide, containing approximately 6.7 g urea) and separated by electrophoresis. The results were visualized by phosphorimaging using a Typhoon 8600 (GE Healthcare).
3.3.5 Identification of the HDV RNA cleavage site

3.3.5.1 HDV RNA cleavage

HDV RNA cleavage reactions were achieved as above with the addition of annealed templates, NTPs, aRNAP, and 10x buffer for 30 min at 69°C. The samples were placed on ice for 1 min then spun down and the RNA was isolated from the remaining sample using the RNA Qiagen RNeasy Mini kits (Qiagen, Inc., Valencia, CA). The final RNA was eluted in RNase-free water and the concentrations evaluated.

3.3.5.2 Tobacco acid pyrophosphatase (TAP) treatment

RNA, RNasin RNase Inhibitor, TAP (Epicentre), and RNase-free water incubated for 1 hr at 37°C, followed by RNA cleanup using the RNA Qiagen RNeasy Mini kits (Qiagen, Inc., Valencia, CA). The final RNA was eluted in RNase-free water and the concentrations evaluated.

3.3.5.3 Ligation

RNA, 10x T4 RNA Ligase Buffer (NEB) TAP-RNA RNA oligonucleotide primer, and T4 RNA Ligase (NEB) were incubated for 1 hr at 37°C, followed by RNA cleanup as above.

3.3.5.4 Kinase primers

Each of the primers, as seen in Table 3.3, were separately combined with 10x Kinase buffer, ATP, Polynucleotide Kinase (NEB) and incubated for 30 min at 37°C, followed by an additional 30 min at 65°C in a thermocycler and left overnight at 4°C.
Table 3.3 Primers used for identification of HDV RNA cleavage site

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1c</td>
<td>CCACAAATCTCTCT</td>
</tr>
<tr>
<td>1f</td>
<td>AGAGCCACTTTTCT</td>
</tr>
<tr>
<td>1bf</td>
<td>ACACCCACTTTTCT</td>
</tr>
<tr>
<td>1bc</td>
<td>CCACCTTTTCTCTCT</td>
</tr>
<tr>
<td>1fx</td>
<td>GCTCTAGAAGAGCCACT</td>
</tr>
<tr>
<td>1cx</td>
<td>GCTCTAGACCACAAATC</td>
</tr>
<tr>
<td>1bcx</td>
<td>GCTCTAGACCTTTTCT</td>
</tr>
<tr>
<td>1bfx</td>
<td>GCTCTAGAACACCCAC</td>
</tr>
<tr>
<td>TAPE</td>
<td>CGGAATTCGGTATTGCG</td>
</tr>
</tbody>
</table>

3.3.5.5 Reverse Transcriptase PCR (RTPCR) (Promega Kit)

The following were combined together in 5x buffer (final 1x concentrations 12 mM Hepes pH 8.0, 7.5 mM MgCl, 0.12 mM EDTA, 0.6 mM DTT, 12 mM KCl, 1 mM BME), dNTPs, MgSO₄, AMV Reverse Transcriptase, TF/DRNAP, RNA templates, and primers 1c, 1f, 1bf, 1bc (Table 3.3). The PCR tubes were placed in a thermocycler for temperatures and times as follows:

40 cycles steps 1-3,  
1 94°C 30 sec  
2 53°C 1 min  
3 68°C 2 min  
4 68°C 7 min  
5 4°C until removed from thermocycler.

3.3.5.6 PCR

Primers included were 1c, 1f, 1bf, and 1bc, as seen in Table 3.3. PCR reaction containing primers, dNTPs, MgCl, 10x Buffer, KOD DNAP, and templates.
Temperatures and times were as followed:

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4</td>
<td>94-95°C</td>
<td>2 min</td>
</tr>
<tr>
<td>2</td>
<td>94-95°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>3</td>
<td>53-55°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>4</td>
<td>72°C</td>
<td>1 min</td>
</tr>
<tr>
<td>5</td>
<td>72°C</td>
<td>2 min</td>
</tr>
<tr>
<td>6</td>
<td>4°C</td>
<td></td>
</tr>
</tbody>
</table>

The PCR products were purified on a 1.5% agarose 0.5x TBE gel, followed by gel-extraction and purification using a Qiagen mini-prep kits (Qiagen, Inc., Valencia, CA), and the DNA was eluted in RNase-free water and the concentrations were evaluated.

**3.3.5.7 PCR-2**

The primers included were 1cx, 1fx, 1bfx, and 1bcx, as seen in Table 3.3. PCR reaction containing primers, dNTPs, MgCl, 10x Buffer, KOD DNAP, and templates. Temperatures and times were as followed:

<table>
<thead>
<tr>
<th>Cycle</th>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-4</td>
<td>94-95°C</td>
<td>2 min</td>
</tr>
<tr>
<td>2</td>
<td>94-95°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>3</td>
<td>53-55°C</td>
<td>30 sec</td>
</tr>
<tr>
<td>4</td>
<td>72°C</td>
<td>1 min</td>
</tr>
<tr>
<td>5</td>
<td>72°C</td>
<td>2 min</td>
</tr>
<tr>
<td>6</td>
<td>4°C</td>
<td></td>
</tr>
</tbody>
</table>

The PCR products were purified on a 1.5% agarose 0.5x TBE gel, followed by gel-extraction and purification using a Qiagen mini-prep kits (Qiagen, Inc., Valencia, CA), and the DNA was eluted in RNase-free water and the concentrations were evaluated.
3.3.5.8 Digestion, ligation, and transformation

The DNA was digested with XhaI (Promega) and EcoRI (NEB) in buffer 2 (NEB) at 37°C for 2.5 hrs, 65°C for 20 min, followed by ligation using Mighty Mix (TAQARA) following the manufacture’s protocol. Transformation of the ligated product into Mach-T1 cells onto LB plates with Xgal and IPTG, were left at 37°C overnight.

3.3.5.9 Sequencing and evaluation of results

Purified and concentrated plasmids were sent to the Genomics Core Facility at University Park in The Huck Institutes of the Life Sciences at The Pennsylvania State University. The results were returned as computer files which were analyzed by sequence alignment for the cleavage site. I used Vector NTI to align the sequences to the original RNA and the vector DNA. The cleavage sites were identified by the interface between the original RNA and the vector DNA.

3.3.6 In vitro RNA extension assays

In vitro RNA extension assays were used to test the ability of P. furiosus RNAP to extend the cleaved RNA. These experiments were completed using 25 mM 5’-end labeled scaffolds combined with ATP (100 µM), UTP (100 µM), GTP (100 µM), CTP (100 µM) and P. furiosus RNAP (50 nM) in buffer. Unless otherwise specified, these assays were carried out in 10 µl volumes. Each reaction was stopped with an equal amount of 2x stop buffer (10 M Urea, 200 mM EDTA, and a trace amount of bromophenol blue) and loaded (2 µl per lane) on a 15 % polyacrylamide gel, (30 % 19:1 acrylamide:bisacrylamide, containing approximately 6.7 g urea) and separated by electrophoresis. The results were visualized by phosphorimaging using a Typhoon 8600 (GE Healthcare).
3.4 Results

3.4.1 Purification of *P. furiosus* RNAP

The purifications of RNAP for the HDV analysis were completed by both Dr. Akira Hirata and myself. The RNAP purification process detailed here was our most recent method which has been primarily used by Kevin Baker. The results described here and displayed in Figure 3.5 are for a *P. furiosus* RNAP purification completed by Kevin Baker. The *P. furiosus* wet cell mass was resuspended in 400 ml TGEDM 0.2 M KCl and lysed and a pellet formed following the Polymin-P and overnight in ammonium sulfate precipitations. The pellet from centrifugation was resuspended and loaded on a Heparin-Sepharose column and eluted (Figure 3.5a and b). Elutions 4-7, about 5 ml each were combined and dialyzed overnight. The following morning the solution containing RNAP was passed over a SourceQ column and the protein was eluted with a linear gradient and a SDS-PAGE was used to analyze the fractions (Figure 3.5c and d). Concentrated peak fractions were loaded on a size-exclusion column and the RNAP eluted in primarily two to three fractions, approximately 2.5 ml each (Figure 3.5e and f). The purification by Kevin Baker was compared to the final gel of a *P. furiosus* RNAP purification completed by myself (Figure 3.5g).
Figure 3.5 RNAP purification chromatograms and SDS-PAGE coomassie blue stained gels

The chromatogram and results ran on a SDS-PAGE stained with Coomassie blue for each. The arrows mark RpoA'/RpoB. A and B. The Heparin-Sepharose column and gel with lanes 1-7 as x1, x2, A3-A7 with the marker same as in G in lane 8. C and D. SourceQ column and gel in lanes 1-7 elutions 4-10 and the marker as in G in lane 8. E and F. Size-exclusion column and gel in lanes 1-5 elutions 3-6 and 9 and the marker as in G in lane 6. G. A previous *P. furiosus* RNAP purification that I had completed.
3.4.2 Binding assays of *P. furiosus* RNAP to HDV RNA

To evaluate whether HDV RNA was able to form a stable complex with aRNAP, the HDV AG RNA oligonucleotide was used for an EMSA. HDV1 RNA was comprised of a single 49 nucleotide RNA strand with hairpin loop and dinucleotide bulge (Figure 3.3). HDV RNA was 5’-end $^{32}$P radio labeled and was purified by removing unincorporated [$\gamma$-$^{32}$P] ATP using G-25 Sephadex quick spin columns. Purified labeled HDV RNA was annealed at 90°C for 5 min and the temperature was slowly decreased (0.02°C/sec) back to 21°C. Annealing was necessary to verify that the RNA oligonucleotides have a uniform correct secondary structure. The re-annealed RNA was combined with *P. furiosus* RNAP at room temperature for 15 min, loaded on Native-PAGE, and visualized by phosphoimaging.

The RNA was completely shifted upon the addition of excess RNAP and the RNA-RNAP complex on the Native-PAGE, suggesting that the aRNAP makes a stable homogeneous complex with HDV RNA (Figure 3.6). Additional experiments were completed to evaluate the binding of aRNAP with HDV2, HDV23 and HDV3 RNA. For each of these RNA the above labeling and annealing were completed prior to incubation with aRNAP. The Native-PAGE showed binding of aRNAP with HDV2 RNA and a combination of HDV2 and HDV3 RNA (Figure 3.6).
Figure 3.6 HDV RNA-RNAP complexes
For both A and B labeled RNA, and RNA bound to RNAP was present in figures. A. Binding assays with constant HDV RNA with RNAP in increasing concentrations. RNA was as in figure 3.3. B. Binding assays with constant HDV and RNAP in increasing concentrations. RNA was as in Table 3.2.

3.4.3 HDV RNA cleavage and extension site

A series of methods were combined together to determine the location that aRNAP binds to the HDV RNA and the site of RNA cleavage. These methods, used sequence analysis (completed using Vector NTI) to compare the sequences made from RNA cleavage, reverse transcription and cDNA cloning into plasmids using the original
HDV1 RNA and vector DNA. The cleavage sites were identified by the interface between the original RNA and the vector DNA and the site of RNA cleavage was determined to be at one of three sites (Figure 3.7). The most prevalent site was in the middle of the dinucleotide bulge. The other two sites were before the dinucleotide bulge, and in the base pair before the dinucleotide bulge, however both were less frequently used for cleavage than the first site (Figure 3.7c).

**Figure 3.7 Identifying HDV RNA cleavage sites**

A. Simplified methodology for isolation of RNA cleaved by RNAP. RNA, same sequence as shown in B, is shown as an outline. RNAP is shown in red. B. HDV RNA sequence with potential cleavage sites shown by arrows (color matches C). C. Sequencing results of DNA from reverse transcription in A.
3.4.4 *In vitro* RNA extension of HDV RNA using *P. furiosus* RNAP

The detailed *in vitro* RNA extension assay with *P. furiosus* RNAP in figure 3.8 shows the results of extension of RNA for both HDV1 and HDV1b. This assay expanded upon RNA cleavage around the dinucleotide site (the arrow or in red in Figure 3.8 on HDV1) and extended RNA based on the nucleotides added. In each lane aRNAP was combined with HDV RNA (1-6 HDV1 or 7-12 HDV1b), in the presence of Mg$^{2+}$ and the listed nucleotides. Lane 1 showed that HDV1 was capable of cleaving the RNA and added one G nucleotide. In lane 2 the addition of A and G allowed the addition of nucleotides back to the original 46 nucleotide location. The third lane contained A, G, and U which extended the RNA to the original 49 nucleotide location. In lanes 4 and 5 contained G and C or G and U neither extended RNA. In the final lane with HDV1 as a template, G, C, U, and A were all added to the reaction. There was a weak product present in the smear slightly larger than the addition of one G also a product ~53 nucleotides long. There were several darker products within the smear in lane 6. It is unclear why there are extension products longer than 53 nucleotides. These results suggest that aRNAP can use RNA as a template for RNA extension.

Lanes 7 thru 12 used HDV1b as a template for RNA extension. Lane 7 contained only G, and there are several weak products between 34 and the 49 nucleotide marker length and a darker product around 49 nucleotides long. In lane 8 the addition of A and G showed a primary product shorter than in lane 8. The following lane 9 contained A, G, and U and showed a 49 nucleotide length product and some faint shorter products. In lane 10 there was a product of the same length as in lane 7, along with a shorter product. In lane 11 only one strong product was visible, and the same length strong product was visible in the smear in lane 12. Also in lane 12 was one other strong product along with several other weak products within the smear, similar to lane 6.

It was unexpected that there were short cleavage products visible with HDV1b figure 3.8 in lanes 7-12, and that the extension products were not similar to those of HDV1 (Figure 3.8 lanes 1-6). A possible explanation for the occurrence of these products could be an altered form of HDV1b folding. The secondary structure of HDV1b
was predicted by mFold (Owczarzy, R. et al., 2008), but possibly not all of the RNA was in this form. To evaluate this further in vitro extension assays were performed with annealed (and 5’-end labeled for Figure 3.9b) HDV1b RNA (Figures 3.9c and d). For figure 3.9b this experiment was completed using 5’-end labeled HDV1b RNA to evaluate cleavage by aRNAP in the presence of 7.5 mM Mg$^{2+}$. Lanes 2-4 were the controls (Figure 3.9b), and the result shown in lane 1 was of a cleavage product closer in length to 49 than to 34, suggesting the HDV1b RNA was not cleaved near the dinucleotide bulge.

The results of the cleavage assay answered part of the question, however there was still the question about the HDV1b extension products. To evaluate the extension of HDV1b RNA, an assay using aRNAP and HDV1b RNA was combined with ATP (100 μM), UTP (100 μM), ([α-32P] GTP (10 μCi), in a buffer with 7.5 mM Mg$^{2+}$. This assay was different from that of figure 3.8 which made it difficult to compare directly between the two. In figure 3.9c there was a smear in lane 5 (α-32P-GTP (10 μCi) and ATP (100 μM)) with a strong dark product and several smaller products in the smear. Also in lane 5 there was an additional faint longer product (purple tilde). Lane 6 (α-32P-ATP (10 μCi) GTP (100 μM)) contained three products 34- (blue box), 49- nucleotides long, and one in between (red circle), about the cleavage length shown in figure 3.9b. Lane 5 suggested that aRNAP cleaves the HDV1b RNA indiscriminately and attached α-32P-GTP. Lane 6 however suggested that HDV1b RNA was in multiple folds, possibly the expected fold and at least one unknown fold. The third product from the top of the gel in lane 6 was almost the expected length of cleavage, so it either cleaved and was unable to add more nucleotides or it was another fold that already added all the nucleotides it could.

Figure 3.9d continued to explain the extension of HDVb RNA and these results observed the same assay as figure 3.9c however with HDV2b RNA. In lane 7 the same smear occurred, seemed more related the way the assay was set up than the actual RNA. The RNA product was about the length of the cleavage product in figure 3.9b. The only product in lane 8 was slightly shorter than the product in lane 7. As mentioned above for HDV1b, the secondary structure of HDV2b was also predicted by mFold (Owczarzy, R. et al., 2008), but possibly not all of the RNA was in this form.
Figure 3.8 Detailed in vitro RNA extension of HDV RNA with *P. furiosus* RNAP

Lanes 1-6 used HDV1 and lanes 7-12 used HDV1b as a template RNA. Location of RNA cleavage by aRNAP (red arrow). Each of these experiments followed the extension assays with the nucleotides as labeled in each lane. Assay was assembled with 25 mM scaffolds combined with ATP (100 µM), UTP (100 µM), [α-³²P] GTP (10 µCi), CTP (100 µM) as shown above the gel) and *P. furiosus* RNAP (50 nM). The assay was run at 68°C for 30 min followed by stop treatment and loaded on gel.
Figure 3.9 In vitro RNA extension of HDV RNA with P. furiosus RNAP

For each of the gel results (B-D) lane 1 contains size markers. The assays were run for 30 min at 68°C prior to stopping and loading on the gel. A. HDV RNA templates HDV1b and HDV2b. The sequence difference between was shown in blue. B. HDV1b RNA cleavage by P. furiosus RNAP in the presence of Mg\(^{2+}\). HDV1b RNA was 5’-end labeled with γ\(^{32}\)P-ATP and purified. Reactions were carried out in the presence of buffer with and without Mg\(^{2+}\) and with or without αRNAP. Arrows indicate products for full length RNA (R) and cleaved RNA (C). Lanes 2-4 were the controls, lane 2 was in the presence of Mg\(^{2+}\) without aRNAP, 3 was the reverse, and 4 was without both. C. HDV1b RNA extension in the presence of α-\(^{32}\)P-GTP and ATP, or α-\(^{32}\)P-ATP and GTP, along with 100 µM UTP, buffer with 7.5 mM MgCl\(_2\), 50 nM P. furiosus RNAP, and 25 nM HDV RNA. Product (P) was labeled by an arrow. Additional symbols were used for explanations in the text. D. Is the same as in C, but HDV2b RNA was used to replace HDV1b RNA for the template RNA.

3.4.5 X-ray crystal structure work on aRNAP-RdRP

The three structures that I proposed to solve in this analysis were 1) pre-cleavage, 2) post-cleavage and 3) RNA extension (Figure 3.10). I planned to prepare binary
complex crystals including αRNAP and HDV RNA without magnesium or NTPs to capture the pre-cleavage state (Figure 3.10 top). Soaking magnesium into the preformed binary complex crystals will advance the complex to the post-cleaved state (Figure 3.10 middle), while also adding ATP and GTP will further advance the complex to RNA extension state (Figure 3.10 bottom).

**Figure 3.10 Model for HDV RNA cleavage**
Model for cleavage of HDV RNA in three stages: 1) pre-cleavage, 2) post-cleavage and 3) RNA extension. The RNAP is in magenta and RNA is in black folded into a double stranded secondary structure with a dinucleotide bulge during pre-cleavage. With the addition of Magnesium cleavage will occur and the RNA will splice. The addition of nucleotides will allow extension of the RNA following the HDV RNA as a template.

At the time I initiated this project there was no reproducible or competent *in vitro* HDV RNA replication or transcription system using Pol II (Fu, T.-B. and Taylor, J. 1993). Unfortunately, while I was working on these assays and developing a protocol for crystallization and structure determination, Patrick Cramer’s group published the paper “Molecular basis of RNA-dependent RNA polymerase II activity” (Lehmann, E. *et al.*, 1993).
This paper detailed a Pol II-RNA-RdRP with short sequences of RNA similar to HDV. The structures discussed in this paper showed RNA in the active site. Thus, I discontinued my work on this RdRP project.

3.5 Discussion

In this analysis, I discovered that archaeal DNA-dependent RNAP is able to synthesis RNA using RNA as a template. The aRNAP is able to bind to and transcribe a short segment of a RNA oligonucleotide having a HDV genome sequence, making it an amenable candidate for X-ray crystallography studies.

The HDV RNA template including a potential RNAP binding site is comprised of a single 49 nucleotide RNA strand with a 5 nucleotide hairpin loop and a dinucleotide bulge. Based on in vitro assays I was able to show that the aRNAP forms a stable complex with the model HDV RNA. To determine the function of the aRNAP on RNA I carried out extension assays in a variety of conditions. Adjusting the concentration of Mg$^{2+}$ showed cleavage of the RNA. Through further evaluation of the cleavage site, I was able to determine three possible locations for RNA cleavage by aRNAP as well as how often these cleavage sites occurred. As shown in figure 3.7, the most often cleaved site around the dinucleotide bulge aligns with previously suggested digestion sites. The flexibility caused by the dinucleotide bulge could allow for the cleavage. Thus, in the presence of Mg$^{2+}$, the enzyme cleaves the RNA around the dinucleotide bulge, but it is unclear if this cleavage is able to form a free 3’-RNA end in the active site allowing for RNA extension.

In vitro RNA extension assays were performed to establish the ability of aRNAP to cleave the RNA around the dinucleotide bulge to form a free 3’-RNA end and carry out RNA-directed RNA extension, without additional factors. The assays in figure 3.8 were able to show this, however it seems that HDV1 was cleaved but the addition of G only was added non-specifically. The addition of nucleotides GA and GAU to HDV1 was as expected. The product from the addition of nucleotides GAUC showed the expected length as well as a smear. The HDV1b results were expected to be similar; however in
figure 3.8 lanes 7-12 there were small cleavage products between lengths 49 and 34 nucleotides long, and the main products were about 49 nucleotides in length. So a second assay was used to further analyze the HDV1b and HDV2b RNA (Figure 3.9). It seemed both cleavage and extension occurred, however not to the expected lengths if the aRNAP was binding to the RNA in the predicted folds. Thus, not all of the HDV1b and HDV2b were folded as predicted by mFold (Owczarzy, R. et al., 2008). The predictions from mFold for HDV1 and HDV2 were not even close to the figures drawn in this dissertation (Owczarzy, R. et al., 2008). The idea for HDV1 and HDV2 RNA came from the HDV AG103, which was also where the structures came from.

At the time, this was the first display of a Pol II-like RNAP transcribing a physiologically relevant RNA using RNA as a template. This suggested that aRNAP would be a possible model system to investigate HDV RNA-directed RNA synthesis in vitro and be useful in a strategy to provide high resolution X-ray crystal structures of RNAP-HDV RNA complexes.

Unfortunately, after several crystallization conditions were set up, Patrick Cramer’s group published a paper detailing the structure of an RdRP using Pol II and short RNAs simulating HDV RNA (Lehmann, E. et al., 2007). This paper used RNA templates derived from HDV, using terminal stem-loops which showed incorporation of the subsequent nucleotide. Each of these structures shows RNA in the active site. However, these structures do not show the dinucleotide bulge, cleavage intermediates, or the product HDV RNA. Additional structures would help show the entire picture of RNA cleavage and synthesis necessary for the HDV life cycle.

3.6 Acknowledgments

I thank Dr. Akira Hirata for providing the RNAP. This work was supported by The Pew Scholars Program in the Biomedical Sciences and supported in part by NIH grant GM071897.
3.7 References


Chapter 4

Inhibition of An Archaeal RNA Polymerase Capable of Intrinsic and Transcription Factor S Enhanced RNA Cleavage

4.1 Abstract

Transcription by DNA-dependent RNA polymerase (RNAP) involves an intricate system of regulators to produce accurate RNA from DNA. RNA synthesis occurs during transcription elongation, while archaeal RNAP (aRNAP) incorporates nucleotides following the template DNA strand. This process is not foolproof and incorrect bases can be introduced. The removal of the introduced mismatches occurs by an aRNAP proofreading mechanism, which are intrinsic and transcription factor S (TFS) stimulated RNA cleavage. Both RNA synthesis and cleavage reactions occur in a single aRNAP catalytic site (Sosunov, V. et al., 2003).

TFS, TFIIS, and bacterial homologues (GreA and GreB) all contain the evolutionally conserved dipeptide Asp-Glu in the zinc finger. Mutation studies of this catalytic dipeptide in eukaryotes and bacteria has shown these residues to be critical for RNA cleavage, and suggests that the catalytic residues in TFIIS and Gre factors are involved in coordinating the RNAP active site metals, which effects RNA cleavage. Currently, no biochemical data is available that evaluates the role of the Asp-Glu dipeptide in archaeal TFS.

My study demonstrates that aRNAP is capable of intrinsic and TFS enhanced RNA cleavage of a transcription elongation complex (TEC) assembled on a DNA/RNA scaffold. In addition, I show that it is possible to adjust the conditions to inhibit RNA cleavage in order to capture the cleavage mechanism for crystallization, with the goal of determining the X-ray crystal structure of the proofreading TEC, with or without TFS. In one of the modifications a phosphorothioate bond derivative of the RNA in the DNA/RNA scaffolds enables inhibition of RNA cleavage by the introduction of sulfur
atoms to replace nonbridging oxygen atoms. The modified RNA in the DNA/RNA scaffold also keeps the TEC in the backtracked form and may help form crystals with homogenous complexes. Additional modifications include altering the metals at the active site and lowering the temperature of the reaction. The final adjustment made was the preparation of the TFS double mutant, TFS[D90A/E91A], using site-directed mutagenesis of Asp90 and Glu91 each to Ala. TFS was able to cleave the backtracked RNA in a stepwise manner; however, TFS[D90A/E91A] did not result in any products for cleavage of the backtracked RNA.

4.2 Introduction

4.2.1 Proofreading

Transcription by DNA-dependent RNA polymerase (RNAP) involves an intricate system of regulators to produce accurate RNA from DNA. During transcription elongation, archaeal RNAP (aRNAP) incorporates nucleotides to synthesize RNA following the template DNA strand. This process is not foolproof, incorrect bases can be introduced, and removal of the mismatches occurs by an aRNAP proofreading mechanism. Both RNA synthesis and cleavage reactions occur in a single aRNAP catalytic site (Figure 1.13; Sosunov, V. et al., 2003).

Two proofreading mechanisms used by aRNAP are intrinsic and transcription factor S (TFS) stimulated RNA cleavage. Several RNAPs have the ability to intrinsically cleave RNA, including viral RNAPs (Hagler, J. and Shuman, S., 1993), bacterial RNAPs, and eukaryotic RNAPs. RNAP from Methanococcus thermolithotrophicus was found to be capable of intrinsic RNA cleavage with a promoter-initiated immobilized stalled elongation complex in the presence of metal (Mg\(^{2+}\) or Mn\(^{2+}\)); and studies with *M. thermolithotrophicus* TFS, which contains the C-terminal domain of the eukaryotic homologue TFIIS (Figure 4.1), showed stimulated RNA cleavage similar to TFIIS. In addition, TFS reactivates transcription of an arrested elongation complex, caused by misincorporated nucleotides, in RNA that is moderately backtracked or not at all (Lange,
In some cases backtracking of RNA occurs, and intrinsic RNA cleavage by RNAP is possible when RNA is backtracked by only a couple of bases. However, when RNAP backtracks a greater distance (Figure 4.2), the presence of a TFIIS homologue is necessary for RNA cleavage in the transcription elongation complex (TEC) (reviewed by Wang, D. et al., 2009).

TFIS, TFIIS, and bacterial homologues (GreA and GreB) all contain the evolutionally conserved dipeptide Asp-Glu in the zinc finger. TFIIS mediated cleavage is more than 100-fold faster than intrinsic RNA cleavage \textit{in vitro} and is probably the primary \textit{in vivo} mechanism (Wang, D. et al., 2009).

\textbf{Figure 4.1 TFIIS domain orientation and TFIIS-RNAP structure}
A. Domain orientation numbering for eukaryotic \textit{Saccharomyces cerevisiae} TFIIS. The white is domain I, domain II is in green, yellow is the linker region, and in orange is domain III (Kettenberger, H. et al., 2003). B. The RNAP-TFIIS ribbon diagram is composed of the 12 Pol II subunits in silver, a pink sphere for the active site metal ion. TFIIS is depicted as in A, and zinc ions as cyan spheres. Domain III of TFIIS is inserted into Pol II with the two acidic hairpin residues near the active site metal. The left view is from the side and the right view is from the front (views similar to Gnatt, A.L. et al., 2001; Armache, K.J. et al., 2003; Cramer, P. et al., 2000; and Cramer, P. et al., 2001; Images modified from Kettenberger, H. et al., 2003).
Mutation studies of the catalytic dipeptide present in the zinc finger of eukaryotes (Asp261 and Glu262 in TFIIIS) and bacteria (Asp41 and Glu44 in GreA and GreB) showed these residues to be critical for RNA cleavage. This suggested that the catalytic residues in TFIIIS and Gre factors are involved in coordinating the RNAP active site metals, affecting RNA cleavage. Currently, no biochemical data is available evaluating the role of the Asp-Glu dipeptide in archaeal TFS. A better understanding of the proofreading mechanisms is essential, as mutations during transcription increase the potential of proteins to aggregate, which is seen in several diseases including
Huntington’s and Alzheimer’s (Roy, H. and Ibba, M., 2006). Additionally, TFIIS was found to possibly be a good target for cancer studies (Hubbard, K. et al., 2008).

### 4.2.3 Prior structural summary

Structural studies of the aRNAP TEC alone, and in complex with TFS, are crucial for understanding the proofreading mechanism of aRNAP. X-ray crystallographic studies of multi-subunit RNAPs have provided abundant insights into the RNAP structure-function relationships. Many structures have been solved for eukaryotic and bacterial core RNAPs (reviewed in Cramer, P., 2002) and several X-ray crystal structures of aRNAP have also been solved including *Sulfolobus solfataricus* at 3.4 Å resolution and (*Sulfolobus shibatae* (Korkihin, Y. et al., 2009)). Both aRNAP structures show strong similarities with eukaryotic RNAP II (Hirata, A., et al., 2008; Korkihin, Y. et al., 2009). Additionally, several RNAP II TEC crystal structures have been determined at moderate resolutions by the Kornberg and Cramer groups, along with the structures of GreB, the RNAP II-TFIIS, and RNAP II-TFIIS-DNA/RNA complexes (Brueckner, F. et al., 2009). Several different backtracked-Pol II complexes of different lengths (Wang, D. et al., 2009). The high resolution structure, from *Thermus thermophilus* shows detailed coordination of the metals and nucleic acids in the active site (Vassylyev, D.G. et al., 2007a). This structure was solved with a scaffold composed of three oligonucleotides, both template and non-template DNA, and RNA. The scaffold TEC avoids many complications, such as a low efficiency of transcription to the desired elongation location. Also, promoter-initiated transcripts do not allow for nascent RNA that is shorter than 9-12 nucleotides due to abortive initiation. Structures of the archaeal TEC and RNAP-TFS-DNA/RNA with the RNA 3’ OH in the active site, and with the RNA backtracked in the TEC remain unsolved (Kettenberger, H. et al., 2004)
4.2.3 Experimental summary

Structural and biochemical studies are necessary to understand the binding of TFS to the aRNAP and the coordination of the metals and amino acids involved in RNA cleavage. Therefore, I set out to study intrinsic and TFS enhanced RNA cleavage by aRNAP with a synthetic scaffold. I demonstrate that it is possible to inhibit intrinsic aRNAP RNA cleavage for TECs assembled with DNA/RNA scaffolds with several methods. Additionally, I constructed a TFS variant, TFS[D90A/E91A], which is unable to stimulate TFS enhanced aRNAP cleavage. These conditions will enable further biochemical studies and crystallization for structural determination of nucleolytic RNA cleavage by aRNAP.

4.3 Methods

4.3.1 Purification of RNA polymerase

Protein purified T. kodakarensis RNAP was a gift of Drs. Thomas J. Santangelo and John Reeve at Ohio State University. I was analyzing RNAPs from several different organisms all at once, and I focused on T. kodakarensis RNAP for this study so I could compare the results in the presence or absence of the T. kodakarensis TFS I had available. The T. kodakarensis RNAP was stored in a buffer containing MgCl₂ which, was added to the metals listed for each experiment with this RNAP. All other RNAPs were purified as specified in Chapter 3.

4.3.2 Purification of recombinant TFS

T. kodakarensis TFS was over-expressed in E. coli BL21 (DE3) – CodonPlus cells (Stratagene). The protein was expressed in a pre-warmed 4L flask containing 1.5 L Luria-Bertani media (LB) with 100 µg/ml ampicillin at 37°C until OD₆₀₀ reached 0.4-0.6. The flasks were then moved to room temperature without shaking for 10 min followed by
adding 0.5 mM IPTG to start induction. The flasks were transferred to a shaker at 22°C for ~18 hrs and cells were harvested by centrifugation.

The cell lysate was heated at 55°C for 20 min, followed by centrifugation. The supernatant was applied to a 5 ml HiTrap Q FF column (GE Healthcare) equilibrated with buffer B (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM β-Mercaptoethanol, 0.3 M NaCl, 1 mM DTT, 5 % glycerol, 10 µM ZnSO$_4$). The flow through fraction was collected and applied to a size exclusion Superdex75 (26/60) column (GE Healthcare) equilibrated with buffer C (50 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 10 mM β-Mercaptoethanol, 0.2 M NaCl, 1 mM DTT, 5 % glycerol, 10 µM ZnSO$_4$). Peak fractions containing TFS were collected, and buffer was exchanged with buffer D (12 mM Hepes-NaOH pH 8.0, 0.1 mM EDTA, 1 mM β-Mercaptoethanol, 12 mM KCl, 0.6 mM DTT, 5 % glycerol) during concentration using a 5 kDa MWCO centrifugal concentrator (VivaScience). Samples were stored at 4°C.

The TFS double mutant, TFS[D90A/E91A], was prepared using site-directed mutagenesis of Asp90 and Glu91 to Ala following a commercial mutagenesis protocol (Stratagene). The plasmid containing TFS was PCR-amplified with primers (5’ AGGGCTGGAGCTGCACCGAGCACGA 3’ and 5’ TCGTGCTCGGTGCAGCTCCAGCCCT 3’) containing a GACGAG-GCTGCA mutation followed by digestion with DpnI (NEB) to eliminate any template plasmid. The digested sample was PCR purified (Qiagen) and transformed into Mach-T1 cells. The plasmids were sequenced for the correct mutation, then transformed into BL21 (DE3) RIPL cells. The same expression and purification method was used for TFS[D90A/E91A] as TFS above, with the exception that approximately 6 g of cells were harvested for TFS[D90A/E91A].

4.3.3 Oligonucleotides

Several synthetic oligonucleotides were used in this study including those that were all based around one primary scaffold shown in figure 4.3 (tec, tec_BT2, tec_BT3, tec*, tec_BT2*). These oligonucleotides were ordered from IDT and Dharmacon. The
DNA/RNA scaffolds were assembled by annealing RNA with DNA templates at 90°C for 5 min and the temperature was slowly decreased (0.02°C/sec) to 21°C in a thermocycler (UNO II or Tgradient, Biometra). The non-template DNA was added to the DNA/RNA hybrid to complete the scaffold. When needed, RNA oligonucleotides were 5'-end labeled with [γ-32P] ATP and polynucleotide kinase (PNK, NEB) and were purified from unincorporated [γ-32P] ATP using G-25 Sephadex quick spin columns (GE Healthcare) prior to scaffold assembly.

**Figure 4.3 Scaffolds involved in this study**

**A.** Structure of ttEC showing RNA/DNA scaffold with RNA, template DNA, and non-template DNA in yellow, red, and blue respectively (Modified from reference Vassylyev, D. *et al.*, 2007a). **B-E.** DNA/RNA scaffolds with same color coding as in A. **B.** TEC scaffold. **C.** Backtracked form of TEC. **D.** TEC with RNA modified to contain phosphorothioate bonds (*) to inhibit nucleolytic cleavage. **E.** Backtracked TEC with modified RNA.

<table>
<thead>
<tr>
<th>Scaffolds</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>B tec Scaffold</td>
<td>GAGUCUGCCGGCGCGCG CGCGCGCGCTTGCGGTCTGTCCC-5'</td>
</tr>
<tr>
<td></td>
<td>AACGCCAGACAGGG-3'</td>
</tr>
<tr>
<td>C tec_bt2 Scaffold</td>
<td>GAGUCUGCCGGCGCGCG CGCGCGCGTTTGCGGTCTGTCCC-5'</td>
</tr>
<tr>
<td></td>
<td>AACGCCAGACAGGG-3'</td>
</tr>
<tr>
<td>D tec* Scaffold</td>
<td>GAGUCUGCCGGCGCGCG CGCGCGCGCTTGCGGTCTGTCCC-5'</td>
</tr>
<tr>
<td></td>
<td>AACGCCAGACAGGG-3'</td>
</tr>
<tr>
<td>E tec_bt2* Scaffold</td>
<td>GAGUCUGCCGGCGCGCG CGCGCGCGTTTGCGGTCTGTCCC-5'</td>
</tr>
<tr>
<td></td>
<td>AACGCCAGACAGGG-3'</td>
</tr>
</tbody>
</table>

### 4.3.4 Electrophoretic mobility shift assay (EMSA)

Electrophoretic mobility shift assays EMSAs were performed by combining RNAP with scaffolds containing 5'-end labeled RNA in the presence of buffer E (12 mM Heps-NaOH pH 8.0, 0.12 mM EDTA, 12 mM KCl, 0.6 mM DTT) for 30 min at room temperature (approximately 20-24°C). The samples were loaded on a 4-15% native-
polyacrylamide gel (PhastGel™ Gradient, GE Healthcare), separated by electrophoresis (PhastSystem, GE Healthcare), and visualized with phosphorimaging by Typhoon 8600 (GE Healthcare), or by coomassie staining.

**4.3.5 Nucleolytic Assay**

Nucleolytic assays were assembled by combining 25 nM 5’ end-labeled RNA/DNA scaffold with 50 nM RNAP in buffer E (12 mM Hepes pH 8.0 room temperature, 12 mM KCl, 0.6 mM DTT, 0.12 mM EDTA). The reactions were incubated in a thermocycler (UNO II or Tgradient, Biometra) for times, temperatures, and with additional factors as indicated. Unless otherwise specified, these assays were carried out in 10 µl volumes. Each reaction was stopped with an equal amount of 2x stop buffer (10 M urea, 200 mM EDTA, and a trace amount of bromophenol blue) and loaded on a 15% polyacrylamide gel, (30% 19:1 acrylamide:bisacrylamide, containing approximately 6.7 g urea) and separated by electrophoresis. The results were visualized by phosphorimaging on a Typhoon 8600 (GE Healthcare).

**4.3.6 Extension of RNA cleaved products**

RNA extension assays were completed for intrinsically and TFS enhanced cleaved RNA. The nucleolytic assays were completed, 11 µl of the sample was removed, and stopped to be used as a control. To the 9 µl remaining, 1 µl of NTPs (2 mM each: ATP, UTP, CTP, and GTP) was added and incubation was repeated for the times and temperatures indicated in the figures. These assays were visualized as the nucleolytic assays above.
4.4 Results

4.4.1 Purification of TFS and TFS double mutant

The expression and purification of TFS was completed for both biochemical and crystallization purposes therefore the protein needed to be both very pure and active. The cells containing tfs and the double mutant were expressed and harvested. The cells were lysed and purification was completed for both TFS and TFS[D90A/E91A]. The lysed cells were heat treated and spun down. The supernatant was loaded on a Q column and the flow through fraction was collected and applied to the size exclusion Superdex75 (26/60) column (Figures 4.4 and 4.5). Peak fractions containing TFS were collected, the protein was concentrated, and buffer was exchanged for crystallization. TFS and the TFS double mutant protein samples were stored at 4°C.
Figure 4.4 Purification of TFS
The chromatogram and results ran on a SDS-PAGE stained with Coomassie blue for each. The black arrows mark TFS. The blue is the UV absorbance at 280 nm, the red columns at the bottom indicate fractions, and brown represents measured conductivity. A sampling of the collected fractions, C6, C9, C11, D12, and E9 were analyzed by 8-25% SDS-PAGE to show the purity of TFS in each fraction.
Figure 4.5 Purification of double mutant TFS[D90A/E91A]

The chromatogram and results ran on a SDS-PAGE stained with Coomassie blue for each. The black arrow marks the double mutant. The blue is the UV absorbance at 280 nm, the red columns at the bottom indicate fractions, and brown represents measured conductivity. A sampling of the collected fractions were analyzed, B9, C9, C11, D3, E7, E9, E12, and F4 by 8-25 % SDS-PAGE to show the purity of TFS[D90A/E91A] in each fraction.

4.4.2 Archaeal RNAP can bind a DNA/RNA scaffold to form the TEC

Scaffolds have been used for authentic TECs of bacterial and eukaryotic RNAPs, which are active to extend RNA. These scaffolds are also useful for solving X-ray crystal structures. The synthetic scaffolds used in this study were designed to be similar to the scaffold used to determine the X-ray crystal structure of the ttEC. I combined a set of template and non-template DNA with the 5’-end labeled RNA oligonucleotide to generate a scaffold to mimic the elongation complex. To evaluate the ability of archaenal
RNAP to interact with the DNA/RNA scaffolds (Figure 4.6), I carried out EMSAs. I evaluated EMSAs from two perspectives. The first, keeping the DNA/RNA scaffold constant and varying RNAP, and the second was to do the reverse (keep RNAP constant, vary the scaffold). I used 5’-end labeled RNA in the scaffold with native-PAGE to visualize the results of the first perspective, and coomassie stained SDS-PAGE for the second. The results of the SDS-PAGE showed a faint smear (difficult to see) with RNAP only; however in the presence of either scaffold, turned into a single product (Figure 4.6c). In the case of the native-PAGE, each of the scaffolds were incubated with the RNAPs in increasing amounts and the results showed that the aRNAP made complexes with the scaffolds in possible alternative forms. This was suggested by the extra products visible in the binding assays (Figures 4.6a and 4.6b). In the future it would be helpful to know if any of the complexes are resistant to heparin. Also evaluated was an alternative scaffold with a backtracked overhang of three nucleotides, and scaffold tec_bt3, as with tec_bt2, showed additional products (Figure 4.6d). Thus, tec_bt3 may also form possible alternative complexes with aRNAP.
Figure 4.6 Evaluation of binding affinity between aRNAP and scaffolds
For A and B with labeled scaffolds as in figure 4.3 and RNA bound to RNAP was present in all figures, and C is a coomassie stained gel shift assay. A. EMSA with constant scaffold with *P. furiosus* RNAP in increasing concentrations. Dark blue square shows possible alternate RNA-aRNAP bound form. B. Same as in A using *S. solfataricus* RNAP. C. Coomassie stained EMSA with *S. solfataricus* RNAP constant and increasing concentrations of DNA/RNA scaffold as shown above gel. D. Same as in A for an EMSA, however using *T. kodakarensis* RNAP with tec_bt3 scaffold.
When RNAP carries out RNA cleavage, the DNA/RNA hybrid backtracks (Lange, U., 2004) producing 3’ unpaired nucleotides which are directed into the secondary channel (Kettenberger, H. et al., 2003). To mimic a backtracked RNAP complex, I designed scaffolds consisting of RNA with either two or three non-complementary nucleotides and tested an EMSA for binding to RNAP. As with the tec scaffold, both backtracked scaffolds (tec_bt2 and tec_bt3) showed two complexes that formed when bound to RNAP (Figure 4.6a and d).

4.4.3 RNAP can induce intrinsic RNA cleavage in backtracked TEC

To test whether RNAP was capable of intrinsic RNA cleavage, I incubated the backtracked DNA/RNA scaffolds with aRNAP in the presence of 10 mM magnesium. The RNA was separated by denaturing polyacrylamide gel, displaying cleaved RNAs. In the case of tec_bt2, containing 18 nucleotides with a two nucleotide overhang, the product was a 16 nucleotide RNA (Figure 4.7 lanes 7-9). The 16 nucleotide cleavage product was also the result when starting from tec_bt3, which contained 19 nucleotides with three backtracked nucleotides (Figure 4.8). These results indicated that RNAP can cleave the RNA at the junction between the DNA/RNA hybrid and the 3’ overhang.
Figure 4.7 Intrinsic RNA cleavage by RNAP

**A.** Scaffold tec_bt2, RNA containing 18 nucleotides with a two nucleotide overhang, the product was a 16 nucleotide RNA. The experiment was carried out for 60 min at 70°C with 50 nM *T. kodakarensis* RNAP and 25 nM scaffold. Possible impure starting RNA (purple dot) and resulting cleavage product (red asterisk).

**B.** Using tec_bt2, the schematic of the cleavage site. The DNA template in red, DNA nontemplate in blue and the RNA in yellow with the 5’-end 32P labeled.
**Figure 4.8 Intrinsic RNA backtrack cleavage by RNAP**

A. Scaffold tec_bt2, RNA containing 18 nucleotides with a two nucleotide overhang, the product was a 16 nucleotide RNA. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk). The experiment was carried out for 60 min at 70°C with 50 nM *T. kodakarensis* RNAP and 25 nM scaffold. B. Scaffold tec_bt3, RNA containing 19 nucleotides with a three nucleotide overhang, the product was also a 16 nucleotide RNA. Experiment was carried out in the same time, temperature and RNAP conditions as in A and used the same amount of scaffold as in A. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk). C. The schematic of the cleavage site. The DNA template in red, DNA nontemplate in blue and the RNA in yellow with the 5’-end ³²P labeled.

To optimize *T. kodakarensis* RNAP to intrinsically cleave RNA, I evaluated several *in vitro* conditions. Experiments showed that a RNAP without metals could not cleave RNA (Figure 4.7 lane 6); however the addition of Mg²⁺ promoted intrinsic RNA cleavage (Figure 4.7 lanes 7-9). Other metals, including Mn²⁺, and Ca²⁺, (10 mM each), were tested in place of Mg²⁺ with varying results. Of all the metals tested Mn²⁺ was the only one able to replace Mg²⁺ in these conditions to promote intrinsic RNA cleavage. Further studies showed that decreasing the metal concentrations for Mg²⁺ and Mn²⁺ to 7.5...
mM resulted in increased truncated products for Mn\(^{2+}\) than for Mg\(^{2+}\). However, when the concentration of Mn\(^{2+}\) was increased (greater than 50 mM), RNA cleavage was non-specific and appears to be a result of degradation by the metal. Increasing the concentration of Mg\(^{2+}\) from 10 mM to 50 mM resulted in slightly more truncated product; conversely the truncated product was no longer present with 100 mM Mg\(^{2+}\) (Figure 4.9 lane 3).

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mg [mM]</td>
<td>10</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

**Figure 4.9 Increasing the concentration of magnesium alters the RNA cleavage products**
Magnesium concentration listed above the gel. The (25 nM) tec_bt2 scaffold was used for this assay with 50 nM *T. kodakarensis* RNAP. Arrows pointing to starting 18 nucleotide and 16 nucleotide cleavage product. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk).

**4.4.4 Intrinsic RNAP RNA cleavage was inhibited at lower temperatures**

I evaluated a temperature gradient to determine the optimal temperatures for intrinsic RNAP RNA cleavage activity. The temperature gradient from 50 to 77.1°C showed that the optimal temperature for intrinsic RNA cleavage by *T. kodakarensis* RNAP was around 70°C (Figure 4.10). At 50°C very little cleavage was detected and the amount of cleavage product increased over the temperature range of 50-72.4°C. At
77.1°C the cleavage pattern showed nonspecific RNA degradation and a decrease of the full length RNA and the cleavage product (Figure 4.10 lanes 19-21). The higher temperature and metal concentration are probably the cause of the RNA degradation, not the RNAP (Hethke, C. et al., 1999).

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature °C</td>
<td>50</td>
<td>53.7</td>
<td>58.1</td>
<td>62.9</td>
<td>67.6</td>
<td>72.4</td>
<td>77.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time in Minutes</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
<td>10</td>
<td>20</td>
<td>60</td>
</tr>
</tbody>
</table>

A

B

Figure 4.10 Evaluation of temperature gradient for intrinsic RNA cleavage by RNAP

Temperature evaluation with arrows pointing to starting 18 nucleotide and 16 nucleotide cleavage product. The 5'-end 32P labeled RNA and the temperature and times listed above the gel. The (25 nM) tec_bt2 scaffold was used for this assay with 50 nM T. kodakarenensis RNAP and 50 mM Mg2+. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk).

4.4.5 Different conditions observed can prevent intrinsic RNA cleavage by RNAP

To solve the structures of the TEC during different stages of proofreading the RNA cleavage process must be prevented. Thus, in vitro assays were used to find conditions capable of inhibiting RNA cleavage. The previous results and results from the
metal examination showed that Ca\(^{2+}\) did not promote intrinsic RNA cleavage (Figure 4.11 lanes 2-4). However, this does suggest that Ca\(^{2+}\) could be a possible candidate for a metal replacement for Mg\(^{2+}\) during crystallization. A competition study in the future using Ca\(^{2+}\) and Mg\(^{2+}\) with the scaffolds to reveal if the RNA cleavage products present in reactions containing Mg\(^{2+}\) (10 mM), could be reduced with the addition of increasing amounts of Ca\(^{2+}\) would be useful. These findings, if they suggest calcium is capable of replacing the magnesium in the RNAP active site, would allow Ca\(^{2+}\) to be another alternative for crystallization.

<table>
<thead>
<tr>
<th>Lane</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>CaCl(_2) [mM]</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Lane</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>MgCl(_2) [mM]</td>
<td>0</td>
<td>10</td>
</tr>
</tbody>
</table>

**Figure 4.11 Calcium cleavage reactions**

On the left side are the results of a cleavage assay using calcium in place of magnesium. The magnesium results are shown on the right as a comparison. The 18mer RNA (tec_bt2) was used for both assays with metal concentrations shown above both assays. The (25 nM) tec_bt2 scaffold was used for this assay with 50 nM *T. kodakarensis* RNAP for 60 min at 70°C. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk).
Additional *in vitro* assays were carried out to test the ability of RNA modified with phosphorothioate bonds (Figure 4.12) to inhibit intrinsic cleavage of the backtracked RNA. Similar modifications were used to optimize a stable tertiary complex for the ttEC in structural studies. Internal phosphorothioate modified RNA was shown to exhibit resistance to cleavage by proteins involved in nuclease digestion (Vassylyev, D. *et al.*, 2007a). The EMSA showed that DNA/RNA scaffolds containing modified RNA, tec* and tec_bt2*, were able to bind RNAP (Figure 4.6). These scaffolds were each analyzed for RNA cleavage in nucleolytic assays containing RNAP and magnesium. The results of these assays showed diminished cleavage products for the scaffolds consisting of modified RNA compared to the unmodified RNA (Figure 4.12), demonstrating that the internal phosphorothioate modification in the RNA of the DNA/RNA scaffolds prevents intrinsic RNA cleavage.
Figure 4.12 Modified internal RNA alters intrinsic RNA RNAP cleavage phosphorothioate bonds to inhibit intrinsic tec_bt2*

A. The results of a cleavage assay using a phosphorothioate bond modified at internal RNA sites (tec_bt2* in place of tec_bt2). The (25 nM) tec_bt2 scaffold was used for this assay with 50 nM *T. kodakarensis* RNAP for 60 min at 70°C. The concentration of Mg^{2+} was as indicated above each lane. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk). B. The unmodified RNA as a comparison. C. The schematic of the cleavage site. The DNA template in red, DNA nontemplate in blue and the RNA in yellow with the 5'-end 32P labeled. The modifications are shown as a star.

4.4.6 TFS was able to enhance RNAP RNA cleavage

Previous studies have shown that TFS from *M. thermolithotrophicus* is capable of stimulating RNAP RNA cleavage with promoter-initiated transcripts in a manner similar to TFIIIS (Lange, U., 2004; Hausner, W., 2000). To evaluate the ability of *T. kodakarensis* TFS to enhance aRNAP RNA cleavage of a scaffold assembled TEC, I incubated aRNAP, DNA/RNA scaffold tec_bt3, and TFS in buffer E with Mg^{2+}. Addition of TFS to the *in vitro* nucleolytic cleavage assays resulted in a different cleavage pattern than the intrinsic RNAP cleavage (Figure 4.13). The products for intrinsic cleavage were present,
along with several extra products (Figure 4.13). Note that the TFS cleavage products occurred in a time dependent manner (Figure 4.13).

4.4.7 TFS dipeptide mutation prevents stimulation of RNAP RNA cleavage

Mutation studies of the dipeptide Asp-Glu in the zinc ribbon of TFIIS domain III and in the N-terminal coiled-coil tip of Gre showed that these acidic residues coordinate the RNAP active site metals (Jeon, C. et al., 1994). To test this property in TFS, I used site-directed mutagenesis to engineer the *T. kodakarensis* double mutant TFS[D90A/E91A]. In the future it will be necessary to examine the EMSA and competition studies to show that TFS[D90A/E91A] is bound to the TEC and that it is able to replace TFS in the *in vitro* assays. TFS[D90A/E91A] was tested for inhibition of RNA cleavage with backtracked DNA/RNA scaffolds. I repeated the nucleolytic assays comparing TFS with TFS[D90A/E91A], as was visualized previously *T. kodakarensis* TFS cleaved the backtracked RNA in a stepwise manner; however, TFS[D90A/E91A] did not result in any products for cleavage of the backtracked RNA (Figure 4.13).
Figure 4.13 TFS dependent RNA cleavage

A. Scaffold tec_bt3, RNA containing 19 nucleotides with a three nucleotide overhang, the product was a 16 nucleotide RNA. For lanes 1-5 independent of TFS, and for lanes 6-10, 250 nM TFS is added to the assay with 50 nM RNAP and 25 nM tec_bt2 at 70°C and buffer E. Arrows show TFS additional cleavage products on right side of gel. Impure starting RNA (purple dot) and resulting cleavage product (red asterisk). B. Comparison of TFS and TFS[D90A/E91A] using tec_bt3 following same protocol as in A. Lanes 1 and 2 contain evaluations with TFS and variation of TFS. The assay was completed with 50 nM RNAP and 25 nM tec_bt3 at 70°C for 30 min. Lane 3 is intrinsic RNAP cleavage (RNAP, buffer E, scaffold, Mg^{2+}). Lanes 4 and 5 are markers of the 19 nucleotide and 16 nucleotide RNA without buffer or Mg^{2+}.

4.5 Discussion

This study demonstrates that aRNAP is capable of intrinsic and TFS enhanced RNA cleavage of a TEC assembled with a DNA/RNA scaffold. In addition, I show that it is possible to adjust the conditions to inhibit RNA cleavage to capture the cleavage mechanism for crystallization, with the goal of determining the X-ray crystal structure of
the proofreading TEC, with or without TFS. DNA/RNA scaffolds were used to mimic different stages of the elongation complex and can be used for crystallization, unlike promoter-initiated transcripts which are often too long for uniform crystallization. The tec and tec* scaffolds imitate a TEC with the RNA 3’OH in the active site and the scaffolds tec_bt2, tec_bt2*, and tec_bt3, mimic the backtracked TEC. Both of these situations have been implicated in proofreading mechanisms. Authors studying aRNAP proofreading have shown data that does not favor the model where backtracking is required for cleavage induction (Lange, U. and Hausner, W., 2004) as suggested by authors studying bacterial and eukaryotic systems (Nudler, E., 1999; Cramer, P. et al., 2001; Gnatt, A.L. et al., 2001). This was concluded based on the result of assays preventing misincorporation of the wrong nucleotides and cleavage by TFS induced on a stalled elongation complex. As the conditions forcing incorrect nucleotide incorporation occurred, the cleavage events could not be increased, the authors concluded the primary reason for cleavage induction was the stalled elongation complex (Lange, U. and Hausner, W., 2004). This does not actually visualize the shift of translocation that occurs during elongation, thus, with the controversy over the TEC during proofreading, I have engineered scaffolds to mimic both models.

Pol II requires metal cofactors for intrinsic RNA cleavage. I have shown that this is also true of aRNAP TEC assembled on a scaffold. Intrinsic T. kodakarensis RNAP RNA cleavage in the presence of 10 mM Mg$^{2+}$ resulted in several cleavage products, including a 16 nucleotide, and a slightly shorter product. The variation (as seen in many of the figures 4.7 - 4.13, as the lavender dot, also seen in the marker in the impure starting RNA) in original transcripts could be from initial impure RNA synthesis, which could translate into the extra shorter cleavage product. In the case of 100 mM Mg$^{2+}$ it has the appearance of inhibiting RNA cleavage, however at high temperatures and high metal concentrations RNA is degraded (Hethke, C. et al., 1999). Thus, magnesium still needs to be replaced in the active site to prevent cleavage from occurring. The use of manganese in place of magnesium, also resulted in cleavage products, and I evaluated a variety of metals with T. kodakarensis RNAP, which were found incapable of stimulating the RNA cleavage including Ca$^{2+}$, Co$^{2+}$, Zn$^{2+}$, and Ni$^{2+}$. Since magnesium prefers a 2.1 Å distance
to coordinate six ligands in an octahedral geometry, substitution of \( \text{Mn}^{2+} \) for \( \text{Mg}^{2+} \) may be possible because manganese has an interatomic distance of less than 3.0 Å and is capable of coordinating six ligands. However, as an example, calcium may not replace magnesium because the closest two \( \text{Ca}^{2+} \) ions have been shown to be is 3.8 Å (Weis, W.I. et al., 1992), and \( \text{Ca}^{2+} \) ions can coordinate up to nine ligands, which may stabilize different transition state geometries than those with \( \text{Mg}^{2+} \), thus preventing intrinsic RNA cleavage (Figure 4.14).

\[
\text{Figure 4.14 Calcium as a catalytic metal ion}
\]
Model to show Calcium in the structural active center in place of magnesium. The 4 Å distance shown in the middle with the 6 active center geometry.

\textit{T. kodakarensis} typically grows in a range of temperatures from 60°C to 100°C with optimal growth at 85°C. Previous \textit{in vitro} assays with \textit{T. kodakarensis} were completed at 85°C. Thus, I evaluated a temperature gradient and determined the optimal temperature for intrinsic \textit{T. kodakarensis} RNAP RNA cleavage activity to be around 70°C (Figure 4.10). There was minimal cleavage at temperatures around 50°C and additional assays did not show cleavage products. The results showed the reaction could be left at 22°C for 1 hr without intrinsic RNA cleavage occurring, but crystallization trials are often left at 4°C (never tested) to 22°C for weeks or months at a time. Therefore, additional methods to inhibit cleavage are necessary for crystallization. The results from testing the metals that promote RNA cleavage, prompted further evaluation
of the ability of metals to inhibit intrinsic RNA cleavage. Replacing magnesium in the active site with particular metals may also enable the use of these metals for phase determination in solving the structures.

One further way to prevent intrinsic RNA cleavage is to modify the scaffolds. Therefore, I used a phosphorothioate bond modified RNA in the DNA/RNA scaffolds to examine inhibition of RNA cleavage (Figure 4.12). The modification introduced two or three sulfur atoms to replace nonbridging oxygen atoms. The structural determination of bacterial ttEC used this modification of the RNA in the DNA/RNA scaffold to prevent nucleolytic cleavage and backtracking of the RNA (Vassylyev, D. et al., 2007a). This study also showed that over time with the TEC in a pre-translocated state, nucleolytic cleavage occurred, which resulted in heterogeneous complexes in the crystal and was prevented by the modified RNA (Vassylyev, D. et al., 2007a). The chemically synthesized RNA oligonucleotide with the phosphorothioate modifications exists in a mixture of enantiomeric isomers with sulfur atoms replacing one of the two nonbridging oxygen atoms of the phosphate backbone. Both of the modified scaffolds, tec* and tec_bt2* were able to bind to RNAP based on the results of an EMSA. I examined the ability of the modification to prevent intrinsic RNAP RNA cleavage and found it to be successful with complete inhibition of RNA cleavage (Figure 4.12). If only one oxygen atom is involved, the mix of isomers should allow for only a partial prevention of RNA cleavage instead of complete inhibition. Thus my results suggest that both of the nonbridging oxygen atoms may be involved in the reaction (Figure 4.15). The modified RNA in the DNA/RNA scaffold will keep the TEC in the backtracked form, prevent RNA cleavage, and may help form crystals with homogenous complexes.
Figure 4.15 Model for the modified RNA in the DNA/RNA scaffold in the TEC-backtracked form
Proposed mechanism for multi-subunit RNAP with metals coordinated by the phosphate backbone. The two metal ions shown as MeA and MeB with arrows designating the direction of the electron density to be transferred. The red S are the replaced sulfur atoms for the non-bridging oxygen atoms. The x suggests that the electrons are not transferred and thus RNA cleavage does not occur.

I conducted experiments to evaluate the proofreading ability of TFS to stimulate RNAP RNA cleavage. A different RNA cleavage pattern was revealed during *T. kodakarensis* TFS enhanced RNA cleavage. As before, with the intrinsic RNA cleavage, several cleavage products were visible; however, with *T. kodakarensis* TFS several additional cleavage products occurred (Figure 4.13). These additional products, which occurred in a time dependent manner (Figure 4.13), were similar in intensity with the intrinsic cleavage products, and did not follow a pattern of nonspecific RNA degradation. TFS may enhance RNA cleavage beyond just removing the backtracked bases, but it may also continue to promote RNA cleavage. Previous analysis of TFIIS showed that some of the products released by TFIIS-stimulated RNA cleavage of both arrested and artificially stalled elongation complexes were different than those released by intrinsic RNA cleavage (Wang, D. *et al*., 2009). The authors suggested that TFIIS stimulates both the site and rate of cleavage (Wang, D. *et al*., 2009). In the case of TFS, the additional cleavage may be because the assay was completed in the absence of NTPs, preventing the
continuation of transcription elongation as cleavage occurred, allowing for further RNA cleavage. However, cleavage products, if in the presence of nucleotides, may or may not result in the positioning of the 3’OH in the active site, successfully or ineffectively enabling transcription elongation to resume. This could be tested in a future experiment by evaluating the presence or absence of extension of these products in a chase experiment. A similar reaction was carried out on a promoter-initiated transcript with aRNAP and TFS from *M. thermolithotrophicus*, which resulted in several active complexes post intrinsic and TFS enhanced RNA cleavage, as well as a 10 nucleotide RNA that was not extended in the presence of unlabeled nucleotides (Hausner, W., 2000).

*Thermus thermophilus* RNA polymerase showed a single nucleotide intrinsic RNA cleavage product, and with the addition of GreA the cleavage products released were dinucleotides only. Similar with my data in figure 4.13 these authors also showed progressive RNA cleavage and speculated that GreA may have the ability induce conformational changes in the RNAP to promote the backtracked conformation (Kashkina, E. *et al.*, 2006). In addition *Methanococcus thermolithotrophicus* TFS induced RNAP RNA cleavage was found to remove trinucleotides from the 3’ end prior to extension of RNA (Lange, U. and Hausner, W., 2004). The eukaryotic homolog of TFS, TFIIS, also mediates RNA cleavage with predominantly a dinucleotide product released by Pol II (Izban, M.G. and Luse, D.S., 1993). Thus, the results shown in figure 4.13 in lanes 7-10 where the cleavage is progressive and removes more nucleotides than intrinsic RNAP RNA cleavage.

These results overall suggest that intrinsic and TFS stimulated RNA cleavage have two different mechanisms of cleavage. From the results of figure 4.13 lanes 2-5 during intrinsic RNA cleavage aRNAP is still able to cleave RNA with a 2- or at least 3-mer overhang. However when TFS is introduced an altered cleavage patterned occurred, most likely the cleavage of one additional nucleotide (Figure 4.13 lanes 7-10). In future experiments it would be necessary to have markers for each nucleotide. Thus, the cleavage pattern suggests an altered conformation. This was shown to be the case in the structural results of superimposition of the backtracked-Pol II-TFIIS complex with the
backtracked pol II alone structures (Wang, D. et al. 2009); however in this overlay of structure the authors determined the need for an altered conformation because of a steric clash between TFIIS hairpin loop with the RNA of the backtracked scaffolds, the longer the backtracked RNA the more conflict occurred. Thus rearrangement is necessary, based on the disordered +2 site in the crystal structure (Wang, D. et al., 2009). They also found that wt TFIIS cleaved backtracked complexes RNA one nucleotide beyond where it was backtracked, for example if it was at two it would cleave at three (Wang, D. et al., 2009). Also similar to the TFS cleavage pattern in **figure 4.13 lanes 7-10**.

Past studies suggest that the TFIIS zinc ribbon Asp-Glu dipeptide is necessary for TFIIS enhanced RNA cleavage in eukaryotes (Jeon, C., Yoon, H. and Agarwal, K., 1994) and this mutation was used for the X-ray crystal structure of Pol II-TFIIS-DNA/RNA to inhibit RNA cleavage (Kettenberger, H., Armache, K.J. and Cramer, P., 2004); however, this mutation had not previously been tested with archaeal TFS. I engineered the dipeptide mutation in TFS[D90A/E91A] with site-directed mutagenesis, and the assays involving both TFS and TFS[D90A/E91A] were completed with identical conditions to rule out any additional contributing factors. The results of the assays involving TFS[D90A/E91A] showed only intrinsic RNA cleavage products. In the future, there remains a need to test the results of the EMSA and competition studies to rule out the possibility that the absence of RNA cleavage by TFS[D90A/E91A] is because of insufficient binding. This confirmation would allow TFS[D90A/E91A] to be a good candidate to determine the structure of RNAP-TFS-DNA/RNA, as the mutations will inhibit TFS enhanced RNA cleavage, but still conserve the structural features of the transcription factor.

Overall, these biochemical assays indicate that aRNAP is able to cleave RNA intrinsically and with the aid of TFS. I have shown through EMSAs that aRNAP can bind to DNA/RNA scaffolds with the 3’OH in the active site as well as backtracked scaffolds. Based on **in vitro** extension assays aRNAP is able to elongate RNA on DNA/RNA scaffolds (tec). I determined by nucleolytic assays that aRNAP can cleave RNA in backtracked DNA/RNA scaffolds (tec_bt2, tec_bt3). Purification of TFS was also shown as well as TFS mediated RNA cleavage by aRNAP. The results of TFS stimulated RNA
cleavage were much different from the intrinsic aRNAP cleavage. This showed that there were differences possibly between the mechanism of intrinsic and TFS-stimulated RNA cleavage by RNAP. My results agree with those of TFS homologs, and I propose that TFS aids in a conformational change in the active site to accommodate cleavage of longer backtracked RNA. However, to test this further, longer backtracked RNA scaffolds would need to be tested and a backtracked DNA/RNA-aRNAP-TFS complex X-ray crystal structure would help to explain the details. In order to enable crystallization, I have shown that it is possible to inhibit RNA cleavage to observe the TECs in different stages to be able to solve the X-ray crystal structures of the proofreading TEC, both with and without TFS. My results suggest that reactions containing a combination of different metals, lower temperature, modified RNA in the DNA/RNA scaffold, and a TFS variant will prevent both intrinsic and TFS enhanced RNA cleavage and enable the formation of crystals with uniform complexes for structure determination.

Throughout all of the biochemical studies, crystallization conditions were also evaluated to solve the X-ray crystal structures of the archaeal complexes of the aRNAP-DNA/RNA and RNAP-TFS-DNA/RNA, each with the RNA 3’ OH in the active site and with backtracked RNA.

4.6 References


Chapter 5

Summary and Future Directions

5.1 Introduction

This final chapter provides a summary of discoveries presented in this dissertation along with preliminary data and future directions for whoever continues this project. The projects described in Chapters three (Biochemical Characterization of RNA-Dependent RNA Transcription by Archaeal RNA Polymerase) and four (Inhibition of An Archaeal RNA Polymerase Capable of Intrinsic and Transcription Factor S Enhanced RNA Cleavage) are not going to be continued so they will not appear in this chapter; however, the following sections will contain an account of what is known based on my work and previous studies as well as the questions that remain for Chapters 2 (Crystal Structure of an Archaeal Homolog of the RNA Polymerase II Transcription Elongation Factor Spt4/5), parts of Appendices A (Growth and Handling of Pyrococcus furiosus Spt4/5 Crystals), B (Structure-based Biochemical Characterization of Pyrococcus furiosus Spt4/5) and C (Cryo-EM). Additionally, future directions are also presented here in the hope of giving the next person(s) continuing this project a place to start.

5.2 Summary and future directions for an archaeal homolog of the RNA polymerase II transcription elongation factor Spt4/5

The archaeal Spt4/5 factor with homologs in eukaryotes (Spt4/5) and bacteria (NusG) is involved in regulation of transcription elongation (Hirtreiter, A. et al., 2010; Guo, M. et al., 2008; Herbert, K.M. et al., 2010). Eukaryotic Spt4 and Spt5 have been studied both structurally and biochemically, suggesting the binding region of the N-terminal NusG (NGN) domain of Spt5 with Spt4 is conserved in archaea (Guo, M. et al.,
2008). In Chapter 2, I reported the X-ray crystal structure of the complete *P. furiosus* Spt4/5 complex solved at 1.8 Angstrom resolution.

This is the first complete Spt4/5 X-ray crystal structure containing Spt4 and the Spt5 N-terminal, linker and C-terminal domains. This structure contains all Spt4 zinc and NGN binding domains (β1-β4 and α1) and Spt5 Kyprides-Ouzounis-Woese (KOW). The Spt4 contacts Spt5 through all domains, the linker, KOW, and NGN; thus Spt4 may restrict the location of the KOW domain with respect to the NGN domain of Spt5. This suggests that the Spt5 linker between NGN and KOW is rigid and presents the KOW domain in a fixed position relative to the NGN domain. A sequence alignment of homologues revealed a bacterial specific insert in *Aquifex aeolicus*. Analyzing the structurally demonstrated minimal contacts between the bacterial insert and the rest of the *A. aeolicus* NusG structure (PDB ID 1NPR Knowlton, J.R, et al., 2003; Surface model comparison) and comparing it with the *P. furiosus* Spt4/5 structure suggests that the bacterial specific insertion is not a counterpart of Spt4 and thus cannot fix the orientation of the KOW domain to the NGN domain. However, this information lost a lot of strength when the rigidness of the Spt5 linker was called into question. Thus, future experiments are needed to determine if the protein-protein interactions in the crystal packing caused the linker to be rigid or if the structure is rigid.

Preliminary experiments carried out to evaluate Spt4/5 are described in Appendix B include site-directed mutagenesis to express and purify mutant Spt5 proteins as well as wild type (wt) Spt4 and Spt5 proteins. The mutant Spt5 proteins contain a combination of Y45A and, or E78A. I used nickel pull-down assays, and gel electrophoretic mobility shift assays (EMSA), to analyze the affinity of RNAP for Spt4/5, including the mutant proteins. Preliminary results show binding between *P. furiosus* Spt4/5 with aRNAP in the absence or presence of a DNA/RNA scaffold. Additionally, mutation studies of Spt4 and Spt5 double mutant Y45A and E78A resulted in decreased affinity in a nickel pull-down assay with aRNAP compared to the wild type Spt4/5 and aRNAP. Each of these preliminary studies needs to be repeated for duplication as well as for quantification.
There are several additional experiments that can be carried out to understand the Spt4/5 complex in archaea. Some of these experiments include footprinting, knock-out mutants, and chIP-chip to locate genomic DNA positions.

5.2.1 Further examine the structural interaction between Spt4/5 and aRNAP

The individual structures of Spt4/5 and aRNAP provide the framework to examine the Spt4/5-aRNAP complex structure. The cryo-EM reconstruction with *S. sulfataricus* RNAP followed by *P. furiosus* RNAP and extra density assigned to Spt4/5. The location of Spt4/5 is between the RNAP claw and in the elongation model, Spt4/5 prevents the DNA from exiting the RNAP.

X-ray crystal structures of both the Spt4/5-aRNAP and Spt4/5-aRNAP-DNA/RNA would help understand the interaction in more detail. More specifically these structures will aid in answering the following questions: first, does Spt4/5 interact with DNA/RNA as predicted in the cryo-EM model, and if so just the RNA or DNA? Second, is the structure rigid because of protein-protein interactions in the crystal packing or is the Spt5 linker in the *P. furiosus* Spt4/5 crystal structure rigid?

An additional possible structural method to be used to determine if the linker is rigid or not is small- and wide-angle X-ray scattering (SAXS/WAXS) and time resolved WAXS. SAXS uses X-rays at low angles to collect data about macromolecules such as shape and size, and even pore sizes. However, the fold of helices and sheets can be determined by WAXS (Hirai, M. et al., 2002). Some advantages of using SAXS/WAXS compared to X-ray crystallography is that no crystals are necessary and the sample can be in solution, therefore minimal sample is necessary and can range from 14 kDa to hundreds of kDa (Chen, L. et al., 1996; Svergun, D.I. et al., 1997; Montelione, G.T. et al., 2000). Thus, it would be possible to remove the crystal packing and observe *P. furiosus* Spt4/5 without the extra constraints. A disadvantage of SAXS/WAXS to X-ray crystallography is the loss of information. However, the low resolution from SAXS/WAXS can be overcome by combining the results with another structure method such as NMR or X-ray crystallography. Thus, solving the structure with SAXS/WAXS to
complement the *P. furiosus* Spt4/5 X-ray crystal structure (Chapter 2, Figure 2.4), and comparing the Spt5 linker region to determine if both structures are rigid, will show if the Spt5 linker is the same in solution as it is in crystal packing. Therefore, if the structures are different, then it plausible to admit at this point that the rigidness of the Spt5 linker of the *P. furiosus* Spt4/5 crystal structure is an effect of crystal packing. However, if the structures are similar or the same, then the crystal packing does not affect the linker and the next assay to study would be time resolved WAXS (TR-WAXS) with the addition of aRNAP. TR-WAXS could be used to examine if the Spt5 linker changes as it binds aRNAP.

### 5.2.2 Follow-up to preliminary experiments

Preliminary nickel pull-down assay results showed binding between *P. furiosus* Spt4/5 and the aRNAP in the absence of a DNA/RNA scaffold (Figure B.6). To understand more about the impact of the Spt4/5-aRNAP interaction, I would determine their $K_d$ in the future using gel electrophoretic mobility shift assays (EMSA). In addition I would evaluate the affinity of Spt4 with Spt5 mutants for RNAP and compare these $K_d$ values with the wild type Spt4/5 factor. All of these EMSAs would be completed on large enough scale native gels to visualize the results, not Phast gels, and with binding of Spt4 to Spt5 prior to RNAP interaction.

The second part of the preliminary experiments displayed Spt4/5 association with the elongation complex formed on a fluorescently 5’-end labeled-RNA with DNA scaffold (Figure B.7). The controls for this assay need to be established as well as the annealing of the DNA/RNA scaffold. Controls and verifying correct scaffold annealing in the future should be the first things examined by an EMSA.

In the assay displayed in Figure B.7 there was remaining scaffold at the bottom of the gel, making it is difficult to explain if the visible shift is due to RNAP binding Spt4/5 only or uptake of fresh scaffold. This could be from two possibilities since I showed RNAP could associate with Spt4/5 without nucleotides. The order of binding Spt4/5 compared to the radioactively labeled scaffold is important. First, if RNAP binding
Spt4/5 enables RNAP to bind additional scaffold, or, second if the shift is due to previously formed transcription elongation complex (TEC) binding to Spt4/5. Further studies are necessary to evaluating the order of binding. If this shows that the TEC forms first then this could be eliminated by concentrating scaffold with the RNAP first then bind with Spt4/5 and the scaffold at the bottom will no longer be present.

Once a DNA/RNA scaffold and all of the controls (RNA only, DNA/RNA only, Spt4/5 only, RNAP only, DNA/RNA with RNAP, and DNA/RNA with Spt4/5) are established for this assay, it is important to repeat the assay in figure B.7. I also suggest completing assays to evaluate the $K_d$ between Spt4/5 and the aRNAP-DNA/RNA (TEC). Then initiating the assay for the Spt4 with Spt5 mutations binding to the aRNAP-DNA/RNA (TEC) to also determine the $K_d$ values for comparison with Spt4 and wild type Spt5. This will quantifiably answer the question of if any of these Spt5 mutations interrupt binding to RNAP and if the DNA/RNA scaffold has any effect on this interaction.

5.3 References


variation. II. A model of the ribosome and its RNA at 3.5 nm resolution. *J. Mol. Biol.* 271, 602–618.
Appendix A

Growth and Handling of *Pyrococcus furiosus* Spt4/5 Crystals

Crystal growth

The *Pyrococcus furiosus* Spt4/5 crystals analyzed in this study were grown under several different conditions. First, the process was evaluated for optimal growth using vapor diffusion which is displayed in figures A.1 and A.2. For this process, I used 24 well trays each containing six wells across. At the opening of each well, I applied a silicone-based grease. The grease covers almost the entire circle of the well; however, the circle is not completed, and a small space is left for air to escape when the (siliconized) cover slide was slowly angled towards the gap. The grease was added using a syringe fitted with a pipette tip that was cut in order to widen the end. This tip was then fastened to the syringe by Parafilm. The plunger of the syringe was removed to allow addition of grease, and once replaced was used to add the grease to the wells.
**Figure A.1 Hanging drop crystallization technique**

The drop, in aqua, was formed by 1-2 µl of protein solution combined with 1-2 µl of reservoir solution, in light blue, and the solutions were mixed together. The glass cover slide was inverted to complete the air tight seal to the top of the well containing the reservoir solution.

**Figure A.2 Sitting drop vapor diffusion crystallization technique**

The drop, in aqua, was formed by 1-2 µl of protein solution combined with 1-2 µl of reservoir solution, in light blue, and the solutions were mixed together. The glass cover slide was inverted to complete the air tight seal to the top of the well containing the reservoir solution.

The reservoir solution was pipetted into each of the wells by following the procedures described below. Following this, 1 to 2 µl of protein solution and 1 to 2 µl of reservoir solution was combined in the well. The cover slide was inverted and placed atop the greased well which then slowly sealed the opening of the well. The theory
behind this method is that time will allow the reservoir to pull water from the solution in the drop to create an equilibrium between the drop and the reservoir. The vapor diffusion method applies to the hanging or sitting method. The result is that the sample in the drop will concentrate, thus generating a supersaturation of the protein in the drop.

**Reservoir preparation**

The optimized crystals of *P. furiosus* Spt4/5 were grown in a reservoir solution of 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25% (w/v) PEG 3,350. Thus, crystals were set up with a gradient of reservoir solutions ranging from solution A (0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 20% (w/v) PEG 3,350) to solution B (0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 30% (w/v) PEG 3,350). This was completed by adding 100% of solution A in the first well and decreasing the amount of A across the six wells from 100 to 0%, while increasing B from 0 to 100% in stepwise increments. An example would be going from left to right, in the first well 100% of solution A and 0% solution B, continuing with the second 80% A and 20% B, in the third well 60% A and 40% B, in the fourth well 40% A and 60% B, in the fifth well 20% A and 80% B, and in the final well 0% A and 100% of solution B. The resulting final total volume of solution contained in each well is identical. Once all the solutions in a row were added, they were mixed completely by pipetting up and down. An alternative method was that only solution C (0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25% (w/v) PEG 3,350) was used in all the wells. Crystals were grown in either 48 or 24 well trays with 200 µl or 1 ml reservoir solutions respectively.

**Drop preparation**

Hanging drop vapor diffusion (Figure A.1):

The drops were prepared by combining reservoir with protein solution using 1-2 µl of each. Using the lid of a 24 well crystal tray cleaned with ethanol I arranged a circular cover slide with the clean side facing up. Only one cover slide was prepared at a time in order that the drops would not dry out. The protein solution was prepared by
gently tapping the bottom of the tube with my finger followed by briefly spinning the mixture in a bench centrifuge. The protein solution was added to the middle of the cover slide. Using the same pipette tip, if possible, the desired amount of the first reservoir was added to the protein solution and gently pipetted up and down to mix the solutions together. In addition some wells were set up using seeding conditions. The seed stock (described above) was added for 10% of the final solution. For example, 1.35 µl of each of the protein and reservoir solutions were added and mixed together, then 0.3 µl of seed stock was added and mixed in. The dilution of the seed stock depends on optimizing by crystallization trials. Using a tweezers, the cover slide was flipped over and inverted over the top of the first well (Figure A.1). Slowly the cover slide was angled toward the gap in the grease to allow the air to escape and close the cover slide over the top of the well. This process was repeated for all of the wells on the tray.

Sitting drop vapor diffusion (Figure A.2):

The differences between setting up a sitting and a hanging drop tray are slight. The primary differences exist in the location of where the drop is set up and mixed. The protein and wells are prepared the same and the reservoir solution is pipetted into the base of each of the lower wells. However in sitting drop trays, each well contains an upper well, in which I was very careful not to leave any reservoir sample. This could provide too much reservoir sample to the crystal. Then the protein was added to the upper well, along with the reservoir solution and mixed. The cover slide was inverted over the top of the well and sealed as explained above.

Cryoprotection details

Several advantages of using a cryoprotectant for crystals were previously discussed in section 2.4.3. These include a reduction in radiation damage, an increase in the lifetime of the crystal, and protection for storage or shipping. Crystals of P. furiosus Spt4/5 formed, in drops, over a reservoir of 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25% (w/v) polyethylene glycol (PEG) 3,350 were used for the evaluation of various cryo conditions. Two such conditions used for cryocrystallography, require that the crystals be
transferred stepwise from mother liquor (0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25 % (w/v) polyethylene glycol (PEG) 3,350) to 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 30 % (w/v) polyethylene glycol (PEG) 3,350, or an alternative condition where the same steps can be repeated followed by 0-20 % glycerol.

**Transferring crystals**

The cover slide of the desired crystal(s) was inverted and removed from the well in order that the crystals could be placed under the microscope. The drops dry out very quickly; therefore, the cover slide was not removed from the well until I was immediately ready to proceed to the following steps. In addition, to protect the crystals several (5-10) µl of mother liquor was added to the drop. I gently pipetted up and down the solution to release the crystals from the drop and the film surrounding the drop, if it was necessary. The glass trays were further cleaned using pressurized air to remove dust and particles. In the first well of the glass tray I pipetted 50 µl from the reservoir. Focusing the microscope and using a pipette, I removed the crystals from the drop and place them into the 50 µl of reservoir in the glass tray. This process was repeated until all of the usable crystals were transferred to the glass tray. I found that using the pipette rather than a loop at this stage significantly reduced the damage to the crystals. All of the crystals were then left in the mother liquor until the ready to transfer them through all of the conditions. This also resulted in much greater reduction of damage to the crystals.

After identifying the best crystals to be frozen, 50 µl of the mother liquor (fresh 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 25 % (w/v) polyethylene glycol (PEG) 3,350) was pipetted into the second well of the glass tray. In the third well was the final condition towards the cryoprotectant (50 µl of 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 30 % (w/v) polyethylene glycol (PEG) 3,350) and in the next well was the final step, with the same solution again but only using 25 µl. The final amount of solution in the well of the glass tray is 25 µl, or less if the crystals are difficult to capture in the loops, this will reduce the movement of the crystal and to prevent placing the pin in the cryoprotectant when collecting the crystal. The crystals are transferred by pipette using the rapid method
of transfer (approximately 30-60 sec in each condition). This needs to be done as soon as possible and can take up to several seconds. If a longer time is taken, approximately minutes, it was observed that the crystals start to crack and melt and in increased PEG percentages.

The second cryoprotectant condition is to increase to 30 % PEG 3,350 in the same 5 % increment followed by 0-20 % glycerol, also in rapid 5 % increments. In this condition similar cracking and melting of the crystals was observed if the crystals were left in the solutions for several minutes. Crystal damage also occurred even in 25 % glycerol. Thus, crystals are rapidly transferred (50 µl) in the same manner as above to the first well of the glass tray into the reservoir solution, followed by fresh mother liquor and then 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 30 % (w/v) polyethylene glycol (PEG) 3,350. The transfer of crystals is continued by rapid transfer in 5 % increments from 5 to 20 % glycerol (with 0.1 M Bis-Tris pH 5.5, 0.2 M NaCl, and 30 % (w/v) polyethylene glycol (PEG) 3,350 in 50 µl for all of the conditions except the final transfer in which the same solution is used a second time for which only 25 µl is used).

Immediately, the crystal is captured in a cryo-loop which is slightly larger than the crystal, and the entire loop is submerged in liquid nitrogen. After submerging the cryo-loop in liquid nitrogen it is placed inside a previously labeled cryo-tube and pre-chilled in the liquid nitrogen. Finally, the cryo-tube is sealed, placed on a cane and stored in a Dewar. These steps are then repeated for each crystal which is to be used to examine the best cryo-conditions and crystals for the Spt4/5 structure.

**Home source evaluation of crystals**

Several crystals were evaluated for diffraction on the home source at The Pennsylvania State University. While at The Pennsylvania State University I used two different goniometers. The first set up we used was for a RaxisIV++ detector with a goniometer facing downwards. I used a different method to mount the crystals for this detector. The home source detector I use for the *P. furiosus* Spt4/5 crystals is a Saturn 944+ CCD detector with an AFC-11, 4-axis, partial χ goniometer (Figure A.3). This
goniometer head is facing upwards and requires a different method to mount the crystals. I removed the cryo-tube from the Dewar and removed it from storage (from a cane) I placed it a Dewar for transfer. I used forceps to transfer the cryo-tube containing the crystal to a large mouth Dewar and held the cryo-tube under the fresh liquid nitrogen while using a wand I attached the magnet to the pin and removed the loop from the cryo-tube. The loop remained under the surface of the liquid nitrogen while the cryotongs were cooled to the temperature of the liquid nitrogen. Within the cryotongs is a space for the pin and loop to fit in the metal block in the cold protecting the crystal away from the air to keep the crystal cool. The cryotongs were opened and the loop was collected inside of the metal space. Removing the tongs from the liquid nitrogen in a single smooth, but quick motion, I transferred the crystal to the goniostat. Then I immediately verified that the crystal base was firmly seated, and checked that the crystal was in the liquid nitrogen gas flow.

If needed, annealing of the crystal was completed while the crystal was still mounted on the goniometer, using a plastic credit card to block the liquid nitrogen stream for several sections. The card was removed and the liquid nitrogen flow was restored to normal and the crystal was allowed to re-freeze. The diffraction should have improved.

**Figure A.3 Data collection of *P. furiosus* Spt4/5 crystals**

This image was from data collected at the home source while screening the Spt4/5 crystals used for the synchrotron data collection. The + at the center is where the beam stop is located.
Details regarding X-ray crystal structure determination

The electron density map for structure determination of *P. furiosus* Spt4/5 was calculated using phases from molecular replacement, and was further improved using the density modification program Resolve (Terwilliger, T.C., 2000). The resulting electron density map had several deviations from the molecular replacement solution that included the Spt5 C-terminal KOW domain, indicating that model bias was effectively removed by density modification. Traditionally a resolve map tends to reduce model bias using map-likelihood phasing. There were five steps in a single cycle of map-likelihood phasing and the first was to start with a set of phases to calculate an electron density map (Terwilliger, T.C., 2000). Second, the map was evaluated for probability distributions for the electron density in the solvent and protein regions. Third, the log-likelihood first and second derivatives were calculated in respect to each individual point for the electron density of the map (Terwilliger, T.C., 2001). Fourth, with a FFT-based algorithm, the first and second derivatives of the log-likelihood in respect to the structure factors were calculated. Fifth, in regards to each reflection $k$ difference log-likelihood in the phase of the map was then estimated (Terwilliger, T.C., 2001). This method allows the original phases to become the probabilities that were used to estimate the new structure factors of the next cycle. These in turn were used to calculate a new phase founded entirely on a series of derivatives based on structure factors.

Resolve (Terwilliger, T.C., 2000) used prime-and-switch phasing to remove model bias. This primed with the first round using all available phase information, that included the model, to initiate map-likelihood phasing. After this point the switch occurs and the initial biased phase information that was needed to get the prime and switch started was disregarded and from this point forward the procedure was independent of the model phase information. Thus, the cycles carried out from the map-likelihood phasing were unbiased, with each round independent from the others, and by the final round showed no signs of the original model’s phases.
References


Appendix B

Structure-based Biochemical Characterization of *Pyrococcus furiosus* Spt4/5

Introduction

The transcription elongation factor Spt4/5 in archaea interacts with RNA polymerase (RNAP) to stimulate elongation and has homologues in eukaryotes, Spt4/5, and bacteria, NusG (Hirtreiter, A. *et al.*, 2010; Guo, M. *et al.*, 2008; Herbert, K.M. *et al.*, 2010). In [Chapter 2](#) I present the structure I solved of the *Pyrococcus furiosus* Spt4/5 complex at 1.8 Angstrom resolution. This is the first complete Spt4/5 X-ray crystal structure containing Spt4 and the Spt5 N-terminal, linker and C-terminal domains. This structure contains all the characteristic Spt4 zinc ion and N-terminal NusG (NGN) binding domains and Spt5 Kyprides-Ouzounis-Woese (KOW). Spt4 contacts Spt5 through all domains, the linker, KOW, and NGN; thus Spt4 may restrict the location of the KOW with respect to the NGN domain of Spt5. This suggests that the Spt5 linker between NGN and KOW is rigid and presents the KOW domain in a fixed position relative to the NGN domain.

[Chapter 2](#) provides a detailed structural analysis of *P. furiosus* Spt4/5; however no biochemical studies exist to explain the structure. Thus, the intent of the following sections will be to detail preliminary data from experimentation used to explain the Spt4/5 structure. Over the past years studies have tested binding of RNAP to NusG in bacteria and Spt5 in eukaryotes. Several of the studies have also tried to locate binding locations on NusG or RNAP. Now it is possible to use the structural details of Spt4/5 in archaea to biochemically analyze the interaction with RNAP.

**RafH:**

In bacteria NusG has been analyzed in parallel with RfaH. RfaH regulates expression of a few operons in *Escherichia coli* and has distinct N and C-terminal
domains similar to NusG. RfaH decreases pausing and increases elongations rate and reduces termination.

Figure B.1 Comparison between the structures of RfaH and NusG

Based on molecular modeling of the N-terminal domain of both RfaH and NusG, competition for binding for the elongating complex could occur. The binding site was suggested to be the β'-subunit clamp helices. This study was followed up when the structure of RfaH was solved at 2.10 Å resolution. The authors suggested the hydrophobic cavity of the N-terminal domain of RfaH binds RNAP. The C-terminal domain is packed up against the hydrophobic region, however, NusG crystal structure structures show a flexible C domain with one form with the N-terminal hydrophobic region exposed the other partially protected (Knowlton, J.R. *et al*., 2003; Steiner, J.T. *et al*., 2002).

At the start of my study there were no archaeal studies, however recently Cramer’s and Finn Werner’s groups solved the X-ray crystal structure of the archaeal Spt4/5N-terminal domain. In addition, they also found that archaeal RNAP may bind the Spt4/5 at the clamp coiled-coil motif by the N-terminal hydrophobic domain (Hirtreiter, A. *et al*., 2010).
**Spt5 Mutations**

There is a remarkable structural similarity between NusG, RfaH, and Spt5 (Artsimovitch, I. and Landick, R., 2002; Belogurov, G.A. *et al.*, 2009; Guo, M. *et al.*, 2008). The potential RNAP binding site on NusG or Spt5 in the hydrophobic pocket was evaluated by biochemical assays involving mutations of NusG. NusG variants in *Escherichia coli* to single residue, alanine rich, or mini sequences, or deletion sequences were prepared (Richardson, L.V. and Richardson, J.P., 2005). Of these, no specific residue was found to be responsible for binding with RNAP, however a region was responsible. A more recent assay was completed by evaluating *E. coli* NusG variants by isolating dominant-lethal substitutions when overexpressed. This study found several NusG variants that had decreased stimulated elongation rates compared to wild type; however when the amount was increased, variants were able to regain stimulation of the elongation rate. Thus, two mutations F65L and Y68H, were suggested to be defective in binding the elongation complex (EC) (Mooney, R.A. *et al.*, 2009).

**Prior structural and biochemical studies**

My complete X-ray crystal structure of the Spt4/5 complex was solved at 1.8 Å resolution (Figure B.3). The structure shows that the Spt4/5 forms a heterodimer. Using the solved structure, possible binding sites to RNAP were determined and Spt5 was targeted for mutational studies.
Figure B.2 Sequence alignment of Spt5 from several organisms

Alignment of the N-terminal region of Spt5. Residues were aligned to search for potential mutations sites to test for RNAP binding. The symbols under the alignment are: star for identical, colon for conserved substitutions, and period for semi-conserved. The two choices used in this study were designated by arrows, Y45A and E78A. Numbering based on P. furiosus Spt5. Pfu - P. furiosus; Sso - Sulfolobus solfataricus; Mja - Methanocaldococcus jannaschii; E. coli; Yeast; and Human. Alignment was completed using ClustalW (Larkin, M.A. et al., 2007).
Figure B.3 X-ray crystal structure of the Spt4/5 complex at 1.8 Angstrom resolution with mutation sites
Spt4 (amino acid residues 2-61) in magenta at the top of the image with zinc, in grey, coordinated by four cysteines in yellow. Spt5 (amino acid residues 2-148) shown in green. In darker green the NGN domain, with the linker (83-87) in light green in the middle, and the KOW domain in bright green on the left. The two mutations used in this study, shown as sticks are Y45A (red) and E78A (blue), are displayed on the Spt5NGN domain on the right side of the structure. Image prepared using PyMOL.

Experimental summary

To biochemically analyze the binding between archaeal Spt4/5 and RNAP three Spt5 mutations were screened for disruption in binding. Variants of these mutations were constructed by site-directed mutagenesis and expressed and purified as was the wild type (wt) Spt5 protein. Pull-down assays were used to evaluate the interaction with RNAP and results were compared to the wild type to evaluate the extent of the effect on binding. The final assay used was a gel electrophoretic mobility shift assay, or EMSA, to further evaluate the affinity of RNAP on a scaffold for Spt4/5.
Materials and Methods

Cloning recombinant Spt4 and Spt5 variants

The DNA coding sequences of *spt5* and *spt4* from *Pyrococcus furiosus* were PCR subcloned into pET21a and pET15b, respectively.

Site-directed mutagenesis

Spt5 mutants were generated using Quick change site-directed mutagenesis to evaluate binding with RNAP. Oligonucleotide primers were designed to anneal to the plasmid encoding *spt5* cloned into NdeI-BamHI digested pET21a, to contain the desired mutations. Primers, *spt5* plasmid, NTPs, and *PfuTurbo* DNA polymerase were combined in 10x *Pfu* buffer in a thermocycler. The conditions for the reactions varied by the primers involved. The reactions enabled the DNA polymerase to extend and incorporate the mutations into nicked circular nascent DNA. After PCR, *Dpn* I was added and the tubes were incubated for 1 hr in at 37°C to digest the original DNA. The samples were transformed into Mach-T1 competent cells, plated on Luria-Bertani media (LB) with ampicillan, and incubated at 37°C overnight. The colonies were then incubated overnight in LB with antibiotics at 37°C in a shaker. Using a Qiagen min-prep kit, plasmids were isolated and were screened by sequencing for correct incorporation of the desired mutations.

Expression and purification of Spt4 and Spt5 variants

DNA plasmid containing either the *spt4* or *spt5* gene was transformed into BL21(DE3) codon-plus RIPL (Invitrogen, Co., Carlsbad, CA). The proteins were each expressed in separate 250 mL flasks containing 50 mL LB medium at 37°C until OD$_{600}$ reached 0.4-0.6. The flasks were moved to room temperature without shaking for 10 min followed by 0.5 mM IPTG induction. Expression of Spt4 protein required the addition of
10 µM ZnSO₄ at the time of induction. The flasks were transferred to a shaker at 22°C for ~18 hrs and harvested by centrifugation.

**Individual purification of Spt4 and Spt5**

The purification of the complex occurred immediately following centrifugation of the cells. The pellets containing Spt4 and Spt5 were each resuspended in 10 ml lysis buffer (10 mM Tris-HCl pH 8.0 at room temperature, 15 % glycerol, 10 mM BME, 10 µM ZnSO₄, and 1 mM phenylmethylsulfonyl fluoride) and were kept separate. The cells were lysed by sonication followed by a heat treatment at 50°C for 20 min. The lysed cells were centrifuged (13,000 rpm, 15 min) at 6°C to remove cell debris and supernatant was used for following purifications.

Spt4:

After centrifugation, the supernatants containing Spt4 were loaded onto a 1 ml Ni-NTA (Qiagen) column equilibrated with wash buffer (TGMZ, 0.3 M NaCl, and 0.01 M Imidazole). The column was washed two times with 5ml of wash buffer (TGMZ, 0.3 M NaCl, and 0.05 M Imidazole). The Spt4 proteins were eluted in approximately 0.5 ml fractions by TGMZ containing 0.3 M NaCl and 0.5 M Imidazole. The peak fractions were analyzed by Bradford as in Chapter 2.

Spt5:

The Spt5 containing supernatant was evaluated for conductivity and diluted to 0.05 M NaCl with TGMZ buffer, which was applied to a 1 ml SP-FF column (GE Healthcare) equilibrated with binding buffer (TGMZ 0.05 M NaCl). The column was washed with binding buffer and Spt5 was eluted from the column with TGMZ and 1 M NaCl. The peak fractions were analyzed by Bradford as in Chapter 2.

**Spt4/5 complex preparation**

Cloning, expression, and purification were completed as in Chapter 2.
Purification of *P. furiosus* RNA polymerase

Purification was as completed in Chapter 3.

Preparation of transcription elongation complex using a DNA/RNA scaffold

The DNA/RNA scaffolds (Figure B.4) were assembled by annealing the 22-mer FAM-labeled RNA with template DNA at 76°C for five min, followed by slowly cooling to room temperature. The non-template DNA was added to the DNA/RNA hybrid for 30 min at room temperature to complete the scaffold.

**Figure B.4 DNA/RNA scaffold**
A. Sequence of DNA/RNA scaffold used in study. The RNA in is red, the DNA template (DNA_T) in dark blue, and the DNA non-template (DNA_NT) in light blue. B. A schematic diagram of the scaffold. DNA template (T) and Non-template (NT) are shown as well as the fluorescently labeled (FAM) in green RNA.
Ni-NTA bead pull-down assay

For each assay 10-15 µl of 5 % suspension of Ni-NTA Magnetic Beads (Qiagen) was equilibrated with wash buffer (TGMZ 0.3 M NaCl and 0.01 M Imidazole). Between each incubation the beads were separated from the solution by placement of the tube near a magnet for 30-60 sec. The solution was removed and fresh solution could be added, mixed, and incubated before placement near the magnet again. Prior to incubation with the beads, Spt4 was incubated with Spt5 for 1 hr, followed by 1 hr of incubation with RNAP. Proteins of interest were incubated with the beads for 1 hr. The unbound sample was removed and the beads were incubated with wash buffer. After four washes, the samples were separated from the beads by elution buffer (TGMZ, with 0.3 M NaCl, and 0.5 M Imidazole). Purity and binding partners were assessed by SDS-PAGE.

Gel electrophoretic mobility shift assay (EMSA)

Spt4/5-aRNAP was assembled by incubating purified *P. furiosus* RNAP with two molar equivalents of purified Spt4/5 complex for 60 min at room temperature. Spt4/5-aRNAP was combined with annealed DNA/RNA scaffolds (2 pmol/reaction) containing the 22-mer FAM-labeled RNA assembled as above. The reactions were incubated for an additional 60 min at temperatures indicated and analyzed by native gel electrophoresis. Following a protocol similar to that used by Dr. Anamika Missra in Dr. David S. Gilmour’s laboratory at The Pennsylvania State University, native gels were run in buffer (RB - containing 50 mM Tris-HCl pH 8.5, 0.38 M Glycine, and 5 mM MgCl2). The 4 % acrylamide gels were composed of 8 ml 30 % 29:1 acrylamide:bisacrylamide solution, 12 ml 5x RB, 1.5 ml 100 % glycerol, and polymerized with 150 µl 25 % ammonium persulfate (APS) and 55 µl TEMED. A very little bromophenol blue dye was added to each reaction to aid in loading the gels. The gels were run at 50-100 volts for 3-4 hrs at 22°C, and were assessed with a phosphorimager by a Typhoon 9410 (Ge Healthcare). In addition to fluorescence detection, the gels were silver stained for protein visualization.
Results and Discussion

Purification of *P. furiosus* RNAP and Spt4/5 complex and variants

The aRNAP purification is as displayed in Chapter 3. The purification of the wild type Spt4/5 was as before, however for these biochemical studies the Superdex75 column was not used for the final step in purification. Thus, the individual Spt4 and Spt5 and variant proteins were slightly less pure for preliminary experiments.

Spt4/5 association with RNAP

The nickel pull-down assay used during this study was to evaluate binding of His-tagged proteins to the metal ion, in this case nickel is attached to the nitilotriacetic acid (NTA) (Figure B.5). *P. furiosus* RNAP which does not contain a His-tag, thus it should not be bound by the Ni-NTA Beads and in the pull-down assay. This was shown to be the case, in lane 5 of figure B.7 as RNAP was not present in the elution. However there was an additional band around 15 kDa. I ran a follow-up Phast gel of the washes. The results showed some residue after the last wash. This suggests that there is possibly some remaining RNAP after the four washes that were used. There are several things that could be concluded from this. First, that further washes are necessary in the future. Second, the beads can only be washed so well. Third, silver staining is very sensitive. The last lane of the primary gel showed the elution of the wild type Spt4/5 from the beads as a control (Figure B.7).
Figure B.5 Pull-down assay binding predictions
A. In blue is the his-tag chain, bound to the nickel ion. The Nitrilotriacetic acid (NTA) in red, is shown bound to a side of a tube (Qiagen - Ni-NTA Magnetic Agarose Beads Handbook 12/2001). B and C. Models of the RNAP bound to the Spt4/5 complex. The His-tag is attached to Spt4, which is coordinated to the nickel ion, attached to the NTA. C. Difference from B shows pre-formation of Spt4/5 complex before incubation with Ni beads.
Figure B.6 Binding interaction between Spt4/5 variants and RNAP
Wild type Spt4/5 and aRNAP
Figure B.7 Binding interaction between Spt4/5 variants and RNAP

Lanes 1-5 all contain assays with wild type *P. furiosus* aRNAP. The first lane also contains *P. furiosus* Spt4/5. The second lane contains Spt4 combined with Spt5 variant (mutations Y45A and E78A). The third lane also has Spt4 with the Spt5 variant with mutation Y45A. In the fourth lane is Spt4 with Spt5 variant containing mutations E78A. And in the fifth and sixth lanes are the controls with the wild type RNAP only, and in the final lane is Spt4 with Spt5, respectively.

The results of the binding assay between the Spt5 variants and RNAP indicated that the single mutations decreased binding of Spt4/5 to RNAP; however, the Spt5 double mutant eliminated association to RNAP. The double mutant of Y45A and E78A, shows almost no association with RNAP. These preliminary results show an additive effect of Spt5 mutations on association with RNAP. To further test this, more sensitive assays, including in vitro transcription assays, which could show more detailed results are necessary. In addition, structural studies featuring the association between Spt4/5 and RNAP would go a long way to explaining this interaction.

**Spt4/5 association with the elongation complex**

These mutation studies have shown a direct association with RNAP in the absence of nucleic acids. However, it is also necessary to determine how Spt4/5 interacts with
RNAP in the presence of DNA/RNA to understand transcription elongation. For this study I introduced a scaffold composed of DNA template and non-template and fluorescently labeled-RNA. The RNA is tracked through the EMSA to evaluate the association of first RNAP with the scaffold, then with Spt4/5. The results showed binding of the scaffold with RNAP as displayed in figure B.6 and B.7. The binding enhanced with increased RNAP concentration until about 15 pmol, however in higher concentrations there was not much change. Thus, to see changes in Spt4/5 association with transcription elongation complexes (TEC), I evaluated EMSAs in reaction conditions using 10 pmol RNAP which showed binding with the scaffold however not saturation with the scaffold. This allowed for changes to occur in the TEC. Since there was remaining scaffold in the reaction at the bottom of the gel, it was difficult to explain if the visible shift was due to uptake of fresh scaffold. This could be from two possibilities since I showed RNAP could associate with Spt4/5 without nucleotides. First, if RNAP binding Spt4/5 enables RNAP to uptake additional scaffold, or, second if the shift was due to previously formed TEC binding Spt4/5. Further studies evaluating order of binding could help rule this possibility out. Also competition assays with other scaffolds would rule out if Spt4/5 was binding to RNAP by recognizing specific TEC sites.
Figure B.8 EMSA
EMSA of increasing RNAP then with constant RNAP with increasing Spt4/5 In each reaction 2pmol of scaffold was added followed by the amount of RNAP listed then Spt4/5 complex as described in 3.5.7. Purple asterisk indicates possible impurity with RNAP that can be removed by concentration.

There is association of Spt4/5 with the TEC on a DNA/RNA scaffold. However, I have also shown that archaeal Spt4/5 does not associate with RNAP through the RNA, as it can bind RNAP without nucleic acids present. A similar result was seen by Cramer and Werner’s groups with (Hirteiter, A. et al., 2010). I identified two Spt5 residues Y45 and E78, involved in association with RNAP binding. However, it is likely that a Spt5 region is responsible for association with RNAP. This region may even be flexible, to move with RNAP while it is involved in transcription elongation. The involvement of Spt4/5 in elongation with RNAP will probably be a flexible mechanism that allows the Spt5 N-terminal to associate with the RNAP, while allowing the KOW to interact with additional factors to regulate transcription elongation, with the linker allowing for give between them.
References


Cycle Is Suggested by X-ray Crystallography and Supported by Site-Directed Mutants. *Biochem.* 42(8), 2275-2281.


Vita

Brianna Joy Klein

EDUCATION
Ph. D., Biochemistry. The Pennsylvania State University, PA. KS Murakami, Major Professor, Expected Fall 2010.

RESEARCH EXPERIENCE
Graduate Assistant, Department of Biochemistry and Molecular Biology. The Pennsylvania State University. University Park, PA. Dr. K Murakami, Major Professor, 2004-present.
Summer Student Intern, Pacific Northwest National Laboratory, Richland, WA, N Isern, Supervisor, Summer 2004.
Summer Student Intern, Pacific Northwest National Laboratory, Richland, WA, Dr. MA Kennedy, Supervisor, Summer 2003.
Laboratory Assistant, Department of Forest Science, Oregon State University. Corvallis, OR. Dr. R Meilan and Dr. SH Strauss, Supervisors, 1998-2002.

PUBLICATIONS

MANUSCRIPTS IN PREPARATION

PRESENTATIONS
Oral presentations

Poster presentations

TEACHING EXPERIENCE
Completed Course in College Teaching, Pennsylvania State University, Summer 2006.
BiSc4, Biological Science, Human Body: Form and Function. M Troyan, Instructor.
Fall 2005.