A STRUCTURAL STUDY OF BACTERIAL RNA POLYMERASE
CONFORMATIONAL CHANGES DURING TRANSCRIPTION ELONGATION

A Thesis in
Biochemistry, Microbiology and Molecular Biology

by
Brittany Warner

© 2012 Brittany Warner

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2012
The thesis of Brittany Warner was reviewed and approved* by the following:

Katsuhiko Murakami  
Associate Professor of Biochemistry and Molecular Biology  
Thesis Advisor

Paul Babitzke  
Professor of Biochemistry and Molecular Biology  
Director of the Center of RNA Molecular Biology

David Gilmour  
Professor of Biochemistry and Molecular Biology

Tae-Hee Lee  
Assistant Professor of Chemistry and the Huck Institutes of the Life Sciences

Craig E. Cameron  
Paul Berg Professor, Department of Biochemistry and Molecular Biology  
Associate Department Head for Research and Graduate Education

*Signatures on file in the Graduate School
Abstract

DNA-dependent RNA polymerase (RNAP) is the enzyme responsible for the initial step of gene expression in all domains of life. This process of transcription is composed of three stages; initiation, elongation and termination. The subject of this work is transcription elongation, which is characterized by addition of an incoming nucleotide followed by forward translocation of the RNAP to incorporate the next nucleotide. The X-ray crystal structure of the bacterial transcription elongation complex (TEC) has been solved to 2.5 Å and provides important details concerning the architecture and mechanism of transcription elongation and RNAP translocation [1].

This thesis describes ongoing work to elucidate the mechanism of transcription elongation, with particular attention being paid to important mobile elements of the bacterial RNAP, the bridge helix (BH) and trigger loop (TL). These elements of RNAP have been shown to play a role in the nucleotide addition cycle and translocation, though specifics concerning their cycle movements as well as order of events are not currently defined. Additionally, there are currently three major models describing RNAP translocation; the power stroke model, Brownian ratchet model, and a shelf/core ratcheting model. We hypothesize that bacterial RNAP translocates according to the shelf/core ratcheting model, which is the only model that accounts for the tremendous force RNAP is capable of generating during elongation. The BH/TL movements, as well as the various models of RNAP translocation will be discussed here and considered with respect to new experiments conducted using Raman crystallography and time-resolved X-ray crystallography. This combination of approaches is effectively used to view a translocation event in real time, and thereby provide direct evidence for the mechanism and order events during transcription elongation.
# Table of Contents

List of Figures...................................................................................................................... vi
List of Abbreviations ........................................................................................................ vii
Acknowledgements ........................................................................................................... viii

Chapter 1: Introduction........................................................................................................ 1
  1.1 Architecture of the Bacterial Transcription Elongation Complex .................................. 1
  1.2 Power Stroke and Brownian Ratchet Models of Transcription Elongation .................. 3
  1.3 Shelf/Core Ratcheting: A Third Model of Transcription Elongation ............................... 5
  1.4 The Bridge Helix and its Involvement in Transcription Elongation ................................. 8
  1.5 The Trigger Loop and its Involvement in Transcription Elongation ............................... 11
  1.6 Bridge Helix/Trigger Loop Movements in the Shelf/Core Ratcheting Model of Transcription Elongation ................................................................................. 14
  1.7 Raman Crystallography and Time-Resolved X-ray Crystallography ............................. 15
  1.8 Specific Aims of this Work ........................................................................................... 19

Chapter 2: Materials and Methods .................................................................................. 21
  2.1 Materials ..................................................................................................................... 21
  2.2 Fermentation of *Thermus thermophilus* cells ................................................................ 21
  2.3 Purification of Endogenous Core RNA Polymerase from *Thermus thermophilus* .......... 22
  2.4 Crystallization of Thermus thermophilus RNA polymerase transcription elongation complex ................................................................. 24
  2.5 Time-Resolved X-ray crystallography and Raman crystallography ............................... 25
  2.6 *In crystallo* RNA extension assay .............................................................................. 25

Chapter 3: Results .............................................................................................................. 27
  3.1 Culturing of *Thermus thermophilus* ........................................................................... 27
  3.2 Purification of Endogenous Core RNA Polymerase from *Thermus thermophilus* .......... 27
  3.3 Crystallization of *Thermus thermophilus* RNA polymerase transcription elongation complex .................................................................................. 30
  3.4 Raman Crystallographic Studies .................................................................................. 32
  3.5 Time Resolved X-ray Crystallography ......................................................................... 33
  3.6 *In crystallo* RNA Extension Assay ........................................................................... 35

Chapter 4: Discussion ........................................................................................................ 38

References ......................................................................................................................... 44
Appendix A: Purification of RNAP from Pathogen-Related Bacterial Species.................................................48

Appendix B: Determining the Binding Location of the $\delta$ Subunit on $B. subtilis$ RNAP.................................79
List of Figures

Figure 1: Transcription Elongation Complex Scaffold .................................................................2
Figure 2: Structure of *Thermus thermophilus* Transcription Elongation Complex ...........................2
Figure 3: *Thermus thermophilus* TEC Structural Details ................................................................3
Figure 4: Power Stroke and Brownian Ratchet Models of Transcription Elongation .......................5
Figure 5: Structure of *T. thermophilus* EC in complex with Gfh1 ..................................................6
Figure 6: Shelf/Core Ratcheting Model of Transcription Elongation .............................................7
Figure 7: Conformations of the Bridge Helix in Various RNAP Structures .....................................9
Figure 8: Model of TL Movements During Translocation .............................................................11
Figure 9: Model of Nucleotide Addition Cycle Involving Conserved TL Histidine Residue .............12
Figure 10: Raman Crystallography Experimental Setup ................................................................18
Figure 11: Purification Scheme for *T. thermophilus* Core RNAP ..................................................27
Figure 12: Chromatogram of SP-Sepharose Purification of RNAP .................................................28
Figure 13: Purity of *T. thermophilus* RNAP after SP-Sepharose Column Purification ....................29
Figure 14: Final Steps of *T. thermophilus* Core RNAP Purification .............................................30
Figure 15: *T. thermophilus* TEC Crystals ....................................................................................31
Figure 16: Scaffold Used for *T. thermophilus* TEC Crystallization ...............................................32
Figure 17: Sample Raman Spectra Demonstrating Nucleotide Addition in *crystallo* .....................33
Figure 18: Sample Diffraction Image .............................................................................................34
Figure 19: *T. thermophilus* TEC in complex with GMPCPP ........................................................35
Figure 20: Crystals Grown Using Fluorescent Labeled RNA .........................................................36
Figure 21: Nucleotide Incorporation Assay ...................................................................................37
Figure 22: In solution vs. in *crystallo* Extension Assay ...............................................................37
List of Abbreviations

RNAP – RNA polymerase
BH – Bridge helix
TL – Trigger loop
TH – Trigger helix
Tth – *Thermus thermophilus*
TEC – Transcription elongation complex
NTP – Nucleotide triphosphate
EC – Elongation Complex
Stl – Streptolydigin
Acknowledgements

I would first especially like to thank Dr. Murakami for all of his support and guidance, as well as my committee members for their support and helpful suggestions. I would like to give a big thank you to all past and present Murakami lab members, particularly Ritwika Basu, without whom I may not have kept my sanity and Kevin Baker, who taught me how to MacGyver everything in the lab. Thank you to Hement and Neela Yennawar at the Penn State X-ray Core Facility for all of their help and advice regarding my crystals over the years. Thank you to the staffs at the Macromolecular Diffraction Facility at the Cornell High Energy Synchrotron Source and the Berkeley Center for Structural Biology for assisting diffraction data collection. I owe tremendous thanks to Mark Signs at Kim Martin at the Penn State Shared Fermentation Facility, who both greatly assisted my work as well as provided many laughs along the way. Thank you as well to Shigeyuki Yokoyama and Shun-ichi Sekine at Riken for all of their advice and assistance with this project. Finally, I thank the NIH for financial support.

Thank you as well to my mom, dad, Joe and all of my friends for their love and support. And to Larry, for all that you do to keep me smiling.
Chapter 1: 
Introduction

DNA-dependent RNA polymerase (RNAP) is a key enzyme responsible for gene expression in all organisms. In bacteria, this remarkable enzyme is capable of recognizing promoter DNA sequences, including the -35 and -10 elements, melting DNA, initiating synthesis of and elongating RNA chains, and finally transcript release upon reaching a termination signal [2]. The multi-subunit bacterial core enzyme, with a subunit composition of $\alpha_2\beta\beta'$ and a molecular mass of about 400 kDa, is able to transcribe RNA de novo with DNA as a template. Transcription is divided into three stages: initiation, elongation and termination. Transcription at specific promoters is possible only when the core enzyme binds with an additional polypeptide, $\sigma$-factor, to form the RNAP holoenzyme during initiation phase [3]. Once the RNA has been extended 13-15 nucleotides, the complex isomerizes into the transcription elongation complex (TEC), characterized by a 12-14 base pair transcription bubble with a 7-9 base pair RNA-DNA hybrid [1]. During elongation, nucleotides are added at a rate of 25-100 nt s$^{-1}$ until the complex dissociates and the new transcript is released during the termination stage [4]. The interesting and dynamic stage of transcription elongation is the focus of this work. The X-ray crystal structure of the TEC has been solved [1], which provides important details concerning the essential process of RNAP translocation and transcript elongation. This structure lays the foundation for the work described in this thesis.

1.1 Architecture of the Bacterial Transcription Elongation Complex

At the early stage of transcription, RNAP releases multiple short RNA oligonucleotides, a process called abortive initiation. Once the nascent RNA transcript reaches approximately 13 nucleotides in length, the RNAP is able to escape from the promoter and continue elongation. Elongation occurs in the 5'→3' direction, with incoming nucleotides added iteratively to the 3' end of the growing RNA transcript. The X-ray crystal structure of the Thermus thermophilus (Tth) elongation complex was published in 2007 and provided important insight into its architecture [1]. This TEC structure consists of the Tth core RNAP, 14 base pairs of downstream DNA, a 9 base pair
RNA-DNA hybrid and 7 single-stranded nucleotides of the RNA transcript. (Figure 1). The TEC in this crystal structure is in the post-translocated state, with the position for the next nucleotide, termed the $i + 1$ site, open and the acceptor template DNA base available for binding to the incoming nucleotide.

Figure 1. Transcription Elongation Complex Scaffold. This scaffold was used to crystallize the *Thermus thermophilus* TEC. The G-C content of the hybrid and the downstream DNA allowed for greater stability, and hence better-diffracting crystals [1].

Various important regions of the Tth TEC can be localized using this structure. The main channel, the region between the two largest subunits of RNAP, $\beta$ and $\beta'$, houses the RNA-DNA hybrid as well as the double-stranded downstream DNA. The elongating RNA chain is threaded out of the active site via the RNA exit channel. A third channel, the secondary channel, leads directly to the catalytic active center of the core enzyme and is the proposed entry site for incoming NTPs (Figure 2). The TEC structure supports this model, as the RNAP main channel is blocked, while the open and accessible secondary channel could accommodate substrate [5].

Figure 2. Structure of *Thermus thermophilus* Transcription Elongation Complex. Surface representation of the Tth TEC highlights the three channels present in core RNAP; RNA exit channel, main channel, and secondary channel, the proposed entry site for incoming NTPs. In this figure, the $\beta$ subunit is blue, $\beta'$ is pink, $\alpha$ and $\omega$ subunits are gray, RNA is orange, template DNA is green, and the non-template DNA is yellow. The double strand of DNA shown in the
image on the right is the downstream DNA. PDB accession number 2O5I. All structural graphics were prepared using Pymol (DeLano Scientific, Palo Alto, CA).

A long alpha-helical region of the \( \beta' \) subunit termed the bridge helix (BH) is highlighted in Figure 3. Composed of \( \beta' \) residues 1066-1103 (in Tth numbering), this region spans the main channel, passing close to the active center where catalysis occurs. A region that cannot be seen in Tth TEC structure, due to its flexibility, is the trigger loop (TL). Located adjacent to the catalytic active center and BH, the TL is capable of transitioning into a more ordered trigger helix structure in the presence of substrate. Both of these regions will be discussed in more detail in subsequent sections.

Figure 3. Thermus thermophilus TEC Structural Details. The coloring scheme in this figure is the same as in Figure 2, with the addition that the \( i \) site is magenta and the bridge helix, which is part of the \( \beta' \) subunit, is highlighted in red on the top left. The \( \alpha \) and \( \omega \) subunits have been omitted for simplicity. The bridge helix passes near the active center of the polymerase, spanning the main channel. The RNA passing through the RNA exit channel and double stranded downstream DNA are also shown.

1.2 Power Stroke and Brownian Ratchet Models of Transcription Elongation
The TEC is characterized by much higher stability than the initiation complex, with the polymerase remaining bound to the nucleic acids [5]. Despite this stability, RNAP must be able to break and re-form nucleic acid contacts to translocate forward along the template and synthesize RNA chains [1]. Various models have been proposed to explain how RNAP accomplishes the process of translocation. The first model, termed the power-stroke model, was proposed based upon X-ray crystal structures solved of the single-subunit T7 phage RNAP elongation complex (EC) in complex with a nonhydrolyzable nucleotide analog or the pyrophosphate (PPi) product [6]. This model, as summarized in Figure 4a, proposes that the forward translocation of the RNAP EC is driven by irreversible PPi release. In the crystal structure solved with incoming NTP or with PPi, mobile elements of T7 RNAP, the O helix and fingers domain, exhibit a closed conformation. In the absence of NTP, a conformational change has occurred that causes these domains to become more open. Therefore, it was proposed that release of PPi causes a conformational change, which pushes the RNA-DNA hybrid forward to allow translocation [6]. In this manner, translocation is coupled with PPi release.

The power stroke model is well supported by the T7 RNAP EC structures; however, it does not appear to be true for multi-subunit RNAPs. A second model, the Brownian ratchet model, correlates well with biochemical and biophysical studies of multi-subunit RNAP, most of which have been conducted with *Escherichia coli* (*E. coli*) RNAP. This model differs from the power stroke model in that PPi release is not tightly coupled to translocation; instead, translocation relies more heavily upon NTP binding at the active site. As shown in Figure 4b, the Brownian ratchet model posits that RNAP is in a random reversible thermal equilibrium, interconverting between pre- and post-translocated states. In the pretranslocated state, the newly incorporated nucleotide has occupied the A site, \((i + 1)\) site, the nucleotide insertion site. In the shift to the posttranslocated state, this nucleotide moves into the P site, \(i\) site or product site, making room in the A site for addition of the next nucleotide. Binding of the NTP to the A site biases the complex towards forward translocation [7].
Figure 4. Power Stroke and Brownian Ratchet Models of Transcription Elongation. A) This scheme describes the power stroke model in which translocation is tightly coupled to pyrophosphate (PPi) release. In this model, the chemical energy derived from release of PPi is converted into the mechanical energy necessary for the RNAP to translocate. This model correlates well with structural data of single-subunit T7 RNAP. B) This scheme describes the Brownian ratchet model of translocation, more appropriate for multi-subunit RNAP systems. In this model, RNAP reversibly interconverts between pre- and post-translocated states until NTP binding the post translocated state biases the complex toward catalysis and forward translocation.

In contrast to the structural evidence for the power stroke model in T7 RNAP, there exists biochemical and biophysical data in support of the Brownian ratchet model for multi-subunit RNAP as well as for T7 RNAP. A series of biochemical experiments demonstrated that halted elongation complexes are capable of sliding their active sites reversibly along a strand of DNA [8, 9]. This verified the model in which RNAP reversibly interconverts between a pre- and post-translocated state. Additionally, it was shown that the affinity of NTP for RNAP was reduced when roadblocks were placed in the way of the elongating RNAP [10]. It is thought this is caused by a decreased frequency of RNAP in the post-translocated state, to which NTP could bind and thereby enable translocation. Perhaps most convincing, however, is a biophysical study conducted on E. coli RNAP [11]. In this study, the authors showed via a single-molecule force-clamp transcription assay that a greater hindering force was needed to hold back RNAP at increasing concentrations of NTP. This result enables the uncoupling of PPi release to translocation, providing more support for a Brownian ratchet model for transcription elongation.

1.3 Shelf-Core Ratcheting: A Third Model of Transcription Elongation

Interestingly, biophysical studies have shown that bacterial RNAP in vitro, under conditions of saturating [NTP], is capable of producing a force of about 14 pN. This is
particularly impressive, given that it is a greater force than that produced by the molecular motors myosin and kinesin [12]. Despite the evidence in favor of a Brownian ratchet model, the random thermodynamic motion it postulates is a major factor in translocation does not account for this large force. A third potential model, however, may allow for a more actively translocating RNAP capable of generating such a force. This model of shelf-core ratcheting was proposed in 2011 based upon a structure of Tth RNAP EC in complex with the transcription inhibitor Gre factor homolog 1, Gfh1 (Figure 5, [13]. Gfh1 belongs to the Gre family of elongation factors. These proteins, which include GreA and GreB of *E. coli*, are capable of assisting cleavage of RNA by RNAP in backtracked and arrested complexes [14]. Gfh1 differs from these elongation factors in that it inhibits transcription by blocking the secondary channel, preventing NTPs from entering the active site [15]. Gfh1 is characterized by a long coiled-coil domain that inserts into the secondary channel, reaching to the active site. The tip of the coiled-coil domain blocks the would-be binding site for the NTP β-γ phosphate groups [15].

![Figure 5. Structure of *T. thermophilus* EC in Complex with Gfh1.](image)

In this figure, the β subunit is blue, β’ is pink, RNA is orange, and template DNA is green. The only nucleic acids present in this structure are that of the RNA-DNA template strand hybrid. Gfh1 is shown as a purple surface model which reaches into the active site via the secondary channel. The α and ω subunits have been omitted for simplicity. The bridge helix and trigger loop domains of the β’ subunit have been highlighted in red and cyan, respectively. This figure was prepared using PDB accession number 3AOH as reported in [13].

The structure of Tth RNAP EC in complex with Gfh1 is in itself interesting; however, this structure also revealed a unique conformation of RNAP that had never been
previously observed. In previously published bacterial and eukaryotic EC structures, RNAP is tightly bound to the nucleic acids, having what appear to be strong interactions with the RNA-DNA hybrid and the rest of the EC scaffold. However, RNAP should need to loosen or break these contacts to translocate along the DNA, calling into question the idea that RNAP remains in this tight state throughout translocation. In the RNAP EC-Gfh1 structure, RNAP is observed in what the authors describe as “ratcheted state” as opposed to the previously described tight state [13]. This ratcheted state is not induced by the binding with Gfh1; in fact, Gfh1 does not fit into the secondary channel of RNAP unless it is already enlarged as part of the ratcheted state. Therefore, it is thought that this ratcheted conformation exists naturally, and Gfh1 binds to and locks RNAP in this state, inhibiting transcription.

![Diagram](https://via.placeholder.com/150)

**Figure 6. Shelf-Core Ratcheting Model of Transcription Elongation.** In this figure, adapted from Tagami *et al.*, 2010, the β’, or shelf module, is pink, the β, or core module, is in blue, and the α dimer is represented in gray. The bridge helix is shown in red. The tight state, shown on the left, represents the conformation of RNAP in previously solved EC structures. The ratcheted conformation shown in the EC-Gfh1 structure, is shown on the right. In this conformation, the shelf and core modules are ratcheted with respect to each other, causing the straight bridge helix to bend towards the main channel in the middle. Additionally, in this conformation, the main and secondary channels are enlarged.

Interestingly, in the ratcheted state, as shown in Figure 6, the BH has a kink to accommodate the shift of the core and shelf modules with respect to each other. This kink would overlap with the space in which the incoming NTP would bind with the template DNA base [16]. In this way, by blocking the NTP binding site as well as by occluding the secondary channel, Gfh1 is capable of inhibiting transcription. To test whether this ratcheted state reflected a relevant and existing conformation of RNAP, the authors performed disulfide-bonding and photo-crosslinking experiments. In the
disulfide-bonding experiments, the authors mutated two Tth RNAP residues to cysteines, one in an α subunit and one in the β’ subunit. In the tight conformation, these cysteines would not be close enough to each other to form a disulfide bond; however, they could form a disulfide bond in the ratcheted state conformation. After incubating this double mutant in oxidizing conditions, they observed disulfide bond formation between the α and β’ subunits both with and without Gfh1 and EC scaffold [13]. To support these experiments, the authors also introduced a photo-crosslinker, p-Benzoyl-L-phenylalanine (pBpa), which targets methionine residues, into an α subunit of Tth RNAP. They then mutated a β’ residue to methionine which was close enough to the location of the pBpa to form a crosslink in the ratcheted state. Again in this experiment, bond formation was observed between the pBpa and methionine residue, indicating RNAP was in the ratcheted state both with and without Gfh1 as well as EC scaffold [13]. Based upon this X-ray crystal structure and biochemical evidence, it appears that RNAP may naturally undergo ratcheting of the core and shelf modules with respect to each other and lends support for this being the method in which RNAP translocates during elongation.

1.4 The Bridge Helix and its Involvement in Transcription Elongation

There are two conserved mobile regions of RNAP postulated to be involved in translocation; the bridge helix (BH) and trigger loop or helix (TL or TH). The BH, highlighted in red in Figures 3, 5 and 6, is a long alpha-helical coil that spans the main channel of RNAP. The BH has been observed in various conformations in X-ray crystal structures solved of bacterial RNAP and yeast Pol II, providing support for its involvement in translocation via conformational changes. A sampling of these conformations are shown in Figure 7, with the BH highlighted in blue and the TL in green. The TL will be discussed in subsequent sections.
Figure 7. Conformations of the Bridge Helix in Various RNAP Structures. This figure demonstrates the different conformations of the BH found in various X-ray crystal structures of RNAP. The bridge helix is highlighted in blue, the trigger loop or helix in green, and the NTP in red. (A) PDB number 1IW7, *T. thermophilus* (*Tth*) RNAP holoenzyme. The bridge helix has a kink and is not perfectly straight, while the trigger loop is highly disordered. (B) PDB number 2O5J, *Tth* EC in complex with a nonhydrolyzable nucleotide analog. In this structure, the BH does not have an unraveled or kinked portion, but is bent towards the NTP. The TL has also ordered to form a TH. (C) PDB number 2E2J, yeast Pol II EC in complex with a nonhydrolyzable nucleotide analog. The BH in this structure is perfectly straight, while the TL is so disordered and flexible that its structure could not be solved. (D) PDB number 2E2H, yeast Pol II EC in complex with the substrate, GTP. In this structure, the BH is unraveled, but in a different position along the alpha helix than in (A). Additionally, the TL is disordered. [17-19]

Based upon the structure in Figure 7A, Vassylyev, et al. proposed a model by which the BH could trigger translocation. The overall shape of the BH in the Tth holoenzyme and in the yeast Pol II EC (panels A and C of Figure 7) is the same; the major difference lies in the folding in the center portion of the BH. In the yeast pol II structure, the BH is completely straight and maintains the α helix, while two residues in the middle of the BH are “flipped-out” in the bacterial holoenzyme structure [17, 19]. It was proposed that the transition between the straight BH and the BH with these flipped-out residues drives forward translocation. The process begins with the BH in a completely α-helical conformation, as seen in the yeast Pol II structure (Figure 7C). After phosphodiester bond formation, two residues in the center of the BH, Asp 1090 and Ser 1091, flip out, forming interactions with the \( i+1 \) nucleotide that could sterically drive translocation forward by pushing the \( i+1 \) nucleotide into the \( i \) site [17].

This flipping out of residues in the middle of the BH has only been observed in X-ray crystal structures of bacterial core and holoenzymes without nucleic acids or substrates [17, 19, 20]. In contrast, the bacterial EC structure with a nonhydrolyzable
ATP analog and yeast pol II structures show completely straight BHs, or a BH that is bent or displaced towards the NTP substrate (Figure 7, [18, 19]). It is possible that these differences are artifacts of crystallization, with differences in scaffolds, substrates, and crystallization solutions for example, resulting in different conformations of the BH. To investigate this possibility, and the hypothesis that BH movements enable translocation, additional studies have been pursued to more conclusively determine the biologically relevant conformations of the BH. One important tool is the antibiotic streptolydigin (Stl), which inhibits various essential processes of RNAP, including transcription initiation, elongation and pyrophosphorolysis [21, 22]. The X-ray crystal structure of Tth RNAP holoenzyme in complex with Stl demonstrated that this antibiotic interacts with and causes changes in the conformation of the BH [23]. These BH conformational changes are not accompanied by any other global changes in the structure of RNAP. In the structure of bacterial RNAP without Stl, the BH is bent, but with Stl, the BH is completely straight, as it is often observed in eukaryotic Pol II structures. Based upon the existence of structures in which the BH is straight and others in which it is bent, it has been proposed that translocation may be driven by the cycling of the BH between straight and bent conformations [24, 25]. The Stl-RNAP structure supports this hypothesis, with the idea that Stl inhibits RNAP by trapping the BH in a straight conformation, blocking this cycling necessary for translocation [23].

In a detailed biochemical study published in 2008, Tan et al. demonstrated that certain mutations of the BH could result in “superactive” RNAPs [26]. In this study, the authors replaced 17 residues in the BH one at a time with each of the other 19 amino acids. In so doing, they found that mutations destabilizing the completely α-helical conformation of the BH caused a dramatic increase in the specific activity of the RNAPs [26]. In fact, the authors concluded that cycling of the BH including a straight conformation may not even be necessary for translocation in the in vitro archaeal system they utilized. The aforementioned structural data, complemented by these mutagenesis studies, all point toward an involvement of BH conformational changes in the process of translocation.

1.5 The Trigger Loop and its Involvement in Transcription Elongation
The trigger loop (TL), another mobile element of RNAP, is located adjacent to the BH just inside the secondary channel. As shown in Figure 7, various conformations of the TL (in green) have been observed. The TL is often in a highly flexible, open conformation, and thus an X-ray crystallographic electron density map for this region did not appear in many structures of bacterial and eukaryotic RNAP (Figure 7 A, C, D). In some cases however, such as the structure of bacterial RNAP EC in complex with a nonhydrolyzable nucleotide analog, the TL folds into a more ordered helical conformation, termed the trigger helix (TH, Figure 7B). The TH interacts extensively with the BH, potentially stabilizing the BH conformation essential for translocation by forming a three-helical bundle [18]. Based upon the comparison between these two structures, Vassylyev et al. postulated that substrate binds RNAP with an open TL, followed by folding of the TL into the TH to trigger catalysis. After release of the PPi product, the TH would unfold to a disordered flexible TL to begin the process again [18].

**Figure 8. Model for TL Movements during Translocation.** This figure, simplified and adapted from Vassylyev et al., 2007, explains their hypothesis of NTP addition. The BH is shown in dark purple, the TL/TH in green, the template DNA strand in yellow, RNA in red, β’ in pink and β in blue. The NTP addition cycle is characterized by first, the NTP binding RNAP with an open TL conformation, followed by ordering of the TL to form the TH. The TH then binds to and properly orients the NTP, triggering catalysis. Following pyrophosphate release, the TH again becomes disordered, reforming the TL [18]. The structures represented by the cartoons
on the top right and bottom left have been solved, while the intermediates of this process have not [1, 18].

The TL has been shown to interact directly with nucleotide substrates via the base, ribose ring, and the β-phosphate [19]. This extensive interaction with the incoming substrate is suggested to be involved in substrate selection. Via the TL, RNAP is able to read out the chemical nature of the NTP, distinguishing rNTPs from dNTPs and discriminating against purine-purine and pyrimidine-pyrimidine mispairing [19]. Additionally, the proximity of a highly conserved TL histidine residue (His1242 in Tth β’) with the β-phosphate of the NTP, as shown in the structures, correlates with involvement in catalysis. At a relatively short distance of about 3.5 Å, a protonated imidazole group of the histidine residue would be capable of withdrawing an electron from the phosphate, facilitating the SN2 attack of the RNA 3’-OH group, enabling phosphodiester bond formation (Figure 9). The subsequent release of pyrophosphate is then proposed to disturb the interaction with this critical histidine residue, releasing the TL and enabling translocation [19]. Molecular dynamics simulations have also demonstrated a strong coupling between pyrophosphate release and TL opening [27].

Figure 9. Model of Nucleotide Addition Involving Conserved TL Histidine Residue. This figure, prepared using ChemSketch (ACD Inc, Toronto, Ontario, Canada) demonstrates the potential utility of the protonated imidazole group of a conserved TL histidine residue in withdrawing electron density from the β-phosphate. This would facilitate the SN2 attack of the RNA 3’-OH group which leads to phosphodiester bond formation. B: represents a general base that extracts the proton from this 3’-OH, enabling this reaction.
Despite the numerous structural clues, the exact function of the TL remains unknown. It is not required for binding of the incoming NTP at the active site, and interestingly, despite its high level of evolutionary conservation, deleting the TL in E. coli and Tth RNAPs does not inactivate these enzymes. Deletion or severe mutation of the TL, however, does dramatically slow down nucleotide addition, without a similar decrease in NTP affinity or the ability of RNAP to distinguish between NTPs and 2’dNTPs [18, 20, 28, 29]. Additional evidence provided details of TL/TH function when the question was asked of which is necessary for efficient loading of substrate; folding of the TL to form the TH, or the movement of the TL while it remains unfolded [18]. To investigate this question, the authors created TL mutants that would prevent the TL→TH folding and then checked the rate of nucleotide addition in vitro. They found that preventing the TH folding slowed nucleotide incorporation 10,000-fold with only a minimal effect on substrate binding [18]. It is therefore possible that NTP is capable of binding RNAP, but requires folding of the TL into the TH to properly align the NTP for efficient catalysis. The necessity for this transition is again emphasized by the existence of antibiotics that work by binding RNAP in such a manner as to prevent the TL→TH transition, including stl in bacterial RNAP and α-amanitin in yeast Pol II [18, 28, 30].

Based upon the investigations described above, it appears evident that the TL is an important mobile element involved in translocation. However, due to the large number of various TL/TH conformations observed in the X-ray crystal structures, and a lack of information concerning intermediate conformations during translocation, the specific function and mechanism of the TL remains unknown. It has been suggested that the interaction between the TL and NTP substrate allows for more efficient catalysis, or the TL could be stabilizing a transition state along the pathway of translocation. However, the lack of timescale information concerning the various steps prevents the discrimination between these two hypotheses [29]. The purpose of my work presented in this thesis is to use new methodologies to overcome this problem and directly observe a translocation step in real time.

1.6 Bridge Helix/Trigger Loop Movements in the Shelf-Core Ratcheting Model of Transcription Elongation
The structural and biochemical evidence collected to date implicates the BH and TL movements in the process of translocation. Therefore, any model of translocation should account for some movement or specific interactions involving the BH and TL of RNAP. The newly postulated shelf-core ratcheting model of elongation, derived from the structure of the Tth EC bound to Gfh1 [13] does implicate these important domains. As shown in Figure 6, the ratcheting of the shelf and core modules with respect to each other is thought to cause the BH to kink, though the kink is interestingly in a different position than was observed in previous bacterial EC structures [15, 16, 18]. This kinked BH contributes to inhibition of the EC by blocking access of the DNA template base to the catalytic site. Simultaneously, the conformation of the TL also changes, in that it assumes an open conformation and moves toward the active site.

It was proposed that this ratcheted intermediate state of translocation was trapped via interaction with Gfh1, and therefore, these conformations of the BH and TL are naturally occurring snapshots of their movements throughout translocation [16]. This model also supports the hypothesis that translocation by one base pair occurs when the kink in the BH pushes the nucleic acids along by one nucleotide [16]. In tandem with solving the structure of the EC bound to Gfh1, the structure of the EC alone was also solved and RNAP was found in the tight state. Based on these two snapshots, the Yokoyama laboratory is able to put forth a model of transcription elongation in which RNAP begins tightly bound to the nucleic acids. As the shelf and core modules ratchet with respect to each other, the BH kinks, relaxing the binding between RNAP and nucleic acids. This kinking of the BH also may force the movement of RNAP by one nucleotide, allowing translocation. At this stage, the RNAP would “reverse-ratchet,” allowing the next template DNA base to enter the active site and enable entering and binding of the incoming NTP. In this manner, elongation would proceed. Additionally, this model remains applicable for a Brownian ratchet model of translocation, in which RNAP is in an equilibrium between pre- and post-translocated states, and NTP binding the post-translocated state drives translocation forward (Figure 4). It is possible that the shelf-core ratcheting occurs spontaneously, and thus this Brownian ratchet model could be occurring in tandem with the shelf-core ratcheting elongation process. Information concerning intermediates along the pathway of translocation would be invaluable in
discriminating between or melding these models to attain a full picture of how RNAP moves and works.

1.7 Raman Crystallography and Time-Resolved X-ray Crystallography

To date, structural biology, genetics, biochemistry and biophysics have contributed the vast majority of the information currently held about RNAP and the process of transcription elongation. There are various limitations to all methods used to this point, however, which prevent the attainment of a specific, definitive model for how RNAP moves along nucleic acids to transcribe DNA into RNA. More specifically, X-ray crystallography has been the primary tool for studying the models of translocation and associated conformational changes of RNAP mobile elements, the TL and BH. One problem with this is that there are many different structures solved of bacterial and eukaryotic RNAP apo-enzymes, in complex with various nucleic acid scaffolds, as well as bound to various transcription factors, inhibitory peptides such as Gfh1 and antibiotics. This large number of structures which have provided much of our high resolution information concerning RNAP and the mechanism of elongation have all been solved using crystals grown in different conditions, and having different packing of molecules inside the crystal. It may be suggested then, that these different crystallization conditions and packing of molecules actually force the RNAP and RNAP complexes into conformations that do not exist in solution [31]. Additionally, many crystals grown in different conditions provide structures with different conformations of RNAP. The question then arises of whether the crystallization process dictated the RNAP conformation. It has been suggested that discrepancies in TL conformations and nucleic acid contacts between their bacterial and other eukaryotic RNAP structures may be explained by differences in crystallization conditions, scaffolds, and the resolution limit of the structures [18].

Another problem is that traditional X-ray crystallography is not time sensitive. Therefore, researchers can only achieve viewing snapshots of the enzyme or transcription at certain points. It is not possible to make conclusions about intermediates along the reaction pathway using only scattered snapshots. Therefore, using traditional X-ray crystallography, it is not possible to determine the order of events or the characteristics of
reaction intermediates along the pathway of transcription elongation. Biochemical and biophysical means may often overcome this problem, providing more information concerning reaction pathways and intermediates. However, these experiments often use unnatural substrates or conditions. While these experiments provide a tremendous amount of important data concerning this essential enzyme and its process, they are not able to give a complete time-resolved picture that does not involve unnatural substrates. Therefore, this thesis uses two methods that in tandem, may allow us to observe the motion of RNAP during elongation over time using natural substrates. These methodologies are Raman crystallography and time-resolved X-ray crystallography.

Raman spectroscopy is able to trace enzymatic events in real time. Raman is a type of vibrational spectroscopy that is capable of measuring very small changes in molecular geometries as well as rates of chemical events. These events include substrate binding and conformational changes [32]. Raman spectroscopy experiments can be completed in crystallo to achieve data with high signal to noise ratios, as the concentration of protein inside a crystal is much higher than can often be achieved in solution [32]. The major advantage of using this technique to study RNAP translocation is that the conformational changes and movement of RNAP can be observed in real time, in a single protein crystal, in a non-invasive manner. Real-time observation of these BH and TL conformational changes is not possible using the previously described biochemical and structural biology experiments that have provided the bulk of information concerning translocation thus far.

Raman scattering is an inelastic process in which a photon interacts with a molecule, enabling an energy transfer and subsequent change in the wavelength of the light. A small portion of molecules excited by photons relax back to a different vibrational energy level than their ground state. This leads to Raman scattering; a compensatory frequency shift of the photon. If the molecule’s final energy state is higher than its initial state, the scattered photon will shift to a lower frequency, known as Stokes Raman scattering. If the molecule’s final energy state is lower than its initial state, the scattered photon will shift to a higher frequency, called anti-Stokes Raman scattering. Raman spectroscopy takes advantage of this scattering effect, as the vibrational
information from the spectra taken can be directly correlated with the chemical bonding and conformations of molecules that are interacting with the photons [39].

A standard Raman crystallography experimental setup consists first of crystals grown in or transferred to a hanging drop crystallization setup. A laser beam is then focused on the crystal in the drop and the illuminating source of the microscope at which this experiment occurs is blocked. The backscattered light then goes back through the microscope and is reflected by optical filters, resulting in a Raman spectral image at an attached charge-coupled device (CCD) photon detector. The detector image is then projected onto a computer screen for viewing [32]. One requirement of this system is crystal size; due to focal volume constraints, the minimum crystal dimension that can be analyzed is 30 µm [32]. Another requirement is time; at least 1 minute is required to record a Raman spectrum, and therefore only slow reactions on the minute timescale can be studied [33]. To observe the chemical and vibrational changes that occur upon ligand binding, ligand is carefully injected into the liquid surrounding the crystal. Another Raman spectra can then be obtained after ligand addition, and difference data can be calculated to characterize the chemical changes that occurred [32]. Utilizing this type of experiment, ligand binding or other chemical changes can be monitored over time using natural substrates. Raman crystallography can also be used in tandem with X-ray crystallography to characterize both the static and dynamic features of a particular chemical or enzymatic process [33].
Figure 10. Raman Crystallography Experimental Setup. This diagram, from Carey and Dong, 2004, delineates the setup for a Raman experiment, and the method in which a Raman spectrum is obtained [32].

For the purposes of the work described in this thesis, Raman crystallography is able to provide valuable information concerning the state of the crystal after introduction of the ligand. The data it provides gives X-ray crystallographers interested in doing time-course studies an idea of the time required for a particular event to occur in crystallo. For example, for the aim of this work, it is necessary to determine the time required for the nucleotide to be added to the elongating RNA transcript and the RNAP to translocate forward in the constraints of the crystal. Many various signals, which will be discussed in subsequent sections, inform researchers that the reaction has occurred after a certain amount of time in the crystal, such as appearance of pyrophosphate signal, decrease in triphosphate signal, etc. Therefore, Raman crystallography is able to provide the necessary time information for highly efficient time-resolved X-ray crystallography experiments.

X-ray crystallography is generally thought to provide static information about structures and protein conformations. However, the method of time-resolved X-ray crystallography is able to provide dynamic information concerning reaction mechanisms and protein conformational changes. This method is possible as long as the protein is active inside the crystal, the reaction within the crystal does not disrupt the crystal lattice,
and the conformation of molecules remain reasonably homogeneous within the crystal so that the intermediate structures can be solved [34]. The reaction to be studied can be initiated in the crystal by either photochemical or diffusion—based methods. Photochemical triggering of a reaction in a crystal works if the reaction is light-sensitive, or if the substrate can be “caged” by a light-sensitive protective group, released upon excitation by photons [35]. Diffraction data is then collected, and various intermediate structures can be determined, providing researchers with a type of movie portraying the reaction of interest. If photochemical means are not possible, the substrate can enter the crystal by diffusion to trigger the reaction, with structures determined over the course of a reaction that is slowed by occurring in the crystal. As long as the reaction itself is substantially slower than the rate of diffusion of substrate into the crystal, then heterogeneity in the crystal, which prevents X-ray structure solving, should not be a problem. In this way, elusive reaction intermediates can be observed inside a crystal with the natural substrates [34].

Raman spectroscopy used in combination with time-resolved X-ray crystallography is highly beneficial. For example, if a Raman spectra revealed that the reaction of interest in the crystal was complete after 10 minutes, then the crystals for structure solving by X-ray crystallography could be frozen for data collection after soaking in nucleotide at time points leading up to and including 10 minutes. In this way, the X-ray crystallographic experiments are focused and will be guided by the information provided by the Raman spectra.

1.8 Specific Aims of this Work

The aim of this work is to directly observe the process of transcription elongation for the very first time, notably without the introduction of unnatural nucleotide substrates. It will be possible to define intermediates along the pathway of translocation, using the information provided from Raman experiments and the structures solved at different time points using X-ray crystallography. The order of events may also be defined; for example, whether the nucleotide binding precedes or follows folding of the TL into the TH. The many discrepancies between the various conformations of the BH that have been reported will be addressed, with the necessary conformations and conformational
changes of the BH being observed with respect to nucleotide addition. Lastly, the
different models of translocation will be tested in this work; the translocation process will
be observed in crystallo in real time, to confirm and/or disprove the various models,
including Brownian ratchet and shelf/core ratcheting mechanisms.
Chapter 2: Materials and Methods

2.1 Materials

All oligonucleotides for the preparation of elongation complex scaffold were purchased from Integrated DNA Technologies (Carlsbad, CA) and Dharmacon (Lafayette, CO). All DNA and RNA were HPLC purified. Thermus thermophilus growth media components were purchased from VWR (Radnor, PA). All purification columns and resins used were purchased from GE Healthcare Bio-Sciences Corp and BioRad (Piscataway, NJ and Hercules, CA, respectively). Vivaspin centrifugal concentrators were purchased from Vivaproducts (Littleton, MA). Polyethylene glycol 8000 and 1,6-hexanediol were obtained from Hampton Research (Aliso Viejo, CA). All other buffers and chemicals were of the highest grade available.

2.2 Fermentation of Thermus thermophilus Cells

Glycerol stocks of Thermus thermophilus HB8 cultures harvested at an OD\textsubscript{600} of 0.4 were created, containing 15% glycerol and flash frozen in liquid nitrogen for storage at -80°C. These stocks were used for all future inoculations of seed cultures for large-scale growth. 6 L of growth media (four 1.5 L cultures in 4 L flasks) were used as seeds for each 240 L culture. Growth media consisted of 8 g/L peptone, 4 g/L yeast extract, 4 mM NaOH, and 34.5 mM NaCl. These were autoclaved for 15 min at 121°C, after which time, the following were added to the media: .35 mM CaCl\textsubscript{2} and 0.4 mM MgSO\textsubscript{4}. These salt solutions were filtered through a 0.4 µm filter before addition. Each 1.5 L culture was inoculated with 1 mL of glycerol stock, and allowed to grow with shaking at 60°C for 26 hours.

A 300 L BioService Fermentor was filled with growth media as well as a standard antifoam agent, saving the addition of CaCl\textsubscript{2} and MgSO\textsubscript{4} until after the fermentor was sterilized. CaCl\textsubscript{2} and MgSO\textsubscript{4} were then added to the growth media, to the same concentrations as in the seed culture, via sterilized addition ports. Seed culture was finally added to the fermentor using sterilized addition ports; the 240 L culture was then allowed to grow for 16 hours at 65°C. Finally, cells were harvested using a Sharples AS-
16 continuous flow centrifuge. 900 g of wet cell paste were collected, and 100 g aliquots were stored at -80°C.

2.3 Purification of Endogenous Core RNA Polymerase from Thermus thermophilus

In a typical purification, 100 g of frozen cells were resuspended in 300 mL of lysis buffer (40 mM Tris-HCl, pH 8 at 4°C, 100 mM NaCl, 10 mM EDTA, 10 mM 2-mercaptoethanol, 1 mM benzamidine, 0.1 mM PMSF, 0.5 µg/mL leupeptin, 0.1 µg/mL pepstatin). A Waring Blender was used to aid in cell resuspension; the blender was then washed with an additional 100 mL of Lysis Buffer, and added to the lysate. Cells were lysed by four passages over an Emulsiflex C3 homogenizer at 20,000 psi. After 30 min, benzamidine and PMSF were added to the lysate to 1 mM and 0.1 mM, respectively. The lysate was then clarified by centrifugation using a Sorvall SS34 rotor for one hr at 17,000 rpm at 4°C.

Glycerol was added to the supernatant to a concentration of 5% (v/v). A 10% stock solution of polyethyleneimine was then added dropwise to a concentration of 0.5% to precipitate nucleic acid and nucleic acid-bound proteins. This mixture was allowed to stir on ice for 30 min before the pellet was recovered by 20 min of centrifugation in the SS34 rotor at 17,000 rpm at 4°C. The pellet was then resuspended and washed with 200 mL of wash buffer (40 mM Tris-HCl, pH 8 at 4°C, 200 mM NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 1 mM benzamidine, 0.1 mM PMSF) and again recovered by centrifugation for 20 minutes at 17,000 rpm at 4°C. This wash step was repeated with an additional 200 mL of wash buffer. Finally, the RNA polymerase (RNAP) was recovered by resuspending the pellet in 100 mL of extraction buffer (40 mM Tris-HCl, pH 8 at 4°C, 1 M NaCl, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 1 mM benzamidine, 0.1 mM PMSF) and centrifuging for 15 minutes at 4°C and 17,000 rpm. This extraction step was repeated with another 100 mL of extraction buffer, and the supernatants from both extractions were combined (200 mL total). Ammonium sulfate powder was then gradually added to 45% concentration (53.3 g ammonium sulfate, ground to a powder, for 200 mL supernatant). This suspension was allowed to slowly stir overnight at 4°C.
Precipitated proteins were collected by centrifugation for 1 hr at 4°C and 17,000 rpm. To remove all traces of polyethyleneimine, the pellets were resuspended in 200 mL TGED (10 mM Tris-HCl, pH 8 at 4°C, 5% glycerol, 1 mM EDTA, 1 mM DTT) with 60% ammonium sulfate (36 g ammonium sulfate per 100 mL TGED). Proteins were again recovered by centrifugation for 15 min at 4°C and 17,000 rpm. The pellet was then resuspended in 1 L of TGED to reduce the conductivity of the sample to below 0.05 M NaCl. This suspension was allowed to stir for 15 min at 4°C.

The sample was then applied to a 250 mL SP Sepharose Fast Flow resin packed in an XK 50 column and equilibrated in TGED. The sample was applied at 4°C using a peristaltic pump at 10 mL/min. The column was then connected to an AKTA FPLC system at 4°C and washed with two column volumes (CV) of TGED at 10 mL/min. The column was then washed with 4 CV of TGED + 0.2 M NaCl. RNAP was then eluted from the column with a 6 CV gradient of 0.2-0.5 M NaCl in TGED. Fourteen-mL fractions were collected, and core RNAP eluted at approximately 0.25 M NaCl. RNAP-containing fractions, as judged by 10% SDS-PAGE gels, were pooled and diluted to reduce the conductivity to 0.1 M NaCl. The core RNAP was then applied to 50 mL of Heparin Sepharose 6 Fast Flow resin packed in an XK 26 column and equilibrated in TGED + 0.1 M NaCl. After loading the sample, the column was washed with 3 CV of TGED + 0.1 M NaCl followed by 3 CV of TGED + 0.2 M NaCl. The protein was subsequently eluted by washing the column with TGED + 0.6 M NaCl, collecting the peak in 5 mL fractions. The RNAP was then concentrated to 5 mL using a Vivaspin 20 Centrifugal Concentrator with a 10,000 kDa molecular weight cutoff.

This 5 mL sample was then applied to a HiLoad 16/600 Superdex 200 prep grade column equilibrated in TGED + 0.2 M NaCl. The column was washed with 1 column volume of TGED + 0.2 M NaCl at 1 mL/min. The elution was collected in 5 mL/tube and fractions were analyzed by 10% SDS-PAGE. Pure RNAP-containing fractions were pooled. This pool was then concentrated and exchanged to storage buffer (20 mM Tris-HCl, pH 7.7, 1 mM EDTA, 10 mM 2-mercaptoethanol, 5% glycerol, 500 mM NaCl) using a Vivaspin 2 Centrifugal Concentrator with a 10,000 kDa molecular weight cutoff. RNAP was concentrated in storage buffer to 35 mg/mL, flash frozen with liquid nitrogen,
and stored at -20°C in 10 µL aliquots. A maximum of 9 mg of pure core RNAP was prepared from 100 g of cells.

2.4 Crystallization of Thermus thermophilus RNA polymerase Transcription Elongation Complex

The elongation complex scaffold consisted of template DNA 5’-TCACTACCACAAGCTACGAGAGCGCCGG-3’, nontemplate DNA 5’-TCGTAGCTTGTGGTAGTGAAGA-3’, and RNA 5’-GCCAGCCGCGGCUC-3’. A second set of template and nontemplate oligonucleotides were also used so as to potentially improve diffraction of the crystals by increasing the number of GC base pairs relative to AT base pairs. The second template DNA sequence consisted of 5’-TCACTACCACAAGCTACGCGAGCGCCGG-3’ and the second nontemplate DNA strand consisted of 5’-GCector Template TTGTGGTAGTGAAGA-3’, with the modifications underlined. Each nucleic acid was dissolved in filtered Millipore water to a concentration of 1 mM.

The elongation complex mixture was prepared on ice, with template DNA, nontemplate DNA, and RNA added to 25 µM and RNA polymerase to 20 µM. Additional storage buffer was added so that the final concentration of NaCl in the mixture was 100-150 mM. The mixture was then incubated in a Biomera Tgradient thermocycler at 70°C for 5 minutes, followed by slow cooling of the mixture to 20°C over the course of 30 minutes. This cooling was achieved by setting the thermocycler to reduce the temperature by 2°C/min.

Crystallization was achieved by the hanging-drop vapor diffusion method using a reservoir solution consisting of 50 mM Tris-HCl, pH 8.5, 30% (w/v) 1,6-Hexanediol, 120 mM Ca(OAc)_2 and a range of 1-5% PEG-8k. To crystallize, 1 µL of the pre-assembled elongation complex mixture and 1 µL reservoir solution were mixed for each hanging drop. The tray was incubated at room temperature, with crystals appearing within 24 hr. Crystals tended to vary in shape and size, but the best diffracting crystals, which were chosen for further study, had at least 200 x 75 µm dimensions and were bipyramidal in shape. Crystals achieved their maximum size after 2-3 days. Crystals are flash frozen in liquid nitrogen after 20 sec of soaking in reservoir solution containing 5% ethanol as a
cryoprotectant. 15-30 sec of soaking in cryoprotectant were acceptable, but longer soaking dried the crystal causing it to no longer diffract.

2.5 Time-Resolved X-ray Crystallography and Raman Crystallography

The Raman crystallography work described here was completed by the Carey laboratory at Case Western Reserve University in Cleveland, Ohio. For the initial Raman crystallography studies of this system, crystals are taken from the tray and soaked in a 5 µL drop of the reservoir solution with 120 mM Mg(OAc)$_2$ instead of Ca(OAc)$_2$ for 1 hr to exchange the metal. This high concentration of magnesium is inhibitory for the transcription reaction [19], and therefore, the crystal was subsequently soaked in the same reservoir solution, but with 10 mM Mg(OAc)$_2$ for 2 hr. Spectra of the virgin crystal were obtained during this 2 hr soak. Next, 5 mM $^{13}$C$/^{15}$N-labeled GTP (Sigma-Aldrich, St. Louis, MO) was soaked into the crystal for 40 min during which time another spectra was taken. These $^{13}$C$/^{15}$N- labeled nucleotides were used to differentiate Raman signals against the DNA bases. Finally, the crystal was transferred to a 5 µL drop containing the reservoir with 10 mM Mg(OAc)$_2$ in place of the calcium, and a final “soak-out” spectra was taken for 87 min. This work was completed at room temperature.

The time-resolved X-ray crystallography was completed with a similar setup, with the crystals being soaked for one hour in the same reservoir containing 120 mM Mg(OAc)$_2$ instead of Ca(OAc)$_2$, followed by soaking for another hour in a 10 mM Mg(OAc)$_2$ reservoir solution to reduce the concentration of magnesium. The crystals were then soaked for various times, up to 3 hr, in nucleotide soaking solution which consisted of the reservoir (with 10 mM Mg(OAc)$_2$ for the metal) and 2 mM of each nucleotide to be added. Before freezing, the crystals were transferred to the same nucleotide soaking solution with the addition of 5% ethanol for 20 sec before freezing in liquid nitrogen. Crystal diffraction was tested at the Penn State home source x-ray core facility as well as the Cornell High Energy Synchrotron Source and the Advanced Light Source at UC Berkeley.

2.6 In crystallo RNA Extension Assay
To directly observe RNA extension in the crystal, the elongation complex scaffold RNA with a 5’-fluorescein label was purchased from Dharmacon, Inc. (Lafayette, CO). *T. thermophilus* elongation complex crystals were then produced using this fluorescently-labeled RNA, following the same protocol as described above, with care being taken to reduce exposure of the crystallization plate to light. After 24 hr, crystals of the same shape appeared, under the same crystallization conditions as described earlier. These crystals were then soaked for one hour in the reservoir solution with a substitution for 120 mM Mg(OAc)$_2$, followed by two hours in reservoir solution containing 10 mM Mg(OAc)$_2$. Finally, crystals were soaked for 1.5 hr in the reservoir solution with 10mM Mg(OAc)$_2$ and 2 mM GTP, to incorporate the nucleotide into the crystals. After 1.5 hours, crystals were added directly to 5 µL of 2x RNA gel loading buffer, composed of 95% formamide, 0.025% bromophenol blue, 20 mM EDTA, and 0.2% SDS. As a control, the nucleotide addition reaction was also completed in solution, by assembling the TEC as previously described, using 100 pmol fluorescent RNA, template DNA, and nontemplate DNA, and 80 pmol of core RNAP. Subsequently, MgCl$_2$ was added to 10 mM and GTP to 2 mM, followed by 10 min of incubation at 60°C. 5 µL of this solution was then directly mixed with 5 µL of 2x RNA gel loading buffer. These reactions were then loaded on a 15% denaturing polyacrylamide gel and ran at 2000V for 3 hr. 5’-fluorescein absorbs at 494 nm and fluoresces at 520 nm and therefore, the gel was visualized on a Typhoon imager (GE Healthcare) using these parameters.

This assay was also used to determine if crystals grown with Mg(OAc)$_2$ will undergo RNA cleavage in the crystal, or if they can be used for future Raman and X-ray crystallographic experiments. *Thermus thermophilus* EC was assembled using the fluorescent labeled RNA according to the previously described protocol. This was then used to set crystallization trays using reservoir solution composed of 120 mM Mg(OAc)$_2$, 50 mM Tris-HCl pH 8.5, 30% 1,6-hexanediol, and a gradient of 2-5% PEG 10,000. The best crystals of about 250 µm in the longest dimension, generally arose in 4% PEG 10,000. These crystals were then soaked in reservoir containing 10 mM Mg(OAc)$_2$ for one hr, followed by soaking in reservoir containing 2 mM GTP for 90 min. Crystals were then picked and directly placed into 1x RNA gel loading buffer, described above, which stopped further nucleotide addition.
Chapter 3: Results

3.1 Fermentation of Thermus thermophiles Cells

A large amount of yellow precipitate formed after 16 hr of growth, which prevented accurate measurements of the OD₆₀₀. However, 16 hr growth is not long enough for the cells to reach stationary phase. The growth phase is critical, as I have found that RNAP purified from stationary-phase cells often has many impurities bound which are difficult to remove. Therefore, it is important for the growth times to be carefully monitored to prevent overgrowth and subsequent difficulties in RNAP purification.

3.2 Purification of Endogenous Core RNAP from Thermus thermophilus

The purification scheme for obtaining highly purified *T. thermophilus* RNAP is provided in Figure 11. Purification from 100 g of cells yielded approximately 9 mg of core RNAP.

![Figure 11. Purification Scheme for Thermus thermophilus Core RNAP.](image)

This scheme summarizes the steps for purification of core RNAP used in these studies. Following this protocol, highly purified, endogenous RNAP can be obtained, which is important for growing high-quality crystals. This procedure can be completed in 4 days.

The SP-sepharose column was useful for separating the core RNAP from holoenzyme. This separation is of particular importance, as it is imperative to remove all sigma factors to ensure a homogeneous population of core RNAP. This homogeneity is
important for preparation of better crystals, which are more likely to yield good quality Raman signals and X-ray diffraction data. Additionally, using a 250 mL column in this first step is helpful, as it allows for a faster flow rate for sample loading and maximum recovery of RNAP. After resuspending the ammonium sulfate pellet in TGED buffer, the conductivity of the sample remained high (equivalent to about 1 M NaCl). This sample could not be dialyzed into low salt buffer, as it precipitated quite heavily. Therefore, the sample must be diluted to approximately 1 L before injecting into the SP-sepharose column, and a large column with a faster flow rate reduced the time needed for this step. An elution profile of the SP-sepharose column is provided in Figure 12.

![Chromatogram of SP-Sepharose Purification of RNAP](image)

**Figure 12. Chromatogram of SP-Sepharose Purification of RNAP.** This profile describes the first column purification step for obtaining *T. thermophilus* RNAP. In this profile, the green line indicates the increasing concentration of NaCl from 0.2-0.5 M. The brown diagonal line gives the increasing conductivity of the column eluent, and the arrow points to the core RNAP-containing peak. These peak fractions were collected and used for the next steps of purification.

After eluting proteins from the SP-sepharose column, the RNAP is already relatively pure as judged by Coomassie staining of an SDS-PAGE gel (Figure 13). These peak fractions were combined and loaded onto a heparin column for further purification.
Figure 13. Purity of *T. thermophilus* RNAP after SP-Sepharose Column Purification. This sample 10% SDS-PAGE gel shows the relative purity of RNAP after the SP-sepharose column. M, molecular weight size marker in kDa. The control lane utilized a *T. thermophilus* core RNAP from a previous preparation. The various subunits of RNAP are labeled next to the peak fraction lane. The lanes in the center of the gel were the wash fractions, which only contained impurities too low in concentration to be visualized by SDS-PAGE.

Heparin columns are frequently used to purify nucleic acid-binding proteins, as heparin is a nucleic acid mimic. After loading the sample onto the column, the column was washed in low salt buffer followed by protein elution in 0.6 M NaCl buffer. The RNAP after this step was relatively pure, as judged by SDS-PAGE (Figure 14).

This heparin eluate was then applied to a Superdex-200 gel filtration column, a common final step for RNAP purification which separates aggregated proteins and contaminants from the RNAP. Following purification, RNAP was concentrated and stored at -20°C in storage buffer for future use. After months of storage in this manner, Tth RNAP was still capable of binding the elongation complex scaffolds and forming high-quality crystals.
Figure 14. Final Steps of Tth Core RNAP Purification. The 10% SDS-PAGE gel given in (A) is an example of the purity of the protein following final purification steps. At this stage, the protein appears to be free of contamination, with only perhaps minor degradation of the large subunits. Again in this gel, the M represents the marker lane with sizes provided in kDa, and the C represents a control lane of previously purified RNAP. (B) Chromatogram of the Superdex-200 column, with the RNAP-containing peak highlighted with an arrow. These peak fractions were stored for future crystallization.

3.3 Crystallization of Thermus thermophilus RNA polymerase Transcription Elongation Complex

As mentioned in the Materials and Methods, the Tth EC was crystallized in 120 mM Ca(OAc)$_2$, 30% 1,6-hexanediol, 50 mM Tris-HCl, pH 8.5, and a gradient of 1-5% PEG 8000. The optimal concentration of PEG-8k is narrow and may depend on the PEG lots, nucleic acid lots, as well as the temperature and humidity of the lab, and therefore a gradient of 1-5% PEG was tried for each crystallization trial. Typically, crystals arose between 3-4% PEG-8k after 1 day, and required 3 days to reach their maximum size. Crystals could also be obtained by growth with PEG 10k, and often yielded bigger crystals. These bigger crystals are suitable for Raman crystallography and in crystallo transcription experiments, but they tended not to diffract to high resolution for X-ray crystallographic studies. Sample crystals are shown in Figure 15.
Crystals within the drops varied in shape and size; however, the highest quality crystals are more than 100µm in the longest dimension. These crystals are then soaked for one hr in the same crystallization solution, but with the substitution of 120 mM Mg(OAc)$_2$ instead of Ca(OAc)$_2$. This soaking step was capable of exchanging the metal from calcium to magnesium to later proceed with the transcription reaction in crystallo. This relatively high concentration of magnesium has been shown to inhibit the transcription reaction [19] and it therefore needed to be reduced. The crystals were transferred to the same crystallization solution containing 10 mM Mg(OAc)$_2$ to reduce the magnesium concentration. The crystals were soaked in this solution for one hr before being transferred to a nucleotide soaking solution. The elongation complex scaffold sequence used in this study was adapted from one provided by the laboratory of Dr. Shigeyuki Yokoyama at Riken, Japan (Figure 16). Based upon this sequence, the next three nucleotides to be added to the elongating transcript were: GTP, CTP, and finally GTP. To incorporate the first nucleotide to the transcript, crystals were soaked in 30% 1,6-hexanediol, 50 mM Tris-HCl, pH 8.5, 3% PEG8000, 10 mM Mg(OAc)$_2$, and 2 mM GTP. To incorporate the first three nucleotides, the crystals were soaked in the same

Figure 15. *T. thermophilus* TEC Crystals. These panels demonstrate typical crystals used for the Raman and time-course X-ray crystallographic studies. These particular crystals were grown in 50 mM Tris-HCl, pH 8.5, 120 mM Ca(OAc)$_2$, 30% 1,6-hexanediol and 3% PEG 8000. The crystals circled in black most likely diffract well; they have a perfect diamond shape, and therefore should be tested. The crystals circled in red have grown into other crystals or contain another imperfection. These crystals should be avoided when making selections.
solution, with the addition of 2 mM CTP. The crystals were lastly soaked in their final soaking solution containing 5% ethanol for 20 sec for cryoprotection.

![RNA Template DNA Non-template DNA](image)

**Figure 16. Scaffold Used for T. thermophilus TEC crystallization.** The next three nucleotides to be incorporated to the 3’ end of the RNA are GTP, CTP and finally GTP, and their template DNA bases are highlighted in green.

### 3.4 Raman Crystallographic Studies

Preliminary studies have been completed using these crystals and soaking conditions by the laboratory of Dr. Paul Carey at Case Western Reserve University. Their preliminary experiments have shown that, within two hr of soaking in GTP, the first nucleotide needed for extension of the RNA transcript (Figure 16), the reaction occurs, RNA is extended by one nucleotide, and product pyrophosphate is formed. In one set of experiments, for example, they determined that after 10 min of soaking in labeled GTP, the signal corresponding to triphosphate was still high, indicating that the nucleotide addition reaction had not yet occurred. After 29 min, the signal for triphosphate had decreased by a factor of 2, while a new signal corresponding to alpha helix formation, arose and remained for the next 6 min. This is presumed to be the TH. Signals for pyrophosphate, the product of the reaction, also arose. After 46 min, the signal shifted to one that corresponded with a beta sheet or disordered structure, potentially the TL. All of the spectra which provided these results had been normalized by buffer subtraction. A sample of these results, provided by the Carey lab, is provided in Figure 17.
Figure 17. Sample Raman Spectra Demonstrating Nucleotide Addition *in crystallo*. This spectra, figure provided by Yianna Antonopoulous, Case Western Reserve University, was taken after 70 min of the crystal soaking in a stabilization solution lacking nucleotides, called a “soak out”. In this solution, the excess unreacted GTP was washed off the crystal, and a signal for triphosphate disappeared, as shown by the black arrow. A few peaks were also present in the pyrophosphate region (region near 1024 cm\(^{-1}\)) indicating the reaction had completed. Additionally, the peak at 1659 cm\(^{-1}\) potentially indicated the presence of an alpha helix, presumably the TH, though more experimentation is necessary to solidify this inference.

3.5 Time-Resolved X-ray Crystallography

Various *T. thermophilus* TEC crystals have been screened and used for data collection at both the Macromolecular Diffraction Facility at Cornell High Energy Synchrotron Source (MacCHESS, Ithaca, NY) and the Advanced Light Source at Lawrence Berkeley National Laboratory (ALS, Berkeley, CA). At these beamlines, the crystals diffract to approximately 5.5 Å, with a small number capable of diffraction higher than 4 Å, both with and without nucleotide soaking. An example of a diffraction image taken at MacCHESS is presented in Figure 18.
Figure 18. Sample Diffraction Image. The crystal used for this diffraction image was soaked in GMPCPP for 10 minutes. It diffracted to 5.4 Å at the CHESS facility. The rings give the resolution limit in each region.

It is also promising for the completion of this work that crystals soaked in nucleotide retain their ability to diffract to moderate resolution. This resolution of 4-5 Å will be sufficient to observe the conformational changes of interest, particularly any bending of the bridge helix, and folding of the trigger loop/unfolding of the trigger helix. Additionally, data was collected at CHESS from a crystal that had been soaked in the nonhydrolyzable nucleotide analog GMPCPP for ten min. Diffraction data was collected to 5.4 Å, and the structure, with space group P4₁2₁2, was solved. The unit cell dimensions were 151.2 x 151.2 x 525.6 Å. R_work for this data was 25.9% while R_free was 32.6%, after completing rigid body refinement to position the nucleic acids into the structure. This structure showed the presence of the GMPCPP in the active site, bound to either the insertion or pre-insertion site. Interestingly, the TL residues were too flexible to provide any electron density showing their position, indicating it had not yet folded into the TH (Figure 19).
**Figure 19. T. thermophilus TEC in Complex with GMPCPP.** In this figure, the blue density is the 2Fo-Fc map, while the red density is the unbiased Fo-Fc map, providing the density in this structure that differs from the reference model. The reference model used was the *T. thermophilus* elongation complex structure published by the Vassylyev laboratory. The template and nontemplate DNA are in green and purple, respectively, while the RNA is in pink. The red density signifying the GMPCPP, bound to either the preinsertion or insertion sites, is also highlighted. The bridge helix can be placed into the density; however, the trigger loop, as evidenced by the dashed green line, is not present. Metal A is also shown in yellow, to be coordinated by the cyan DFDGD motif, but metal B cannot be placed into this density.

### 3.6 In crystallo RNA Extension Assay

A fluorescence-based assay was utilized to visualize extension of the RNA in the crystal, and to check for RNA cleavage inside EC crystals grown in Mg(OAc)$_2$. The Tth EC was assembled using 5'-fluorescein labeled RNA, and crystals were grown in the standard reservoir solution containing Mg(OAc)$_2$ as the salt. Crystals arose after 24 hr that were the same general shape and size as those grown without fluorescent-labeled RNA (Figure 20).
Figure 20. Crystals Grown Using Fluorescent-Labeled RNA. The crystals shown in panel A were grown using TEC complex with 5'-fluorescein labeled RNA. The yellow hue of these crystals, imparted by the RNA, can be seen in comparison with standard unlabeled TEC crystals shown in panel B.

After 3 days, these crystals were incubated in 2 mM GTP for 90 min. The resulting denaturing polyacrylamide gel is shown in Figure 21. This gel demonstrates two important facts; 1) Crystals grown with Mg(OAc)$_2$ in the buffer do not cleave RNA and 2) the RNA is being extended with the addition of GTP in crystallo. This first piece of information will simplify our subsequent experiments, enabling the first soaking step to exchange the metal from Ca to Mg to be skipped. Instead, crystals can be grown directly in Mg and then soaked only once to reduce the concentration of Mg to 10 mM. Finally, and most importantly, this experiment demonstrates in a simple manner that nucleotide incorporation is possible inside a crystal. This confirms the feasibility of the goals of this project. Additionally, this fluorescent-labeled RNA extension assay demonstrated that monitoring the transcription process in crystallo is more efficient than in solution. Figure 22 demonstrates that in solution, unextended RNA remains, as observed by the large amount of original-length RNA transcript remaining after incubation with GTP for one hr at room temperature. The crystallization process, however, is capable of separating active and inactive RNAPs, as well as unextended RNA from the active EC complexes, allowing the active RNAPs to form a homogenous population inside the crystal. Therefore, when I incubate the crystal with GTP for one hr at room temperature, I did not see original 14mer RNA (Figure 22). This confirmed that the protein crystal was an excellent system to use for the study of transcription.
Figure 21. Nucleotide Incorporation Assay. This 15% denaturing polyacrylamide gel directly shows incorporation of GTP to the TEC inside the crystal. Lane 1 contains a control: the *T. thermophilus* TEC in solution, incubated with GTP. Lane 2 gives the TEC crystals, after 90 min incubation with GTP, and contains the same 15mer extended RNA shown in lane 1. Lane 3 contains crystals that were never soaked in nucleotide, demonstrating these crystals grown in Mg do not contain RNA shortened by hydrolysis. Finally, lane 4 shows the original 14mer RNA.

Figure 22. In solution vs. in crystallo Extension Assay. In this 15% denaturing polyacrylamide gel, lanes 1 and 2 show the incorporation of GTP to the TEC inside the crystal, and in solution, respectively. In both cases, a 2.5 fold excess of RNAP with respect to nucleic acid scaffold was utilized and incubation with GTP occurred at room temperature for 1 hour. Lane 3 gives the original 14mer RNA. As can be seen, crystallization served as a purification step to isolate a homogeneous population of active TEC, evidenced by the much smaller proportion of unextended RNA as compared with lane 2.
Chapter 4: Discussion

The purpose of this work is to investigate the mechanism of transcription in bacteria, determining the order of events in the nucleotide addition cycle, and specifically defining the necessary conformational changes of RNAP, nucleic acids and the nucleotide during the process of transcription. Importantly, this is the first study of this process in which we can directly observe transcription in real time, without unnatural substrates. In this work, the motion of RNAP can be observed in a single crystal, in real time at high resolution, to provide a complete picture of the process including elusive intermediates. The combination of Raman crystallography with time-resolved X-ray crystallography allow for efficient investigation of these questions, without introducing RNAP inhibitors or unnatural nucleotide substrates. While previous studies have provided a wealth of important information addressing these questions, it will be necessary for the complete understanding of this process to watch the progression of the reaction in real time at high resolution. This study also provides insights into the transcription mechanism of archaeal and eukaryotic RNAPs since the basic mechanism of transcription is highly conserved from bacteria to humans.

The crystallization of Tth TEC using the DNA/RNA scaffold sequence provided by the Yokoyama Laboratory is fairly straightforward and requires only 24 hours to obtain TEC crystals. Fermentation of Tth cells using a 250 L reactor at the Pennsylvania State University Fermentation Facility and the purification scheme of core RNAP have been optimized to yield about 9 mg of RNAP from 100 g of cells, which is enough for preparing many crystals. This Tth TEC system, therefore, is ideal for investigating the transcription process in crystallo, as different nucleotide soaking conditions can be tested with new crystals grown each day. For Raman crystallography experiments, larger crystals are better than smaller ones, since the purpose of crystallization is to enrich the macromolecule and make a homogeneous population of molecules, providing excellent Raman signals. However, in the case of X-ray crystallographic structure determination, the wide variety in shape and size of crystals, even from the same batch of protein and crystallization condition, as shown in Figure 15, makes selecting adequate crystals
slightly more difficult. After careful analysis of the diffraction qualities of these crystals, I found that perfect diamond shaped crystals diffract best, and are the most stable when soaked in nucleotide-containing solutions. Other shapes, such as those circled in red in Figure 15, as well as smaller crystals under 75 µm generally do not diffract as well. Some crystals can grow to as much as 500 µm in the longest dimension, and these tend not to diffract well due to their instability in nucleotide soaking solutions, and the relative ease with which these crystals crack. The crystals must be handled delicately, especially the larger ones, as their pointed ends tend to crack off easily. Given this information, when selecting crystals for further study, I use perfect diamond shape crystals with no obvious cracks or defects that have dimensions of approximately 200 x 50 x 50 µm. Despite the ease with which this complex is crystallized, reproducibility remains an issue. More work must be completed to ensure that consistently high quality crystals can be obtained and used to collect high-resolution data. Ideally, data could be consistently collected at 4 Å at least, to aid with structure solving and enable observation of the mobile elements of interest.

The protein crystal is an excellent system in which to study this process of transcription, in some ways a better system than in solution. First, the highly concentrated environment in the crystal more closely represents the concentration inside the cell; such high concentrations are not often possible in solution. In the crystal, macromolecules are almost entirely homogeneous and are highly active, moreso than in solution, as shown in Figure 22. The process of crystallization appears to isolate a homogeneous population of active enzyme, and therefore, it is a highly efficient system to use for studying such enzymes. The reaction speed inside the crystal is also much slower than in solution, enabling the reaction to be followed by Raman crystallography as well as for intermediates to be trapped and analyzed by X-ray crystallography. Lastly, completing this study in crystallo allows for the solving of high-resolution structures, which are very important for the complete understanding of the transcription mechanism.

Our collaborators at Case Western Reserve University, Prof. Paul Carey and graduate student Yianna Antonopoulos, have completed the preliminary Raman crystallography work for this project. Their results are highly promising, as they have shown evidence of the generation of pyrophosphate, the product of a nucleotide addition
event, after soaking nucleotide into crystals. RNAP is active inside the crystal, and it will therefore be possible to use this methodology to trace the course of this reaction with respect to time, utilizing all natural substrates and wild-type proteins. It seems the nucleotide addition cycle does occur more slowly in the crystal, which will be ideal for monitoring the Raman signal and having adequate time to freeze crystals at specific time points to trap intermediates and solve their crystal structures. Preliminary results indicate that about 30 min were required for the completion of nucleotide addition in the crystal based on the time course completed in the Raman crystallography experiments. In the future, more time course work must be completed to determine the time necessary for RNAP conformational changes to begin, and for the reaction to complete upon release of pyrophosphate followed by translocation. Additionally, while the results do indicate formation of an alpha helical region during GTP soaking, near perfect Raman spectral overlaps are required to make definitive claims about what chemical species are present at what time during the reaction. These nearly perfectly reproducible results have not yet been attained, and further experimentation is needed to confirm the folding of an alpha helix, presumably the TH, during the course of the reaction.

I originally decided to soak crystals in nucleotide-containing solution for approximately 10 min, based on previous results from my laboratory, which demonstrated nucleotide incorporation in N4 phage RNAP crystals within 10 minutes. An interesting result was obtained by solving the x-ray crystal structure of the Tth TEC after soaking the nonhydrolyzable nucleotide analog GMPCPP. It was found that GMPCPP was in the proper location at the RNAP active site, but may not be in the insertion site, and the TH was not present (Figure 19). The insertion site is the binding site for the next nucleotide to be incorporated, and in this position, it is in the proper orientation for binding. Another nucleotide binding site has been proposed, by the Kornberg laboratory, in which the NTP is bound in an incorrect orientation. This is termed the preinsertion site, and the nucleotide must rotate to bind the insertion site in preparation for incorporation [36]. It cannot be specifically determined which orientation the GMPCPP is bound using this structure because of the lower resolution (5.4 Å).

Binding of the nucleotide in the RNAP active site without formation of the TH could indicate two things; 1) the reaction *in crastallo* will require a longer time to
complete and/or 2) that folding of the TL into a TH may not be essential for proper alignment and binding of the nucleotide. This latter possibility partially correlates with a study in which mutation of the TL to prevent the TL → TH transition did not completely eliminate substrate binding and catalysis of the reaction [18]. Therefore, while folding of the TL into an alpha helix does enhance transcription and lend increased specificity to the substrate selection, it may not be completely required for transcription elongation. It is also possible that the nucleotide may bind to the preinsertion site first, requiring formation of the TH to properly orient the nucleotide in the insertion site. Lastly, the nucleotide could bind the insertion site without the folding of the TL, followed by subsequent TH formation and catalysis.

If the preliminary data is confirmed, and the TL → TH folding does occur after completion of nucleotide addition, then it is possible that the formation of the TH is involved in translocation as well as catalysis of the reaction. Perhaps the nucleotide is capable of binding the insertion site and the reaction is catalyzed while the TL is present, followed by folding of the TL to form the TH, which can work in concert with the BH to push the RNAP forward along the template DNA strand. Again, however, more time course information is required to make any inferences about the order of events in a translocation step. Upon receiving Raman data concerning the amount of time required for maximizing certain intermediates, including product formation and any conformational change such as alpha helix formation. In the future, the same soaking experiment must be completed to solve the intermediate X-ray crystal structures. In this manner, the order of events of nucleotide addition and translocation can be defined, and the different models of translocation can be analyzed in light of real time data obtained without the use of interfering biochemical and biophysical probes.

Other work that must be completed in the future includes improving the X-ray crystal structure. Thus far, the resolution using these crystals cannot be improved reproducibly beyond about 5 Å. This causes difficulty in placing the nucleic acid sequence accurately into the electron density. One proven way to combat this problem would be to incorporate a 5-bromo uridine into the RNA, which will provide an anomalous signal that enables accurate placement of the nucleic acid scaffold into the electron density map [37]. Additionally, the fluorescence-labeled RNA can be used for
TEC crystallization and nucleotide incorporation can be monitored by fluorescence after running a denaturing gel. This is in parallel to the Raman crystallography experiments used to determine the time necessary for completion of the reaction. Crystals produced with the fluorescent-tagged RNA (Figure 20) can be soaked in nucleotide solution and the reaction stopped by dissolving of crystals in RNA gel loading solution after various amounts of time. A time-course gel can then be completed to determine the time necessary for extension of the RNA transcript. This assay should also be optimized, however, so as to be completed with only a single crystal. This experiment can also be used to determine if extension of the RNA transcript by three nucleotides is possible inside the crystal. By soaking crystals in GTP and CTP together, followed by visualization on a gel, incorporation of the next three nucleotides of the transcript should be observed (see Figure 16). Currently, there is a problem of precipitation of the soaking solution containing GTP and CTP. This problem can be avoided, however, by sequentially soaking the crystals in nucleotide once a more definite idea of the time required for nucleotide incorporation is obtained. This variety of experiments will enable the answering of important questions concerning this essential process of transcription elongation, particularly translocation, in the bacterial system.

This work is important for our basic understanding of the essential process of transcription in bacteria. Additionally, bacterial and eukaryotic RNAPs are both multi-subunit, as opposed to the single subunit RNAPs found in viruses, chloroplasts and mitochondria. Therefore, knowledge gained by the study of the movement of a multi-subunit RNAP in bacteria can potentially be applied to the eukaryotic system. Perhaps most importantly, a thorough investigation of this fundamental process in bacterial systems will someday play a role in the development and understanding of antimicrobial therapeutics research by providing new targets for antibiotics or better defining current ones. This is of particular importance given the increasingly more common and dangerous problem of antibiotic-resistant bacterial infections. For example, tuberculosis (TB), which is caused by the pathogen *Mycobacterium tuberculosis*, has been treated successfully with a combination of antibiotics including rifampicin, which acts by inhibiting bacterial RNAP. It is currently estimated that one-third of the world’s population currently carries a dormant form of *M. tuberculosis* which can become
activated at any time, usually when the immune system is under duress in situations such as an HIV infection or with malnutrition. The complicated and long-term treatment plans have resulted in a lack of proper treatment and management of TB. This has resulted in the emergence of antibiotic-resistant strains, such as MDR-TB (multi-drug resistant) and XDR-TB (extensively drug-resistant TB), which have high mortality rates and do not respond to the available drugs. As of 2009, an estimated 40,000 people had XDR-TB in 58 different countries [38]. A manner in which to combat this expensive and dangerous problem is to develop new antibiotics that target essential macromolecules of bacterial activity. Bacterial RNAP is an ideal target, as evidenced by the previous success of rifampicin, an RNAP inhibitor, against TB. Other antibiotics, including myxopyronin and lipiarmycin, also work by inhibiting bacterial RNAP. Bacterial RNAP is essential for gene expression, permitting efficient antibiotic treatment, and the sequences of RNAP subunits in bacteria are not highly conserved with sequences of eukaryotic RNAPs, permitting therapeutic selectivity. A thorough understanding of RNAP and how it works is necessary for the development of much-needed novel antibiotics.
References


Appendix A. Purification of RNAP from Pathogen-Related Bacterial Species

Introduction

Given its size (core enzyme: ~380 kDa; holoenzyme: ~420 kDa), structural studies of bacterial RNAP have been pursued by X-ray crystallography [1-4]. The overall structure of the bacterial RNAP core enzyme, which is the catalytic part of the RNAP, resembles a “crab claw” shape with a 27 Å wide cleft for DNA binding [4]. The enzyme’s active site is located on the back wall of the channel, where essential Mg$^{2+}$ ions are bound. The structure reported in 2002 of T. aquaticus RNAP holoenzyme made great contributions to the understanding of transcription initiation in bacteria [2]. In particular, this structure elucidated the contacts between core RNAP and the promoter specificity subunit, $\sigma$. In addition, it proposes that the process of abortive initiation is caused by a steric clash between the elongating RNA chain and the $\sigma_3-\sigma_4$ linker, which was found to occupy the RNA exit channel. Both of these structures have significantly enhanced our knowledge concerning RNA polymerase function and the mechanism of transcription in bacteria.

To date, the only bacterial species from which adequate RNAP crystal structures have been solved are the thermophiles T. aquaticus [2, 4] and T. thermophilus [3]. Both of these are Gram-negative species [5, 6]. These species are distantly related to common Gram-positive human pathogens such as Staphylococcus aureus and Staphylococcus epidermis, as shown on phylogenetic trees (Figure 1). The more commonly used model bacteria, Escherichia coli, is also Gram-negative and located distant to these pathogens, which is further exemplified by the amino acid sequence identities of the two largest RNAP subunits (Table 1). Thus, neither T. aquaticus nor E. coli can provide an adequate model system for studying RNAP from species such as S. aureus.
Figure 1. Phylogenetic tree of bacterial species, adapted from Ogunseitan (2007) Microbial Diversity. The groups of bacterial species described in the text are highlighted by boxes. Structures have been solved using RNAPs from the thermus genus, boxed in red. Species of interest are boxed in blue and green, and are much more closely related to the model species being used here.
Table 1. Amino Acid Sequence Identities (%) of the Two Largest Bacterial RNAP Subunits. A) Gives amino acid sequence identities for species that can be used as a model for the study of tuberculosis. Currently, the structural work has been completed using *T. aquaticus* and much of the genetic and biochemical work has used *E. coli*. This are both shown to be much poorer models than *M. smegmatis*. B) Gives amino acid sequence identities for species that can be used as a model for the study of Gram-positive pathogens such as *S. aureus*. Again in this table, it can be seen that *B. subtilis* and *G. stearothermophilus* would be much better structural models.

The first goal of this project is to obtain a structure of RNAP from a pathogen-related species, using the model systems *Bacillus subtilis* and *Geobacillus stearothermophilus*. This goal is important to achieve for two major reasons; a high-resolution structure from a pathogen-related species can be used for structure-based drug design, enabling the creation of new antibiotics, and details regarding this essential enzyme and its function in these Gram-positive species is important for our overall knowledge concerning these species with which humans frequently interact.

This goal is important due to the current and increasing problem of antibiotic resistant pathogenic bacteria. MRSA (Methicillin-resistant *Staphylococcus aureus*) is a...
common bacteria that lives on the skin and in the noses of humans, but has acquired mutants which render it resistant to the commonly used antibiotics, including penicillin and amoxicillin. If these microbes enter the bloodstream through a cut or other injury, it can result in serious or potentially life threatening system infections. In 2005, 18,650 people in the U.S. died from a MRSA infection, greater than the number that died of AIDS [7]. A number of antibiotics currently in use work by inhibiting bacterial RNAP, rifampicin and lipiarmycin, for example, and therefore, study of this essential enzyme using an appropriate model is needed.

To obtain this first goal of solving a bacterial RNAP structure from a Gram-positive model species, I have chosen *B. subtilis* as the model organism for the following reasons: 1) fast growth allows for the collection of greater biomass from which to extract highly purified endogenous RNAP; 2) it is genetically tractable, allowing for modifications such as introducing a His6-tagged β’ subunit-encoding gene by homologous recombination which allows for simpler and more complete purification; 3) *B. subtilis* is the model Gram-positive organism and as such, its RNAP is extremely well-characterized; 4) perhaps most importantly, *B. subtilis* RNAP intrinsically lacks many non-conserved regions, which tend to be more flexible; an advantage for producing better-diffracting crystals. Therefore, this study will not only provide the first structure of an RNAP from a Gram-positive bacterial species closely related to important pathogens but its smaller size may enable the collection of higher resolution data. After I obtain high-resolution structures of the *B. subtilis* core and holoenzymes, I will be able to structurally compare Gram-positive and Gram-negative RNAPs. Though the active site regions are highly conserved between these two species, notable differences between the two RNAPs have been observed in biochemical studies. Of particular interest to myself is the differences in specific DNA contacts with the core enzyme; *E. coli* and *B. subtilis* RNAPs have been shown to differ in terms of promoter recognition and utilization, with the *B. subtilis* holoenzyme preferring different promoters from *E. coli* even when it uses the *E. coli* σ70 [8]. This suggests that there are core-specific differences in promoter recognition between the two enzymes, and I will investigate these differences structurally. *B. subtilis* RNAP was also shown to be less prone to abortive transcription than *E. coli*, and responded to hairpin-dependent pause sites and certain termination
signals differently [8]. Biochemical studies suggest that the inability of \emph{B. subtilis} RNAP to recognize hairpin-dependent pause signals is due to differences in the $\beta$ subunit [8]. A high-resolution structure of the core enzyme would provide a starting point for determining the cause of these differences. The holoenzyme structure will enable comparison of $\sigma$-factor contacts with the core enzymes to further elucidate the observed differences in transcription regulation. In addition, the current bacterial holoenzyme structures lack one region of the $\sigma$-factor, $\sigma_{1,1}$, due to its flexibility [2, 3]. It is possible that in this new system, and by utilizing the smaller RNAP, I may obtain holoenzyme crystals containing $\sigma_{1,1}$, which will help elucidate its role in transcription initiation.

Additionally, the potential for successfully crystallizing RNAP and obtaining high-resolution structures increases with the number of different bacterial RNAPs being used. Therefore, I have also decided to use the core and holo RNAPs from the species \emph{Geobacillus stearothermophilus}, a Gram-positive thermophilic species. Crystallization of RNAP from thermophilic species has been more successful, with structures published of \emph{Thermus thermophilus} and \emph{Thermus aquaticus} RNAPs, due to the temperature stability of the enzyme. Finally, cell growth and subsequent RNAP purification using the species \emph{Thermus aquaticus} will also be described in this appendix. Though this species is not closely related to pathogens, working with its RNAP may improve the likelihood of obtaining atomic-level resolution structures, which are necessary for structure-based drug design.

The pathogen \emph{M. tuberculosis} is of interest for the second goal of this project. Despite its former treatability with the antibiotic rifampicin, the World Health Organization maintains that tuberculosis remains second only to AIDS in terms of death by a single infectious agent worldwide. 1.4 million people died from TB in 2010 alone, and multi-drug resistant TB (MDR-TB) has been found worldwide. It is thought that one-third of the world’s population carries dormant TB, which can activate depending upon factors such as overall health and wellness. For instance, AIDS patients are also at a high risk of developing TB [9]. MDR-TB is resistant to the most important first-line drugs that have been used in the treatment of TB; isoniazid and rifampicin. However, of more concern is the recent emergence of an extensively-drug-resistant (XDR-TB) strain, treatment of which is currently limited in terms of options and success. It is estimated
that about 5% of MDR-TB cases are actually XDR-TB, which presents a challenging future picture in which new drugs must be developed to combat these resistant and dangerous bacteria [9].

As shown in Table 1, *Mycobacterium smegmatis* is a much more adequate model system in which to study the pathogen *M. tuberculosis*. Therefore, I have also undertaken developing protocols for the growth of and RNAP purification from *M. smegmatis*. The subsequent sections of Appendix A will address the protocols associated with these various bacterial species I have worked with.

**Methods**

**A 1. Fermentation of Bacillus subtilis Cells**

The MF1 strain of *B. subtilis*, which contains a His$_6$-tag at the C-terminus of the β’ subunit, was provided by Professor Fujita at the University of Houston (Fujita and Sadiae 1998). A previously prepared glycerol stock of this culture, prepared from log phase cells suspended in sterile glycerol to 15%, was used to streak an LB plate containing 10 µg/mL kanamycin for selection. This plate was incubated overnight at 37°C. One isolated colony was selected from the plate and used to inoculate 50 mL LB (1% tryptone, 0.5% yeast extract, 0.5% NaCl) containing 10 µg/mL kanamycin. This seed culture was incubated with shaking at 37°C overnight before being used to inoculate the 100 L culture. The 100 L fermentation culture consisted of 2xYT media, which consists of 1.6% tryptone, 1% yeast extract and 1% NaCl as well as the 10 µg/mL kanamycin. This enhanced media is not necessary for growth of *B. subtilis*, it will grow in the LB media formulation described above, but the growth was fast using the 2xYT media. The media was autoclaved using the Standard Operating Procedures of the Penn State Shared Fermentation Facility prior to inoculation with seed culture. During fermentation, the pH was maintained at 7 using H$_3$PO$_4$ and NaOH, the pressure was maintained at 3.0 psi, air flow was kept at 100 slpm, temperature maintained at 37°C and finally, the agitation was maintained at 200 rpm.

Growth was monitored by OD$_{600}$ hourly measurements, and under these conditions, it was found that beginning cell harvest after 5 hours, at an approximate OD of 4.0, enabled a late log-phase growth harvest, which is optimal. The maximum OD
attainable under these conditions was 5.5. Approximately 800g of cell paste was collected and stored at -80°C in 50 and 100 g aliquots.

A2. Purification of B. subtilis Core RNAP

The RNAP of this strain, as mentioned above, contains a His6-tagged RNAP, enabling affinity purification. The procedure described in this section is outlined in Figure 2. Beginning with a 50 g cell pellet, I first broke up the pellet while it was still frozen, and then resuspended it in 200 mL Lysis Buffer (50 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.3 M NaCl) with the addition of 5 mM 2-mercaptoethanol (BME) and 2 mM phenylmethanesulfonylfluoride (PMSF). Cells can be lysed by either sonication or by 4 passes through an EmulsiFlex-C3 high pressure homogenizer (Avestin, Inc, Ottawa, Ontario, Canada). Next, the lysate was clarified by centrifugation for 30 min at 17,000 rpm and 4°C. If the lysate is not completely clear afterwards, the supernatant was transferred to new centrifuge tubes and the centrifugation repeated. The supernatant was then applied to 25 mL Ni-NTA superflow resin (Qiagen, Valencia, CA) pre-equilibrated in Lysis Buffer. After collecting the flow through, the column was washed with three column volumes (CV) of High Salt Wash Buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 1 M NaCl, 5 mM BME and 2 mM PMSF) followed by 2 CV Low Salt Wash Buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF). Finally, proteins were eluted with 100 mL of Elution Buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 250 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF). Two 50 mL fractions were collected, and generally, about 20 mg of protein are eluted from 50 g starting cell paste. At this point, the protein was ready for Bio-Rex 70 purification, which is capable of removing sigma factors from the core RNAP (Anthony et al. 2000). The success of this purification step depends greatly on pre-equilibration of the Bio-Rex 70 resin. This was achieved by following the equilibration procedure as described in the manual. 10 g of resin were suspended in at least 100 mL of TGED / 0.1 M NaCl buffer (10 mM Tris-HCl, pH 8 at 4°C, 5% glycerol, 0.1 mM EDTA and 1mM DTT in addition to the salt). The suspension is stirred at 4°C for 30 min, after which time the pH is checked and adjusted to pH 8 as necessary. The suspension stirs another 30 minutes,
then the buffer is decanted off and replaced with fresh TGED / 0.1 M NaCl buffer. The entire procedure is then repeated until the pH does not change upon the addition of fresh buffer, generally requiring at least 2 hr.

The Ni-NTA elution was dialyzed for 4 hr at 4°C into TGED / 0.1 M NaCl with the buffer replaced to fresh buffer after 2 hr. If necessary, this dialyzation step can be completed overnight. Next, the dialyzed protein is applied to the equilibrated Bio-Rex 70 resin and is allowed to stir gently for 1 hr at 4°C. The resin is then packed into an XK 16/20 column and washed with TGED / 0.1 M NaCl until the conductivity stabilizes. Finally, the protein can be eluted via a gradient of 0.1 – 0.7 M NaCl over 10 CV. Alternatively, the column can be washed with 2 CV of TGED / 0.2 M NaCl and the protein eluted with TGED / 0.6 M NaCl. A sample results from this step is shown in Figure 3. However, I have found that even with extensive equilibration of the BioRex resin, RNAP is present in the flow through and wash fractions. These fractions can be pooled and reapplied to the Bio-Rex 70 column for maximum yield. Finally, all RNAP-containing fractions were pooled and diluted as necessary to attain conductivity equivalent to 0.2 M NaCl. The sample was lastly applied to a pre-equilibrated, prepacked 5 mL HiTrap Q-sepharose HP column which was then washed with 2 CV TGED / 0.2 M NaCl. The final purified core RNAP was then eluted via a gradient of 0.2 – 0.7 M NaCl over 10 CV. 2 mL fractions were collected. These fractions were judged to be pure by SDS-PAGE, shown in Figure 4. This final Q-sepharose column step is essential for removing one particular protein contaminant which often appears in the early steps of B. subtilis RNAP purification; phosphotransbutyrylase. Additionally, if the cells used were harvested during the end of late log phase or stationary phase, an additional polypeptide, δ factor, will be bound to the RNAP. This final Q-sepharose column step also removes the δ factor, leaving highly purified core RNAP.

After purification from 50 g cells, approximately 8 mg core RNAP can be obtained. These should be used immediately for crystallization trials; however, if storage is necessary, the best method is to buffer exchange the fractions using a Vivaspin concentrator, or dialyze them into RNAP storage buffer (10 mM Tris-HCl, pH 8 at 4°C, 50% glycerol, 0.1 mM EDTA, 1mM DTT, 100 mM NaCl, 10 mM MgCl₂) and store at -20°C.
Figure 2. Purification Scheme of *Bacillus subtilis* Core RNAP. This scheme, as outlined in the previous section, describes the purification protocol for the core RNAP. This procedure can be completed in 1-2 days and about 8 mg of pure core RNAP can be obtained from 50 g of starting material.

![Diagram of purification scheme](image)

Figure 3. Bio-Rex 70 Purification Step of *B. subtilis* Core RNAP. These chromatography profiles and SDS-PAGE results demonstrate the outcome of the Bio-Rex 70 purification step following nickel column purification. Panel A gives the elution profile which included a gradient from 0.1 – 0.7 M NaCl in TGED buffer. As can be seen in the 8-25% polyacrylamide gel on the right, core RNAP was present in the flow through (lane 1) and 0.1 M NaCl wash fraction (lane 2)
as well as in the peak fraction (lane 3). The β and β’ subunits appear as one dark band, since they are both very similar in molecular weight (about 143 kDa) and cannot be resolved. The α subunit is 35 kDa. The M lane in both gels refers to molecular weight marker, which gives the sizes in kDa. In panel B, the flow through and wash fractions from the previous BioRex-70 column step have been re-applied to the BioRex 70 resin. After this second application and repeat of the same elution profile, the core RNAP binds the column almost entirely, with very little shown in the flow through. Instead, the protein elutes in the peak fractions labeled in the chromatography profile and the gel as 2-5. This peak occurred at about 0.25 M NaCl. As can be seen on the gel, low molecular weight impurities remain with the RNAP after this step, making a final purification step using a Q-sepharose column necessary.

Figure 4. Final Purification Step of B. subtilis Core RNAP. This figure gives the Q-sepharose column chromatography profile and resultant SDS polyacrylamide gel that has been silver stained. Lanes 1 and 2 give the Q column flow through and wash fractions, respectively. The remaining lanes correspond to the labeled peaks on the profile. The major peak fraction, corresponding to lanes 5-7, show a very clean core RNAP which has been purified of the contamination remaining after the Bio-Rex 70 column. The profile shows a gradient of 0.2 – 0.7 M NaCl, and core RNAP again eluted at around 0.3 M NaCl.

A3. Expression and Purification of B. subtilis σ^A

Recombinant σ^A, the housekeeping sigma factor of B. subtilis, was used to prepare holoenzyme for this study. I used the pNG590 construct, which encodes wild-type σ^A, to PCR amplify sigA, the gene encoding σ^A and insert it into the pSUMO vector. This vector is useful for the expression of insoluble proteins, since the SUMO tag it encodes enables increased solubility. Additionally, this vector enables quick and efficient purification of the target protein via the His tag it encodes. Lastly, the His and SUMO tags attached to the target protein can finally be cleaved using Ulp1 digestion, resulting in a completely purified target protein free of affinity tags. Additionally, B. subtilis σ^A is toxic in E. coli, and expression using the pSUMO system prevents that
toxicity. Therefore, this system was used for expressing *B. subtilis* sigA. After creating the sigA-SUMO construct, I transformed it into C41(DE3) cells for expression, using kanamycin as a selection marker. I then inoculated 3 mL LB media containing 50 µg/mL kanamycin with one isolated colony. I incubated this culture at 37°C until it became cloudy, and then added the 3 mL of culture into 1.5 L of LB + kan. I incubated this culture at 37°C until it reached an OD{sub}600 of 0.4, then induced with isopropyl-β-D-thiogalactoside (IPTG) to 0.25 mM. After 3 hr of induction at 37°C, cells were harvested at stored at -80°C for future use.

The first purification step was to resuspend the cells in 50 mL of Lysis Buffer (50 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.3 M NaCl) with the addition of 5 mM 2-mercaptoethanol (BME) and 2 mM phenylmethanesulfonylfluoride (PMSF). Cells can be lysed by either sonication or by 4 passes through an Emulsiflex-C3 high pressure homogenizer (Avestin, Inc, Ottawa, Ontario, Canada). Next, the lysate was clarified by centrifugation for 20 min at 17,000 rpm and 4°C. The supernatant was applied using a peristaltic pump to a pre-equilibrated 5 mL Qiagen Ni-NTA Superflow cartridge column. The column was then rinsed with an additional 15 mL of lysis buffer followed by 4 CV wash with high salt buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 1 M NaCl, 5 mM BME and 2 mM PMSF) and 2 CV low salt buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF). Finally, the protein was eluted with elution buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 250 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF), collecting three 5 mL fractions. A total of 25 mg of protein is often eluted from the nickel column after these cell growth conditions. Next, the eulate is digested using the enzyme Ulp1; 1 µg of Ulp1 is added for every 500 mg of protein, and the sample is incubated for 1 hr on ice. If the Ulp1 stock is old or is not working well, the reaction can alternatively be incubated for 2 or more hours at 30°C, as necessary.

After Ulp1 digestion, the sample can be diluted or dialyzed as necessary to reduce the concentration of imidazole to 20 mM. In this way, the sample can then be applied to a second nickel column, which has been equilibrated in the low salt buffer, described above. In this case, the flow through is collected, since the 6x-His affinity tag has been cleaved by the Ulp1 enzyme, leaving un-tagged σ⁵ which will not bind the nickel
column. Typically, 12 mg of purified σA protein can be obtained from this procedure. The procedure summary and sample gels are provided in Figures 5 and 6, respectively.

**Figure 5. Purification Scheme of Recombinant B. subtilis σA.** This scheme, as outlined in the previous section, describes the purification protocol for housekeeping σ factor, σA. This procedure can be completed in 1-2 days and about 12 mg of pure protein can be obtained from 1.5 L of culture.

**Figure 6. Purification of Recombinant B. subtilis σA.** In the 4-15% SDS-PAGE gel labeled A, the result of the first nickel column purification step is shown. M designates the molecular weight marker lane, with units of kDa. Lane 1 is the flow through of the first nickel column. Lanes 2-4 are the elution fractions, which contain the σA-SUMO fusion protein. Lane 5 contains the result after Ulp1 digestion, where the SUMO has been cleaved from σA. Gel B gives the result of the second nickel column, with FT designating the flow through, which contains purified σA free of SUMO and Ulp1.

*A4. Reconstitution of B. subtilis RNAP Holoenzyme*
Reconstituting the *B. subtilis* RNAP holoenzyme from the core RNAP and $\sigma^A$ is straightforward and simplified by the presence of the 6xHis-tag on the core RNAP. First, a three-fold molar excess of $\sigma^A$ was incubated with core RNAP for 30 min at 30°C. A conductivity meter was used to make sure that the conductivity was equal to or less than 0.1 M NaCl so that this sample can be immediately applied to a 1 mL Qiagen Ni-NTA cartridge column equilibrated in low salt buffer. The holoenzyme will bind to the nickel resin, while excess $\sigma^A$ will flow through. Next, elution buffer, described in the previous section, was applied to the column to recover holoenzyme in three 1 mL fractions. An example of a *B. subtilis* holoenzyme purification result is provided in Figure 7.

![Figure 7. Purification of *B. subtilis* RNAP Holoenzyme.](image)

In this figure, M stands for molecular weight marker in kDa. The holoenzyme was reconstituted by mixing core RNAP with a 3 fold molar excess of $\sigma^A$ and incubating for 30 min at 30°C. The sample was then applied to an equilibrated nickel column, and the flow through (FT lane) was collected. Holoenzyme was recovered in this fraction most likely because the column capacity was low. Lanes 1-5 contain elution fractions, which contain reconstituted holoenzyme, in some cases with minor contaminating proteins. The $\sigma^A$ lane provides the control for presence of $\sigma^A$ in the elution fractions.

### A5. Crystallization Trials with *B. subtilis* RNAP

For crystallization of *B. subtilis* RNAP core and holoenzymes, the protein was concentrated to 10 mg/mL. There has been no major progress in the crystallization of this protein thus far at this concentration, and therefore future trials should likely include screens using a higher concentration of RNAP, starting with 20 mg/mL at least. Additionally, this protein generally precipitates in traditional crystallization screening solutions, and therefore, diluted screens or different screens that contain lower concentrations of precipitant should be attempted.
The buffer of the core or holoenzyme being crystallized was first exchanged to standard crystallization buffer using 500 µL Vivaspin protein concentrators with a molecular weight cutoff of 10,000 kDa. The crystallization buffer consists of 50 mM NaCl, 0.1 mM EDTA and 10 mM Tris-HCl, pH 8 at 4°C. The core and holoenzymes have been subjected to crystallization screens using the Qiagen JCSG+ kit at both room temperature and 4°C. They have also been sent to the Hauptmann-Woodward High Throughput Screening Facility in Buffalo, NY, which screens over 1500 different crystallization solutions. Early crystal trials yielded crystals of a protein contaminant, PTB, which I later found could be removed by the addition of the Q-sepharose column. The majority of the crystallization screens had been completed prior to this, and therefore in the future, the crystallization trials should be repeated with the fully purified protein. PTB crystallizes in many different conditions, and so all the early hits I found are mostly likely PTB and not RNAP. I also used the same fork junction DNA used for the *T. aquaticus* RNAP holoenzyme- fork junction DNA complex structure (Murakami et al). I incubated a 1.5 molar excess of fork junction DNA with holoenzyme on ice for 10 minutes and then set JCSG+ screening trays, without success. I also attempted to cocrystallize poly-L-lysine and poly-L-arginine with core RNAP. The idea is that these molecules would serve as polylinkers to facilitate crystallization. I used 10 mg/mL of protein that was 2000 as well as 10,000 fold in excess of the polylinker. I also tried adding 1 mM and 5 mM spermine and spermidine to the RNAP before crystallization to achieve the same goal of facilitating crystal packing. These conditions did not work, however, as I only saw heavy precipitation using the JCSG+ screening kit. Additionally, the high-throughput screens that have been attempted using core and holoenzymes have generally yielded salt crystals.

**A6. Fermentation of *Geobacillus stearothermophilus* Cells**

Strain 10 of *G. stearothermophilus* (Gst) was obtained from the Roe laboratory at the University of Oklahoma. I have prepared glycerol stocks of this strain which can be used for subsequent large-scale growths. 2 L of seed culture were first prepared by inoculating each 1 L 2xYT culture with a 1 mL glycerol stock directly. The cultures grew overnight at 60°C with shaking. They were then used to inoculate 100 L of 2xYT
media which consists of 1.6% tryptone, 1% yeast extract and 1% NaCl, in a bioreactor that had been previously sterilized according to standard operating procedures. During growth, the pH was not maintained by addition of acid and base as in the *B. subtilis* fermentation. Cell harvest began after 5 hours of growth when the culture had reacted an OD$_{600}$ of 1.6. This corresponded to the late-log phase, which is the optimal time for cell harvest. Under these conditions, about 600 g of cell paste can be collected.

A7. Purification of *G. stearothermophilus* Core RNAP

First, 100 g of cells were resuspended in 400 mL of Lysis Buffer (50 mM Tris pH 8 at 4°C, 10 mM MgCl$_2$, 2 mM EDTA, 1 mM DTT, 1 mM BME, 23 mM NaCl, 10% glycerol). After the addition of PMSF to 1 mM, cells were then lysed using a high pressure homogenizer or a microfluidizer. After lysis, the sample was centrifuged at 17k rpm for 45 min at 4°C. The supernatants were combined into a large beaker, to which Polymin P was added dropwise to a final concentration of 0.5%. This solution was then allowed to stir an additional 30 min at 4°C. The precipitate was collected by centrifugation for 20 min at 14k rpm using a FiberLite F14S 6x250y rotor. The pellets were then washed and resuspended in enough TMED + 0.4 M NH$_4$Cl (10 mM Tris pH 8 at 4°C, 0.1 mM EDTA, 1 mM DTT and 10 mM MgCl$_2$) to re-fill the centrifuge tubes. They were then centrifuged for 15 min at 17k rpm. The supernatants were poured off and the pellet was resuspended again in TMED + 0.4 M NH$_4$Cl. The centrifugation step was repeated to collect the pellets. The pellets were then resuspended in TMED + 1 M NH$_4$Cl and incubated on ice for 5 min before centrifugation for 15 min at 17 k rpm. The supernatant of this centrifugation was saved for later purification. This elution step can be repeated to ensure complete recover of proteins.

Next, ammonium sulfate powder (AS) is gradually added at a concentration of 36 g AS per 100 mL sample. After slowly adding AS powder, I allowed the sample to stir an additional 30 min at 4°C and then allowed it to sit overnight at 4°C. The next day, I collected the precipitate by centrifugation for 45 min at 17 k rpm. I then resuspended the pellet in 100 mL of chilled TGED + 60% AS (36 g AS + 100 mL TGED) and centrifuged again for 20 minutes at 17 k rpm to recover the pellet. This step is included to remove all traces of Polymin P from the preparation, which may hinder future purification steps. I
then suspended the pellet in 50 mL of TGED and dialyzed it for 4 hr into TGED buffer, replacing the TGED with fresh buffer halfway through. After this time, the conductivity meter showed that the concentration of salt was low enough to continue with the next step. Alternatively, the sample could be dialyzed overnight into TGED, or diluted with TGED to reduce the salt concentration to 0.1 M NaCl. Once this is complete, the sample was applied to a pre-equilibrated 50 mL Heparin FF column that I assembled previously using resin purchased from GE Healthcare. Proteins were eluted over a 10 CV linear gradient of 0.1 M – 0.6 M NaCl. 5 mL fractions were collected. Core RNAP elutes at around 0.45 M NaCl. A sample result from this step is provided in Figure 9. The RNAP-containing fractions are then diluted to reduce the concentration of NaCl to 0.1 M and loaded onto pre-equilibrated 1mL HiTrap Q column. The 0.1 M – 0.6 M NaCl linear gradient was repeated on this column, with 1 mL fractions collected during a 20 CV gradient. Core RNAP again eluted around 0.45 M NaCl. A sample result from this final step is shown in Figure 10. Approximately 12 mg core RNAP can be obtained from 100 g starting material.

Figure 8. Purification Scheme for Gst Core RNAP. The purification of Gst core RNAP yields on average 12 mg of RNAP from 100 g starting material. This entire purification requires about 3 days to complete.
Figure 9. Heparin Column Purification of Gst Core RNAP. The arrows on the chromatography profile on the left correspond to the fractions collected for SDS-PAGE analysis on the right. This 10% polyacrylamide gel shows that RNAP elutes later in the heparin column 0.1-0.6 M NaCl gradient. The red arrow corresponds to #23 in the gel on the right, containing core RNAP. Lane FT shows the heparin column flowthrough, and lane I shows the input before heparin column purification. As can be seen, this purification step works well to isolate RNAP from the dirty ammonium sulfate precipitation product.

Figure 10. Q Column Purification Step of Gst Core RNAP. This is a summary of the result of the final purification step of Gst core RNAP. The profile on the left shows that two major peaks elute from the Q-sepharose column program, which consisted of a 0.1-0.6 M NaCl linear gradient. Peak I, shown on the SDS-PAGE result on the right, contains contamination, while peak II contains pure core RNAP. The core RNAP eluted around 0.45 M NaCl.

A8. Expression and Purification of G. stearothermophilus $\sigma^A$ and $\sigma^A\Delta$1.1

To set up crystallization trials with G. stearothermophilus (Gst) holoenzyme, I expressed and purified the housekeeping sigma factor, $\sigma^A$ as well as a mutant lacking sigma region 1.1, which is flexible enough that it may interfere with crystallization. To complete this, I first prepared Gst genomic DNA. During the 100 L fermentation of Gst,
I collected a small sample of the culture and centrifuged it to collect a log phase cell pellet. I used this pellet to purify genomic DNA according to the following protocol:

Next, I used two primers, which can be found in the laboratory database, to amplify $\sigma^A$ plus a 400 bp surrounding region using 300 ng genomic DNA as a template. The cycling conditions for this PCR consisted of 3 min at 95°C, (30 sec 95°C, 30 sec 57°C, 30 sec 72°C) x 30, and finally 5 min at 72°C. Next, the amplified $\sigma^A$ was prepared for cloning into the pSUMO vector by using two additional primers, which I have added to the laboratory database. The cycling conditions for this PCR were 98°C for 15 sec, 59°C for 2 sec, followed by 72°C for 20 sec. This was repeated 30 times, and followed by a final extension for 5 min at 72°C. The product was PCR purified, and then ligated into pSUMO using the guidelines in the pSUMO manual, just as I did with $B. subtilis$ $\sigma^A$.

To prepare Gst $\sigma^A$ Δ1.1, I used 100 ng Gst $\sigma^A$-SUMO as a template for PCR using primer numbers. I then cloned this insrt into pSUMO according to the manual protocols.

For expression of these proteins, I transformed Gst $\sigma^A$-SUMO or Gst $\sigma^A$ Δ1.1-SUMO into BL21(DE3) cells. After overnight growth at 37°C, I inoculated 50 mL of LB + kanamycin with a single colony and allowed it to grow with shaking at 37°C for 2.5 hr. 3 mL of this small culture was then added to a 1 L LB + kanamycin culture and was allowed to grow with shaking for about 6 hr until the OD$_{600}$ reached 0.4. Expression was induced by the addition of IPTG to 0.25 mM and overnight growth at 22°C. Cells were then harvested by centrifugation for 5 min at 5000 rpm using a FiberLite F10-6x500y rotor. Alternatively, expression could occur for 3 hr at 37°C. This same protocol was followed for both Gst $\sigma^A$ constructs.

To purify either of these constructs from the harvested cells, the exact same procedure as in methods section A3 was followed. Sample results from purification are shown in Figure 11.
Figure 11. Purification of Gst $\sigma^A$ and $\sigma^A \Delta 1.1$ using the SUMO system. The SDS-PAGE result on the left gives the purification of wild type $\sigma^A$. Lane 1 gives the $\sigma^A$-SUMO fusion protein which is the elution from the first nickel column purification step. Lane 2 gives the product of Ulp1 digestion of this first nickel column elution. Lane 3 is the elution from the second nickel column step, which means it contains Ulp1 and SUMO proteins, as well as some undigested $\sigma^A$-SUMO. Lastly, lane 4 gives the flow through of the second nickel column, which is purified $\sigma^A$.

The SDS-PAGE result on the right gives the same information, except with the $\sigma^A \Delta 1.1$ mutant. Lane 1 is completely purified Gst $\sigma^A \Delta 1.1$ as a control. Lane 2 is the second nickel column elution, made up of Ulp1 and SUMO. Lane 3 is the cleaved $\sigma^A \Delta 1.1$ -SUMO before the second nickel column step. In this lane, the $\sigma^A \Delta 1.1$-SUMO is not completely digested by Ulp1. Lane 4 is the flow through of the second nickel column, and thus contains only purified $\sigma^A \Delta 1.1$. This is in comparison to lane 5, which contains wild-type Gst $\sigma^A$. Wild-type Gst $\sigma^A$ is 43 kDa, while the region 1.1 deletion mutant is 31 kDa.

A9. Reconstitution of G. stearothermophilus RNAP Holoenzyme

To reconstitute Gst holoenzyme for crystallization trials, a 3-fold molar excess of the $\sigma^A$, either wild-type or the region 1.1 deletion mutant, was incubated with core RNAP for 30 min at 30°C. This sample was then applied to a Superdex 200 column equilibrated in TGED / 0.1 M NaCl. 1 CV of Gel Filtration Buffer (TGED + 0.2 M NaCl) was applied to the column, with 5 mL fractions collected. A sample result is provided in Figure 12.
Figure 12. Reconstitution of Gst holoenzymes. Panel A gives the reconstitution of Gst holoenzyme using the wild type σ^A. As can be seen on the SDS-PAGE result on the right, fractions 9-10 on the chromatographic profile contain pure RNAP wild-type holoenzyme. Panel B gives the reconstitution of Gst holoenzyme using a σ^A deletion mutant, which lacks the region 1.1. This region is highly flexible and may inhibit crystallization. The SDS-PAGE result on the right gives the peak fraction contents (P) of the profile on the left, as well as the purified σ^A Δ1.1 as a control. Both of these samples were created by incubating a three-fold molar excess of the sigma factor with core RNAP for 30 min at 30°C, followed by gel filtration purification.

A10. Crystallization Trials with G. stearothermophilus RNAP

For crystallization of Gst RNAP core and holoenzymes, the protein was concentrated to 10 mg/mL. There has been no major progress in the crystallization of this protein thus far at this concentration, and therefore future trials should likely include screens using a higher concentration of RNAP, starting with 20 mg/mL at least. Additionally, this protein generally precipitates in traditional crystallization screening solutions, and therefore, diluted screens or different screens that contain lower concentrations of precipitant should be attempted. The buffer of the core or holoenzyme being crystallized was first exchanged to standard crystallization buffer using 500 µL Vivaspin protein concentrators with a molecular weight cutoff of 10,000 kDa. The crystallization buffer consists of 50 mM NaCl, 0.1 mM EDTA and 10 mM Tris-HCl, pH 8 at 4°C. The core and holoenzymes with both wild-type and deletion mutant σ^A, have
been subjected to crystallization screens using the Qiagen JCSG+ kit at both room temperature and 4°C. They have also been sent to the Hauptmann-Woodward High Throughput Screening Facility in Buffalo, NY, which screens over 1500 different crystallization solutions. Sample potential crystallization hits are given in Figures 13-15.

**Figure 13. Gst RNAP crystallization trials.** These five images give hits that I have found with the goal of crystallization Gst RNAP. The specific reservoir solution for each image is provided underneath. In all cases, the protein was at approximately 10 mg/mL. Images A and B utilized Gst holoenzyme with a full-length σ^A. The microcrystals in images A and B were used for microseeding into drops that contained a lower concentration of precipitate, however, this did not provide crystals. Images C, D and E utilized Gst σ^A Δ1.1 holoenzyme.
Figure 14. Gst RNAP crystallization trials. As in Figure 13, each condition is written underneath its specific image. Image A was taken after 6 weeks of incubation and used Gst holoenzyme containing $\sigma^A \Delta 1.1$. Image B was taken after 2 months of incubation and used Gst holoenzyme with full-length $\sigma^A$. Images C-E also used Gst $\sigma^A \Delta 1.1$ holoenzyme. Also, I tested image B and the microcrystals did fluoresce.
Figure 15. Gst RNAP crystallization trials. As in Figure 13, each condition is written above its specific image. The top two images contain aggregated protein and oil droplets that do fluoresce. The bottom image contains what appear to be oil droplets and also fluoresce.

*11. Fermentation of Mycobacterium smegmatis Cells*

A *Mycobacterium smegmatis* (Msmeg) strain containing a His$_6$-tagged RNAP was provided by the Keiler Lab at the Pennsylvania State University. This strain requires 50 µg/mL of the antibiotic hygromycin for growth. Msmeg growth media consists of standard LB supplemented with 0.5% dextrose, 0.5% glycerol, 0.05% Tween 80. I inoculated small cultures of Msmeg growth media with isolated colonies of this strain and allowed them to grow for 28 hours at 37°C. I then added glycerol to 15% and stored these glycerol stocks at -80°C for future use. For large-scale fermentation of this species, I inoculated two 1 L cultures of Msmeg growth media +50 µg/mL hygromycin with a 1 mL glycerol stock each. After 23 hours, the OD$_{600}$ of the cultures were 0.041. I then allowed these cultures to grow without shaking at 37°C overnight. This was done to ensure that I did not inoculate the large-scale culture with a stationary-phase seed culture. As a note, when autoclaving the media and materials for growth, the dextrose is always autoclaved as a separate solution, and then added to the final culture before inoculation.
The seed culture was then used to inoculate the large scale (150 L) culture. At the time of inoculate, the seed culture had an OD of 0.167, which corresponds to early log phase. The culture was allowed to grow, without adjustment of pH, for 24 hours at 37°C. 250 g cells were collected from the 150 L culture after 24 hours of growth. It was not possible to be confident in the measurement of the OD at the time of cell harvest, as a standard antifoam agent was added to the culture and may affect OD readings. This antifoam agent was added because the Tween 80 causes foaming and bubbling of the culture which could negatively affect the fermentation.

A12. Purification of Mycobacterium smegmatis Core RNAP

Msmeg core RNAP is purified using the same buffers and procedure as the purification of B. subtilis core RNAP, provided in section A2. The one major difference between the B. subtilis and Msmeg core RNAP purifications is that the BioRex-70 column is able to separate σ^A holoenzyme from core RNAP in the case of Msmeg. It may be because of the phase of growth in which the cells were harvested, but I never see σ^A holoenzyme eluting from the BioRex-70 resin when purifying B. subtilis core RNAP. However, in the case of Msmeg RNAP purification, the BioRex-70 column elutes two different peaks, the first of which contains σ^A holoenzyme and the later peak containing core RNAP. I further purified the core RNAP according to the B. subtilis core RNAP purification. I further purified the holoenzyme-containing fractions using a Superdex 200 gel filtration column, the same procedure and buffers as described in section A9. Sample purification results are given in Figures 16 and 17.

![Figure 16. Purification of M. smegmatis RNAP.](image)
right summarizes the first two steps of Msmeq core RNAP purification; nickel and BioRex-70 columns. M stands for marker lane in kDa. Lane 1 is empty. Lane 2 gives peak III from the BioRex-70 gradient. This contains purified core RNAP. Lane 3 gives a sample from peak II shown on the profile on the left, and this corresponds with σ^A holoenzyme along with some impurities. Lane 4 gives a sample from peak I. Lane 5 gives the flowthrough from the BioRex-70 purification, and as can be seen (similar with the result for B. subtilis purification) some core RNAP is often lost in this step unless the flowthrough is passed back over the BioRex resin. Lane 6 gives the nickel column elution. Lanes 7 and 8 give the low and high salt washes of the nickel column, respectively. Lane 9 is the flowthrough of the nickel column. Lane 10 is the supernatant after lysing and centrifuging the lysed cells. And finally, lane 11 gives the cell lysate.

**Figure 17. Final Purification Step of M. smegmatis core RNAP.** The top profile gives the elution profile for the final Q-sepharose column step of purification of peak II in the BioRex-70 profile in Figure 16. This input was therefore σ^A holoenzyme. As can be seen, 4 major peaks eluted in the profile. These peaks were tested on the SDS-PAGE result in the lower image. Lane 1 gives the input. Lane 2 is a control, showing the migration of M. smegmatis σ^A. Lane 3 is empty, containing the Q column flow through. Lane 4 contains a sample from peak I in the above profile. Lane 5 gives a sample from peak II. Lane 6 is a sample from between peaks II and III. Lane 7 is a sample from peak III, which contains purified σ^A holoenzyme. Finally, lane 8 contains a sample from peak IV, which did not contain protein. The holoenzyme prepared here did contain some contamination or degradation of the largest subunits, as can be seen on this gel.

**A13. Expression and Purification of M. smegmatis σ^A**
The gene encoding the Msmeq housekeeping sigma factor, \( \sigma^A \) also called mysA, was purchased from GeneArt. I ligated this insert into the pSUMO vector, which is an excellent system to use for the expression and purification of bacterial sigma factors, as explained in section A3. The expression and purification of mysA protocols are exactly the same as described in sections A3 and A8. At least 6 mg of purified mysA can be purified from a 1 L culture. Sample results from this purification are shown in Figure 18.

Figure 18. \textit{M. smegmatis} \( \sigma^A \) Purification. This 4-15\% SDS-PAGE result gives the purification steps of \( \sigma^A \) using the pSUMO system. Lane 1 gives the elution from the second nickel column, and is therefore purified \( \sigma^A \). Lane 2 gives the Ulp1 digested \( \sigma^A \)-SUMO before passage through the second nickel column. Lane 3 gives the elution from the second nickel column, which does contain some undigested \( \sigma^A \)-SUMO.

\textit{A14. Reconstitution of M. smegmatis RNAP Holoenzyme} 

To reconstitute Msmeq holoenzyme for crystallization, a three-fold molar excess of mysA was incubated with core RNAP for 10 min at 37\(^\circ\)C [10]. Additionally, since native mysA holoenzyme can be purified along with the core RNAP, I tried a different tactic to obtain a homogeneous population of holoenzyme for crystallization. I took the holoenzyme purified as in section A12, and saturated it with newly purified recombinant mysA. I then incubated this for 10 minutes at 37\(^\circ\)C and purified by a Superdex 200 column equilibrated in TGED / 0.2 M NaCl. I collected 2 mL fractions and eluted a singular sharp peak. These fractions were then directly used for crystallization trials.

\textit{A15. Crystallization Trials with M. smegmatis RNAP}
The crystallization trials I attempted were very similar as with *B. subtilis* and Gst RNAP trials. I generally only tried 10 mg/mL protein concentration, and therefore a higher concentration of protein as well as lower concentration of precipitant in the crystallization buffers should be tried. I set trays using the JCSG+ kit as well as the Cryo1 and Cryo2 screening solutions. I also tried homemade trays as usual for my core RNAP screening trials, which included various pH values of Tris buffer (pH 7-9), 20 mM MgCl$_2$, and ammonium sulfate gradients (35-50%) [4]. I tried the screening kits using both core and holoenzymes. For the holoenzyme crystallization, I also tried a screen using 10-40% PEG (molecular weights 200, 3350 and 8000), 0.1 M HEPES pH 7 along with one of three different salts (0.1 M Li$_2$SO$_4$, (NH$_4$)$_2$SO$_4$, and MgSO$_4$). I also tried diluting the JCSG+ screening kit solutions 6-fold, and using those new solutions for crystallization of holoenzyme. Unfortunately, I had problems with fungus growing in the crystallization drops. It appears that the lower concentration of precipitant results in relatively quick contamination and growth of fungus in the drops regardless of filtering of the solutions before use. I did observe a few hits, but they were found to be salt crystals and in a few cases, a crystalline shape formed from fungus. Thus far then, for the purposes of pathogen-related bacterial RNAP crystallization, the most promising candidate is Gst.

**A16. Fermentation of Thermus aquaticus Cells**

*Thermus aquaticus* (Taq) is species from which RNAP has already been crystallized. To cocrystallize this RNAP with different antibiotics, the growth and RNAP purification procedures in this laboratory also needed to be developed. The main difference between Taq and *T. thermophilus* growth is that Taq cannot grow in enriched media. Therefore, for the growth of this species I used the following media: 0.2% glucose, 0.2% yeast extract, 4 mM NaOH, 0.35 mM CaCl$_2$, 0.2% NaCl and 0.4 mM MgSO$_4$. I first inoculated two 1.5 L cultures using a 1 mL glycerol stock each. After each large-scale growth, I prepare a large amount of glycerol stocks, mixing the culture with glycerol to 15% followed by storage at -80°C. I allowed each of the seed cultures to grow overnight, and then added the entire seed culture to a 240 L bioreactor for fermentation. I allowed this culture to grow at 70°C for 40 hours until the OD$_{600}$ reached...
0.5. After harvest, I collected 280 g of cell paste. Under these conditions, Taq experiences a long lag phase of growth, with no measurable OD until after at least 30 hours of growth. Also, I did not autoclave any of the media or materials used, since the growth temperature is high. When constrained by time, this could be done; however, if time allows, everything should generally be autoclaved to prevent contamination from other thermophilic species.

A17. Purification of Thermus aquaticus RNAP

The RNAP purification protocol for Taq is very similar to the *T. thermophilus* RNAP purification. I first resuspended 100 g of cells in 350 mL Lysis Buffer (40 mM Tris pH 8, 100 mM NaCl, 10 mM EDTA, 10 mM BME, 1 mM benzamidine, 0.1 mM PMSF, 0.5 µg/mL leupeptine and 0.1 µg/mL pepstatin). I broke up this cell pellet using a Waring Blender, then used 4 passes over the Emulsiflex C3 to lyse the cells. Next, I centrifuged for 45 min at 17k rpm and 4°C to obtain the supernatant. I added glycerol to 5% to the supernatant and then precipitated nucleic acids and their associated proteins by adding Polymine P dropwise to 0.5%. I then allowed this suspension to stir for 30 min at 4°C. Next, I centrifuged this sample for 20 min at 14 k rpm. I saved these pellets, but needed to recentrifuge the supernatant, as the pellets resulting from this Taq purification do not pack as well as they do using *T. thermophilus*. After completely recovering the pellets, I washed and resuspended them in Wash Buffer (40 mM Tris pH 8, 200 mM NaCl, 1 mM EDTA, 10 mM BME, 5% glycerol, 1 mM benzamidine and 0.1 mM PMSF). I then repeated this wash step. Next, I resuspended the pellets in Extraction Buffer (40 mM Tris pH 8, 800 mM NaCl, 1 mM EDTA, 10 mM BME, 5% glycerol, 1 mM benzamidine, 0.1 mM PMSF) and then centrifuged for 20 min at 14 k rpm. This extract step was completed twice, and the liquid collected was yellow-brown in color.

Next, I slowly added powdered ammonium sulfate to 45% (160 g for 600 mL sample). I stirred this for 30 min at 4°C, then allowed it to sit overnight at 4°C. The next day, I collected proteins by centrifugation for 45 min at 14k rpm and 4°C. I then resuspended the pellet in 60% saturation of ammonium sulfate, which was cooled before use. This suspension was centrifuged for 20 min at 14 k rpm, and the resultant pellets were resuspended in 1.5 L of TGED buffer. This was allowed to stir for 15 min at 4°C
before being applied to a 50 mL Heparin FF column pre-equilibrated in TGED. After sample application, the column was washed with 3 CV of TGED, followed by washing with TGED / 0.2 M NaCl until the conductivity and UV on the chromatography profile did not change. Next, RNAP was eluted by washing the column with TGED / 0.6 M NaCl until the conductivity and UV profiles did not change. I collected 28 mg of protein from this step. Next, I loaded the elution onto a 1 mL Q-sepharose column to concentrate the sample and further purify. This Q column was pre-equilibrated in TGED / 0.2 M NaCl. I washed the column after sample application with TGED / 0.2 M NaCl, and then eluted with 7 mL of TGED / 0.6 M NaCl. Approximately 14 mg protein eluted from this step as judged by the Bradford assay. Finally, I loaded these 7 mL onto a Superdex-200 column (26/20) which was equilibrated in TGED / 0.2 M NaCl. Proteins were eluted by applying this equilibration buffer and collecting 5 mL/fraction. The 6.5 mg of RNAP found in one peak of the SD-200 elution profile still contained contamination. These RNAP-containing fractions were applied to a Source Q column equilibrated with TGED / 0.2 M NaCl. RNAP was eluted by a linear gradient of 0.2 -0.6 M NaCl over 10 CV, collecting 4 mL / fraction and running the column at 2 mL/min. The two fractions containing the most pure core RNAP were pooled and saved for future use. Due to my strictly selecting the highest-purity RNAP-containing fractions for the next step, as well as other minor technical issues, I only obtained about 2 mg of core RNAP from 100 g cells. This can most assuredly be improved, however, with the usage of new or regenerated heparin resin and more flexibility in selecting fractions for further purification. A sample SDS-PAGE result is provided in Figure 19.
Figure 19. Purification of Taq Core RNAP. This 10% SDS polyacrylamide gel gives the final result of purification of Taq core RNAP, after the Source Q column step. To save time, the initial column chromatography steps were not linear gradients, just washes followed by a higher salt elution step. Therefore, this purification is fast as the product does not need to be checked until after the SD-200 column, or the final step, the Source Q column. In this gel, lanes 2-5 all contain samples from each peak that eluted from the Source Q column. As can be seen, the last two peaks contained pure Taq core RNAP.
References


Appendix B. Determining the Binding Location of the δ Subunit on the B. subtilis Core RNAP

Introduction

This 20.4 kDa polypeptide exists as a subunit of RNAP only from Gram positive species such as B. subtilis and S. aureus [1]. The complete structure of the δ subunit and its binding location on RNAP are unknown, and its physiological function has not been specifically determined. The major clues as to its purpose have derived from biochemical work, a small amount of NMR structural data, and some genomic and proteomic work.

The NMR crystal structure of the B. subtilis δ N-terminal domain (NTD) has been reported and consists of four alpha helices and an antiparallel beta sheet [2]. The C-terminal domain (NTD) could not be solved along with this structure because it is highly acidic and flexible, resulting in disorder that prevents structure solving. Based upon this structural information, it is thought that the well-ordered NTD binds RNA polymerase, while the CTD could potentially serve as a nucleic acid mimic.

Figure 1. NMR Structure of B. subtilis δ NTD. This NMR structure shows the general makeup of the proposed RNAP-binding domain of δ, the NTD. It is composed of four alpha helices and one small antiparallel beta sheet. The CTD structure could not be solved due to its flexibility caused by a high percentage of acidic amino acids. [2].

Despite its small size and not being an essential protein, it is involved in virulence in pathogenic Gram positive species. It has been reported that a certain expression level of δ is required for maintaining virulence in the species Streptococcus agalactiae [3]. S. agalactiae is a human pathogen responsible for aggressive infections against which surgery in combination with antibiotics is often required [4]. It is thought then, that δ is
somehow involved in regulating the expression of virulence-related genes. Another clue as to its function resulted from a study also involving *S. agalactiae*. In this case, the authors showed that δ is required for survival of the species in that it enables resistance against phagocytic killing [1]. In this way, δ is not an essential subunit for the survival of the species under adequate environmental conditions, but in cases where survival is being threatened, δ is essential.

δ has also been studied in the Gram-positive pathogen *Streptococcus mutans*, which is responsible for human dental disease. This presence of δ in this particular species is an indication of its function in itself; this species, since it lives in the human mouth, must be capable of surviving a rapidly changing environment. It can survive rapid changes in factors such as pH, oxidation level, as well as presence or absence of sugar sources [5]. The authors here created an *S. mutans rpoE* knockout mutant, which is the gene that encodes δ. They then investigated the phenotype under normal conditions as well as stressed conditions, and found that δ is required for stress tolerance. Knocking out *rpoE* resulted in extended lag phases during growth, reduced resistance to antibiotics that inhibit RNAP, altered biofilm formation, and general reduction in tolerance levels for stresses such as oxidation level changes and drops in pH [5]. Transcriptome profiling using this mutant demonstrated that 550 genes were differentially expressed between the wild-type and *rpoE* knockout strains, with more than 50% of upregulated transcripts being in noncoding regions [5]. This melds well with the hypothesis that δ is involved in transcript specificity; without δ, RNAP may be more likely to transcribe from weak promoters. In general, the genes affected by deletion of *rpoE* were involved in the stress response in some way, providing critical insight into its function.

*In vitro* transcription assays have shown that the presence of δ subunit on RNAP either increases or decreases the rates of RNA synthesis depending upon the promoter in use and reaction conditions. In this way, it is thought to increase transcriptional specificity. It may also potentially influence the isomerization between the closed complex and transcription-competent open complex [6]. Despite the inhibitory effect δ can have on promoter binding and open complex formation, it can increase the yield of RNA product formation, presumably due to an increased rate of RNAP recycling [7].
that the δ subunit is capable of disassembling RNAP-nucleic acid complexes by displacing the nucleic acid, particularly if it is only nonspecifically bound. This would therefore free more RNAP to transcribe active promoters, thereby increasing transcription [6]. It also supports the idea that the acidic and flexible δ CTD may function as a nucleic acid mimic; it may use this structural feature to dissociate the RNAP-nucleic acid complexes.

Due to its involvement in virulence, investigation of δ subunit is important in terms of its potential as a novel antibiotic target. As a drug target, δ would enable great specificity in targeting bacteria, as there exists no δ homolog in humans. It will also enable a more comprehensive understanding of Gram-positive bacterial transcription. Therefore, it is important to study this RNAP subunit, and one aspect that has not been defined yet is its binding location on RNAP. This information should be determined in order to better learn the function and mechanism of action of δ, as well as to determine ways to disrupt this function in an effort to decrease the virulence of a species.

Materials and Methods

B1. Expression and Purification of B. subtilis δ Subunit

For ease of expression and purification, I PCR amplified the B. subtilis rpoE gene from the pFL31 plasmid obtained from the Helmann Lab at Cornell University. I then ligated this insert into the pSUMO vector, which enables efficient and rapid purification of a protein that will lack any affinity tags by the final step. I had originally attempted to express and purify δ directly from this pFL31 plasmid, but I was never quite able to obtain completely pure protein. Using the pSUMO system, however, I obtain large quantities of highly purified δ. To express δ, I transformed C41(DE3) cells with the pSUMO- δ vector I created. I then inoculated 3 mL of LB + kanamycin with an isolated transformed colony. After 4 hours of growth at 37°C, when the culture became cloudy, I then transferred this into 250 mL of LB containing kanamycin. I allowed this culture to grow at 37°C until the OD$_{600}$ was 0.4, then induced expression with IPTG to 0.25 mM. Growth continued for 3 hrs at 37°C before cells were harvested.

The first purification step was to resuspend the cells in 50 mL of Lysis Buffer (50 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.3 M NaCl) with the
addition of 5 mM 2-mercaptoethanol (BME) and 2 mM phenylmethanesulfonylfluoride (PMSF). Cells can be lysed by either sonication or by 4 passes through an Emulsiflex-C3 high pressure homogenizer (Avestin, Inc, Ottawa, Ontario, Canada). Next, the lysate was clarified by centrifugation for 20 min at 17,000 rpm and 4°C. The supernatant was applied using a peristaltic pump to a pre-equilibrated 5 mL Qiagen Ni-NTA Superflow cartridge column. The column was then rinsed with an additional 15 mL of lysis buffer followed by 4 CV wash with high salt buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 1 M NaCl, 5 mM BME and 2 mM PMSF) and 2 CV low salt buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 20 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF). Finally, the protein was eluted with elution buffer (10 mM Tris-HCl, pH 8 at 4°C, 10% glycerol, 250 mM imidazole, 0.1 M NaCl, 5 mM BME and 2 mM PMSF), collecting three 5 mL fractions. A total of 25 mg of protein is often eluted from the nickel column after these cell growth conditions. Next, the eluate is digested using the enzyme Ulp1; 1 µg of Ulp1 is added for every 500 mg of protein, and the sample is incubated for 1 hr on ice. If the Ulp1 stock is old or is not working well, the reaction can alternatively be incubated for 2 or more hours at 30°C, as necessary.

After Ulp1 digestion, the sample can be diluted or dialyzed as necessary to reduce the concentration of imidazole to 20 mM. In this way, the sample can then be applied to a second nickel column, which has been equilibrated in the low salt buffer, described above. In this case, the flow through is collected, since the 6x-His affinity tag has been cleaved by the Ulp1 enzyme, leaving un-tagged δ which will not bind the nickel column. Typically, 10 mg of purified δ protein can be obtained from this procedure. The procedure summary and sample gels are provided in Figure 2.
Figure 2. Purification of \textit{B. subtilis} \(\delta\) Using the pSUMO System. These two 4-15\% SDS-PAGE gels provide the summary of \(\delta\) purification. The first nickel column product is given on the left, with lanes 1 and 2 being the nickel column high and low salt washes (empty), and lanes 3 and 4 are the elutions. These elution fractions contain the SUMO- \(\delta\) fusion, which runs at approximately 35 kDa. This elution was digested with Ulp1 and then applied to a second nickel column, the result of which is shown on the right. Lane 1 contains purified core RNAP to ensure that \(\delta\) was not contaminating my core RNAP samples. Lane 2 contains the second nickel column flow through, which corresponds to the 20 kDa \(\delta\) protein. Lane 3 contains the elution, which is SUMO/Ulp1 protein. Finally, lane 4 gives the digestion before application to the second nickel column, to ensure that the 32 kDa SUMO- \(\delta\) fusion was completely digested, and it was.

\textbf{B2. Mutation of \(\delta\) for FRET Experiments}

To prepare \(\delta\) for fluorescence resonance energy transfer (FRET) experiments, cysteine residues must be engineered into the sequence of \(\delta\). The fluorescent labels used for labeling \(\delta\) can then be attached via these cysteine residues. The two mutations I made were S42C and D119C, so as to have one engineered cysteine in the NTD and one in the CTD. To accomplish this, I used the protocol from the QuikChange Site Directed Mutagenesis kit as well as Pfu Turbo DNA polymerase (Agilent Technologies). The S42C mutation was a point mutation, and therefore, the only adjustment I made to the kit protocol was cycling the PCR program 12 times, in accordance with kit recommendations. I also used fresh MachT1 cells for transformation. I confirmed presence of the correct\textit{rpoE} by sequencing completed by the Penn State Nucleic Acid Facility. The D119C mutant was slightly different as this was a complete amino acid replacement and not just a point mutation, and so the only different step taken was to cycle the PCR program 16 times, as per the kit manual. This mutant was also confirmed to have the correct sequence by the Penn State Nucleic Acid Facility.

Both of these mutants were expressed and purified in the same manner as the wild-type, as per section B1. The rest of this work was then completed by the laboratory of Professor Tae-Hee Lee at Penn State University. They completed fluorescent labeling of the \(\delta\) protein as well as subsequent FRET experiments. In short, \textit{B. subtilis} RNAP holoenzyme in complex with a fluorescent labeled fork-junction DNA scaffold was affixed to a siliconized glass surface. Labeled \(\delta\) was then washed over the RNAP, and FRET data was collected in order to determine the proximity of the two fluorescent dyes. In this manner, the binding location of \(\delta\) RNAP on could be determined. Using this experimental setup, preliminary data could be obtained which indicated \(\delta\) bound near or
on the β flap-tip helix region of RNAP. This region is located near the RNA exit channel of RNAP, which potentially supports the hypothesis that the δ CTD functions as a nucleic acid mimic, displacing RNA in the RNA exit channel to dissociate the transcription elongation complex. To test this hypothesis, experiments were undertaken to determine if the δ CTD is the main region by which δ binds RNAP. I therefore again used SDM to introduce stop codons after amino acids 92 and 109, creating two different δ mutants lacking varying lengths of the CTD (δS42C92 and δS42C109). I used the same QuikChange kit, and the δ S42C mutant as a template so that a cysteine was already engineered into the NTD. The sequences of these two truncation mutants were also confirmed by Penn State Nucleic Acid Facility Sequencing. The Lee laboratory then repeated their FRET experiments using this truncation mutant, to determine if δ would still bind while lacking the CTD.

B3. Preliminary FRET Results

As mentioned in section B2, the FRET results suggest that δ is binding B. subtilis RNAP near the β flap-tip helix region. This region is highlighted in Figure 3, which gives a summary of two FRET experiments completed; binding of full-length δ S42C, and binding of δ S42C ΔCTD. The region is located near the RNA exit channel, providing access for δ to disrupt RNA-RNAP binding. According to these binding locations, it is possible that δ could be influencing transcriptional specificity, particularly under conditions of stress, by disrupting the RNA transcript-RNAP binding. This would dissociate the complex, releasing RNAP to transcribe from other promoters. Further work must be completed however to investigate the specifics of this binding as well as how δ is capable of specifically inhibiting progression of transcription at certain promoters.
Figure 3. FRET Result for δ Binding Site. This figure, prepared by Jae-Hyoun Lee in the Lee lab at Penn State University, demonstrates the potential binding sites of δ, on RNAP. Though *B. subtilis* RNAP was used in the experiments, the structure used for preparation of this figure was the *T. aquaticus* holoenzyme-fork junction DNA complex (Murakami 2002). Two populations of binding sites were determined using the FRET experiments; a major site, and a minor site which contained a smaller population of δ. The major binding sites are the larger circles and the minor population binding sites are the smaller circle. The red circles correspond to the binding site of the full-length δ, while the light purple circles correspond to the CTD-deletion δ mutant binding site. All 4 of these potential sites are located on the β flap region of RNAP. The β subunit is shown in cyan, with β’ in pink, α and ω in dark gray, and σ in orange. The template DNA strand is in yellow and the nontemplate is in green. This β flap region is located very near to the RNA exit channel, providing a potential mechanism of action of δ.
References


