

The Pennsylvania State University

The Graduate School

Eberly College of Science

**MECHANISTIC INVESTIGATIONS OF NUCLEOTIDYL-TRANSFER
REACTION AND *DE NOVO* INITIATION IN RNA POLYMERASES BY X-RAY
CRYSTALLOGRAPHY**

A Dissertation in

Biochemistry, Molecular Biology, and Microbiology

by

Ritwika Basu

© 2014 Ritwika Basu

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2014

The dissertation of Ritwika Basu was reviewed and approved* by the following:

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

Craig E. Cameron
Eberly Chair in Biochemistry and Molecular Biology

Paul Babitzke
Professor of Biochemistry and Molecular Biology

Andrey S. Krasilnikov
Associate Professor of Biochemistry and Molecular Biology

Scott Showalter
Assistant Professor of Chemistry

Richard Frisque
Professor of Molecular Virology
Interim Head of the Department of Biochemistry and Molecular Biology

*Signatures are on file in the Graduate School

ABSTRACT

Nucleotidyl-transfer reaction is the fundamental reaction catalyzed by nucleic acid polymerases to synthesize RNA or DNA. Nucleotidyl transferases, including DNA and RNA polymerases (DNAP and RNAP) utilize the two-metal catalysis mechanism. However the temporal order of events in the reaction mechanism is not fully understood. Studies so far are based on predictions from biochemical studies and X-ray structural snapshots obtained with the use of non-reactive substrate analogues. This thesis presents a novel time-resolved X-ray crystallography approach to monitor the enzyme motions through high-resolution crystal structures. A soak-trigger freeze X-ray crystallography technique provides an atomic resolution movie of the active site events during the natural course of nucleotidyl-transfer reaction catalyzed by bacteriophage N4 mini-vRNAP. The structures are the first direct evidence of the temporal order of substrate assembly and metal co-factor binding at the active site of single-unit polymerases. The work reveals that the nucleotide substrates bind the active site along with Me^{B} (nucleotide binding metal) and the catalytic metal (Me^{A}) coordination is only temporary, established just before and dissociated immediately after phosphodiester bond formation.

To expand the mechanistic studies to cellular multisubunit RNAPs, X-ray crystal structures of transcription initiation complexes from bacteria *Thermus thermophilus*, have been presented. Bacterial RNAP binds to initiation factor σ , forming the holoenzyme, which is capable of recognizing and binding to promoter DNA, leading to a transcription-ready open complex formation. The holoenzyme initiates RNA synthesis *de novo*, the only step during transcription where RNAP accepts two nucleotide substrates and performs primer-independent phosphodiester bond formation. The molecular mechanism of preorganization of the RNAP holoenzyme-promoter DNA complex in bacteria has been well studied, but there is no direct evidence for the mechanism of *de novo* initiation. I solved a structure of a *de novo* initiation complex that provides the first structural basis for the stable binding of the initiation nucleotide. The structure shows base stacking

interactions between the DNA template at -1 position and the incoming nucleotide at +1, which explains the sequence preference of purine-pyrimidine for the -1 and +1 template positions of many RNAP promoters. Further, I solved the structure of a holoenzyme transcription complex containing a 6-mer transcript that was synthesized *in crystallo*. The structure demonstrates an early step of σ ejection from the RNAP core enzyme when transitioning to a stable transcription elongation complex. The structure shows that the RNA 5'-triphosphate end reaches the σ region of 3.2 causing it to be disordered, suggesting that 6-mer RNA transcription is a first step in this transition.

TABLE OF CONTENTS

LIST OF FIGURES.....	viii
LIST OF TABLES.....	x
ABBREVIATIONS.....	xi
ACKNOWLEDGEMENTS.....	xii
Chapter 1. Introduction.....	1
1.1. Overview.....	2
1.2. Single-unit RNA polymerases.....	4
1.3. Structure of prototype phage RNAP: T7 RNAP.....	6
1.4. Transcription Initiation.....	8
1.4.1. Promoter Binding.....	8
1.4.2. Transcript Initiation.....	9
1.5. Transcription Elongation.....	10
1.5.1. Promoter release and processivity.....	10
1.5.2. Transition to Elongation complex.....	12
1.5.3. Nucleotide Addition Cycle.....	13
1.5.3.1. Substrate Selection at Pre-insertion site.....	13
1.5.3.2. Substrate Binding at Catalytic site.....	15
1.5.3.3. Translocation.....	15
1.6. N4 virion encapsidated RNAP: factor dependent single-unit RNAP.....	16
1.6.1. Structure of N4 vRNAP.....	19
1.6.2. Unique Promoter requirements of N4 mini vRNAP.....	19
1.6.3. Nucleotidyl transfer reaction.....	21
1.7. Multi-subunit RNAP.....	24
1.7.1. Transcription Elongation.....	27
1.7.2. Transcription Initiation.....	30
1.7.3. Nucleotide Addition Cycle.....	33
1.7.3.1. Substrate binding sites.....	34
1.7.3.2. Role of trigger loop and bridge helix in NAC.....	35
1.7.3.3. Substrate Entry.....	35
1.8. Thesis Format.....	36
1.9. References.....	38
 Chapter 2. Watching the nucleotidyl-transfer reaction mechanism in N4 RNAP by time- resolved trigger-freeze crystallography.....	 47
2.1 Abstract.....	48
2.2 Introduction.....	49
2.2.1 Nucleotidyl-transfer reaction.....	49
2.2.2 Time-resolved Raman crystallography.....	52

2.2.3	Time-resolved soak-trigger-freeze X-ray crystallography.....	54
2.3	Experimental Summary.....	55
2.4	Materials and Methods.....	56
2.4.1	Purification of N4 mini-vRNAP.....	56
2.4.2	Purification and preparation of hairpin DNA.....	57
2.4.3	Crystallization of binary complex.....	57
2.4.4	Soak-trigger-freeze X-ray crystallography.....	59
2.5	Results.....	60
2.5.1	Time trace of in crystallo transcription by Raman crystallography.....	60
2.5.2	Structures of transcription intermediates using soak-trigger-freeze X-ray crystallography.....	63
2.5.3	Attempts of <i>in crystallo</i> reaction with Mg.....	69
2.6	Discussion.....	70
2.6.1	Temporal order of nucleotidyl-transfer reaction.....	70
2.6.2	Dual role of catalytic metal in nucleotidyl transfer reaction of A-family polymerases.....	71
2.6.3	Soak-trigger-freeze approach is suitable for tracing enzymatic reactions in four dimensions.....	73
2.6.4	Acid-base mechanism of reaction catalysis.....	74
2.7	Acknowledgements.....	77
2.8	References.....	78

Chapter 3. Structural basis of de novo transcription initiation in bacterial RNAP..84

3.1	Abstract.....	85
3.2	Introduction.....	86
3.3	Material and Methods.....	89
3.3.1	Purification of endogeneous <i>T. thermophilus</i> core RNAP and reconstitution of holoenzyme.....	89
3.3.2	Crystallization of RNAP holoenzyme – promoter DNA complex.....	91
3.3.3	X-ray data collection and structure solving.....	92
3.4	Results.....	93
3.4.1	Design of the X-ray crystallographic experiment to determine the de novo transcription complex.....	94
3.4.2	The +1 NTP binding site in the de novo transcription complex.....	95
3.4.3	+2 NTP binding site in the de novo transcription complex.....	98
3.4.4	Structure of transcription initiation complex containing 6-mer RNA... ..	100
3.5	Discussion.....	104
3.6	Acknowledgements.....	106
3.7	References.....	106

Chapter 4. Summary and Future Directions.....110

4.1	Studies of N4 mini-vRNAP transcription.....	110
4.1.1	N4 mini-vRNAP activation by hairpin-form promoter.....	112
4.1.2	Toward crystal structure determination of N4 mini-vRNAP transcription elongation complex.....	116

4.2	Studies on <i>T. thermophilus</i> RNAP initiation complex.....	120
4.3	Acknowledgements.....	121
4.4	References.....	121

LIST OF FIGURES

Figure 1.1: Organization of T7-like RNAPs.....	5
Figure 1.2: Phylogenetic Relationship of T7 RNAP like enzymes.....	5
Figure 1.3: Structures of single-unit DNA dependent polymerases.....	7
Figure 1.4: Structural motifs for promoter recognition.....	9
Figure 1.5: Comparison of the structures of initiation complex, intermediate complex and elongation complex.....	11
Figure 1.6: Active site views during nucleotide addition cycle.....	14
Figure 1.7 N4 bacteriophage infection and transcription program.....	17
Figure 1.8. Model for in vivo N4 promoter extrusion and recognition.....	18
Figure 1.9: Activation of N4 mini vRNAP by the specific hairpin promoter.....	20
Figure 1.10. Two Metal catalysis mechanism for phosphoryl-transfer reactions.....	21
Figure 1.11: Suite of structural snapshots in nucleotidyl-transfer reaction.....	23
Figure 1.12: Evolutionarily conserved regions of the large subunits form the core of multi-subunit RNAPs.....	25
Figure 1.13: Schematic representation of the RNAP structure in transcription elongation.....	26
Figure 1.14. Structures of <i>T. thermophilus</i> elongation complex complex.....	29
Figure 1.15 Structures of bacterial RNAP holoenzyme.....	31
Figure 1.16 Structure of transcription initiation complexes.....	32
Figure 2.1 Chemical structures of a nucleotide and its non-reactive analogues used for structural studies of polymerases.....	50
Figure 2.2: Basic setup of the Raman microscope.....	53
Figure 2.3 Strategy for time-dependent soak-trigger-freeze experiment.....	55
Figure 2.4 Purification of N4 mini-vRNAP and crystallization of Binary complex.....	58
Figure 2.5 The platform for tracing <i>in crystallo</i> reaction.....	60
Figure 2.6. Representative Raman spectra of PC formation.....	61
Figure 2.7. Summary of Raman Intensity changes vs time upon PC formation.....	63
Figure 2.8 Structural analysis of the TIC-1' active site and comparison to other intermediates.....	65
Figure 2.9 Structural analysis of TIC-2' active site.....	67
Figure 2.10 Structural analysis of TIC-3' active site.....	68
Figure 2.11 Structural analysis of TIC-4' active site.....	69
Figure 2.12 Active site views of the TIC _{Mg} -6' and TIC _{Mg} -7'.....	70
Figure 2.13 A schematic of the transcript initiation.....	71
Figure 2.14 Electron density of Y612 proposing proton transfer.....	75
Figure 2.15 Effect of mutations of Asp 951 on transcription activity.....	76
Figure 2.16 Electron density at the active site of the complexes with D951N mutant at 2 mins and 3 mins.....	77

Figure 3.1 Overall structure of the <i>de novo</i> transcription complex. <i>T. thermophilus</i> RNAP is depicted as a molecular surface model.....	88
Figure 3.2 Purification of <i>T. thermophilus</i> RNAP core enzyme and preparation of holoenzyme by <i>in vitro</i> reconstitution.....	91
Figure 3.3 Crystals of RNAP-promoter DNA complex.....	92
Figure 3.4 Schematic of the <i>de novo</i> transcription and the holoenzyme transcription complexes preparation.....	95
Figure 3.5 Nucleotide binding sites of the <i>de novo</i> transcription complex.....	97
Figure 3.6 Comparison of the <i>de novo</i> transcription complex to holoenzyme – promoter DNA and transcription elongation complexes.....	99
Figure 3.7 The holoenzyme transcription complex containing 6-mer RNA.....	101
Figure 3.8 Comparison of the <i>de novo</i> transcription and holoenzyme 6-mer transcription complexes and active structure.....	102
Figure 3.9 2Fo – Fc electron density map of the holoenzyme transcription complex....	103
Figure 3.10. Model of longer RNA transcript into the RNA exit tunnel explains σ release.....	104
Figure 4.1: Crystals of mini vRNAP:HP3 binary complex.....	112
Figure 4.2: RNAP activation by short hairpin HP3.....	116
Figure 4.3 Assembly of N4 mini-vRNAP elongation complex.....	117
Figure 4.4. Crystals of N4 mini-vRNAP elongation complex.....	119

LIST OF TABLES

Table 2.1. Data collection and refinement statistics of the N4 mini-vRNAP transcript initiation complexes.....	59
Table 3.1: Data collection and refinement statistics of T. thermophilus initiation complexes.....	93
Table 4.1 Crystallization condition for N4 mini-vRNAP short hairpin complex.....	113
Table 4.2 Data collection and Refinement Statistics of N4 mini-vRNAP - short hairpin (HP3) binary complex.....	113
Table 4.3 Crystallization of N4 mini-vRNAP elongation complex.....	118

ABBREVIATIONS

BC- binary complex
BH- bridge helix
CTD- C-terminal domain
ds- double stranded
EcoSSB- E. coli single stranded DNA binding protein
F_c- calculated structure factor
F_o- observed structure factor
Me^A- Catalytic metal or Metal A
Me^B- Nucleotide binding metal or Metal B
N-site- nucleotide binding site
N4 vRNAP- N4 bacteriophage encoded virion encapsidated RNAP
N4SSB- N4 encoded single stranded DNA binding protein
NAC- Nucleotide addition cycle
NMPCPP- nucleotide 5'-[(α , β)-methyleno] triphosphate
NTD- N-terminal domain
dNTP- deoxyribonucleotide triphosphate
rNTP- ribonucleotide triphosphate
NTR- nucleotidyl transfer reaction
P-site- primer site
PC- product complex
PMSF- Phenyl methyl sulfonyl fluoride
Pol I/II/III- eukaryotic RNA polymerase I/II/III
PPi- inorganic pyrophosphate
RNAP- RNA polymerase
RPc- bacterial RNA polymerase closed complex
RPO- bacterial RNA polymerase open complex
Taq- *Thermus aquaticus*
SCI/II- substrate complex I/II
TBE- Tris borate EDTA
TEC- Transcription elongation complex
TGED- Tris glycerol EDTA DTT
TIC- Transcription initiation complex
TL- trigger loop
TLS- translation libration screw refinement
Tth- *Thermus thermophilus*

ACKNOWLEDGEMENTS

I am grateful to numerous people for contributing and supporting me in completing this dissertation. I cannot thank Dr. Katsuhiko Murakami enough for accepting me as his student in the first place and guiding me throughout, despite my lack of any formal training in biochemistry benchwork or structural biology. He patiently taught me most of the lab techniques and structure determination himself. I owe him a lot for whatever scientific research skills that I have acquired in the five years. His tendency for scientific inquiries and aiming at challenging problems went a long way to teach me the skills of persistence and micro-observation that is so essential in a structural biologist.

I owe many thanks to both past and present members of the Murakami Lab. Although I have not met Mike Gleghorn, his thesis work served as a solid foundation for my studies and gave me initial guidelines to approach my research topic. Brianna Klein, Raka Ghosh and Kevin Baker have all been my patient and friendly teachers of lab techniques from the early days of my Ph.D. I am very grateful to the present post-docs Drs. Sung-Hoon Jun and Vadim Molodtsov for being ready to address my smallest and most naive questions and problems that I faced in my research. Lastly, my utmost thanks to dear friend Brittany Warner, with whom I enjoyed a parallel graduate life, together making all the mistakes and also together learning from them. She was a comforting lab-mate during grim days with whom I shared many fun times both in and out of lab.

Next, I have to thank my collaborators in the laboratories of both Prof. Paul Carey at Case Western Reserve University and Prof. Lucia Rothman-Denes at the University of Chicago for being such wonderful people to work with. Yuanyuan Chen in the Carey Lab provided the Raman Crystallography results and helped us with many discussions for our manuscript. Laura Satkamp and Eda Koculi have been very prompt collaborators in the Rothman-Denes lab for supporting us with studies and materials for our projects.

I am extremely grateful to Profs. Craig Cameron and Paul Babitzke for their helpful advice in the progress of my dissertation work. I acknowledge my other committee members Dr. Andrey Krasilnikov and Dr. Scott Showalter for all the fruitful discussions.

I thank the X-ray core facility at Penn State, especially Hemant and Neela Yennawar who have been my X-ray crystallography teachers both in class and hands-on. They have always been available for answering my question and I acknowledge their generous discussions. I thank Mark Signs and Kim at the Penn State Shared Fermentation facility for always putting on a happy face and helping me out with my cell cultures.

Finally, I would like to express my gratitude to my family to have supported me to come so far from home just to achieve my ambitions and keeping me boosted up and sending in blessings to accomplish whatever I had aimed for. My parents have been my primary source of encouragement, teaching me to put a dedicated effort in any undertaking and my elder brother, Suddha, has been my earliest inspiration for developing an interest in science. And, all of this would be impossible without the love and support of my husband, Sunny. He encouraged me to come over to this country and pursue my Ph.D, despite him having to shift his base for me. I thank him for understanding the needs of graduate life and constantly giving me reasons to smile. I dedicate this thesis to my loving family for all the sacrifices they have made to see me complete my Ph.D.

Chapter 1

Introduction: Transcription in single-subunit and multi-subunit RNAPs

Part of this chapter is adapted from

Basu, R. S. and Murakami, K. S. (2013) Bacteriophage RNA polymerases (Chapter 10; pp 237 – 250) in **Nucleic Acid Polymerases** (eds. Murakami, K. S. and Trakselis, M. A.)

Copyright © Springer-Verlag Berlin Heidelberg 2014

1.1 Overview

This thesis compiles structural work that I have done to explore the mechanism of RNA polymerase (RNAP) transcription, a reaction fundamental to life processes. In this pursuit I worked on enzymes from two highly diverse organisms that belong to the two basic types of RNAPs: single-subunit and multi-subunit RNAPs. This chapter provides a detailed background on the two types of RNAPs, its two types and the basic RNA synthesis mechanism. This mechanism defines several aspects of transcription for both RNAP families and thereby puts my work into an appropriate context.

Transcription is not only the first step of gene expression but also a target for gene regulation, both of which are highly important life processes. Being central to life, RNAP is as ancient as the origin of life, encoded even in the genomes of biological macromolecular assemblies such as viruses; and with the evolution of diverse cellular organisms, RNAPs from these different kingdoms maintain equal diversity. RNAPs encoded by viruses and the endosymbiotic products of evolution, such as mitochondria and some chloroplasts, are composed of a single polypeptide chain that largely functions independently. On the other hand, cellular RNAP from the three domains of life are composed of multiple subunits that additionally require transcription factors to form the transcription machinery. The bacterial RNAP core enzyme is composed of 5 subunits, the archaeal enzyme contains 11~13 subunits and eukaryotes have three distinct RNAPs, each of which contain at least 12 subunits. The mechanism of transcription involves a complex coordination of sequential events, including promoter binding, DNA melting, substrate selection, nucleotidyl transfer reaction, RNA translocation, RNA separation from DNA template, *etc.* Due to the large size and complexity of preparing multi-subunit RNAPs, single-subunit bacteriophage RNAPs became a model for structural, biophysical and biochemical investigations of the transcription mechanism. Indeed, the core catalysis mechanism was identified as a universal feature of all phosphoryl-transferase enzymes; i.e. *two metal catalysis* as proposed by Steitz et al. in 1993 (Steitz et al., 1994; Steitz and Steitz, 1993a). Thorough structural and biochemical dissection of bacteriophage T7 RNAP led to a high-resolution structure, providing the first view of the active site

architecture of RNAPs. Many T7 RNAP structures thereafter provided a detailed structural model of the molecular mechanism of the entire nucleotide addition cycle, as well as the key steps of transcription initiation and elongation. These studies will be reviewed in the introduction of this chapter. Next, I will introduce our model system, bacteriophage N4 RNAP. This enzyme has the ability to produce atomic-resolution crystal structures that serve as a platform for reaction mechanism studies, which I exploited in my dissertation work. However this RNAP is biologically unique, not only as an enzyme but also with respect to the transcription program of this bacteriophage. This RNAP is an example of a pre-encapsidated enzyme within the virion particle (N4 vRNAP) and therefore exists before infection of host bacteria. Considering the extensive characterization of the bacteriophage transcription program, N4 vRNAP and its unique mechanism of action merits introduction. Finally, as a foundation for my results I will describe the past knowledge on NTP binding and the nucleotidyl transfer reaction, which were obtained from a series of structural snapshots of N4 vRNAP, leading us to the missing piece that I address in the first part of this dissertation: a real-time structural basis of the nucleotidyl-transfer reaction in single-subunit polymerases.

In the second half of my dissertation studies, I worked on multi-subunit RNAP from bacteria. Although RNAPs catalyze the same chemical reaction for RNA synthesis, some processes of the transcription cycle have differences between single and multi-subunit enzymes. The molecular mechanism of multi-subunit RNAP is more complex due to the participation of distinct motifs from several subunits and external transcription factors. These differences extend to the architecture of the active site in which the participation of essential metals and surrounding motifs in the nucleotide addition cycle (NAC) cannot be explained by analogy with single-subunit RNAP studies. Thus, studies of multi-subunit RNAP have diverged into a separate field swarming with structural and biochemical findings. Accordingly, I will review the current understanding of multi-subunit RNAP and the NAC mechanism in context to bacterial systems, but also using the similarities with eukaryotic RNAP whenever applicable. This background will help pose the question about bacterial *de novo* transcription initiation, that I have addressed in my dissertation.

1.2 Single unit RNA polymerases

Single unit RNAPs were first discovered in T7 phage-infected *Escherichia coli* cells (Chamberlin et al., 1970). It was known that phage infection of host bacterial cells led to redirection of host gene expression towards generation of progeny phage particles, but it was through this study that a previously uncharacterized ‘switching event’ leading to the expression of late bacteriophage genes was attributed to a phage-encoded RNAP. This phage RNAP could recognize promoters on the phage genome and express phage genes using a single polypeptide polymerase of ~100 kDa molecular weight, which is ~4 times smaller than bacterial RNAPs. Phage RNAPs are much simpler than those from bacteria and eukaryotes, nonetheless, the single-subunit T7 RNAP is able to recognize promoter DNA, unwind double stranded (ds) DNA to form an open complex and after abortive initiation, can proceed to processive RNA elongation (Cheetham and Steitz, 2000). The simplicity of T7 phage RNAP made it an ideal model system to study the transcription mechanism and an ideal tool for protein expression system in bacterial cells. Ever since, a multitude of structures were solved that showed the structure of T7 RNAP, the simplest and the most widely used phage RNAP, in different stages of transcription. This enzyme became the reference model for explaining key features of transcription including initiation, translocation, transition from initiation to elongation and processive RNA synthesis.

Single-subunit RNAPs were also discovered in mitochondria (Masters et al., 1987) and chloroplasts (Hedtke et al., 1997), two vital eukaryotic organelles. Sequence alignment of these RNAPs identified 10-12 homology blocks concentrated in the carboxy terminal domains (CTD) while the amino terminal domains (NTD) were variable in sequence and length throughout the family. The sequence bore no similarity to any of the subunits of cellular RNAPs. The homology blocks include four conserved motifs A, B, C, and D/TxxGR (**Figure 1.1**) that encompass the polymerase active center (Cermakian et al., 1997). Moreover, these motifs are also conserved in pol I type DNAPs (Delarue et al., 1990), which suggested a common active site architecture and conserved mechanism for the essential polymerase reaction. Motifs A and C house the universally conserved

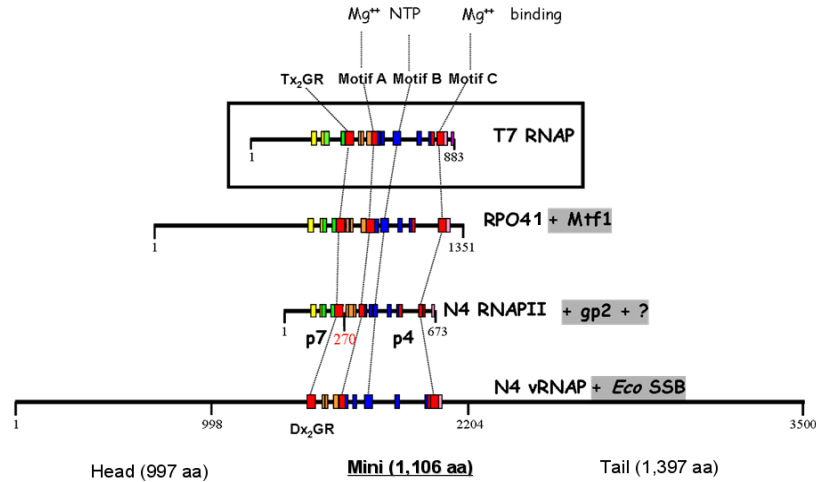


Figure 1.1: Organization of T7-like RNAPs. Four conserved motifs of the T7-like single-subunit RNAP family are depicted by similarly colored rectangles on lines representing the primary sequence of the following related RNAPs: T7 phage RNAP, yeast mitochondrial RNAP (RPO41) and N4 phage RNAPII and N4 virion-RNAP (vRNAP). Factors necessary for *in vivo* transcription are highlighted in grey. The three different regions of N4 vRNAP are labeled at the bottom. (adapted from Kazmierczak et al., 2002)

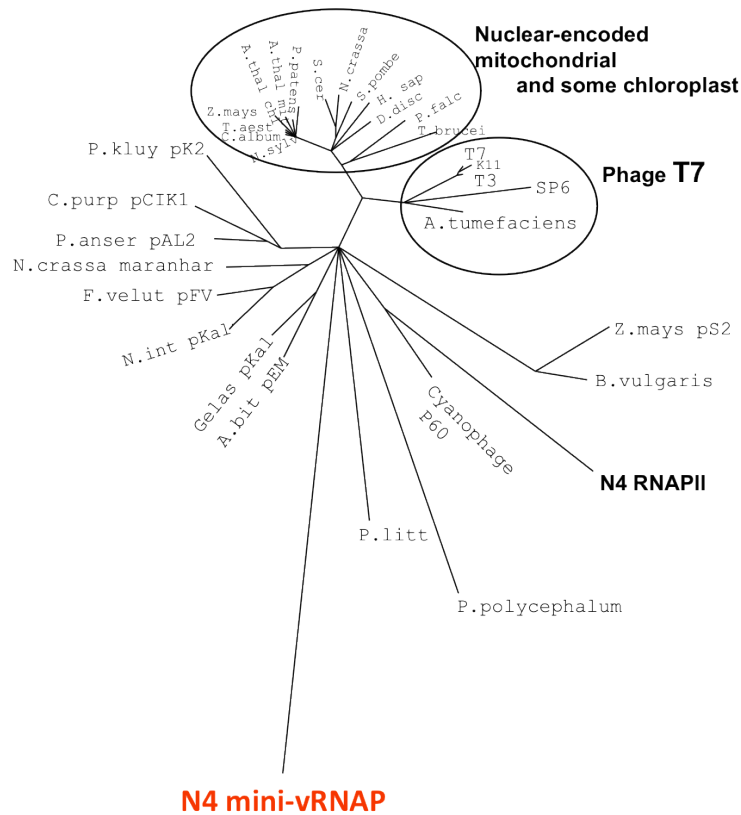


Figure 1.2: Phylogenetic Relationship of T7 RNAP like enzymes. The tree was generated using conserved sequence blocks common to all single-subunit RNAPs. (adapted from Kazmierczak et al., 2002)

aspartates for the two metal catalysis. Bacteriophage N4 vRNAP, our model system, does not have extensive sequence similarity to the phage-related RNAPs; however, it has substantial matches to the four conserved motifs, especially in the functionally important residues. Further, the sequence matches belong to the central polymerase domain of N4 vRNAP called the mini-vRNAP. The flanking additional regions are dispensable for polymerase function and probably account for the N4-specific functions. N4 mini-vRNAP was thus classified as a distant relative of the single-subunit family of polymerases (**Figure 1.2**).

1.3 Structure of the prototype phage RNAP: T7 RNAP

The first crystal structure of RNAP was determined in 1993 from the single-subunit T7 RNAP (Sousa et al., 1993). The domain organization of the T7 RNAP was found similar to the bacterial DNAP I (Arnold et al., 1995) e.g. Klenow Fragment (KF) of *E. coli* Pol I. (Ollis et al., 1985) (**Figure 1.3**). The structure resembles the anatomy of a right hand comprising of Palm, Fingers and Thumb subdomains that are analogously arranged around a DNA binding cleft. In addition, an N-terminal domain (NTD), not seen in DNAP I, constitutes the front wall of the DNA binding cleft, making the DNA binding cleft deeper and narrower in RNAP. The NTD also plays roles in promoter recognition and DNA unwinding for making a transcription-competent open complex.

The palm subdomain (**Figure 1.3**) forms the base of the DNA binding cleft with the fingers and the tall thumb subdomains forming either sidewalls of the channel. Invariant residues from motifs A, B and C cluster around this catalytic cleft. The aspartate residues, conserved in all polymerases, bear the most important catalytic function of chelating two divalent metals, Mg^{2+} in general, at the active site. The catalytic metal A (Me^A) generates the nucleophile at the 3'-end RNA for the S_N2 nucleotidyl-transfer reaction and a nucleotide binding metal (Me^B) stabilizes the charge distribution of the incoming nucleotide at the reaction transition state (Steitz et al., 1994; **Figure 1.10**).

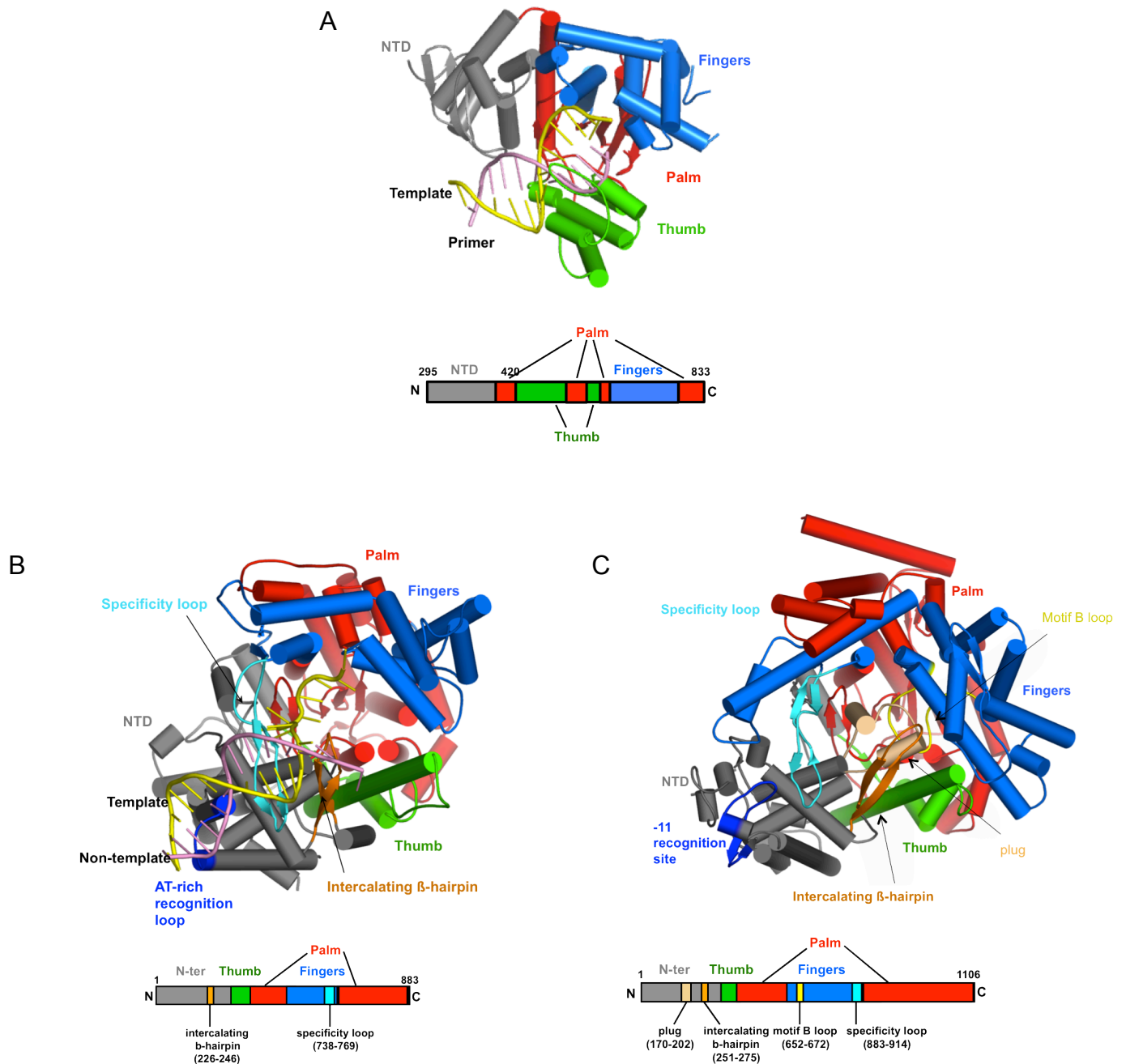


Figure 1.3: Structures of single-subunit DNA dependent polymerases. Right hand-like organization of domains in polymerase domain of KlenTaq DNAP I (**A**), T7 RNAP complexed with promoter DNA (**B**) and N4 mini-vRNAP (**C**) are shown. The same orientation of the structures were obtained by superposing the palm domains. The palm (red), fingers (blue), thumb (green) domains and NTD (gray) are shown as cylinders (α -helix) and arrows (β -strands). Double stranded promoter DNA containing template (yellow) and non-template (pink) in T7 RNAP and the primer (pink)-template (yellow) duplex in KlenTaq reach the active site cleft formed by three β -strands. The active site of N4 mini vRNAP (**C**) is blocked by the plug (wheat) and the motif B loop (yellow). The primary structures of the polymerases are shown below each 3-D structure with the same color code.

1.4 Transcription Initiation

T7 –like RNAPs are intrinsically equipped to recognize and bind to respective promoter sequences. Promoter-bound RNAPs are transcription-ready complexes that can begin transcribing *de novo*, without the aid from any primer, and eventually release the promoter to transform into a continuous elongation complex. Two structures of T7 RNAP promoter bound complex provided the detailed understanding of the initiation events in transcription, which include promoter recognition, template melting and positioning in the active site, and the start of RNA synthesis (Cheetham et al., 1999; Cheetham and Steitz, 1999).

1.4.1 Promoter binding

The T7 promoter sequence is conserved from the -17 to +6 position with a highly AT-rich region centered around -17. The upstream duplex DNA from -17 to -5 bind to the NTD and the DNA bases downstream are melted and a single-stranded template DNA is directed into the active site. T7 RNAP recognizes the promoter through three main interactions (**Figure 1.4**): (i) DNA bases from the major groove are recognized by an antiparallel β -hairpin of the fingers, the specificity loop,; (ii) an AT-rich recognition motif in the NTD recognizes the AT-rich (-17) region by inserting a flexible surface loop into the DNA minor groove; (iii) another motif in the NTD, the intercalating β -hairpin, melts the promoter DNA by inserting a hydrophobic residue to stack with the -5 base pair of the promoter, marking the upstream edge of the transcription bubble. This precise location of the transcription bubble ensures correct positioning of the transcription start site DNA base at the active site. Further precision is employed by an aromatic tryptophan within the specificity loop, which stacks with the -1 base of the template inducing a sharp bend to dip the +1 base into the active site. At this point, RNAP is ready to accept the two nucleotides base pairing with the +1 and +2 template bases to initiate RNA synthesis.

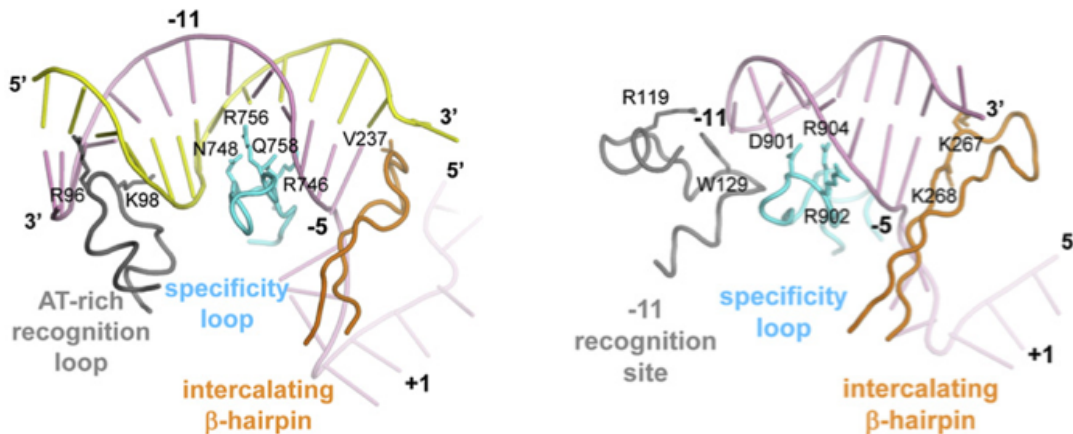


Figure 1.4: Structural motifs for promoter recognition. (A) AT-rich recognition loop (gray), specificity loop (cyan) and intercalating hairpin (orange) in the T7 RNAP binary complex. Template (yellow) and non-template DNA strands (pink) are shown. (B) -11 recognition loop (gray), specificity loop (cyan) and intercalating hairpin (orange) shown to recognize the hairpin promoter (pink) in N4 vRNAP. The figure has been adapted from (Gleghorn et al., 2008).

1.4.2 Transcript Initiation

The structure of T7 RNAP transcribing initiation complex (Cheetham and Steitz, 1999) showed how RNAP positions the template DNA bases at the active site at every subsequent step of NTP addition. In this structure, RNAP was bound to a 17 bp duplex promoter and a 3-mer RNA transcript base-paired with the single stranded tailed template. The presence of RNA in this polymerase structure offered an insight into the basis of rNTP specific RNA synthesis. In contrast to DNAP, a bulky glutamate ‘steric gate’ near the active site is replaced by a glycine in RNAP that makes space for the 2’OH of the incoming NTP. Secondly, a carbonyl group in the active site hydrogen bonds with the 2’OH of the 3’ primer end allowing only a 3’ *endo* ribose conformation of the base. Further, the DNA binding pocket is complementary to the A-form heteroduplex, thus favoring the formation of a DNA/RNA heteroduplex at the stage of RNA extension.

The promoter contacts are maintained while the DNA/RNA heteroduplex accumulates in the active site positioning the growing primer end at the catalytic site. Accordingly, the -1 template base that stacked with the +1 template for proper positioning during *de novo* initiation, assumes a flipped out conformation allowing

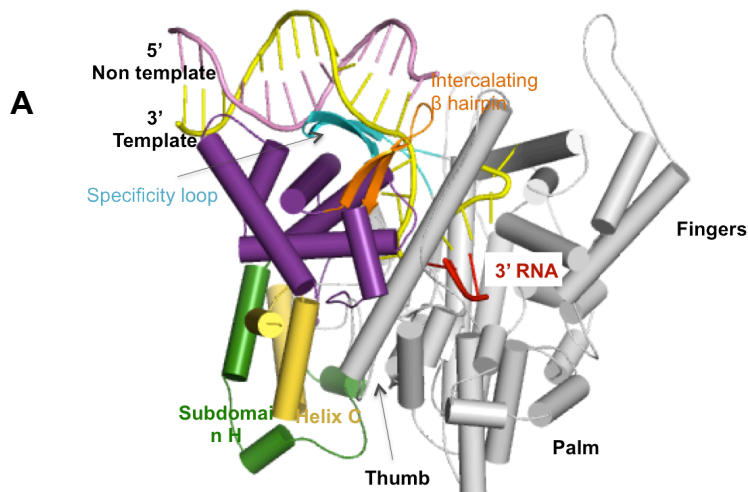
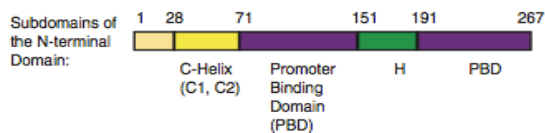
transcript extension to the 3-mer stage. This observation served the first structural evidence for the ‘DNA scrunching’ mechanism during initiation. The primer DNA scrunches into the active site through the initial synthesis phase until it chooses one of the two fates, abortive or productive transcription. During initiation, abortive transcripts are displaced from the template by collapsing the newly-formed bubble due to small, weak DNA/RNA hybrids (Briebe and Sousa, 2001). Alternatively, it could extend the DNA/RNA hybrid and scrunch it until it reached the threshold length of 8 nt, after which the entire complex undergoes a phenomenal conformational change into a processive elongation complex.

1.5 Transcription Elongation

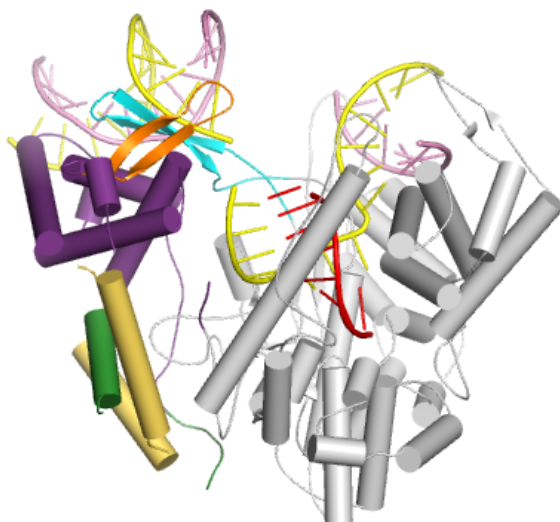
Two structures of T7 RNAP with 8-mer (Tahirov et al., 2002) and 11-mer (Yin and Steitz, 2002) transcripts provide a holistic view of how the features of a transcription elongation complex emerge from the initiation complex after extensive reorganization of the RNAP. A transcribing complex is committed to elongation when it has successfully accommodated numerous concurrent events including promoter release, partial collapse of the bubble, construction of an RNA exit channel, peeling off the 5’ end of the nascent RNA from the template DNA. The threshold length of DNA/RNA hybrid determining this transition to elongation had been only predicted from biochemical observations but in both the transcription elongation structures the nascent RNA forms a heteroduplex with the template DNA of only 8 bps, upstream of which is peeled off the template and directed into a new formed exit channel.

1.5.1 Promoter release and processivity

The NTD, being most closely involved in interaction with the promoter, undergoes major rearrangement. On comparing the RNAP in the initiation and elongation stages, three subdomains in the NTD show independent rearrangement movements. i) A six helix bundle called the promoter binding domain (PBD) undergoes a massive rigid-



B



C

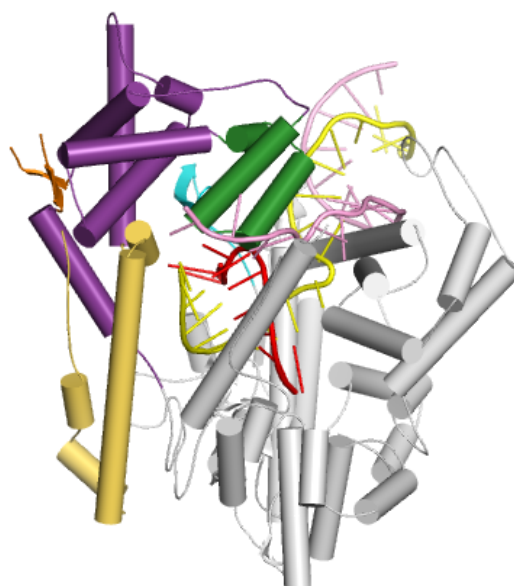


Figure 1.5: Comparison of the structures of initiation complex (A), intermediate complex (B) and elongation complex (C). Helices are represented by cylinders and β -sheets by arrows, while disordered regions are shown in dotted lines. The structures are similarly oriented by superposing the palm domains. Transition involves major conformational changes in the NTD colored as independently moving regions: Promoter binding domain (purple), intercalating hairpin (orange), helix C motif (pale yellow) and subdomain H (green). The CTD (gray) remains mostly unchanged except for movements in the long thumb helix and the specificity loop (cyan). The template strand is yellow, non-template is pink and RNA is red. The downstream DNA is highly bent with respect to the upstream regions to help bubble formation (C). The primary structure of the NTD subdomains is also shown with the same color code (top). These figures have been adapted from (Yin and Steitz 2002)

body rotation of 140° to a position previously occupied by the promoter, thereby destroying its promoter interaction sites. Along with the PBD, the adjacent intercalating β -hairpin, an important motif for promoter melting in initiation, also becomes disordered in this promoter release event. ii) An α -helix (named C-helix) at the N-terminus of the NTD nearly doubles in length by stacking of two smaller helices seen during initiation, thereby forming part of the binding site for DNA/RNA hybrid. This helix protrudes into the region occupied by the PBD in initiation, thus suggesting the concerted nature of the two motions. iii) Subdomain H undergoes extensive refolding into two antiparallel helices, paired with a large translation of 70 Å to the opposite side of the polymerase, forming the rim of the RNA exit channel on one side and interacts with the non-template DNA from its opposite surface (**Figure 1.5 A and C**).

The formation of the RNA exit channel is the most important determinant of transcription processivity. Along with the subdomain H, two important motifs are involved in its formation including the thumb and the specificity loop. The interaction of the channel wall with RNA is only through hydrogen bonding between the phosphate backbone and basic residues of the specificity loop and subdomain H. Processivity is also favored over abortive transcription due to the extensive interactions of the 7 bp heteroduplex with its binding site. Processivity is coupled with the continuous downstream progress of the transcription bubble. At the onset of transcription, the bubble is generated from rotation of the downstream duplex DNA through $\sim 146^\circ$ with respect to the upstream promoter. The template strand also plunges deep into the active site and comes out by bending about 80° at the upstream and downstream end of the bubble.

1.5.2 Transition to Elongation Complex

Transition to the elongation complex from the initiation complex should involve metastable complexes that not only form the basis of abortive cycling but also assume the conformation of an expanded active site to accumulate a growing DNA/RNA hybrid of about 8 bp length (Huang and Sousa, 2000; Temiakov et al., 2000) while still bound to

the promoter. Biochemical studies proposed that the transition was two-steps (Bandwar et al., 2007; Guo et al., 2005; Ma et al., 2005), where RNA extended to about 8 nt followed by major refolding events accompanying synthesis of 9-14 nt (Tang et al., 2008). One of intermediate structures with a 17-mer promoter DNA and 7-mer RNA transcript in a bubble showed the nature of transition between the vastly different initiation and elongation complexes (Durniak et al., 2008). This structure was captured using a mutant (P266L) in a loop connecting the polymerase NTD and C-terminal domains. In the first stage of transition, NTD movements including the PBD and helix C appear to have proceeded halfway, leading to the enlargement of the active site to accommodate the 7 bp DNA/RNA hybrid. Subdomain H of the NTD, however, remains in its initiation orientation (**Figure 1.5 B**). The second stage of transition involves a final movement of the NTD, specificity loop and subdomain H that loses promoter contact, which completes formation of the exit channel and also forces the downstream duplex to its bent position.

1.5.3 Nucleotide addition cycle

In every nucleotide addition cycle, RNAP sieves through the pool of NTPs for the correct substrate through its intricate mechanism of nucleotide selection. The catalytic-competent nucleotide binding N-site elicits the nucleotidyl transfer reaction between the RNA primer 3'-end base at the P-site and the incoming NTP to extend the RNA by one base. The extended DNA/RNA hybrid then translocates upstream relative to the active site, opening the N-site for next round of the cycle. T7 RNAP conducts this harmonized process through fine regulation by elements mainly from the fingers and palm subdomains. Crystal structures of T7 RNAP complexes with the DNA, RNA and nucleotide provide great insights into its nucleotide addition cycle (Yin and Steitz, 2004); (Temiakov et al., 2004).

1.5.3.1 Substrate selection in pre-insertion site

A structure of a ternary elongation complex with a non-hydrolysable ATP analog identified a 'pre-insertion site' for a substrate binding prior to catalytically competent

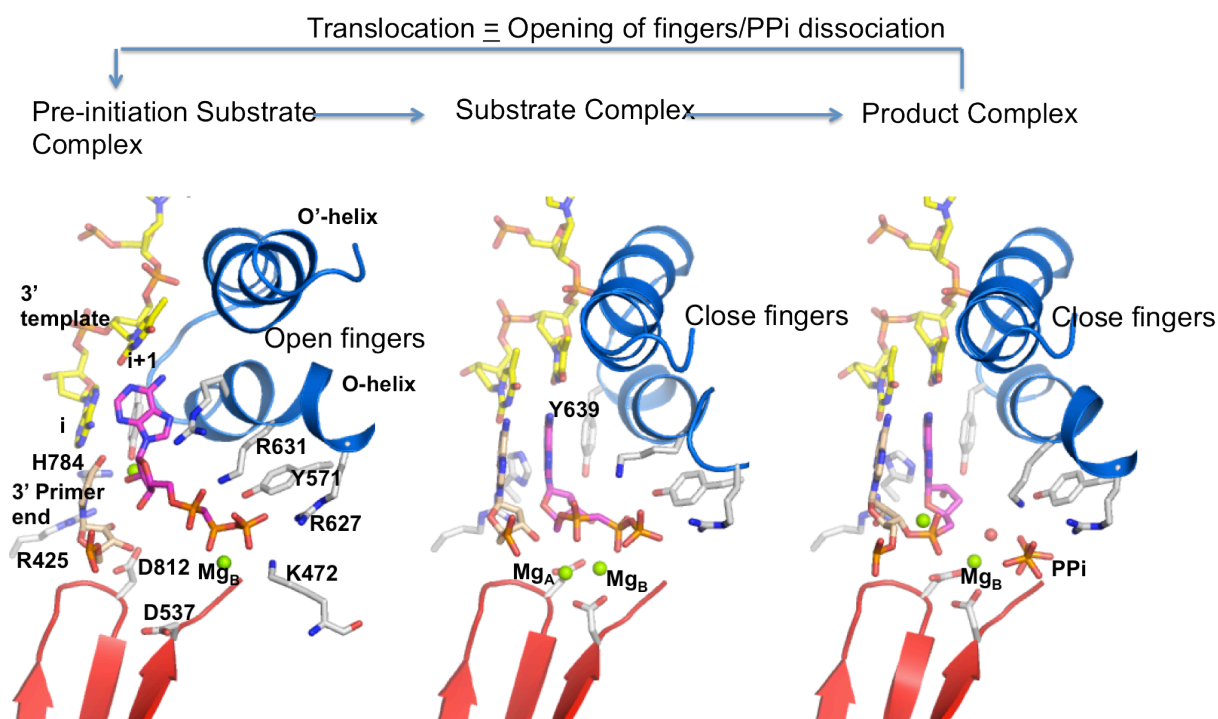


Figure 1.6: Active site views during nucleotide addition cycle. Structures of the preinitiation complex; PDB ID -1S07 (**A**), substrate complex; PDB ID – 1S76 (**B**) and product complex; PDB ID-1S77 (**C**) are aligned similarly by superposing the palm core. Important motifs (ribbons) and amino acid residues (sticks) are shown. Helix O and O' from fingers (blue) and motifs A and C (red) from the palm provide amino acid side chains (white carbon, red oxygen and blue nitrogen atoms) to bind the incoming nucleotide (magenta carbons), magnesium atoms (green spheres) or transcript primer end (light pink carbons). The template DNA is shown as sticks with yellow carbon atoms

nucleotide binding at the ‘insertion’ N-site (Temiakov et al., 2004); **Figure 1.6 A**). The pre-insertion site is linked to the “open” conformation of a conserved O-helix of motif B in the fingers. In this conformation, the templating base faces away from its accepting position and the conserved tyrosine residue at the end the O-helix meanwhile blocks the nucleotide insertion site. The tyrosine hydroxyl group interacts with the 2'-OH of the substrate, forming the primary discrimination of incoming rNTPs against dNTPs. The substrate is bound along the O-helix but is not Watson-Crick paired with the template DNA base, implying that the pre-insertion site is an early fidelity checkpoint.

1.5.3.2 Substrate loading at the catalytic site

Substrate loading from the pre-insertion site to the catalytic insertion site is achieved by “closing of the fingers”, wherein rotation of the O-helix is the most significant. Structure of a pre-catalytic substrate complex trapped with the use of a non-reactive nucleotide analog showed the nature of O-helix movement when substrate was loaded onto the catalytic insertion site (N-site) and thereby defined the ‘closed conformation of the fingers (Yin and Steitz, 2004). The O-helix rotates about a pivot point at its middle, causing the amino-end of the helix to close onto the substrate triphosphate moiety, and is stabilized in this position through positive charge from lysine and arginine residues. Simultaneously, the opposite end of the helix, including the important tyrosine residue, moves away to make space for the base moiety of the incoming substrate. The base specific and ribose discriminating interactions of the pre-insertion site are maintained in the closed conformation to face a final round of fidelity check.

The incoming substrate alignment for catalysis in the active site is maintained not only by the O-helix residues, but also by the accompanying nucleotide binding metal Me^B , which is, in turn, positioned by chelating the conserved active site aspartates. Further, the octahedral coordination of catalytic metal ion Me^A maintains the critical alignment of the reactive groups- 3'OH of primer terminus and 5' αP of incoming substrate. The mechanism of nucleotidyl transfer reaction, which extends one RNA base and produces a pyrophosphate (PPi) by-product, will be described in the **Section 1.6.3**.

1.5.3.3 Translocation

After the addition of a base to the transcript, the final step of the nucleotide addition cycle is the translocation of the DNA/RNA hybrid through one base distance such that the 3'-end RNA positions in the P-site. The structure of a post-catalytic, pre-translocated product complex, isolated right after phosphodiester bond formation but before dissociation of nascent PPi, showed that the phosphodiester bond does not cause

any change in the RNAP or DNA/RNA hybrid (**Figure 1.6 C**). In the post-translocated state, the only difference lies in the dissociation of PPi along with the coordinated Me^{B} , which breaks the interactions with the O-helix residues and thus favors the open state. Moreover, the rNTP discriminator tyrosine at the O-helix, moves towards the heteroduplex and stacks with the primer end base preventing backtracking of the hybrid, while opening the triphosphate site due to the pivoted helix motion. It is thus proposed that translocation of the DNA/RNA hybrid in single-subunit RNAP is coupled with PPi dissociation, which leads to the opening of O-helix for the next round of nucleotide addition, often called the power-stroke mechanism of translocation (Jiang and Sheetz, 1994).

1.6 N4 virion encapsidated RNAP - factor dependent single-subunit polymerase

While the characterization of T7 related RNAP were progressing rapidly, discovery of a unique phage-encoded, virion-encapsidated RNAP (vRNAP) isolated from a lytic coliphage N4 (Falco et al., 1977) added breadth to the studies of single-subunit RNAPs. In contrast to other phage RNAPs, N4 vRNAP molecules are encapsulated within the virion to be injected into the bacterial cell at the onset of infection (**Figure 1.7A**). Bacteriophage N4 (lytic coliphage from Podoviridae), a ds DNA phage was discovered in Genoa, Italy (Molina, 1965; Falco 1977) with a very special infection mechanism and transcription program. General phage infection, like in T7, involves injection of the genomic phage DNA into the bacteria to be transcribed by host RNAP, which is later shut off due to newly transcribed phage gene products (Chamberlin et al., 1970). N4 phage, surprisingly, contains pre-synthesized copies of RNAP encapsidated within the virion, called the N4 virion-encapsidated RNAP (N4 vRNAP), which is injected into the bacterial cell at the onset of infection. This antithesis was concluded based on genetic studies showing that rifampicin (*E. coli* transcription inhibitor) and chloramphenicol (*E.coli* translation inhibitor) treated cells continued to produce early phage gene products, proving that host machinery was not needed at this stage and that there must be present an RNAP present in the virions before infection (Schito et al., 1974). The presence of this RNAP was recently confirmed by Cryo-EM based

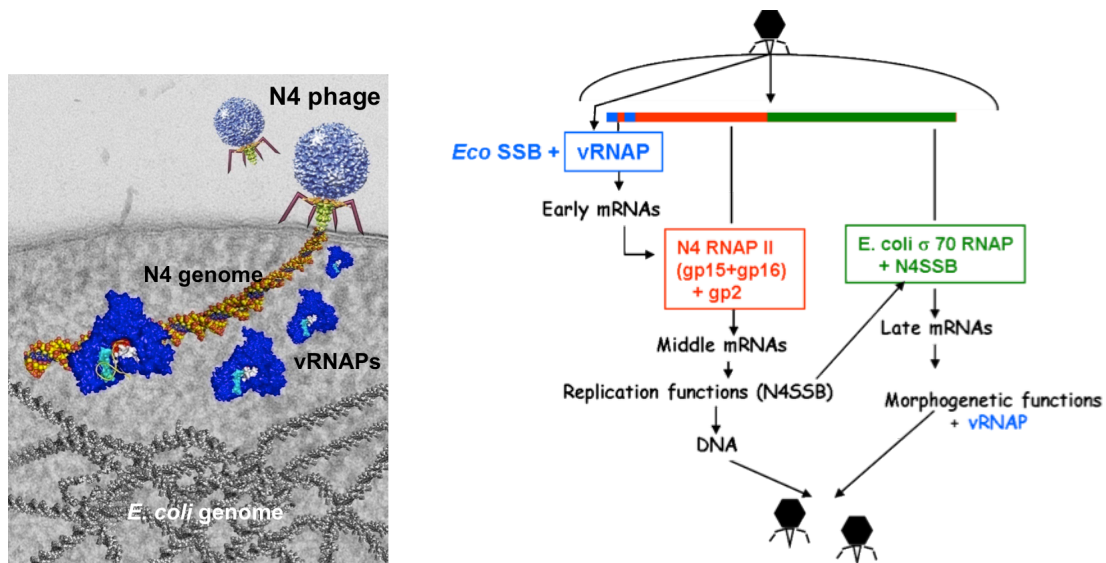


Figure 1.7 N4 bacteriophage infection and transcription program. (A) The cartoon represents the injection of vRNAP (blue) copies with the bacteriophage genome (yellow and grey) into the host cell (gray scale) (Image by K. H. Choi). (B) N4 phage adsorbs on the host cell membrane and injects its genome (multicolored line) and N4 vRNAP. The vRNAP with the help of host *EcoSSB* protein transcribes the early (Class I) mRNA (blue portion) which encodes N4 RNAPII and gp2. These proteins together transcribe the middle (Class II) mRNA (red portion), which encodes for N4SSB. N4SSB then assists Ecoli RNAP to transcribe the late (Class III) mRNAs (green portion) that encode for N4 vRNAP, which in turn is packaged into phage progeny. (Image provided by Lucia Rothman-Denes)

localization of the virion proteins within the capsid. Using S^{35} labeled methionine this study also proved that there are 4 copies of vRNAP per capsid (Choi et. al 2008).

In N4, the early genes, transcribed by the vRNAP, encode for a second RNAP, called N4 RNAP II that takes over transcription of the later genes like other phage RNAPs (**Figure 1.7B**). N4 RNAP II is also special in that it is made of two subunits (p7 and p4; **Figure 1.1**), although the combined sequence contains the same conservation blocks as the single-subunit RNAPs (not discussed in this thesis). Further atypical behavior of this phage is that the host transcription and translation machinery is not shut off even late in infection. Instead, the late genes are actually transcribed by the host RNAP (Rothman-Denes et al., 1972). The unique infection behavior of N4 has special implications on the molecular mechanism of transcription by the vRNAP. After entry into the host, N4 vRNAP needs to be maintained in an inactive state until presented with the special hairpin form N4 promoter so as to maintain its specificity in the host cell. Upon injection of the N4 double stranded (ds) genome, host *E. coli* DNA gyrase and (Dai et al.,

1997; Markiewicz et al., 1992) and single stranded DNA binding protein (*EcoSSB*) (**Figure 1.8**) prepare the N4 vRNAP-specific promoter, which is comprised of a hairpin with 5-7 bp stem and a 3-base loop. (Glucksmann-Kuis et al., 1996; Haynes and Rothman-Denes, 1985). This hairpin promoter is the key to activate the co-injected vRNAP from its erstwhile transcription-inactive state. After transcription initiation, N4 vRNAP depends on *EcoSSB* for transcript elongation too, which helps to displace nascent RNA transcripts from the template (Davydova and Rothman-Denes, 2003) for recycling. Thus the part of the NTD responsible for RNA separation in T7 transcription elongation is absent in N4 vRNAP. Thus, *EcoSSB* is a transcription factor for N4 vRNAP and N4 vRNAP emerged as an important model for studying the structural basis of transcription activation as well as factor-dependent transcription. Much larger than other phage RNAPs, the 320 kDa N4 vRNAP can be divided into three domains and a central polymerase domain of 1100 amino acid (called mini-vRNAP) exhibits transcription

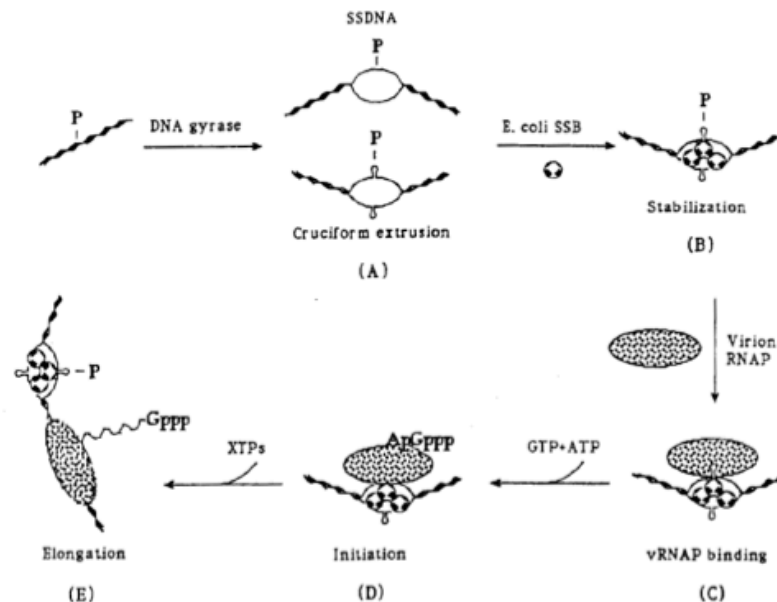


Figure 1.8. Model for in vivo N4 promoter extrusion and recognition. P represents the N4 promoter sequence, which is negatively supercoiled with the help of *Eco* DNA gyrase. Unwinding the genome presents a cruciform structures consisting of the N4 vRNAP promoter (A). *EcoSSB* stabilizes the hairpin promoter by binding to the single stranded region of the complementary strand (B). N4 vRNAP then recognizes the promoter, binds to it (C), initiates transcription (D) and upon elongation releases transcripts with the help of *EcoSSB* (E). (Glucksmann et al., 1992)

initiation, elongation and termination properties identical to full length vRNAP (Kazmierczak et al., 2002).

1.6.1 Structure of the N4 vRNAP

Studies of the structure and function of N4 mini-vRNAP advanced the scope of single-subunit RNAP studies. Despite a low sequence similarity, the N4 mini-vRNAP structure was highly similar to the T7 RNAP structure (Murakami et al., 2008) (**Figure 1.3C**). It reiterated the modular organization of the right-hand like structure of the A-family of DNAP, with the same subdomains surrounding the DNA binding cleft and the same structural motifs. However, the promoter recognition motifs in the fingers and NTD, although similar, has adapted interactions specific for the special hairpin N4 promoter.

The two major differences between the N4 vRNAP and T7 RNAP structures are the presence of a plug module insertion in the NTD of N4 vRNAP and a loop inserted in the middle of motif B (motif B loop). The plug and motif B loop interact with motifs A and C of the palm that covers all catalytically essential amino acid residues of N4 vRNAP, thereby maintaining the apoenzyme in a transcription-inactive state (**Figure 1.9A and B**).

1.6.2 Unique Promoter requirements of N4 mini vRNAP

Although N4 promoters are highly distinct, N4 mini-vRNAP employs the same recognition motifs as in T7 RNAP and are positioned similarly so as to deliver the same functions such as strand separation and template positioning at the active site. However, the nature of these interactions are tailored for the specific promoter sequence and length. N4 vRNAP recognize three promoters P1, P2 and P3, all of which have the hairpin form (5 bp stem and 3 base loop) with subtle sequence differences (**Figure 1.9D**).

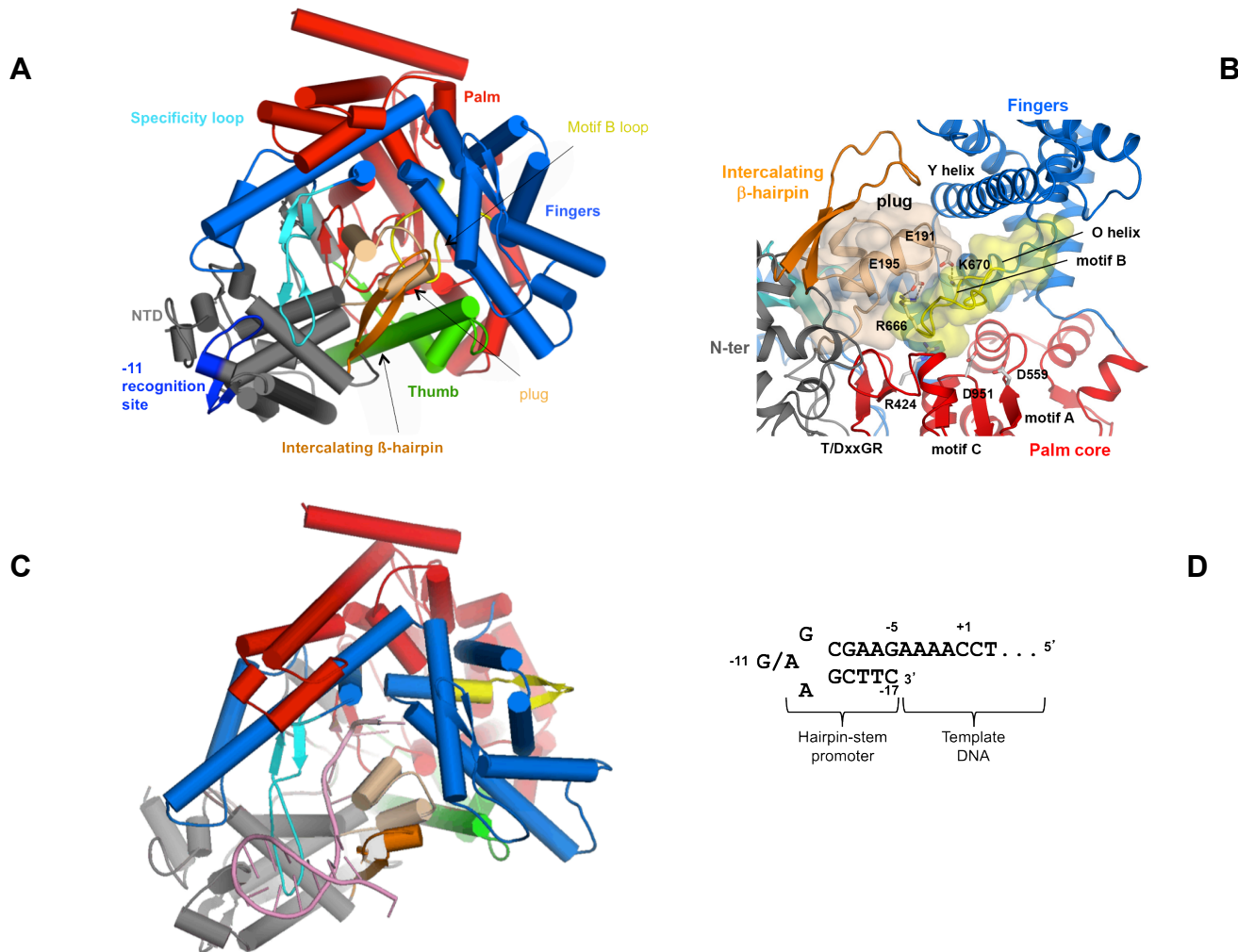


Figure 1.9: Activation of N4 mini vRNAP by the specific hairpin promoter. (A) N4 mini-vRNAP apo structure shows the plug (wheat) and the motif B (yellow) blocking the active site. (B) Surface representations of the plug module and motif B in the apo-enzyme. (C) Binary Complex structure of N4 mini-vRNAP with the promoter (pink) shows the displaced plug and the refolded motif B loop pointing outwards from RNAP surface. (D) Typical N4-vRNAP promoter sequence that forms the hairpin of 3 base head and 5bp stem followed by a single stranded template region.

Analogous features in N4 vRNAP are adapted for specific promoter recognition (**Figure 1.4 B**): (i) The -11 recognition site in the NTD (around the AT-rich recognition site) recognizes the tip of the hairpin loop by base-specific interactions mainly through the stacking of a tryptophan against the guanosine base of the center (-11) of the loop, and interactions of basic residues with the promoter bases -10/-11. (ii) The specificity loop from the fingers contacts the hairpin from the major groove as in T7 RNAP, forming base specific interactions with upstream 8-10 bases. (iii) Intercalating β -hairpin melts the

promoter by employing two lysine residues, in place of valine in T7 RNAP, that form hydrophobic walls stacking against the junction of the double stranded and single stranded regions. (iv) An additional salt bridge interaction between the phosphate backbone of the hairpin stem with lysines from the fingers add to the stability of the complex.

Template positioning is enforced through additional interactions in N4 in comparison to T7 (**Figure 1.4 B**). (i) Stacking between an aromatic ring of specificity loop and head of the promoter hairpin defines the upstream end of the promoter. (ii) The intercalating hairpin defines the junction of the upstream duplex and single stranded template at positions -4 and -5. (iii) A unique four-adenine stretch in the promoter sequence (-4 to -1) serves as a ‘molecular ruler’. (iv) Cation- π interactions of an arginine with the -2 base and salt bridges with the -2/-1 backbone introduce unstacking, thus sharply bending the backbone between bases -2 and -1 for entry into active site (Davydova et al., 2007; Gleghorn et al., 2008a). The backbone between +2 and +3 bends 90° preventing transcription initiation from +2 and downstream.

In addition to the regular promoter binding functions, N4 vRNAP has developed features for activating the RNAP. The bases in the loop (-9 to -12) of the hairpin promoter are stacked against each other to resemble a double stranded helical conformation of DNA. The promoter-induced activation of N4 mini-vRNAP is, precisely, a 25.1° rigid body movement of the plug away from the active site. Simultaneously, the motif B loop refolds outwards from the active site, moving its tip through a vast 32.6 Å (**Figure 1.9 A-C**), which turns the catalytically important residues on the O-helix to face the active site for binding to incoming NTP.

1.6.3 Nucleotidyl transfer reaction

The nucleotidyl-transfer reaction by DNA and RNA polymerases follows a generalized mechanism of two-metal catalysis (Steitz et al., 1994); the two metals being coordinated by two or three conserved aspartic acid residues of the active site. The

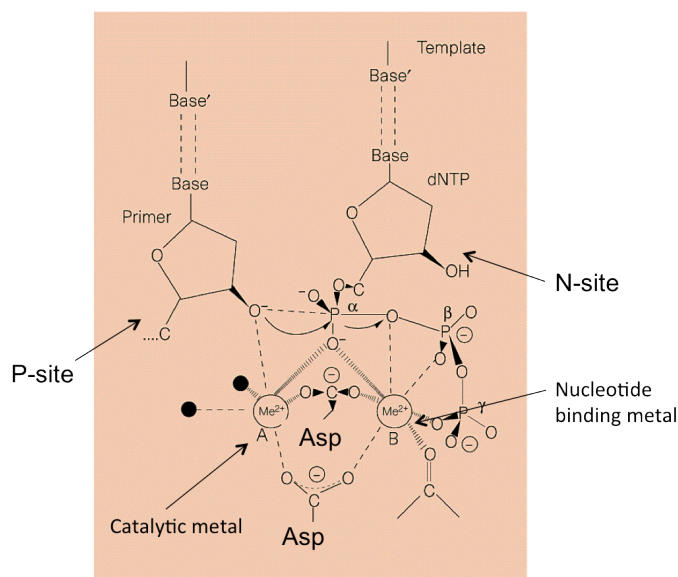


Figure 1.10. Two Metal catalysis mechanism for phosphoryl-transfer reactions. The Schematic shows the 3' primer end at the P-site and incoming NTP bound to the N-site. Catalytic metal (A) is coordinated by two aspartate carboxylates and the 3'OH of the primer end. Nucleotide binding metal (B) is bound to the aspartates and the triphosphate moiety of the incoming NTP. Curly arrows represent the electron transfer mechanism; Deprotonated 3'O⁻ attacks the 5' α P of N-site NTP and releases PPi.

catalytic metal (Me^{A}) is known to work as a Lewis acid to enhance the nucleophilicity of the 3'-oxygen attacking group of the primer 3'-end and initiate the $\text{S}_{\text{N}}2$ attack on the 5'- α P of the incoming nucleotide. The nucleotide binding metal (Me^{B}) is coordinated by the triphosphates of the incoming NTP, which stabilizes the pentacovalent transition state (**Figure 1.10**).

In the case of transcription initiation, two NTP substrates bind to the empty P and N-sites concurrently to form the first phosphodiester bond. Structural snapshot studies (Gleghorn et al., 2011b) of ternary complexes of N4 mini-vRNAP during transcript initiation provided the most updated knowledge of the general mechanism of nucleotidyl transfer reaction in single-subunit polymerases before the start of my work (Chapter 2). While similar structural snapshots had earlier been produced for T7 RNAP in the elongation complex (**Section 1.5.3**), the constituent details were not so completely established. Instead the T7 studies provided only a view of the translocation mechanism following RNA synthesis.

Two forms of pre-catalytic substrate complexes were isolated for N4 mini-vRNAP; both of which contained the two initiating nucleotides at the P and N-sites but the catalytic metal (Me^A) was absent in one of them. These intermediates showed the important, final molecular rearrangements in the active site elicited by the binding of the critical Me^A to allow catalysis. Consistent with the steps in transcript elongation, the O-helix in the initiation complexes was in a closed state, providing its basic residues for stable interactions with the

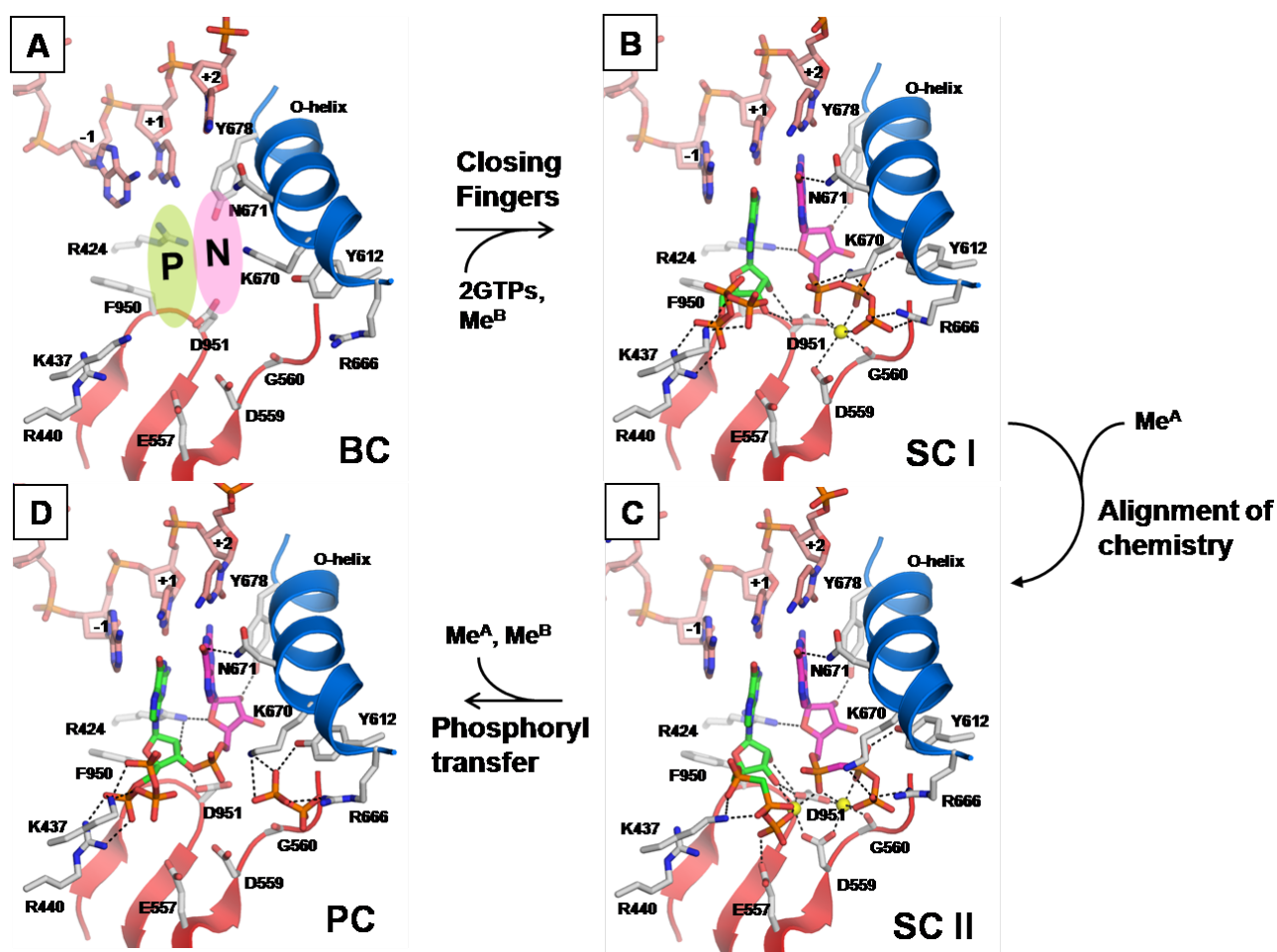


Figure 1.11: Suite of structural snapshots in nucleotidyl-transfer reaction. This figure shows the main chains (ribbon models) of motifs A and C (red) and of O helix (blue), and main and side chains (stick models) involved in nucleotide and metal binding in the BC (A), SCI (B), SCII (C) and PC (D). NTP binding sites, P- and N-sites, are indicated as green and magenta circles in A. DNA template (from -1 to +2, pink) and nucleotides at +1 (green) and +2 (magenta) positions are shown as stick models. Divalent metals (Mg^{2+} or Mn^{2+}) are depicted by yellow spheres. Hydrogen bonds and salt-bridges are depicted by black dashed lines. Amino acid residues discussed in the text are labeled. Structural snapshots of the N4 mini-vRNAP active site in transcription initiation. (Gleghorn et al., 2011)

nucleotide at the N-site. Additionally, the stability of the nucleotide binding at the P-site, an initiation-specific event, could be explained by electrostatic interactions of basic residues of the palm with the triphosphates and the partial base stacking with a purine at the -1 template DNA position. Furthermore, the γ P group of GTP (+1) participates in the Me_A octahedral coordination. Loss of affinity and catalytic activity with GDP (+1) compared with GTP (+1), led to the proposition of substrate-assisted catalysis for the first phosphodiester bond formation mediated by the γ P group and Me^A binding.

In the absence of Me^A , the 3'-O from GTP (+1) and the 5'- α P of GTP (+2) were beyond reacting distance (4.1 Å), and were brought closer only with the binding of Me^A (**Figure 1.11 B and C**). This observation suggested that the catalytic metal binding, which is sensitive to its octahedral coordination requirements, serves as a final fidelity checkpoint where a fine misalignment of the 5'- α P of a mispaired incoming nucleotide disallows the Me^A binding and thereby prevents catalysis. However, this proposition regarding the fundamental mechanism begs confirmation and asks what the temporal order of these binding and rearrangement events were. This is the question that I have addressed in my work, using a time course soak-trigger-freeze crystallographic approach, which I present in Chapter 2. I also collaborated with Prof. Paul Carey's lab to probe the same mechanism using Raman Crystallography, which provided helpful background to design my time-course crystallography experiments. A synopsis of these results have also been provided in Chapter 2.

1.7 Multi-subunit RNAP

Cellular RNAP from all organisms consist of multiple subunits. While bacteria and archaea contain a single RNAP, eukaryotes have three RNAPs to transcribe a separate subset of genes: RNAP I (Pol I) for large rRNA, RNAP II (Pol II) for mRNA and RNAP III (Pol III) for tRNA and 5S rRNA (Cramer, 2002). The bacterial core RNAP is composed of 5-6 subunits, archaeal RNAP comprises 11-13 subunits, eukaryotic Pol I, Pol II and Pol III are, respectively composed of 14, 12 and 17 subunits (Werner and Grohmann, 2011). Having evolved from a common ancestor, the 5 core subunits of bacterial RNAP ($\alpha_2\beta'\beta\omega$) have homologues in all cellular RNAPs. Subunits β' and β

respectively correspond to Rpb1 and Rpb2 in eukaryotic Pol II and RpoA'/RpoA'' and RpoB in archaea, α I and II are Rpb3 and Rpb11 in Pol II and RpoD and RpoL in archaea, and ω is Rpb6 in Pol II and RpoK in archaea. Consequently, the architecture of the main body of RNAP and functions of the subunits in transcription cycle are conserved in all three domains of life. While the largest subunits β' and β form the catalytic subunits consisting of all the architecture necessary for catalysis, nucleic acid scaffold binding and substrate entry, the α subunit dimer forms an assembly platform on which the catalytic unit is built. The first view of the multisubunit RNAP was obtained from the X-ray crystal structure of the *Thermus aquaticus* (*Taq*) RNAP core enzyme (Zhang et al., 1999), which resembled a crab-claw, with the β' and β subunits forming each of the claws. The cleft between the claws is the main channel for double-stranded DNA binding. The downstream DNA runs along the β' side of the cleft until it hits the 'wall' of the cleft, wherein resides the active site. The active site is made of two double psi β -barrel motifs

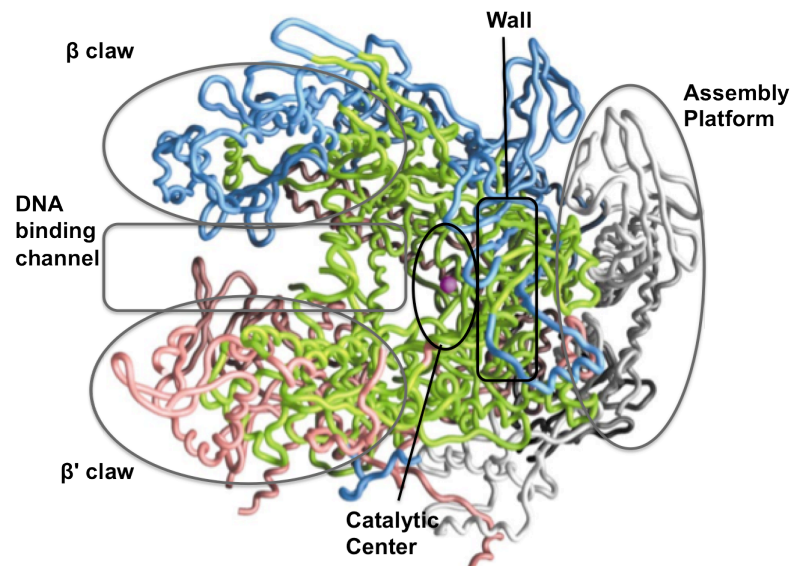


Figure 1.12: Evolutionarily conserved regions of the large subunits form the core of multi-subunit RNAPs. The side view of the active site channel in the *Taq* core RNAP is shown (worm representation of the α C backbone). The α and ω subunits are shown in white, β in cyan and β' in pink except the conserved regions in cellular RNAPs which are shown in green. The catalytic metal at the active site is shown as a magenta sphere. These regions are evolutionarily conserved among bacteria, archaea and eukaryotes. The important functional parts of the RNAP architecture, discussed in the main text, have been marked with grey boundaries and labeled (Figure adapted from (Darst, 2001)).

that is well conserved in all cellular RNAPs (**Figure 1.12**) and includes a DFDGD motif consisting of the active site essential three aspartates for divalent metal coordination. A trigger loop and a bridge helix are essential elements in the nucleotide addition cycle and these separate the main DNA binding channel from the secondary channel.

The three steps of the transcription cycle including initiation, elongation and termination in cellular RNAPs are more sophisticated than the single-subunit RNAPs. Here, each stage of transcription involves the timely participation of specialized subunits or factors that either switch on or off a function specific for each stage in transcription. Exquisite structures of RNAP in the initiation or elongation stages from thermophilic bacteria *Taq* (Murakami et al., 2002a; Murakami et al., 2002b), *Thermus thermophilus* (*Tth*) (Vassylyev et al., 2002; Zhang et al., 2012) and recently from the most studied *Escherichia coli* (Murakami, 2013) and *Saccharomyces cerevisiae* (reviewed in Martinez-Rucobo and Cramer, 2013) have provided valuable information about the

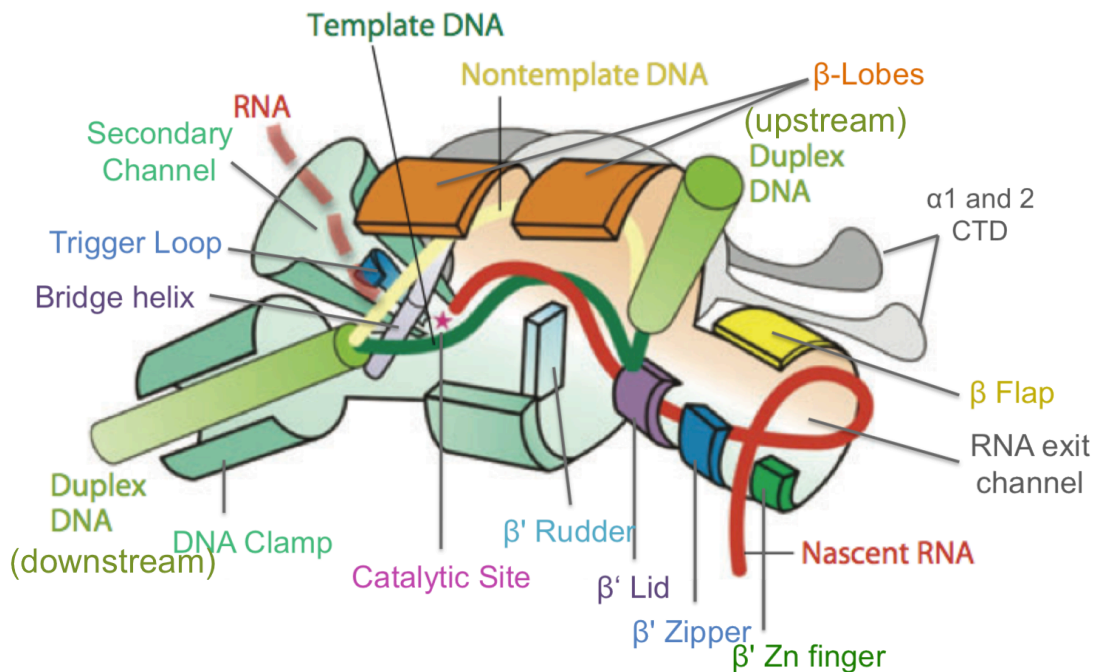


Figure 1.13: Schematic representation of the RNAP structure in transcription elongation. The key structural features and mobile elements during transcription are labeled. Their roles are defined in the main text. RNAP moves the DNA template from right to left. The downstream DNA is bound to the clamp forming the DNA binding site. The trigger loop and bridge helix separate the DNA binding site from the DNA/RNA hybrid binding site, which is upstream of the catalytic center (pink star). The zipper, lid and Zn finger with the β flap form the RNA exit channel. Figure adapted from (Nudler, 2009)

complex orchestration of events in the transcription cycle. Transcription in multi-subunit RNAP is highly regulated through external factors for enhancing and repressing gene expression or modulating the rate of transcription. However, this topic has not been discussed in this thesis.

1.7.1 Transcription Elongation

The bacterial RNAP core enzyme is capable of processive RNA synthesis, which constitutes the productive and most frequented stage of transcription, involving the reiterative addition of nucleotides to the 3'-end of an RNA transcript followed by single-base translocation. In the elongation complex, the core enzyme binds to a nucleic acid scaffold, which includes a double stranded DNA, a transcription bubble of 10-12 nt in length that fits into the main channel and a DNA/RNA hybrid (8-9 bp long). Crystal structures of elongation complexes from *Tth* (Vassylyev et al., 2007a) and *Saccharomyces cerevisiae* (Westover et al., 2004a, b)) define the architectural basis that supports nucleic acid binding, while ensuring the smooth threading of duplex DNA into the active site, thereby maintaining a constant size of the transcription bubble and allowing the unimpeded extrusion of RNA through an exit channel.

Three important nucleic acid contacts that are maintained throughout elongation include 1) a downstream DNA binding site, 2) nascent DNA/RNA hybrid bound in the central channel and 3) the RNA binding site for the new transcript (**Figure 1.13**). (1) The downstream duplex DNA, 9 base pairs down from the $i+2$ position (the numbering ' i ' corresponds to the 3' primer end position and ' $i+1$ ' is the binding site for the incoming NTP) are clamped by a deep cleft of the β' subunit. The RNAP-DNA interface is mostly stabilized by long-distance electrostatic or van der Waals interactions to ensure the frictionless threading of DNA through the clamp for processive elongation. The DNA faces a sharp 90° kink after entry into the active site before which point it is unwound to generate the transcription bubble. This strand separation is caused by base stacking of the $i+2$ site with the fork loop 2, which prevents further entry of duplex DNA and thereby

defines the downstream edge of the transcription bubble. (2) A DNA/RNA hybrid of about 8 bp is stabilized in the main channel between the 3' growing RNA end and an upstream contact with the β' lid loop. Stabilization is again mediated by weak Van der Waals contacts of hydrophobic side chains of the β and β' subunits with phosphate backbones that minimize hurdles in the smooth sliding of the hybrid. The β' lid loop marks the upstream limit for the hybrid by inserting a wedge between the strands at the *i*-9 base to discontinue the hybrid length. The β' rudder contacts the -9, -10 and -11 DNA bases which prevents its reassociation with the RNA. The rudder along with the β fork loop 1, β' lid and β' zipper form a dynamic network of interactions with nucleic acids, which serves as an upstream zip-lock for the DNA/RNA hybrid during the back and forth sliding of the hybrid. Analogously there is a downstream zip-lock, namely the bridge helix that serves as a wall in a backtracked complex forcing strand separation and allowing only single stranded RNA to extrude through the secondary channel. Thus the zip-lock mechanisms at the two ends maintain a largely constant hybrid length of 7-10 base pairs. (3) The nascent RNA transcript separated at the *i*-9 position is stabilized through interactions with the β' lid and the β saddle that form a pore leading to the RNA exit channel, which is lined by the β' zipper, β' Zn finger and β flap (**Figure 1.13 and Figure 1.14**).

When compared to the single-subunit RNAP elongation complex, the DNA-RNA scaffold in the multi-subunit RNAP elongation complex is very similar. The downstream DNA is bent about 80°, with respect to the heteroduplex in both cases. The length of the heteroduplex (8-9 bp) and the number of unpaired template bases between the primer terminus and the downstream duplex are the same, suggesting a common mechanism of translocation and duplex opening, although these processes are mediated by distinct structures in the two systems. On the other hand, the RNA exit channel in the two systems differ considerably. In T7 RNAP, the RNA exit channel is constructed in the elongation stage by major conformational changes of the NTD whereas multi-subunit enzymes already contain an exit channel, which is only blocked by σ factor in bacteria or TFIIB in eukaryote Pol II during transcription initiation.

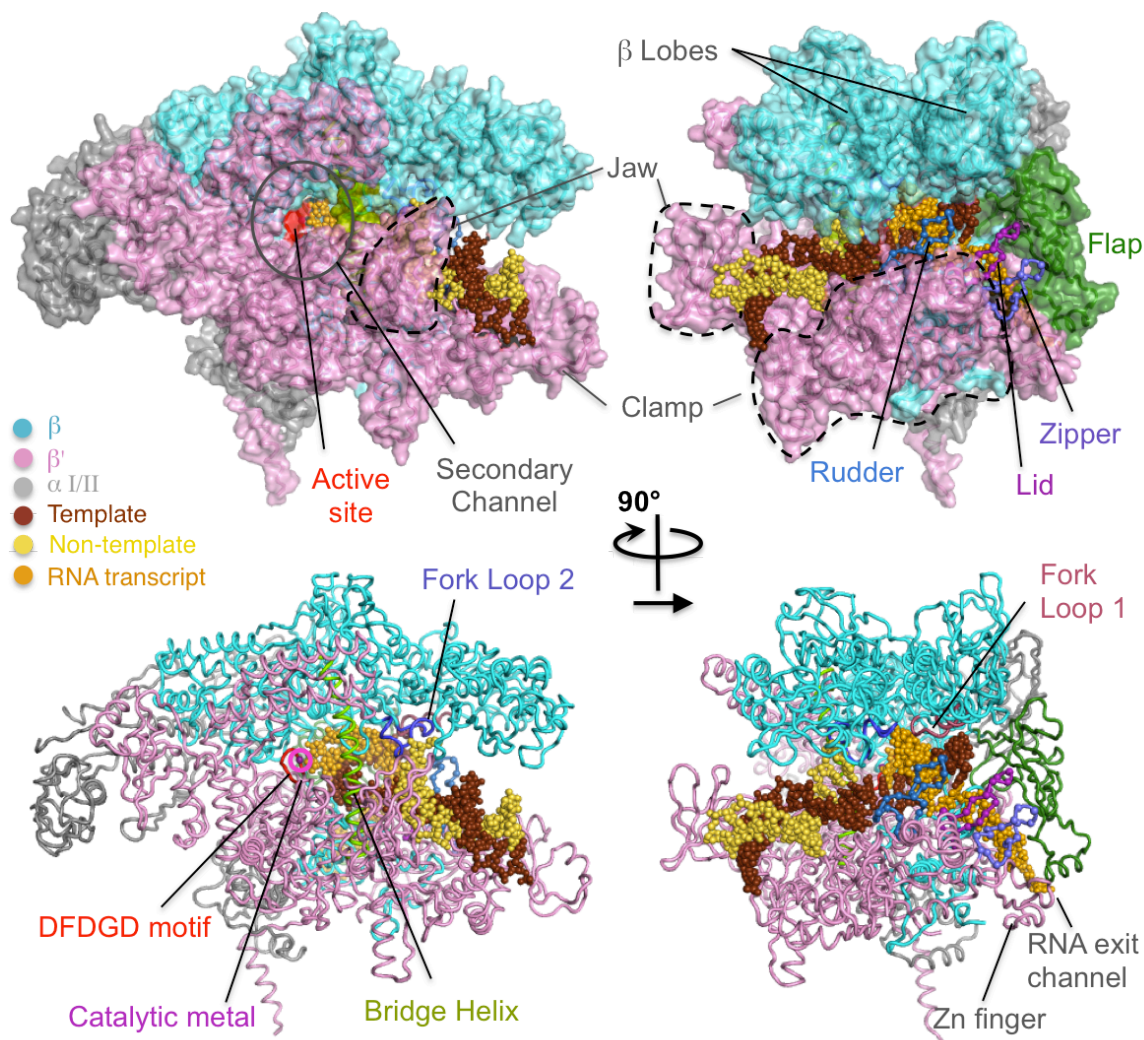


Figure 1.14. Structures of *Tth* elongation complex (PDB ID: 2O5I; (Vassylyev et al., 2007a)). Two views of the *Tth* EC are shown as molecular surfaces (A and B) and the same views are shown with α -carbon backbone as worms (C and D). The nucleic acid strands are shown as spheres. The structure consists of two components: the *Tth* core RNAP subunits are color-coded (β , cyan; β' , pink; and α , ω gray; catalytic Mg²⁺, magenta sphere); and the nucleic acid scaffold (template strand, brown; nontemplate strand, yellow; and RNA transcript, gold). The β' non-conserved domain is omitted for simplicity. Parts of the protein structure discussed in the text are colored (β flap in green, β' rudder in skyblue, β' lid in magenta, β' zipper in violet). The structure on the left shows a view down the secondary channel, showing the path for diffusion of the incoming nucleotide substrate into the active site. The catalytic Mg²⁺ ion (magenta sphere) is just visible to the right of the catalytic motif (red). The structure on the right shows a view perpendicular to the main active site channel, which runs roughly horizontal. For clarity, parts of the β' subunit nearest the viewer are shown as a backbone worm without the corresponding molecular surface. and the upstream duplex DNA has been removed.

1.7.2 Transcription Initiation

Promoter-specific transcription of bacterial RNAP is enabled through an additional initiation factor (σ) that binds to the above described core enzyme, and together is called the holoenzyme. The holoenzyme has the ability to recognize a specific DNA sequence called the promoter, which is the major determinant for initiating transcription of a gene. The σ factor reversibly binds to the core RNAP and the holoenzyme in turn binds to promoter sequences that triggers DNA unwinding and preorganizes the template to initiate transcription. After the transcription machinery is capable of processive elongation the σ factor usually dissociates. Bacterial systems have multiple σ factors. The predominant σ factor is called the housekeeping σ and it is responsible for most of the transcription in cells. The housekeeping σ in *E. coli* is σ^{70} and in other bacteria is σ^A . The housekeeping σ contains several conserved regions that fold into four independent structural domains including $\sigma 1.1$, $\sigma 2$, $\sigma 3$ and $\sigma 4$ (Campbell et al., 2002; Lonetto et al., 1992). The first structure of the bacterial holoenzyme from *Taq* disclosed the basis of σ factor binding with RNAP (Murakami et al., 2002a), which was later confirmed with a higher resolution *Tth* structure (Vassylyev et al., 2002) and completed by the latest structure from *E. coli* (Murakami, 2013). The $\sigma 2$, $\sigma 3$ and $\sigma 4$ domains are spread across the surface of the core RNAP with discrete interaction points (1) $\sigma 1.2$ and 2.2 to the β' coiled coil at the floor of DNA binding channel; (2) $\sigma 2-3$ linker and $\sigma 3$ to the β' zipper, β protrusion, and β flap; (3) $\sigma 3-4$ linker to the RNAP main channel, β saddle, and RNA exit channel; and (4) $\sigma 4$ to the β flap tip, β' ZBD, and β' dock; (5) $\sigma 1.1$ binds to the main DNA binding channel, from where it is excluded upon transition to a stable open complex (RPo). The formation of RPo from the transient DNA bound closed complex (RPc) involves the stable binding of the holoenzyme to the respective promoter elements, melting of duplex DNA, and feeding only the template DNA strand into the active site cleft (**Figure 1.15**). σ region 1.1 is the one domain that drastically changes its position during this transition (Mekler et al., 2002). It is this highly flexible N-terminal 90 amino acids of σ region 1.1, which was obscured from early

DNA (Zhang et al., 2012) define the basis of promoter binding in RPo (**Figure 1.16**). The promoter DNA upstream from the -10 element sits on the surface of holoenzyme and all these promoter elements are recognized by σ factor (not the RNAP core enzyme).

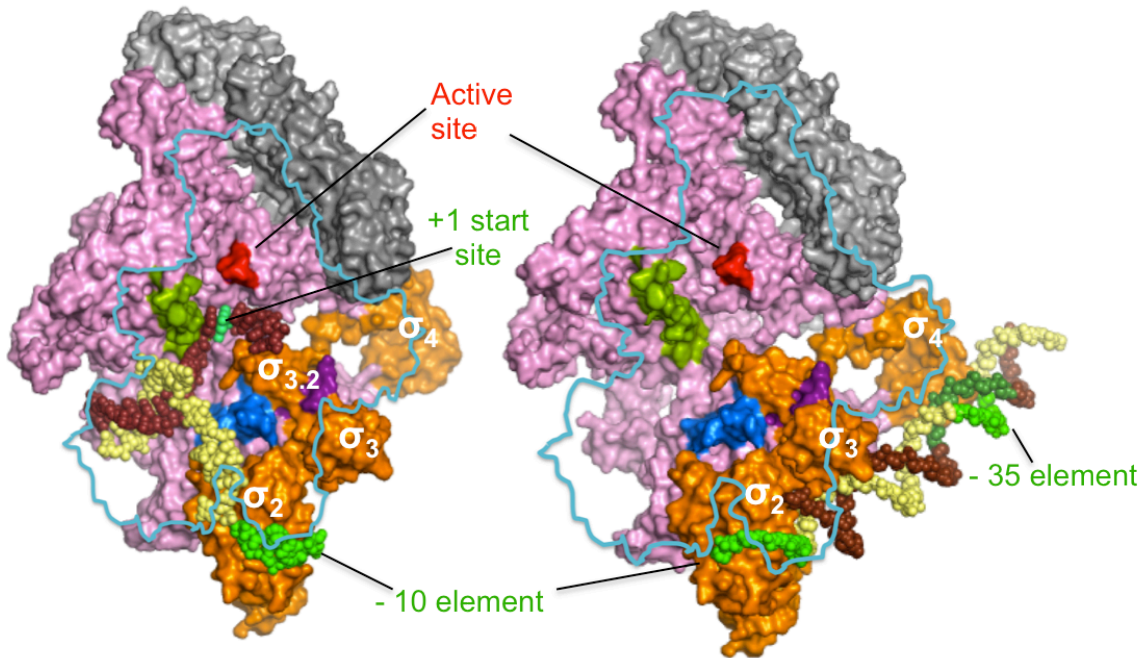


Figure 1.16. Structures of transcription initiation complexes. (A) Structure of the *Tth* holoenzyme (PDB ID: 4G7H) with the promoter -10 element (green) and transcription start site (green). (B) Structure of *Taq* holoenzyme with upstream promoter elements -10 and -35 (green) (PDB ID: 1L9Z). Color code same as Figure 1.15; Template is colored brown and non-template is colored yellow. Obscuring parts of β are removed and only shown as outline. σ domains have been labeled.

After the RNAP – promoter open complex has been established, it is ready to initiate transcription. RNAP initiates transcription without using an RNA primer, which is called *de novo* transcription. This is the only step in the entire transcription process that accepts two nucleotides at the same time, one of them going to the unique P-site or *i*-site and another to the usual incoming nucleotide N-site or *i+1* site. While all subsequent nucleotide addition steps obtain stability from base-stacking with a 3' RNA primer end, what is the extra stabilizing feature in a *de novo* complex that allows nucleotides to bind in catalytic orientation and proceed to reaction? Although the initial pre-organizing steps of transcription initiation has been well studied in multi-subunit RNAPs, this question about the *de novo* transcription mechanism have not been adequately addressed. In

Chapter 3, I discuss my studies on the *Tth* RNAP initiation complex structures to describe the basis of stabilization of a *de novo* transcription complex. Here I also demonstrate the role of σ in this crucial step of transcription initiation.

After the first nucleotidyl-transfer reaction, RNAP is able to transcribe a stretch of 3-9 nts. During this stage the nucleic acid strands accumulate in the active site by scrunching because RNAP is still tethered to the promoter elements (Goldman et al., 2009; Margeat et al., 2006). During the abortive transcription stage RNAP not only maintains its promoter contacts but also has the σ factor bound to it, and therefore transcribes in this late-initiation or early elongation stage as a holoenzyme. At this point, RNAP repeatedly cycles through an iterative process of abortive release of transcripts, which coincides with its attempts to transition from an initiation to an elongation mode, which is characterized by an orchestrated multi-stage process of promoter release and σ release (Murakami and Darst, 2003). After several aborted attempts, the transcription complex successfully transitions to an elongation complex capable of processive transcription, including translocation and simultaneous separation of RNA from the upstream template. While there is evidence of σ bound transcription elongation (Bar-Nahum and Nudler, 2001; Mooney et al., 2005), the transition to processive elongation requires the necessary release of promoter and concomitant conformation changes of σ . Not enough structural basis is available to fully understand the complex steps of the transition from initiation to elongation during abortive transcription. A structure of such a holoenzyme complex has been presented in Chapter 3, which contains a 6-mer RNA and illuminates the initial steps of this transition.

1.7.3 Nucleotide addition cycle in multi-subunit RNAP

At the core of RNA synthesis, lies the central repeating process of the nucleotide addition cycle (NAC), each step of which includes NTP selection and binding, reaction catalysis and translocation. The fundamental chemistry adopts the general two-metal ion catalysis for all polynucleotide synthases and is same as the single-subunit RNAP (Sosunov et al., 2003). On the other hand, other obligatory events of the NAC are

administered by a distinct mechanism in multi-subunit RNAP involving the conserved architecture of the active site, including the trigger loop, bridge helix, DFDGD active site motif, the secondary channel and template DNA. Accumulation of volumes of structural information from transcribing complexes of *Tth* RNAP and Pol II have contributed to the substantial understanding of this process. However, there are still multiple possibilities for some mechanistic steps that sway one way or the other with support from vast amount of kinetic, biophysical and computational studies (Nudler, 2009).

1.7.3.1 NTP entry

The only two routes available for substrate entry from the surface of RNAP into the active site is the main DNA binding channel or the secondary channel (**Figures. 1.13 and 1.14**). Whereas the presence of negatively charged DNA and RNA in the main channel could restrict the free movement of NTPs, the secondary channel is believed to be an attractive pathway. The channel is 20 Å long and 12 Å in diameter, lined with negative electrostatic potential that allows flow of one NTP at a time and may control the rate of NTP diffusion (Batada et al., 2004). Consistent with this model, a secondary NTP binding site, named the entry site (E-site), was observed in a crystal structure of the Pol II elongation complex next to the NTP binding site (Westover et al., 2004a). The E-site is close to the positively charged rim of the secondary channel, and may serve as a regulatory site for transcription activity where the NTP resides longer in the presence of high NTP concentrations to enable substrate selection whereas at low NTP concentrations it might act as a reserve site to constantly feed NTPs. This model is further supported by the competition between NTP binding and a peptide inhibitor MccJ25 binding at the secondary channel (Mukhopadhyay et al., 2004). The alternative theory for NTP entry through the main channel professes that there is a secondary NTP binding site based on observations of two kinetic populations, slow and fast species, of elongating complexes. An allosteric binding site other than the usual *i+1 site*, to the tip of fork loop 2 (near the downstream DNA), was proposed to explain the conversion between these two states (Foster et al., 2001) and thus regulate the transcription rate according to the cellular NTP concentrations. While variable conformations of fork loop 2 in structures of Pol II, *Taq*

and *Tth* RNAPs may support its swinging motion for NTP delivery and forcing translocation (Holmes and Erie, 2003), direct evidence for NTP binding at this site is lacking. An analogous model of NTP paired to downstream template bases $i+2$ and $i+3$ was also proposed (Burton et al., 2005) to define the two coexisting populations as pre- and post-translocated complexes. This model explained how template bound NTP at the $i+2$ and $i+3$ sites pushed forward translocation and accelerated NTP incorporation by the easy flipping of the base-pair. This premise requires the $i+2$ and $i+3$ template DNA to be single-stranded for NTP pairing, which however is not the case in any bacterial RNAP structures. Further, biochemical data suggest that there is only one template base unpaired downstream of the $i+1$ site (Kashkina et al., 2007).

1.7.3.2 Substrate NTP binding sites

The NTP binding step is crucial for substrate selection and transcription fidelity. At the RNAP active site, two temporary NTP binding sites near the catalytically competent NTP insertion site have been observed in transcribing RNAP complex structures, which are considered instrumental in selection of the correct substrate and suggest that the nucleotide binding step is a complex, regulated process involving intermediate steps. One such intermediate is the pre-insertion site that is distinct from but largely overlapping with the insertion site. It was observed with a nucleotide analog (NMPCPP) in a template paired but catalytically incompetent orientation (Kettenberger et al., 2004). A second distal site called the ‘E-site’ at the entry pore of the secondary channel (lined by positive residues). This site was observed in the presence of a mismatched base NTP (Westover et al., 2004a), which as explained above serves as a regulatory site for selection. Whether both of these sites are obligatory in the pathway towards nucleotide binding at the catalytic insertion site is not known. Nevertheless, following this preliminary binding event, NTPs assume a catalytic conformation, which is named as the ‘insertion site’ ($i+1$) wherein it is committed to incorporation.

1.7.3.3 Roles of the trigger loop and the bridge helix in NAC

The change from pre-insertion to insertion site NTP conformation involves a tight synchrony of conformational changes of two critical moving motifs of the β' subunit called the trigger loop and the bridge helix. Based on crystal structures of RNAP elongation complexes, it was proposed that the trigger loop switches its conformations between a disordered loop (Cramer et al., 2001; Gnatt et al., 2001; Kettenberger et al., 2004; Zhang et al., 1999) and a folded helix-turn-helix (Brueckner and Cramer, 2008; Kaplan et al., 2008; Vassylyev et al., 2007b; Wang et al., 2006), coinciding with the absence and presence of the correct incoming nucleotide at the insertion site, respectively. Conversely, after incorporation the translocation of the nascent RNA by one register coincides with the unfolding of the trigger helix to a disordered loop (Malinen et al., 2012).

An unambiguous explanation for the complete translocation mechanism in RNAP is still lacking despite various structures and FRET studies. Several models still exist to account for the translocation step of the NAC. The model with the most support from biochemical and FRET studies is the Brownian ratchet model, which postulates that RNAP, owing to intrinsic kinetic energy, slides up and down the template in Brownian motion and attains forward directionality upon NTP binding. Thus the correct NTP binding serves as a pawl for the sliding RNAP ratchet, hence the name (Bar-Nahum et al., 2005). This binding is assisted by the folding of the trigger loop, secured through direct interactions of its residues with the NTP, which also moves the bridge helix together as a three helix bundle. Although the bridge helix has been observed in straight and bent conformations coinciding with the pre- and post-translocated complexes, its individual role in translocation has not been confirmed. In contrast a concerted movement of the trigger loop and the bridge helix along with NTP binding, or a combined Brownian Ratchet model, maybe proposed.

1.8 Thesis Format

Chapter 2 contains my research on investigating the mechanism of the nucleotidyl-transfer reaction in single-subunit RNAP from bacteriophage N4 using a

novel technique called time-resolved trigger freeze X-ray crystallography (Basu and Murakami, 2013). Viewing enzymes in motion has been an old challenge, especially at meaningful, high, atomic resolution. Tapping the direct atomic view available from X-ray crystallography at a time resolution is an ultimate goal of enzyme mechanistic studies. With this target, I succeeded in establishing a simple soak-trigger-freeze strategy to obtain a time-trace of the reaction events, in the form of high-resolution crystal structures, of a complex enzyme system like RNAP for the first time.

In Chapter 3, I expanded the scope of this dissertation to multi-subunit RNAPs. With the ultimate goal of time-resolved studies on *Tth* initiation complex, I initiated an *in crystallo* transcription experiment with *Tth* RNAP. I was able to obtain a snapshot structure of the pre-catalytic *de novo* transcription complex, which is the first structural evidence for the stabilizing mechanism of *de novo* transcription initiation in multi-subunit RNAPs. Additionally, I present the structure of another initiation complex at the 6-mer transcript stage obtained by *in crystallo* reaction. This structure sheds light on the initial steps of transition from transcription initiation to elongation. This structure is also a setting stage for future time-resolved crystallographic experiments on NAC in multi-subunit RNAP.

Finally in Chapter 4, I will summarize the contribution of my studies and future studies that may be performed using this foundation. I included studies I performed to understand the big picture of N4 transcription, which does not necessarily fit in with the rest of my dissertation. This chapter comprises a structure of the N4 mini-vRNAP with a truncated version of the promoter DNA, which was solved to identify the key structural feature in the promoter element responsible for RNAP activation and secondly, my progress towards the structural determination of the N4 mini-vRNAP transcription elongation complex. In the latter part of Chapter 4, I summarize the contribution of the second part of my thesis studies on the *Tth* initiation complex. Additionally, I will describe the scope of the *in crystallo* transcription system established in Chapter 3 to address the gaps of knowledge in the multi-subunit RNAP NAC and in transcription late-initiation phase.

1.9 References

- Arnold, E., Ding, J., Hughes, S.H., and Hostomsky, Z. (1995). Structures of DNA and RNA polymerases and their interactions with nucleic acid substrates. *Current opinion in structural biology* 5, 27-38.
- Bandwar, R.P., Ma, N., Emanuel, S.A., Anikin, M., Vassylyev, D.G., Patel, S.S., and McAllister, W.T. (2007). The transition to an elongation complex by T7 RNA polymerase is a multistep process. *J Biol Chem* 282, 22879-22886.
- Bar-Nahum, G., Epshtein, V., Ruckenstein, A.E., Rafikov, R., Mustaev, A., and Nudler, E. (2005). A ratchet mechanism of transcription elongation and its control. *Cell* 120, 183-193.
- Bar-Nahum, G., and Nudler, E. (2001). Isolation and characterization of sigma(70)-retaining transcription elongation complexes from Escherichia coli. *Cell* 106, 443-451.
- Basu, R.S., and Murakami, K.S. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J Biol Chem* 288, 3305-3311.
- Batada, N.N., Westover, K.D., Bushnell, D.A., Levitt, M., and Kornberg, R.D. (2004). Diffusion of nucleoside triphosphates and role of the entry site to the RNA polymerase II active center. *Proc Natl Acad Sci U S A* 101, 17361-17364.
- Briebe, L.G., and Sousa, R. (2001). T7 promoter release mediated by DNA scrunching. *The EMBO journal* 20, 6826-6835.
- Brueckner, F., and Cramer, P. (2008). Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol* 15, 811-818.

- Burton, Z.F., Feig, M., Gong, X.Q., Zhang, C., Nedialkov, Y.A., and Xiong, Y. (2005). NTP-driven translocation and regulation of downstream template opening by multi-subunit RNA polymerases. *Biochem Cell Biol* 83, 486-496.
- Campbell, E.A., Muzzin, O., Chlenov, M., Sun, J.L., Olson, C.A., Weinman, O., Trester-Zedlitz, M.L., and Darst, S.A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Molecular cell* 9, 527-539.
- Cermakian, N., Ikeda, T.M., Miramontes, P., Lang, B.F., Gray, M.W., and Cedergren, R. (1997). On the evolution of the single-subunit RNA polymerases. *J Mol Evol* 45, 671-681.
- Chamberlin, M., McGrath, J., and Waskell, L. (1970). New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature* 228, 227-231.
- Cheetham, G.M., Jeruzalmi, D., and Steitz, T.A. (1999). Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature* 399, 80-83.
- Cheetham, G.M., and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* 286, 2305-2309.
- Cheetham, G.M., and Steitz, T.A. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Current opinion in structural biology* 10, 117-123.
- Cramer, P. (2002). Multisubunit RNA polymerases. *Curr Opin Struct Biol* 12, 89-97.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292, 1863-1876.
- Dai, X., Greizerstein, M.B., Nadas-Chinni, K., and Rothman-Denes, L.B. (1997). Supercoil-induced extrusion of a regulatory DNA hairpin. *Proceedings of the National Academy of Sciences of the United States of America* 94, 2174-2179.
- Darst, S.A. (2001). Bacterial RNA polymerase. *Curr Opin Struct Biol* 11, 155-162.

Davydova, E.K., and Rothman-Denes, L.B. (2003). Escherichia coli single-stranded DNA-binding protein mediates template recycling during transcription by bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci U S A* *100*, 9250-9255.

Davydova, E.K., Santangelo, T.J., and Rothman-Denes, L.B. (2007). Bacteriophage N4 virion RNA polymerase interaction with its promoter DNA hairpin. *Proceedings of the National Academy of Sciences of the United States of America* *104*, 7033-7038.

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

Dombroski, A.J., Walter, W.A., and Gross, C.A. (1993). Amino-terminal amino acids modulate sigma-factor DNA-binding activity. *Genes Dev* *7*, 2446-2455.

Durniak, K.J., Bailey, S., and Steitz, T.A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* *322*, 553-557.

Falco, S.C., Laan, K.V., and Rothman-Denes, L.B. (1977). Virion-associated RNA polymerase required for bacteriophage N4 development. *Proceedings of the National Academy of Sciences of the United States of America* *74*, 520-523.

Foster, J.E., Holmes, S.F., and Erie, D.A. (2001). Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell* *106*, 243-252.

Gleghorn, M.L., Davydova, E.K., Basu, R., Rothman-Denes, L.B., and Murakami, K.S. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc Natl Acad Sci U S A* *108*, 3566-3571.

Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B., and Murakami, K.S. (2008). Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* *32*, 707-717.

Glucksmann, M.A., Markiewicz, P., Malone, C., and Rothman-Denes, L.B. (1992). Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. *Cell* *70*, 491-500.

- Glucksmann-Kuis, M.A., Dai, X., Markiewicz, P., and Rothman-Denes, L.B. (1996). E. coli SSB activates N4 virion RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. *Cell* 84, 147-154.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* 292, 1876-1882.
- Goldman, S.R., Ebright, R.H., and Nickels, B.E. (2009). Direct detection of abortive RNA transcripts in vivo. *Science* 324, 927-928.
- Guo, Q., Nayak, D., Brieba, L.G., and Sousa, R. (2005). Major conformational changes during T7RNAP transcription initiation coincide with, and are required for, promoter release. *Journal of molecular biology* 353, 256-270.
- Haynes, L.L., and Rothman-Denes, L.B. (1985). N4 virion RNA polymerase sites of transcription initiation. *Cell* 41, 597-605.
- Hedtke, B., Borner, T., and Weihe, A. (1997). Mitochondrial and chloroplast phage-type RNA polymerases in Arabidopsis. *Science* 277, 809-811.
- Holmes, S.F., and Erie, D.A. (2003). Downstream DNA sequence effects on transcription elongation. Allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem* 278, 35597-35608.
- Huang, J., and Sousa, R. (2000). T7 RNA polymerase elongation complex structure and movement. *Journal of molecular biology* 303, 347-358.
- Jiang, M.Y., and Sheetz, M.P. (1994). Mechanics of myosin motor: force and step size. *Bioessays* 16, 531-532.
- Kaplan, C.D., Larsson, K.M., and Kornberg, R.D. (2008). The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Molecular cell* 30, 547-556.

- Kashkina, E., Anikin, M., Brueckner, F., Lehmann, E., Kochetkov, S.N., McAllister, W.T., Cramer, P., and Temiakov, D. (2007). Multisubunit RNA polymerases melt only a single DNA base pair downstream of the active site. *J Biol Chem* 282, 21578-21582.
- Kazmierczak, K.M., Davydova, E.K., Mustaev, A.A., and Rothman-Denes, L.B. (2002). The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *Embo J* 21, 5815-5823.
- Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Molecular cell* 16, 955-965.
- Lonetto, M., Gribskov, M., and Gross, C.A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. *J Bacteriol* 174, 3843-3849.
- Ma, K., Temiakov, D., Anikin, M., and McAllister, W.T. (2005). Probing conformational changes in T7 RNA polymerase during initiation and termination by using engineered disulfide linkages. *Proceedings of the National Academy of Sciences of the United States of America* 102, 17612-17617.
- Malinen, A.M., Turtola, M., Parthiban, M., Vainonen, L., Johnson, M.S., and Belogurov, G.A. (2012). Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res* 40, 7442-7451.
- Margeat, E., Kapanidis, A.N., Tinnefeld, P., Wang, Y., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2006). Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys J* 90, 1419-1431.
- Markiewicz, P., Malone, C., Chase, J.W., and Rothman-Denes, L.B. (1992). Escherichia coli single-stranded DNA-binding protein is a supercoiled template-dependent transcriptional activator of N4 virion RNA polymerase. *Genes Dev* 6, 2010-2019.
- Martinez-Rucobo, F.W., and Cramer, P. (2013). Structural basis of transcription elongation. *Biochim Biophys Acta* 1829, 9-19.

- Masters, B.S., Stohl, L.L., and Clayton, D.A. (1987). Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. *Cell* *51*, 89-99.
- Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* *108*, 599-614.
- Mooney, R.A., Darst, S.A., and Landick, R. (2005). Sigma and RNA polymerase: an on-again, off-again relationship? *Molecular cell* *20*, 335-345.
- Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R.M., and Ebright, R.H. (2004). Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. *Molecular cell* *14*, 739-751.
- Murakami, K.S. (2013). X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem* *288*, 9126-9134.
- Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* *13*, 31-39.
- Murakami, K.S., Davydova, E.K., and Rothman-Denes, L.B. (2008). X-ray crystal structure of the polymerase domain of the bacteriophage N4 virion RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 5046-5051.
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* *296*, 1285-1290.
- Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* *296*, 1280-1284.

Nudler, E. (2009). RNA polymerase active center: the molecular engine of transcription. *Annu Rev Biochem* 78, 335-361.

Ollis, D.L., Brick, P., Hamlin, R., Xuong, N.G., and Steitz, T.A. (1985). Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* 313, 762-766.

Rothman-Denes, L.B., Haselkorn, R., and Schito, G.C. (1972). Selective shutoff of catabolite-sensitive host syntheses by bacroxyurea pharmino mutation genes coliphages growth I virus replication *Escherichia coli* growth I lysogeny crosses genetic coliteriophage N4. *Virology* 50, 95-102.

Schito, G.C., Giraldi, M., and Toni, M. (1974). Rifampicin-resistant replication of coliphage N4 DNA. *Nature* 251, 340-342.

Sosunov, V., Sosunova, E., Mustaev, A., Bass, I., Nikiforov, V., and Goldfarb, A. (2003). Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase. *Embo J* 22, 2234-2244.

Sousa, R., Chung, Y.J., Rose, J.P., and Wang, B.C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature* 364, 593-599.

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

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* 90, 6498-6502.

Tahirov, T.H., Temiakov, D., Anikin, M., Patlan, V., McAllister, W.T., Vassylyev, D.G., and Yokoyama, S. (2002). Structure of a T7 RNA polymerase elongation complex at 2.9 Å resolution. *Nature* 420, 43-50.

Tang, G.Q., Roy, R., Ha, T., and Patel, S.S. (2008). Transcription initiation in a single-subunit RNA polymerase proceeds through DNA scrunching and rotation of the N-terminal subdomains. *Molecular cell* 30, 567-577.

Temiaikov, D., Montesana, P.E., Ma, K., Mustaev, A., Borukhov, S., and McAllister, W.T. (2000). The specificity loop of T7 RNA polymerase interacts first with the promoter and then with the elongating transcript, suggesting a mechanism for promoter clearance. *Proceedings of the National Academy of Sciences of the United States of America* 97, 14109-14114.

Temiaikov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassilyev, D.G. (2004). Structural basis for substrate selection by t7 RNA polymerase. *Cell* 116, 381-391.

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

Vassilyev, D.G., Vassilyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448, 157-162.

Vassilyev, D.G., Vassilyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007b). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163-168.

Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127, 941-954.

Werner, F., and Grohmann, D. (2011). Evolution of multisubunit RNA polymerases in the three domains of life. *Nat Rev Microbiol* 9, 85-98.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119, 481-489.

Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* 303, 1014-1016.

Yin, Y.W., and Steitz, T.A. (2002). Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* 298, 1387-1395.

Yin, Y.W., and Steitz, T.A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* 116, 393-404.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* 98, 811-824.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* 338, 1076-1080.

Chapter 2

Watching the bacteriophage N4 RNA polymerase transcription by time-resolved trigger-freeze X-ray crystallography

Publications from the work in this Chapter are:

Chen, Y., **Basu, R.**, Gleghorn, M.L., Murakami, K.S., and Carey, P.R. (2011). Time-resolved events on the reaction pathway of transcript initiation by a single-subunit RNA polymerase: Raman crystallographic evidence. **J Am Chem Soc** *133*, 12544-12555.

Basu, R.S., and Murakami, K.S. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. **J Biol Chem** *288*, 3305-3311.

2.1 Abstract

The challenge for structural biology is to understand atomic-level macromolecular motions during enzymatic reaction (Bourgeois and Royant, 2005; Henzler-Wildman and Kern, 2007; Schmidt, 2008). X-ray crystallography can reveal high-resolution structures; however, one perceived limitation is that it reveals only static views. In this chapter I present X-ray crystal structures of real-time, transient intermediates of the transcription pathway using time-dependent trigger-freeze X-ray crystallography. Although time-resolved structural studies is a well explored field in crystallography, it has been used with limited success only in special enzymes, mostly involving specialized equipment. I was able to successfully establish kinetic crystallography of the complex enzymes like nucleotidyl transferases using a simple strategy of soak-trigger-freeze crystallography. This method involves soaking nucleotide and divalent metal into the bacteriophage RNAP-promoter DNA complex crystals to trigger the nucleotidyl transfer reaction and freezing crystals at different time points. In each time-resolved crystal structure, different intensities and shapes of electron density maps corresponding to nucleotides and metals were revealed at the RNAP active site, resulting in the visualization of nucleotide and metal binding and the natural mechanism of phosphodiester bond formation in a time perspective. This study thus provides the temporal order of substrate assembly and metal co-factor binding at the active site of the enzyme and completes our understanding of the two-metal-ion mechanism and fidelity mechanism in single-subunit RNAPs. The nucleotide binding metal (Me^{B}) is coordinated at the active site prior to the catalytic metal (Me^{A}). Me^{A} coordination is only temporary and is established just before and dissociated immediately after phosphodiester bond formation. I exploited the slow enzymatic reaction *in crystallo* to capture these elusive intermediates. These results demonstrate that the simple time-dependent trigger-freeze X-ray crystallography offers a direct means for monitoring enzymatic reactions.

2.2 Introduction

2.2.1 Nucleotidyl-transfer reaction

The phosphoryl transfer reactions including nucleotidyl transfer play a fundamental role in genome maintenance and gene expression (Lassila et al., 2011; Yang et al., 2006b). The nucleotidyl transferases, DNAP and RNAP, catalyze the nucleophilic attack of the 3'-O oxyanion of the primer terminus onto the 5'- α -phosphate of the incoming nucleotide using the two metal catalysis mechanism (Steitz, 1998). These enzymes are characterized by both versatility and specificity of substrate utilization – DNAP uses four different deoxynucleotide triphosphates and RNAP uses four different nucleotide triphosphates – combined with DNA sequence-dependent correct nucleotide addition (Johnson, 2010; Joyce and Benkovic, 2004).

Stopped-flow kinetics (Bermek et al., 2011b; Johnson, 2010; Joyce and Benkovic, 2004; Joyce et al., 2008) and X-ray crystallographic studies of the A-family polymerases, including the *E. coli* DNAP I Klenow fragment (Johnson et al., 2003; Li et al., 1998) and bacteriophage DNAP (Doublie et al., 1998) and RNAP (Gleghorn et al., 2011a; Yin and Steitz, 2004), have elucidated the kinetic scheme of the nucleotide addition cycle and provided structural perspectives on nucleotide selection and incorporation. These studies proposed the presence of at least two prechemistry steps as the nucleotide selectivity checkpoints, including conformational change of the fingers O-helix and another unidentified step preceding phosphodiester bond formation (Johnson, 2010; Joyce and Benkovic, 2004). Thus the proposed reaction scheme consisted of early conformational changes consisting of template rearrangement and 'closing of fingers', followed by tighter NTP and metal binding in an orientation amenable for reaction. However, the unambiguous definition of the mechanism and the temporal order of the steps therein, awaits direct evidence.

X-ray crystal structures provide detailed architecture of the reaction intermediates, shedding light into the nucleotidyl transfer reaction pathway. One of the strategies for trapping intermediates of nucleic acid polymerases is by using substrate analogs (**Figure 2.1**) to prevent forward reaction beyond substrate binding. Ternary complexes of DNA-bound polymerases with a 3'-deoxynucleotide analogue (Doublié et al., 1998; Johnson et al., 2003; Kennedy et al., 2007; Li et al., 1998), which lacks the essential O3' nucleophile for phosphodiester bond formation, have revealed the 'closed' state of the O-helix of the fingers subdomain. This closed state is due to nucleotide binding and it was proposed that the DNA rearrangement plus the closing of O-helix is the early fidelity checkpoint of nucleotide selection (Joyce et al., 2004). However, the 3'-deoxynucleotide analogue is unable to coordinate the catalytic metal (Me^A), which results in misalignment of the reactive groups for phosphodiester bond formation. Another commonly used substrate analogue is the non-reactive guanosine 5'-[(α , β)-methylene] triphosphate (GMPCPP) (**Figure 2.1**), which allows the coordination of the critical Me^A but prevents the forward reaction due to the non-hydrolysable methylene substitution on its triphosphate. Using this analog, intermediates in the T7 RNAP elongation and bacteriophage N4 RNAP initiation pathways were trapped, which helped analyze the pre-chemistry events (Gleghorn et al., 2011a; Temiakov et al., 2004; Yin and Steitz, 2004). These structures showed that substrate and metal binding induced the 'closing' of the fingers domain to engage basic residues on the O-helix to interact with the triphosphates of the incoming nucleotide, and thus provided the structural basis of the primary substrate selection mechanism before binding to the N-site.

Former studies from our lab by Dr. Michael Gleghorn also used GMPCPP to trap an intermediate of transcription initiation, called substrate complex II (SCII). SCII contained all atoms at the active site, including Me^A , Me^B and reactive groups of nucleotides, poised for the nucleotidyl transfer reaction (Gleghorn et al., 2011a). In addition, structures of other intermediates from this study included a pre-chemistry substrate complex (SCI) and a post-chemistry 2-mer product complex (PC). Together, these structures provided a complete set of snapshots during the transcript initiation process from binding of the nucleotide and metals to the phosphodiester bond formation,

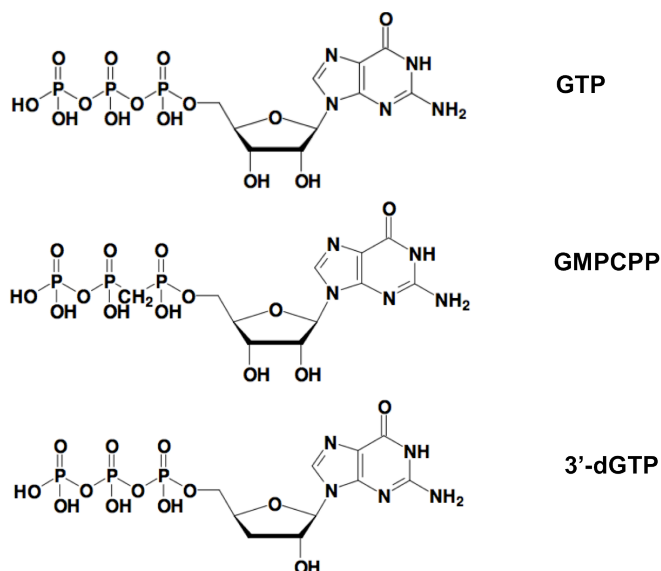


Figure 2.1 Chemical structures of a nucleotide and its non-reactive analogues used for structural studies of polymerases. Figures taken from Jena Biosciences.

based on which a mechanism was proposed (**Figure 2.12**) (described in **Section 1.6.3**). Accordingly, transition from a binary RNAP-promoter DNA complex (BC) to the ternary SCI primarily involved ‘closing’ of O-helix and template DNA rearrangement. Comparison of the two pre-catalytic intermediates SCI and SCII identified critical realignment of the substrates to a catalysis-competent conformation brought about by the addition of catalytic Me^A in SCII. We thus proposed that the Me^A binding is the last step before phosphodiester bond formation. The same conclusion was also proposed by the stopped-flow kinetics of DNAP I (Bermek et al., 2011b). However, there is no direct evidence to confirm this proposed mechanism, which led us to monitor the temporal order of events in the nucleotidyl-transfer reaction.

A two-pronged structure-based kinetics approach was adopted to address this question concerning the nucleotidyl transfer reaction mechanism. Clearly, the previous high-resolution structural studies lacked the time-resolution, which is critical to understand the complete mechanistic pathway of the reaction. The two kinetic approaches used were time-resolved Raman crystallography and time-resolved trigger-freeze X-ray crystallography. The use of both spectroscopic and diffraction techniques on the same

enzyme environment was expected to correlate vibrational transitions to structural transitions, thus eliminating any ambiguity in the reaction mechanism.

2.2.2 Time-resolved Raman crystallography

The kinetic scheme of N4 mini-vRNAP transcription initiation *in crystallo* was investigated by Raman crystallography in our collaborator Prof. Paul Carey's lab in Case Western Reserve University (Chen et al., 2011). Raman crystallography is a technique to record spectroscopy signals arising from a single crystal using Raman microscopy (**Figure 2.2**). Thus, this method captures the Raman scattering events from proteins, nucleic acids and/or nucleotides in the crystal within its focal volume, which if perturbed by the entry of a reaction substrate, is able to monitor enzymatic reactions in terms of the ensuing protein conformational changes, substrate/cofactor binding and/or product formation (Carey, 2006). Raman signals arise from the bond vibrational modes of molecules, and owing to its sensitivity to molecular geometry it has classically been used as an important technique for identification and characterization of molecules and reactions. However, the most advantageous inferences come from measurements of Raman difference spectra, which are able to monitor the smallest perturbations in bond-lengths of molecular species when bound by a ligand or en route a reaction. It is therefore possible to trace events of a reaction in real-time by measuring the difference spectra at successive time-points. This advantage has been exploited in Raman crystallography, wherein changes in an enzyme crystal may be monitored by the simple soaking-in of substrates/cofactors. Moreover, the protein-enriched crystals provide a higher signal-to-noise ratio than in solution (Carey, 2006). The only limitation being that it requires roughly 1 min to acquire a single spectrum, and thus only slow reactions can be monitored.

Raman spectral assignments of enzyme signals are primarily based on the existing Raman literature on small molecules, proteins and nucleic acids. To isolate signals coming from different moieties of the molecules, isotope labeled analogues (*eg.* ^{13}C and/or ^{15}N labeling) may be used for ligands or nucleic acids. However, the lack of site-

directed ‘isotope-editing’ of protein residues may be compensated by guiding information from X-ray crystal structures of enzymes/complexes. The presence of X-ray snapshot structures in the course of a reaction highlight the conformational changes that occur, which serve as a guide for the reliable assignment of spectral changes to individual structural features or even residue side chains (Carey et al., 2011). In most cases, Raman crystallography serves as a complementary technique for enzyme mechanism studies because it adds the missing time-resolution to the rich geometric and architectural information from X-ray crystallography snapshot studies. In this study, I use the temporal information from Raman crystallography to guide my systematic time-resolved X-ray structural study of the N4 phage mini-vRNAP transcription.

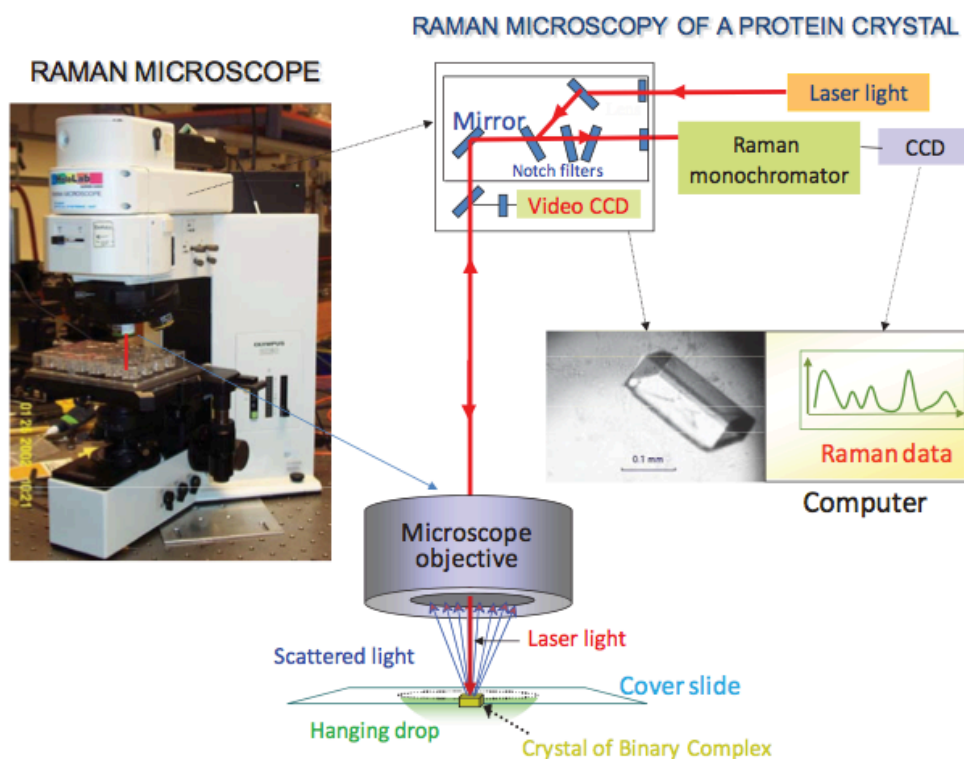


Figure 2.2: Basic setup of the Raman microscope. Both the actual image and schematic representation have been shown. Figure adapted from (Carey, 2006)

2.2.3 Time-resolved soak-trigger-freeze X-ray crystallography

Time-resolved structure determination techniques were developed to film biological processes at atomic or near-atomic resolutions (Bourgeois and Royant, 2005; Hajdu et al., 2000; Henzler-Wildman and Kern, 2007; Schmidt, 2008; Westenhoff et al., 2010). Nuclear magnetic resonance (NMR) spectroscopy has a clear advantage and is able to reveal the time-scale of protein motions with atomic resolution structure (Fraser et al., 2009; Henzler-Wildman et al., 2007). However, NMR is only applicable to only well-behaved, small proteins (upto about 10 kDa) and is not able to monitor the reaction over a time-course. The requirement of X-ray crystallography is the uniform arrangement of molecules throughout the crystal, which poses a demand for uniform triggering of the reaction in kinetic crystallography. Kinetic X-ray crystallography involving Laue diffraction, which is able to collect whole crystallographic data in a single diffraction frame, is able to deliver unprecedented detail of enzymatic reaction (Bourgeois and Royant, 2005; Schmidt, 2008; Schotte et al., 2003); however, it demands on ultrafast triggering of the reaction, largely restricting the application to light-induced reactions. However, most biological reactions are triggered by diffusion and kinetic X-ray crystallography has so far only been successful in tracing simple transitions in small enzymes (Hajdu et al., 2000). Complex reactions such as DNA replication and transcription have been excluded from the scope of time-resolved structural studies. Guided by Raman crystallography data, I was convinced of a slow *in crystallo* transcription reaction and hence attempted to capture structures of reaction intermediates. A simplistic time-resolved soak-trigger-freeze approach was used to build a three dimensional movie of the transcription reaction pathway at atomic resolution, ambient temperatures and without any modifications of enzyme or substrates. Moreover, this approach did not require any specific equipment for X-ray data collection. Hence, this approach promises to be applicable to other enzyme systems and may become a general method to directly visualize biological reactions (Freudenthal et al., 2013; Nakamura et al., 2012).

2.3 Experimental Summary

Transcription initiation complexes were prepared as follows: Binary complex (BC) crystals (Gleghorn et al., 2011a) were sequentially transferred through cryo-protectant solutions and finally soaked in a solution containing 5 mM GTP, 5 mM ATP and 10 mM MnCl₂ in the final cryo-solution (details in **Section 2.5.2**). Crystals were harvested from the soaking solution at specific time points and flash frozen in liquid nitrogen (**Figure 2.3**). For Raman crystallography, crystals were focused under the microscope in the drops they grew in, followed by injection of the substrate mix to trigger the reaction, followed by spectra collection at specific time-points.

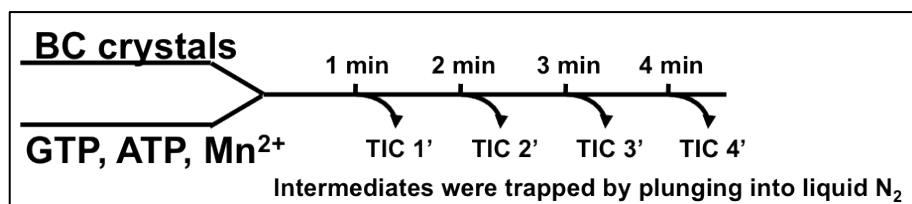


Figure 2.3 Strategy for time-dependent soak-trigger-freeze experiment.

Final concentration, c , of the N4 phage polymerase-DNA complex in a single crystal was calculated using the equation

$$c = n / (v \times N_A)$$

where n is the number of transcription complex molecules per unit cell of the crystal, v is the volume of a unit cell, and N_A is Avogadro's number. The BC crystals contain two RNAP-DNA complexes per asymmetric unit and in the **P2₁2₁2₁** space group each unit cell has 4 asymmetric units. Thus, there are $n=8$ molecules per unit cell. From the unit cell parameters (**82 Å x 112 Å x 276 Å**), v is calculated to be **2.54×10^{-12} nl**. Therefore, the concentration of the complex in the crystal (c) is **5.1 mM**. The volume, V , of a typical crystal was calculated from its dimensions (**25 μm x 100 μm x 125 μm**) = **0.3125×10^6 μm³ = 0.31 nl**.

Calculations also indicated that highly concentrated RNAP-DNA complex (5 mM) in the small volume of crystals (0.3 nl) allowed complete diffusion of nucleotide-

metals to all binary complexes in the crystal before the reaction was triggered. This unique and ideal environment of RNAP *in crystallo* could synchronize the reaction of all molecules and enable the capture of some hitherto undetected and unstable intermediates by freezing crystals at different times during the reaction.

2.4 Materials and Methods

2.4.1 Purification of N4 mini-vRNAP

N-terminal His₆-tag mini-vRNAP was overexpressed in *E. coli* BL21 cells transformed with the expression plasmid pKMK56 (Kazmierczak et al., 2002), which contains an ampicillin resistance gene and an arabinose inducible promoter. Cells were grown in 1 L LB media containing 100 µg/ml ampicillin at 37°C to an A₆₀₀ of ~0.5. Protein expression was induced by adding 20% arabinose to a final concentration of 0.2% and incubated for 2 hrs. Cells were harvested and stored at -80°C. Cells were suspended in sonication buffer (20 mM Tris-HCl, pH 8.0 at 4°C, 0.5 M NaCl, 5 mM imidazole, 5% glycerol and 1 mM PMSF) and lysed by sonication on ice. The lysed cells were spun at 16,000 rpm at 4°C for 30 min. The supernatant was passed through a 5 ml HiTrap chelating column (GE Healthcare) charged with Ni²⁺. The column was washed with 25 ml of sonication buffer, 25 ml wash buffer (20 mM Tris-HCl, pH 8.0 at 4°C, 50 mM NaCl, 10 mM imidazole, 5% glycerol and 1 mM PMSF) and finally His₆-tagged mini-vRNAP was eluted with 25 ml elution buffer (20 mM Tris-HCl, pH 8.0 at 4°C, 50 mM NaCl, 200 mM imidazole and 5% glycerol). The mini-vRNAP in the elution fraction was further purified by chromatography using a 5 ml Q-sepharose column (GE Healthcare) equilibrated with TGED (10 mM Tris-HCl pH 8.0 at 4°C, 5% glycerol, 0.1 mM EDTA and 1 mM DTT) using a linear gradient of 0.1 to 0.6 M NaCl (**Figure 2.4 A**). The proteins were further purified by gel-filtration column chromatography (Superdex200, GE Healthcare) equilibrated with TGED + 0.2 M NaCl (**Figure 2.4 B**). About 3 mg protein could be obtained per liter of culture. The peak fractions were pooled and glycerol was added to a final concentration of 15%. 1 ml aliquots were flash-frozen in liquid N₂ and stored at -80°C.

2.4.2 Purification of hairpin DNA

DNA oligonucleotides were purchased from Integrated DNA Technologies (1 μ M scale). The lyophilized DNA was dissolved in 40 μ l 5X TBE and 360 μ l formamide and heated to 95°C for 5 mins. DNA was purified by fractionation through a 15% denaturing polyacrylamide gel (29:1 acrylamide:bis-acrylamide) containing 8M Urea. The gel was electrophoresed with 1X TBE at 70 W power. UV shadowing was used to mark the DNA band on the gel. The top 2/3rd of the band was cut into small pieces to avoid any truncated oligos from the bottom of the band. DNA was eluted from these gel pieces in three cycles of soaking in NANOpure water, mixing for 12 hours, brief centrifugation and storing the supernatants at 4°C. The DNA solution was dialyzed in a 1000 MWCO membrane against NANOpure water for 24 hrs, filtered through 0.2 μ m filter and lyophilized by speedvac at 60°C. To form the hairpin, the DNA powder was dissolved in annealing solution (5 mM Na-cacodylate, pH 7.4, 0.5 mM EDTA, and 0.2 M NaCl), heated to 95°C for 5 min, followed by cooling to 25°C at a rate of 0.01°C /s. Annealed oligos (~1 mM) were stored at -20°C.

2.4.3 Crystallization of Binary Complex

N4 mini-vRNAP frozen fractions were thawed and their buffer was exchanged with crystallization buffer (10 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.1 mM EDTA) through a 10K MWCO Vivaspin centrifugal concentrators at 4°C at 7000 rpm and concentrated to ~20 mg/ml as measured by a Bradford assay and stored at 4°C until use. BC was formed by mixing mini-vRNAP and DNA (1:1 molar ratio) in the crystallization buffer to a final mini-vRNAP concentration of 10 mg/ml. Equal volume of BC (1 μ l) and crystallization solution (25% PEG 3350, 0.1 M Tris-HCl pH 8.0, 0.21 M dibasic ammonium citrate, 0.09 M ammonium dihydrogen phosphate) were mixed on a glass cover slide and used to setup a hanging drop vapor diffusion crystal tray over a reservoir of the same crystallization solution.

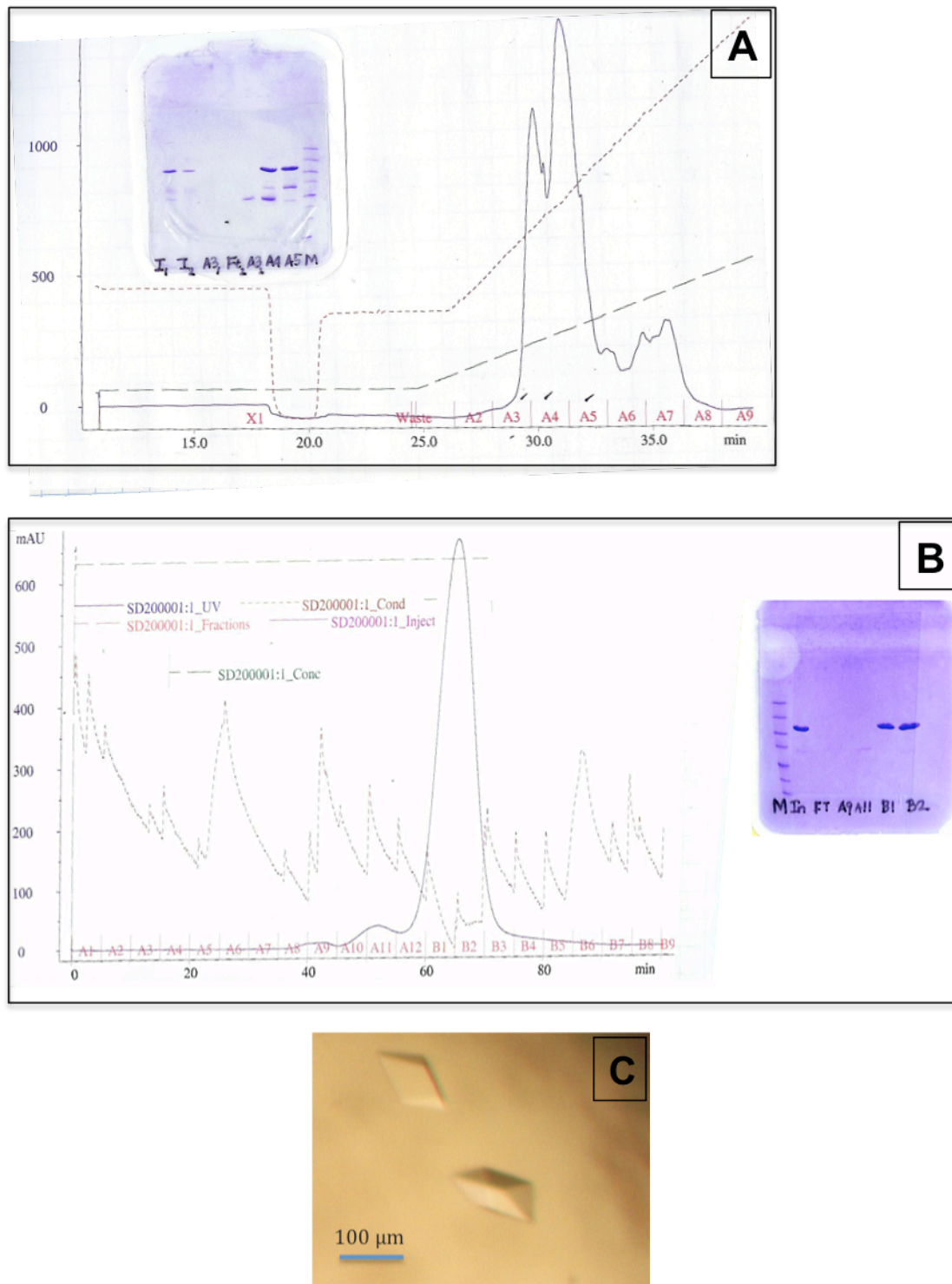


Figure 2.4 Purification of N4 mini-vRNAP and crystallization of BC. UV₂₈₀ chromatograms and Coomassie brilliant blue stained SDS-polyacrylamide gels of fractions from Q-Sepharose column chromatography **(A)** and Gel Filtration SD 200 column chromatography **(B)** are shown. 10 mg/ml N4 mini-vRNAP mixed with 1:1 molar ratio of promoter DNA to form BC, that was crystallized in 25% PEG 3350, 0.21 M Ammonium dibasic citrate, 0.09 M Ammonium dihydrogen phosphate and Tris @pH 8.0 (Gleghorn et al., 2008b) **(C)**.

2.4.4 Soak-trigger freeze X-ray crystallography and structure determination

Transcript initiation complexes were prepared as follows: BC crystals were sequentially transferred through cryo-protectant solutions containing the crystallization solution with increasing concentrations of PEG 3350 (30, 35 and 40 %). Then crystals were transferred to the soak solution containing 5 mM GTP, 5 mM ATP and 10 mM MnCl_2 in addition to the final cryo-solution composition, unless otherwise indicated (TIC_{Mg} and TIC-PPi). Crystals were harvested from the soaking solution at specific time points and flash frozen in liquid nitrogen. Crystallographic data were processed with HKL2000 (Otwinowski and Minor, 1997b) (**Table 2.1**). The BC structure (Gleghorn et al., 2008b) was used as an initial model for the rigid body and TLS refinements by using Phenix (Adams et al., 2010b). Nucleotides were modeled into the respective $F_o - F_c$ electron density maps by Coot (Emsley and Cowtan, 2004). Final coordinates and structure factors were submitted to the PDB with their IDs listed in **Table 2.1**.

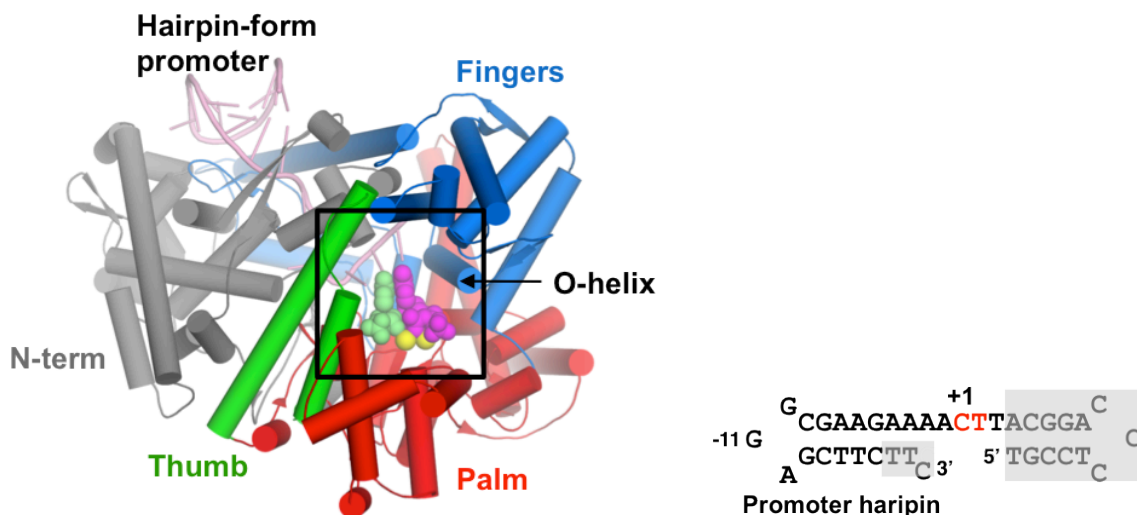
Table 2.1. Data collection and refinement statistics of the N4 mini-vRNAP transcript initiation complexes

Complex PDB code	TIC-1' 4FF1	TIC-2' 4FF2	TIC-3' 4FF3	TIC-4' 4FF4
Data collection				
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions				
<i>a</i> (Å)	82.163	81.908	82.019	82.398
<i>b</i> (Å)	111.500	111.437	111.652	111.779
<i>c</i> (Å)	277.103	275.760	276.145	276.894
Resolution (Å)	50-2.46	50-2.00	50-2.00	50-2.00
<i>R</i> _{sym}	0.116 (0.426)*	0.186 (0.806)*	0.105 (0.546)*	0.127 (0.583)*
<i>I</i> / σ	15.6 (2.9)*	7.4 (1.2)*	15.8 (2.6)*	11.2 (1.4)*
Completeness (%)	95.4 (93.0)*	90.3 (75.4)*	89.0 (79.0)*	95.1 (84.6)*
Redundancy	5.0 (3.6)*	4.4 (3.4)*	3.8 (3.7)*	3.3 (2.4)*
Refinement				
Resolution (Å)	47.8-2.46	47.7-2.00	42.4-2.00	43.9-2.03
No. reflections	87,771	154,842	153,225	156,753
<i>R</i> _{work} / <i>R</i> _{free}	15.3/21.1	20.8/25.5	18.4/21.6	19.1/23.9
R.m.s deviations				
Bond length (Å)	0.007	0.004	0.004	0.006
Bond angles (°)	1.075	0.831	0.820	0.929

Data sets were collected at MacCHESS-F1, Ithaca, NY

*Highest resolution shell is shown in parenthesis

2.5 Results



2.5.1 Time trace of the *in crystallo* transcription by Raman Crystallography

In this study BC crystals of N4 mini-vRNAP-promoter DNA complex were focused on the Raman microscope stage to collect an initial spectra (BC). Isotope (C^{13} and N^{15}) labeled NTP substrates (A^* and G^*) and metal (Mn^{2+}) were injected to the drop containing crystals to trigger the transcription initiation reaction.

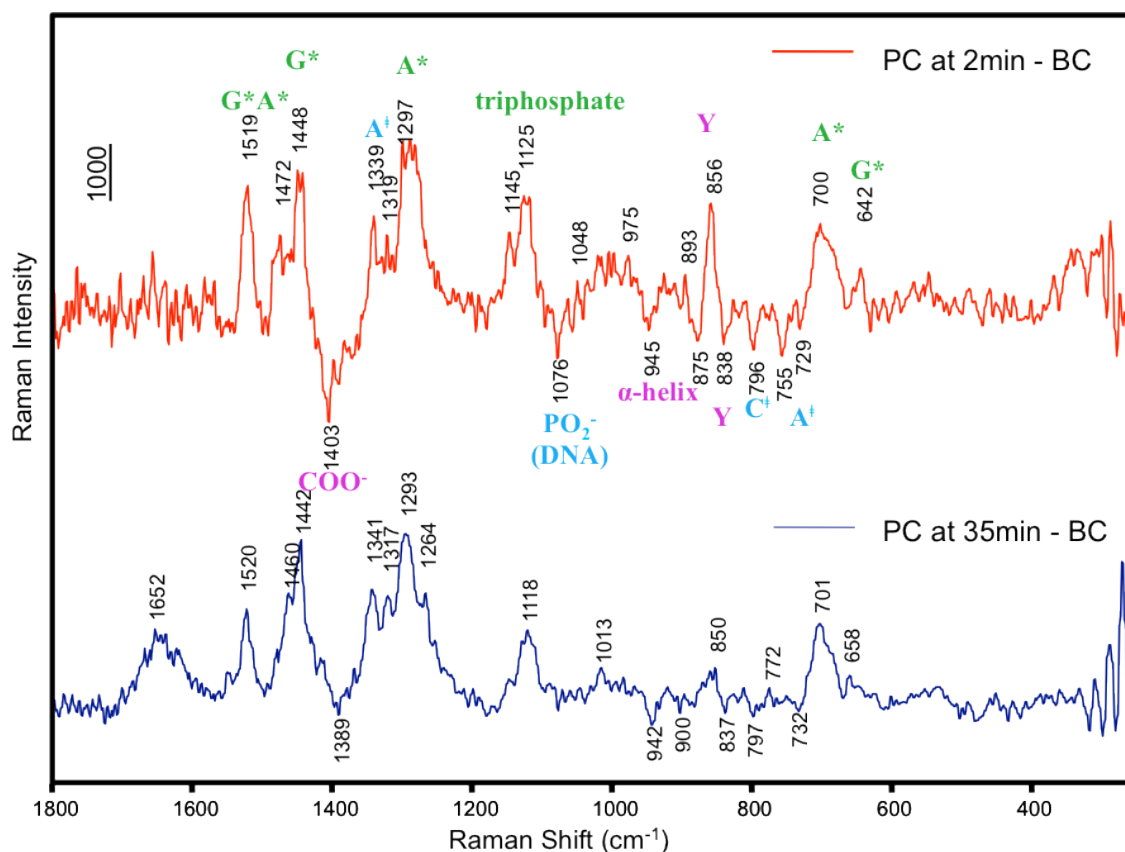


Figure 2.6. Representative Raman spectra of PC formation. Subtracted spectra of PC at 2 mins – BC spectra (red) and subtracted spectra of PC at 35 mins – BC spectra (blue). Peaks assignments are marked in blue for DNA, green for substrate NTP and magenta for protein conformation changes.

After allowing a 1 min time window for equilibration of the substrates into the environment, PC spectra were collected at the 2nd min and every 5 min thereafter. Difference spectra PC (x min) – BC are calculated to highlight the features that arose with the progress of reaction over ‘x’ min. Two representative difference spectra are shown in **Figure 2.6**. In the busy spectral window of 250 to 1800 cm⁻¹, the signals significant to RNAP-DNA complex and the nucleotidyl-transfer reaction were assigned primarily from previous assignments in the literature and narrowed down to specific motifs, residues and cofactors based on the high-resolution snapshot structures of the N4 mini vRNAP transcription intermediates (Gleghorn et al., 2011a). Following are the spectral modes of interest (**Figure 2.6**):

- i) α -helix conformation changes –

- a. a negative feature at 945 cm^{-1} due to an α -helix conformation was used to monitor the ‘closing of O-helix’ as the crystal structure of BC and SCs show that to be the only helical movement.
 - b. A tyrosine doublet at 837 and 853 cm^{-1} was narrowed down to Y678 and Y612 which appear to be the only residues changing their H-bonding patterns during the reaction.
- ii) DNA backbone perturbations were assigned to a Raman feature at 1076 cm^{-1} .
- iii) NTP substrate signals were separated from the template DNA signals with the use of isotope labeled NTPs. A reference spectrum of the substrate alone was used to assign these two signals and also probe the difference.
 - a. Raman feature around 1125 cm^{-1} is assigned to substrate triphosphate only, and its intensity was used to monitor the phospho-diester bond formation.
 - b. Ring modes from A^* and G^*+A^* at 703 cm^{-1} and 1519 cm^{-1} , respectively, were used to quantitate the NTP population.
- iv) Metal coordination signals – The negative feature at 1400 cm^{-1} is attributed to free carboxylic ligands on acidic residues, which shifts upon metal binding due to the formation of the $\text{COO}^- \text{-Mg}^{2+}$ coordination. This change is again totally attributed to D559 and D951, which are the only metal coordinating carboxylates.

The above signal intensities were normalized against an internal standard phenylalanine ring mode at the 1004 cm^{-1} . The intensities were then plotted against time to give a summary plot of the events during reaction **Figure 2.7**. The results revealed that the DNA rearrangement and the O-helix closure are completed immediately after soaking in of nucleotides. Although metal binding starts soon after, their complete coordination occurred at a later stage of the reaction, which coincides with the phosphodiester bond formation, marked by the intensity decay of the triphosphate signal (Chen et al., 2011).

These data revealed that the enzyme remained active *in crystallo* and that the single-turnover nucleotidyl transfer reaction *in crystallo* proceeded substantially slower

(seconds to minutes) compared to the reaction in solution (Gleghorn et al., 2011a). These results provided guidance for designing our soak-trigger-freeze experiments for X-ray crystallography and provided confidence of uniform initiation of reaction in crystals, which is critical for accumulating synchronized intermediates for X-ray structural analyses.

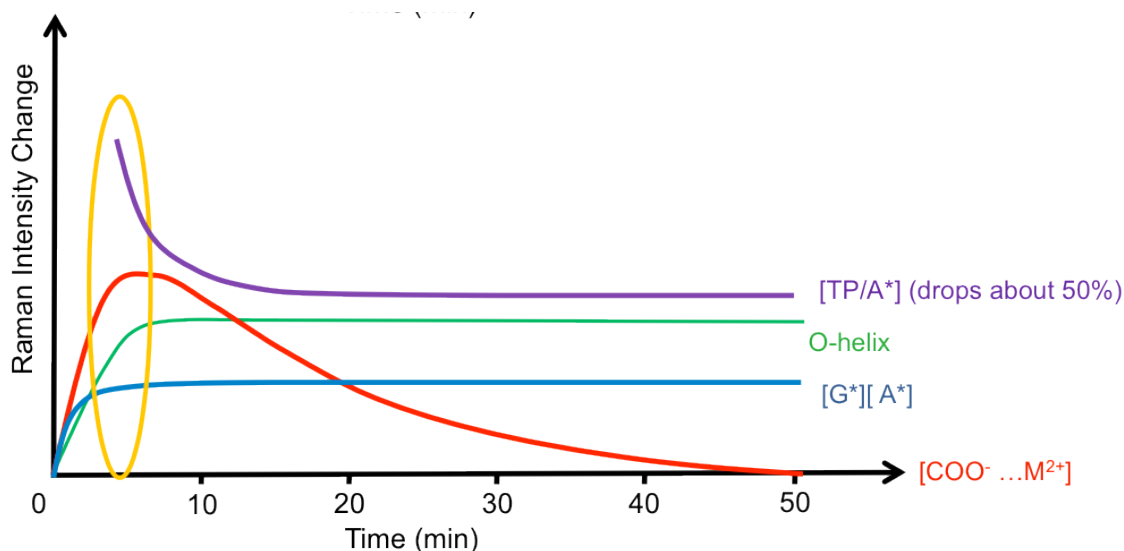


Figure 2.7. Summary of Raman Intensity changes vs time upon PC formation. Time traces of signal intensities are shown for substrate triphosphate (purple), O-helix movements (green), Substrates ATP and GTP binding (blue) and metal coordination of the carboxylate groups of Aspartates (red).

2.5.2 Structures of transcription intermediates using soak-trigger-freeze X-ray crystallography

The BC crystal contains N4 mini-vRNAP bound to its promoter DNA (P2_7c, 36 bases, **Figure 2.5 B**), which has the template DNA bases of C(+1) and T(+2). Addition of GTP and ATP plus divalent cations allows formation of the transcript initiation complex (TIC) accommodating GTP at +1 position (P-site) and ATP at +2 position (N-site) that, upon phosphodiester bond formation, yields the 5'-pppGpA-3' 2-mer RNA and pyrophosphate. The BC and all TIC structures resemble a “cupped right hand” (**Figure 2.5 A**) and the active site located at the bottom of this cup is not involved in crystal packing. The enzyme is active *in crystallo* (Gleghorn et al., 2011a) and the crystal packing prevents a

next round of reaction, which provides an ideal environment for observing the single-turnover reaction. The intermediates were trapped by freezing crystals at different time points – from 1 to 4 min (**Figure 2.3**) after transfer of the BC crystals to the soaking solution containing 5 mM GTP, 5 mM ATP and 10 mM Mn^{2+} . The four structures are indicated as **TIC-1'**, **TIC-2'**, **TIC-3'** and **TIC-4'**. The TIC structures were solved by rigid-body refinement followed by TLS (Translation/Libration/Screw) refinement using the BC structure as the initial model. TLS refinement accommodates flexibility in a structure by partitioning individual chains into multiple segments that are modeled as rigid bodies undergoing vibrational motion. In each TIC structure, different intensities and shapes of F_o-F_c electron density maps, corresponding to nucleotides and metals, were revealed at their active sites. These unbiased F_o-F_c maps allowed true evaluation of the occupancies of nucleotides and metals and the phosphodiester bond formation in the transient intermediates along the course of reaction. In addition, the mobility of individual atoms could be scored from the atomic displacement factor (B-factor).

After 1 min of soaking (TIC-1', **Figure 2.8 A**), the active site shows a discontinuous F_o-F_c map corresponding to GTP, ATP and faint density for Me^B , indicating the presence of nucleotides and metal with partial occupancies. However, the O-helix has almost completed the conformational change and residue Y678 in the helix has swung back to make a space for ATP(+2) binding (**Figure 2.8 B**). The electron density maps for Y678 and DNA template from -1 to +2 positions are not solid (**Figure 2.8 C-b**) that were also observed in the BC structure (**Figure 2.8 C**) and the B-factor of the Y678 side-chain is substantially high (59.5) suggesting ongoing conformation changes of the O-helix. Though the 'closing' of fingers is a common early conformational change seen in all A family polymerases upon substrate binding, the extent of movements is not the same in all. The large rotation of the O-helix of the fingers is more modest in T7 RNAP (Cheetham and Steitz, 1999; Yin and Steitz, 2004) with respect to DNAP I (Doublié et al., 1998; Li et al., 1998) and further subtle in the N4 mini vRNAP (Gleghorn et al., 2011a).

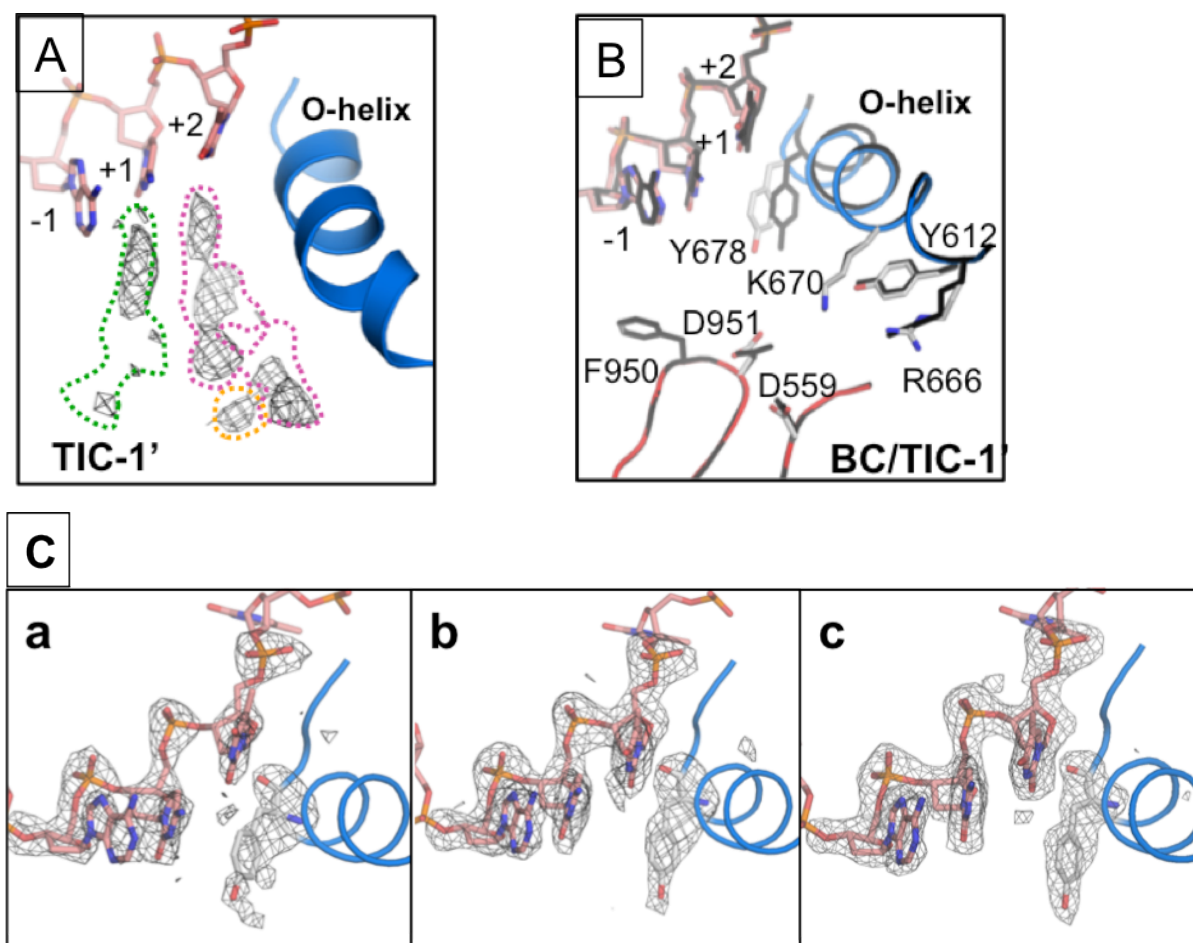
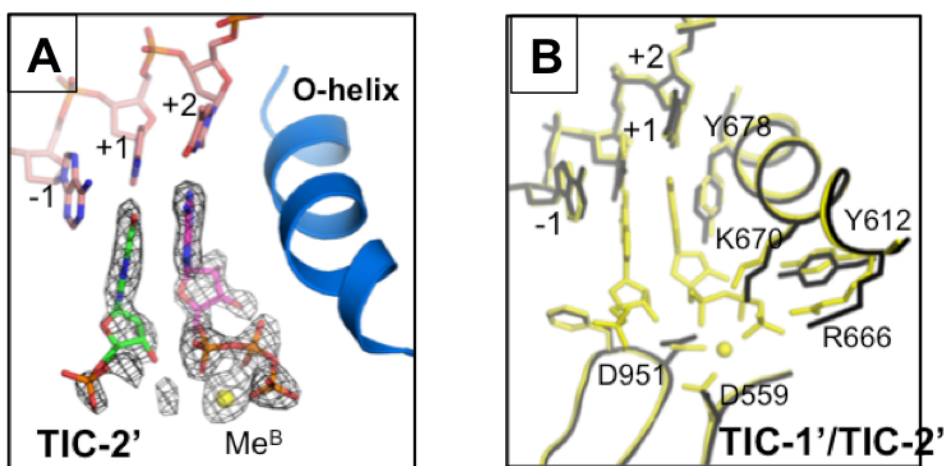


Figure 2.8 Structural analysis of the TIC-1' active site and comparison to other intermediates. The template DNA (pink), O-helix (blue), motifs A and C (red) and important active site residues (white) are shown. **(A)** Electron density (F_o-F_c , sigma cutoff = 4) maps showing nucleotides found in the TIC-1' structure. These maps were calculated using the native amplitudes and the phase derived from the final structures without nucleotides, metals. Positions of GTP(+1), ATP(+2) and Me^B binding sites, which are based on the TIC-3' structure are indicated by green, magenta and yellow dashed lines, respectively **(B)** Conformational changes of RNAP and DNA is shown by Superposition of the BC (black) vs TIC-1' (color). **(C)** Structural transition of template -1 to +2 positions and Y678 are shown. F_o-F_c maps (black net, sigma cutoff = 3) are superimposed on the final models (sticks and tubes) of BC **(a)**, TIC-1' **(b)** and TIC-2' **(c)**.

At 2 min (TIC-2', **Figure 2.9 A-B**), the electron density maps for Y678 and DNA template were well ordered (**Figure 2.8 C-c**) and the Y678 side-chain B-factor was substantially low (22.6), indicating completion of the O-helix movement (**Figure 2.9 B**). Three fingers residues – Y612, R666 and K670 – that play a role in the +2 nucleotide binding also changed their positions (**Figure 2.9 B**). Concomitantly, continuous/complete F_o-F_c maps corresponding to GTP(+1) and ATP(+2) were observed. The electron densities for the β and γ phosphate groups of GTP(+1) were not well resolved due to their

flexibility; thereby GMP was modeled at the +1 position. In addition to nucleotides, a strong F_o-F_c map for Me^B was found at the active site. The F_o-F_c map around the Me^A binding site was weaker than the one for Me^B , suggesting the presence of Mn^{2+} with partial occupancy. The Mn^{2+} coordinating conserved D559 changed its conformation upon Me^A binding (Gleghorn et al., 2011a). Accordingly, the D559 electron density map of TIC-2' (**Figure 2.9 C-a**) was distinct from the ones observed from SCI (coordinating Me^B , **Figure 2.9 C-b**) and SCII (coordinating Me^A and Me^B , **Figure 2.9 C-c**) (Gleghorn et al, 2011), suggesting the presence of alternative conformers of the D559 side-chain and supporting the presence of Mn^{2+} with partial occupancy. After 2 min of soaking, some molecules contain GTP(+1), ATP(+2) and Me^B corresponding to the SCI structure, and some molecules contain GTP(+1), ATP(+2), Me^A and Me^B corresponding to the SCII structure determined in our previous study (**Figure 2.13**) (Gleghorn et al., 2011a). TIC-2' is an intermediate where the two nucleotides and Me^B are loaded at the active site and is waiting for the Me^A binding to form the catalytically competent complex.



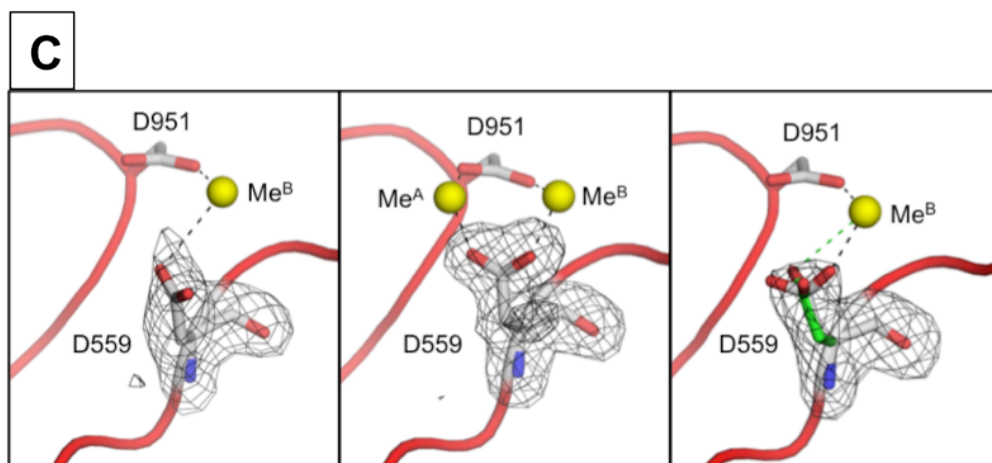


Figure 2.9 Structural analysis of TIC-2' active site. (A) Electron density (F_o-F_c , sigma cutoff = 4) maps showing nucleotides and metals in TIC-2'. (B) Conformational changes between TIC-1' (black) and TIC-2' (yellow) are shown by superposition of the two models. (C) F_o-F_c maps (black net, sigma cutoff = 3) showing D559 superimposed on the final models (sticks, tubes and spheres) of SCI (a), SCII (b) and TIC-2' (c). Two conformers of D559 found in the TIC-2' are shown (green: Me^B coordinating form; white: both Me^A and Me^B coordinating form).

At 3 min (TIC-3', **Figure 2.10**), a strong F_o-F_c map for Me^A appeared at the active site and its intensity was equivalent to the one of Me^B . Except for the D559 side-chain orientation, loading of Me^A did not trigger any other conformational change of RNAP and DNA (**Figure 2.10 B**). After 3 min of soaking, the active site contained GTP(+1), ATP(+2), Me^A and Me^B , corresponding to SCII isolated in our previous study (**Figure 2.13**) (Gleghorn et al., 2011a). However, in this study, we solved this elusive intermediate structure with natural substrates. TIC-3' is a fully reactive, naturally existing enzyme-substrate-cofactor complex, closest to the pentacovalent phosphate intermediate during the nucleotidyl transfer reaction.

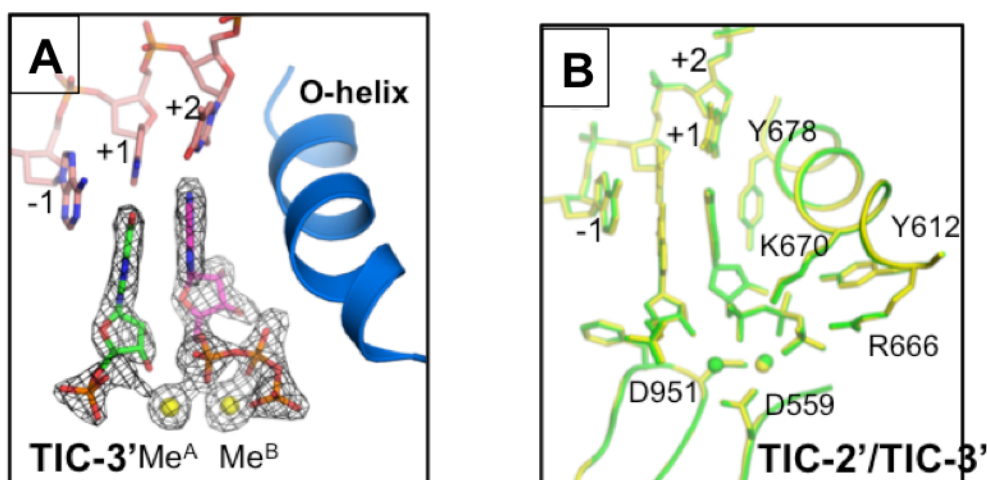


Figure 2.10 Structural analysis of TIC-3' active site. (A) Electron density (F_o-F_c , sigma cutoff = 4) maps showing nucleotides and metals in TIC-3'. **(B)** Conformational changes between TIC-2' (yellow) and TIC-3' (green) are shown by superposition of the two models.

At 4 min (TIC-4', **Figure 2.11**), a strong F_o-F_c map for the +1 and +2 nucleotides was observed, but there was continuous density between O3'(+1) and α P(+2) and a gap between the α P(+2) and β P(+2) groups indicating that this map corresponded to the 2-mer RNA (5'-pppGpA-3'). The pyrophosphate (PPi) was still coordinated by residues K666, R660 and Y612 and was observed with Me^B; however, the F_o-F_c map for Me^A is absent indicating that the Me^A dissociates from the active site immediately after phosphodiester bond formation. Accordingly, D559 rotates its side-chain (**Figure 2.11 B**) to approximately the same position found in the BC and SCI structures suggesting its coordination to single Me^B (Gleghorn et al., 2011a). The temporary presence of Me^A at the active site for only producing the 3'-oxanion is beneficial for the A-family polymerases to prevent a reverse reaction (pyrophosphorolysis). After 4 min of soaking, the active site contained the nascent 2-mer RNA, PPi and Me^B, corresponding to the PC isolated in our previous study (**Figure 2.13**) (Gleghorn et al., 2011a). The high B-factors for PPi (82.1) and Me^B (92.1) suggest their ensuing dissociation from the active site.

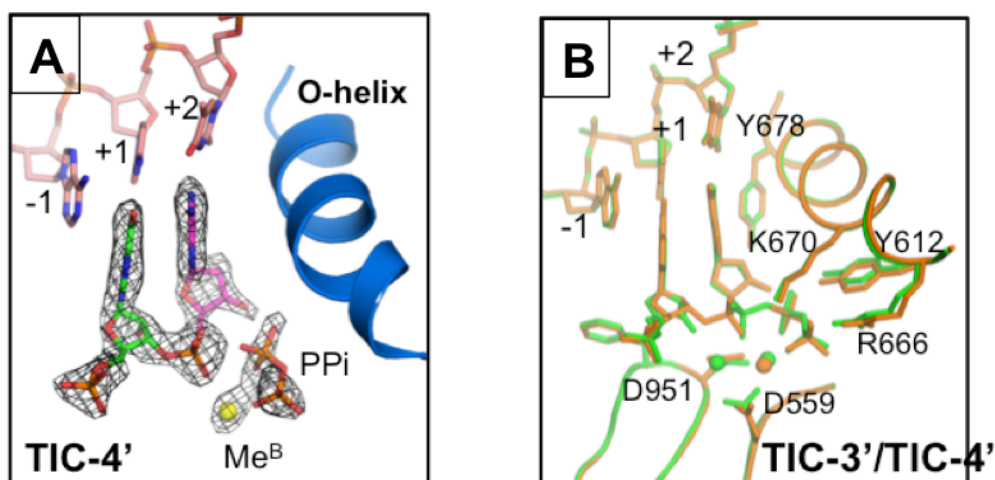


Figure 2.11 Structural analysis of TIC-4' active site. (A) Electron density ($F_o - F_c$, sigma cutoff = 4) maps showing nucleotides and metals in TIC-4'. **(B)** Conformational changes between TIC-3' (green) and TIC-4' (orange) are shown by superposition of the two models.

2.5.3 Attempts of *in crystallo* reaction with Mg^{2+}

The above transcription reaction conditions were established using Mn^{2+} after extensive trials failed to incorporate the physiological metal Mg^{2+} at the catalytic site. This had been a consistent problem in the previous snapshot structures of intermediates because of the crystallization cocktail being used. The citrate ion in the crystallization solution chelates the provided Mg^{2+} and lowered the available concentration of free metal for reaction and catalytic binding. Citrate has posed similar problems in crystals of other nucleic acid polymerases (Garcia-Diaz et al., 2007; Tomita et al., 2006). The best success that I had with *in crystallo* catalysis with Mg^{2+} was that I was able to perform *in crystallo* reaction in these conditions by fully depleting citrate and increasing Mg^{2+} to 30 mM. The identified time window for this reaction occurred between 6 and 7 min, which is slower than the reaction with Mn^{2+} . However, Mg^{2+} binding at the catalytic site A was so transient that I could not trap a natural pre-catalytic complex with Mg^{2+} bound at both the A and B sites. This observation was consistent with the Raman traces, in which the COO-Mg stretching modes appeared much later for SCI formation than PC. Structures of $TIC_{Mg}-6'$ and $TIC_{Mg}-7'$ are shown (**Figure 2.12 A-B**).

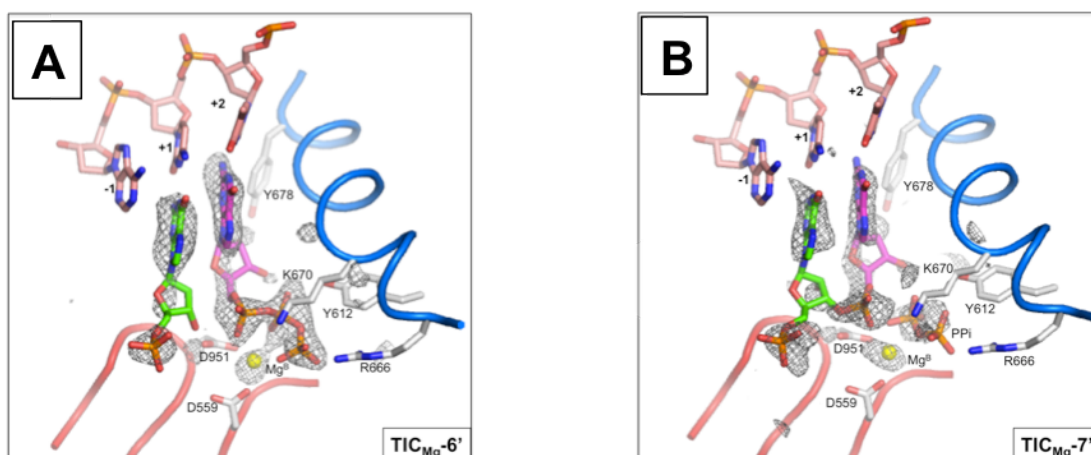


Figure 2.12 Active site views of the TIC_{Mg}-6' (A) and TIC_{Mg}-7' (B). Electron density (F_o-F_c , sigma cutoff = 4) maps show two separate NTPs in (A) while continuous density is seen at the phosphodiester bond in (B). Only density for Mg^B appears in both the structures.

2.6 Discussion

2.6.1 Temporal order in nucleotidyl-transfer reaction

In this study I successfully captured four intermediate structures of N4 phage mini-vRNAP transcription initiation (**Figure 2.13**), extending from the nucleotide/metal binding at the active site to phosphodiester bond formation. Nucleotide loading at the active site triggers the O-helix closing at an early stage of the reaction. After 1 min of soaking (TIC-1', **Figure 2.8**), a weak but distinct F_o-F_c map for Me^B is observed with fragmented +2 nucleotide electron density, supporting a widely-accepted but unproven prediction that Me^B comes to the active site with nucleotide. After 2 min of soaking (TIC-2', **Figure 2.9**), both +1 and +2 nucleotides as well as Me^B binding are completed. However, Me^A binding is only completed at 3 min of soaking (TIC-3', **Figure 2.10**). This is a clear proof that Me^B is loaded to the active site prior to Me^A, which is the penultimate step before phosphodiester bond formation. The presence of Me^A at the active site is only temporary during the reaction. After 4 min of soaking (TIC-4', **Figure 2.11**), the active site possesses a 2-mer RNA, PPi and Me^B but not Me^A indicating that Me^A dissociates from the active site immediately after phosphodiester bond formation. Therefore, leaving of Me^A from the active site is the first step after bond formation and is preparing for the

next round of reaction. The transient and timely presence of Me^{A} at the active site is beneficial not only for the transcript initiation, but may also be beneficial for all nucleotidyl transfer reaction of the A-family polymerase to prevent a reverse reaction (pyrophosphorolysis).

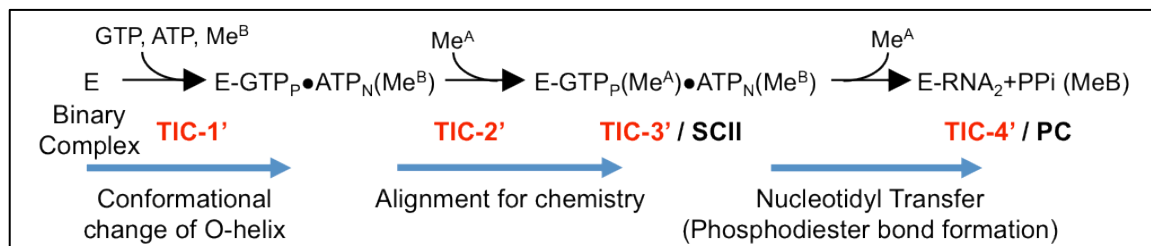


Figure 2.13 A schematic of the transcript initiation. The RNAP and promoter DNA complex is depicted as “E”. Appearance of intermediates, from TIC-1’ to TIC-4’, isolated in this study as well as SCI, SCII and PC identified in the previous study (Gleghorn et al., 2011a) are indicated.

I could establish a kinetic scheme of N4 mini-vRNAP transcription initiation (**Figure 2.13**) by using “trigger-freeze” X-ray crystallography experiment (**Figure 2.3**). The kinetic scheme is based on structural snapshots from multiple crystals; however, this scheme reflects the temporal order of events in a single crystal since it is consistent with results from Raman crystallography experiment, which uses a single crystal to trace the entire reaction. Furthermore, our scheme is consistent with one that was proposed based on fluorescence-based assays and stopped flow kinetics of *E. coli* DNAPI (Bermek et al., 2011b).

2.6.2 Dual role of catalytic metal in nucleotidyl transfer reaction of A-family polymerases

Why does Me^{A} come to the active site just before phosphodiester bond formation? The Me^{A} is a key element of the nucleotidyl transfer reaction and plays multiple roles in catalysis. When the Me^{A} coordination is established, it aligns the reactive groups of 3’-OH(+1) and $\alpha\text{P}(+2)$ and the active site proceeds via the $\text{S}_{\text{N}}2$ reaction. The A-family polymerase is not able to coordinate Me^{A} without a properly oriented 3’-OH and αP groups from two different nucleotides. This requirement satisfies two essential functions

of nucleic acid polymerase, catalytic efficiently and fidelity of nucleotide selection, by using a single atom at the active site. Since the presence of Me^{A} is essential for catalysis, the Me^{A} is a great sensor of Watson-Crick base pairing between the DNA template and nucleotide substrates despite base pairings being 10~12 Å away from the catalytic site. While the A-family polymerase uses O-helix movement as an early check point of nucleotide selection, an additional kinetic checkpoint right before bond formation is able to greatly enhance the fidelity (Joyce and Benkovic, 2004).

Due to highly conserved active site architecture, the reaction scheme of the nucleotidyl transfer reaction of the N4 vRNAP should be universal to all A-family polymerases including bacterial DNAP I (Johnson et al., 2003; Li et al., 1998), bacteriophage DNAP (Doublie et al., 1998), and Poly in eukaryotes involved in the mitochondrial DNA replication (Lee et al., 2009). This reaction scheme may be applicable to other types of polymerase such as X-family DNA polymerases β (Batra et al., 2006) and ϵ (Garcia-Diaz et al., 2007) in eukaryotes, since the Me^{A} coordination in these polymerases also required the presence of a primer 3'-OH and nucleotide substrate with Me^{B} at the active site. In contrast, the order of Me^{A} and Me^{B} binding at the active site is different in multi-subunit RNAP. The Me^{A} is coordinated by three Asp residues in the absolutely conserved DFDGD motif in the largest subunit (Vassylyev et al., 2007b) at the active site without primer 3'-OH or nucleotide. This difference is consistent with the fact that the A-family polymerase carries out only the nucleotidyl transfer reaction whereas the multi-subunit RNAP is capable of not only nucleotidyl transfer but also RNA hydrolysis, which plays a role in transcription proofreading (Cheung and Cramer, 2011).

The mechanism of the phosphoryl transfer reaction is explained by the “two-metal-ion” mechanism, which was proposed two decades ago (Beese and Steitz, 1991; Steitz, 1998; Steitz and Steitz, 1993). This model has been an important framework for explaining many reactions of nucleic acid enzymes including DNAP, RNAP, self-splicing introns, and nucleases (Yang et al., 2006). Our study provided a new basis – chronology of substrate and metal bindings and protein conformational change – for the two-metal-ion mechanism. We demonstrated that the Me^{A} binds after Me^{B} and

nucleotides and does so immediately before the chemistry step (phosphodiester bond formation). The Me^{A} is a Lewis acid to reduce the pKa of the primer 3'-OH and aligns the reactive groups of 3'-OH and α -phosphate to proceed through the $\text{S}_{\text{N}}2$ reaction. The Me^{A} also satisfies the fidelity of nucleotide selection. The Me^{A} loading depends on the presence of the primer 3'-OH group and other ligand positions (Batra et al., 2006; Johnson, 2010) and both reactive groups – 3'-OH and α -phosphate – are involved in the Me^{A} coordination. Therefore, the Me^{A} can be a great sensor to monitor correct base pairing between DNA and nucleotide; once the Watson-Crick base pairs are established, the Me^{A} binds at the active site and completes the reaction. Our structures provide evidence that the Me^{A} binding is the final checkpoint of the nucleotide selection of A-family polymerase.

2.6.3 Soak-trigger-freeze approach is suitable for tracing enzymatic reactions in four dimensions

In each TIC structure, distinct shapes and intensities of $\text{F}_o\text{-F}_c$ electron density maps corresponding to nucleotides and metals were revealed at their active sites, although time of soaking for each TIC preparation differ on the minute time scale. Emerging evidence supports a hypothesis that protein flexible motions play an important role in enhancing the speed of catalysis (Bhabha et al., 2011; Henzler-Wildman and Kern, 2007). In crystals, due to the crystal packing, protein motions are restricted. Thus, the nucleotidyl transfer reaction proceeds at a substantially slower rate (minutes) than in solution (milliseconds to seconds) (Chen et al., 2011; Gleghorn et al., 2011a; Yuan et al., 2011). Moreover, the RNAP-DNA complex is highly concentrated (5 mM) in a small volume (0.3 nl) (**Section 2.2**). This unique environment allows diffusion of nucleotides plus metal to all binary complexes in a short period of time, thereby synchronizing the reaction of all molecules in the crystal to reveal some hitherto undetected intermediates by freezing crystals sequentially and solving their high-resolution structures.

2.6.3 Acid base mechanism of reaction catalysis

The proton transfer mechanism in the nucleotidyl transfer reaction also poses a lingering question regarding the identification of the general base that accepts the proton to generate the 3'O nucleophile and a general acid, which provides a proton to stabilize the PPi leaving group. While a conserved lysine residue at the O-helix was suggested as a general acid in the reaction (Castro et al., 2009), K670 in the N4 mini-vRNAP transcription initiation complex structures does not lie close to the bridging oxygen of the β P of ATP(+2) (**Figure 2.14 A**). However, the -OH group of residue Y612 is at a closer distance from the non-bridging oxygen at the β P (+2) (**Figure 2.14 B**) and shows a continuous electron density. I also solved the structure of an intermediate in equilibrium between the post and pre-catalytic states by soaking crystals in 5 mM PPi in addition to the usual substrates, in an attempt to slow down the forward reaction and PPi release that could make available some more transient intermediates for freeze-trapping. This structure provided an interesting electron density for the nucleotides, continuous from the 3'O (+1) to the 5' α P(+2) and further down to the β and γ P (+2) (**Figure 2.14 D**). While this intermediate may be in fast equilibrium between the two states, it probably provides nearly as much information as the pentacovalent transition state of the nucleotidyl-transfer reaction. In this structure, Y612 also lies closer to β P (+2) oxygen. I thus ask the question if this Y612 was lending its hydrogen to stabilize the negative charge developed on the PPi leaving group, could it serve as the general acid in the reaction? A mutation Y571F at this position of T7 RNAP reduces its catalytic activity (Rechinsky et al., 1995), and in the case of A-family DNAP, histidine found at this position when mutated (H932Y in human mitochondrial DNAP γ) reduced the specificity constant k_{pol}/K_D by ~ 150 fold, which was attributed largely to the slowing down of PPi release (Batabyal et al., 2010). Near neutral pK_a of histidine imidazole group is a more efficient general acid, which might be a reason why DNAP is more than 10 times faster than RNAP for the nucleotidyl transfer reaction. However, such a conclusion should await a higher-resolution structure or complementary studies capable of locating the proton between the Y612 and the phosphate.

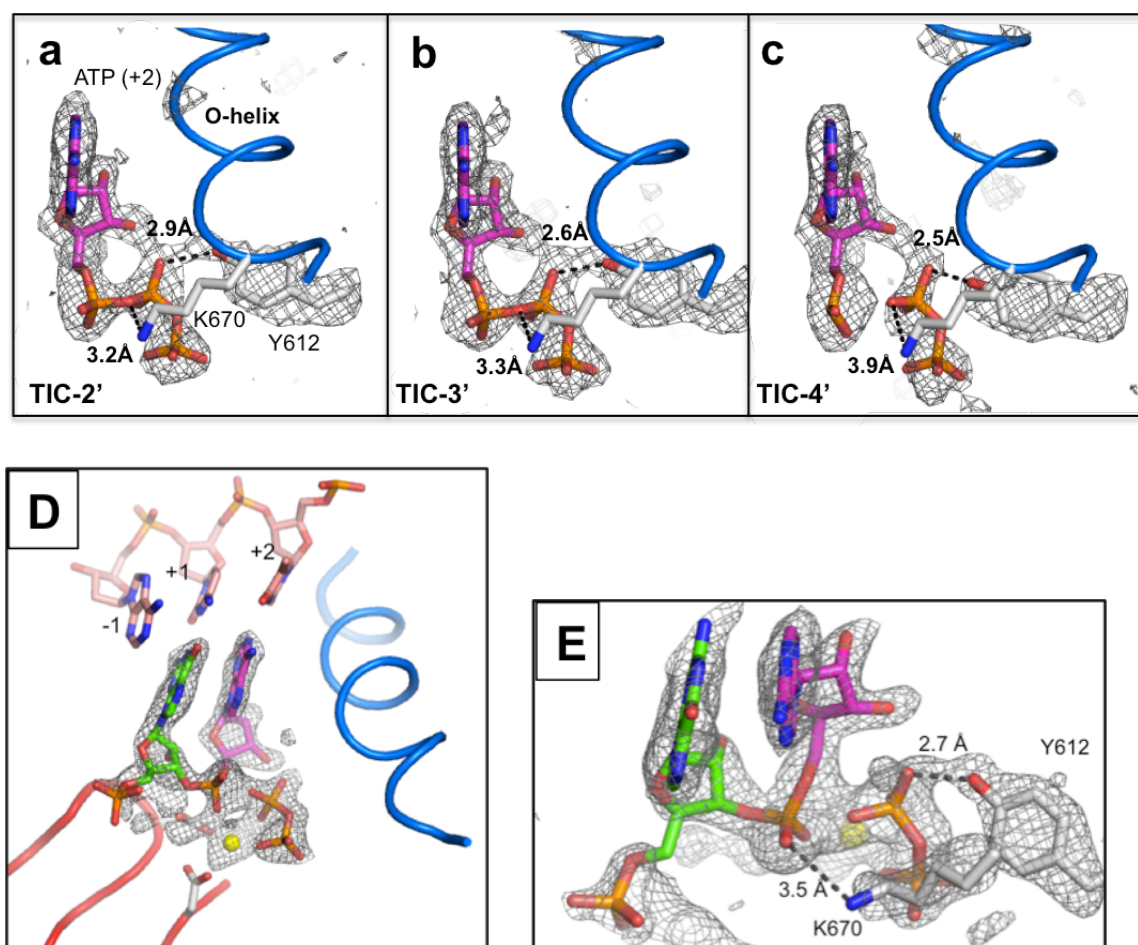


Figure 2.14 Electron density of Y612 proposing proton transfer. F₀-F_c maps (black net, sigma cutoff = 3) of Y612 and ATP (+2) superimposed on final models (sticks: Y612, K670 and ATP (+2)); cartoon: O-helix) of TIC-2' (A), TIC-3' (B) and TIC-4' (C). Distances between atoms (black dotted lines) are labeled. (D) F₀-F_c map of the TIC-PPi complex is shown, wherein the 5'αP(+2) appears like a pentacoordination probably due to an equilibrium of pre- and post-catalytic states. Full density for Me^B is seen with slight density for Me^A. The density has been modeled with a 2-mer RNA product and PPI. (E) F₀-F_c maps of Y612 and K670 have been superposed on the models.

Likewise, there is uncertainty in the identity of the proton acceptor for generation of the attacking nucleophile 3'oxyanion. The possible candidates for the base are bulk solvent, one of the conserved aspartate residues or the ionic oxygen groups at the 5'αP (Florian et al., 2003). To investigate the possibility of the closest aspartate D951 role as a general base, we solved time-resolved structures of D951N mutant RNAP during a reaction cycle and also tested the effect of this mutation on transcription activity (performed in our collaborator Prof. Rothman-Denes' lab).

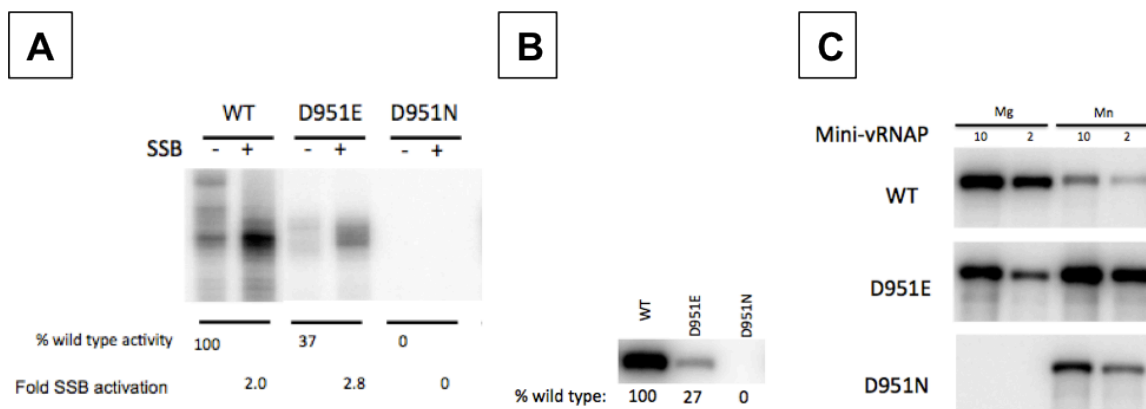


Figure 2.15 Effect of mutations of Asp 951 on transcription activity. Denaturing PAGE gel showing products of Run-off transcription (**A**) and first phospho-diester bond formation (**B**). The effect of Mn substitution for Mg on transcription activity by the mutants have been tested (**C**).

Since D951 is the closest to the 3'OH, it is a more attractive candidate for the proton acceptor, because of its backside position. Going by the general stereochemistry of the S_N2 attack, one of the lone pair of electrons on the attacking Oxygen (3'O) faces towards the target 5'P, while another faces the Me^A for coordination. Thus, it follows that the hydrogen atom of the OH should face to the back, closer to the D951 if obeying the distorted tetrahedral geometry of -OH. Matching with this proposition, transcription activity both for run off (**Figure 2.15 A**) and formation of the first phospho-diester bond (**Figure 2.15 B**) in the mutant D951N in Mg²⁺ was totally abolished. However, transcription was rescued in the presence of Mn²⁺ (**Figure 2.15 C**), which led us to doubt the role of D951 beyond metal coordination. Whereas the strict binding geometry of Mg²⁺ probably disallows its binding to the changed coordination due to Asn residue, Mn²⁺ with a more relaxed requirement binds at position A and catalyzes the reaction.

To investigate this possibility I solved structures of real-time transcription intermediates of mutant D951N complexes with Mn²⁺ (TIC_{N951-2'} and TIC_{N951-3'}). Expectedly, the phosphodiester bond was formed between 2 – 3 mins. The DSC-2' is a pre-catalytic substrate complex showing the presence of two NTPs, but interestingly lacks both metals. This reinforces the importance of the conserved aspartate D951 in

coordination metals, even for the stably bound Me^{B} . Consequently, the β and γ PO_2^- are disordered in the structure as is the Y612 ring. Interestingly though, the reaction proceeds in the next minute, which is evidence for the transient binding of the Me^{A} . Throughout the reaction in both pre and post-catalytic complexes, the orientation of N951 is perpendicular to its usual bidentate coordinating conformation (as in the case of D951), with the CO^- facing the metal binding sites and the NH_2 facing upwards (**Figure 2.16**). The structures show that D951N impairs critical metal binding however delivers catalysis, following which it may be inferred that the major role of D951 is metal binding and metal mediated substrate alignment for proper reaction of correct substrate and that it has no role as a general base for nucleophile generation. However, an ultra-high resolution structure capable of showing proton positions is required to confirm this proposition.

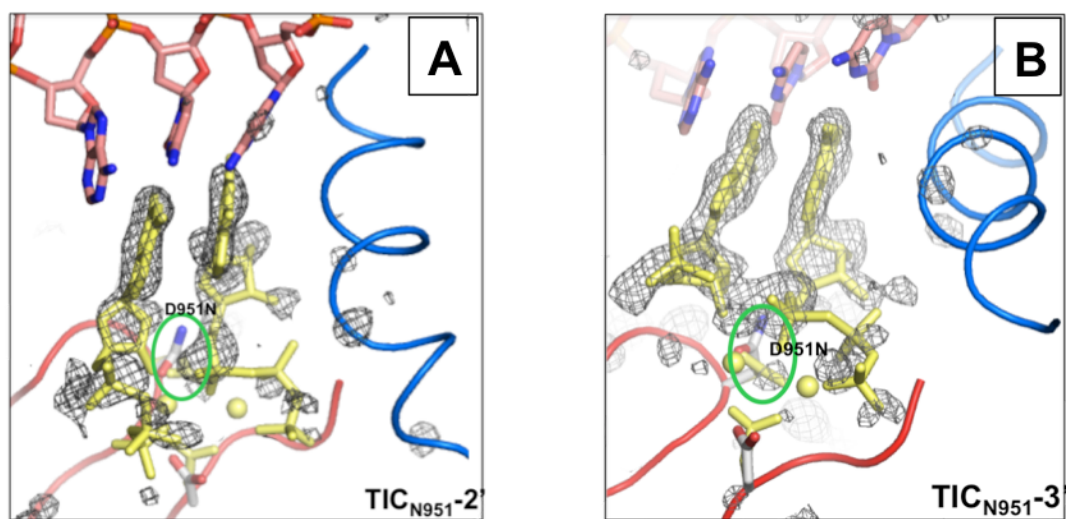


Figure 2.16 Electron density at the active site of the complexes with D951N mutant at 2 mins (A) and 3 mins (B). Fo-Fc density maps have been shown. Color codes are same as previous figures. Models of the NTP and aspartates of the TIC-3' (yellow) has been superposed on the densities to show the position of the densities. The residue N951 has been marked with a green circle, showing the perpendicular orientation of the N951 vs the usual D951 in other transcription complexes.

2.7 Acknowledgements

I thank Yuanyuan Chen from Paul Carey's lab for performing the Raman crystallography studies. I thank Eda Koculi from Lucia Rothman-Denes' lab for performing the

transcription assays for D951N mutant N4 mini-vRNAP and providing me with the mutant constructs for crystal work. I also thank the staff at F1 beamline at MacCHESS for supporting crystallographic data collection. Figures have been made using Pymol.

2.8 References

- Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* *66*, 213-221.
- Batabyal, D., McKenzie, J.L., and Johnson, K.A. (2010). Role of histidine 932 of the human mitochondrial DNA polymerase in nucleotide discrimination and inherited disease. *The Journal of biological chemistry* *285*, 34191-34201.
- Batra, V.K., Beard, W.A., Shock, D.D., Krahn, J.M., Pedersen, L.C., and Wilson, S.H. (2006). Magnesium-induced assembly of a complete DNA polymerase catalytic complex. *Structure* *14*, 757-766.
- Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J* *10*, 25-33.
- Bermek, O., Grindley, N.D., and Joyce, C.M. (2011). Distinct roles of the active-site Mg^{2+} Ligands, Asp882 and Asp705, of DNA polymerase I (Klenow fragment) during the prechemistry conformational transitions. *J Biol Chem* *286*, 3755-3766.
- Bhabha, G., Lee, J., Ekiert, D.C., Gam, J., Wilson, I.A., Dyson, H.J., Benkovic, S.J., and Wright, P.E. (2011). A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science* *332*, 234-238.
- Bourgeois, D., and Royant, A. (2005). Advances in kinetic protein crystallography. *Current opinion in structural biology* *15*, 538-547.
- Carey, P.R. (2006). Raman crystallography and other biochemical applications of Raman microscopy. *Annu Rev Phys Chem* *57*, 527-554.

Carey, P.R., Chen, Y., Gong, B., and Kalp, M. (2011). Kinetic crystallography by Raman microscopy. *Biochimica et biophysica acta* *1814*, 742-749.

Castro, C., Smidansky, E.D., Arnold, J.J., Maksimchuk, K.R., Moustafa, I., Uchida, A., Gotte, M., Konigsberg, W., and Cameron, C.E. (2009). Nucleic acid polymerases use a general acid for nucleotidyl transfer. *Nat Struct Mol Biol* *16*, 212-218.

Cheetham, G.M., and Steitz, T.A. (1999). Structure of a transcribing T7 RNA polymerase initiation complex. *Science* *286*, 2305-2309.

Chen, Y., Basu, R., Gleghorn, M.L., Murakami, K.S., and Carey, P.R. (2011). Time-resolved events on the reaction pathway of transcript initiation by a single-subunit RNA polymerase: Raman crystallographic evidence. *J Am Chem Soc* *133*, 12544-12555.

Cheung, A.C., and Cramer, P. (2011). Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature* *471*, 249-253.

Doublie, S., Tabor, S., Long, A.M., Richardson, C.C., and Ellenberger, T. (1998). Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* *391*, 251-258.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.

Florian, J., Goodman, M.F., and Warshel, A. (2003). Computer simulation of the chemical catalysis of DNA polymerases: discriminating between alternative nucleotide insertion mechanisms for T7 DNA polymerase. *J Am Chem Soc* *125*, 8163-8177.

Fraser, J.S., Clarkson, M.W., Degnan, S.C., Erion, R., Kern, D., and Alber, T. (2009). Hidden alternative structures of proline isomerase essential for catalysis. *Nature* *462*, 669-673.

Freudenthal, B.D., Beard, W.A., Shock, D.D., and Wilson, S.H. (2013). Observing a DNA polymerase choose right from wrong. *Cell* *154*, 157-168.

- Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Pedersen, L.C., and Kunkel, T.A. (2007). Role of the catalytic metal during polymerization by DNA polymerase lambda. *DNA repair* 6, 1333-1340.
- Gleghorn, M.L., Davydova, E.K., Basu, R., Rothman-Denes, L.B., and Murakami, K.S. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc Natl Acad Sci U S A* 108, 3566-3571.
- Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B., and Murakami, K.S. (2008). Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* 32, 707-717.
- Hajdu, J., Neutze, R., Sjogren, T., Edman, K., Szoke, A., Wilmouth, R.C., and Wilmot, C.M. (2000). Analyzing protein functions in four dimensions. *Nat Struct Biol* 7, 1006-1012.
- Henzler-Wildman, K., and Kern, D. (2007). Dynamic personalities of proteins. *Nature* 450, 964-972.
- Henzler-Wildman, K.A., Lei, M., Thai, V., Kerns, S.J., Karplus, M., and Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450, 913-916.
- Johnson, K.A. (2010). The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim Biophys Acta* 1804, 1041-1048.
- Johnson, S.J., Taylor, J.S., and Beese, L.S. (2003). Processive DNA synthesis observed in a polymerase crystal suggests a mechanism for the prevention of frameshift mutations. *Proc Natl Acad Sci U S A* 100, 3895-3900.
- Joyce, C.M., and Benkovic, S.J. (2004). DNA polymerase fidelity: kinetics, structure, and checkpoints. *Biochemistry* 43, 14317-14324.

- Joyce, C.M., Potapova, O., Delucia, A.M., Huang, X., Basu, V.P., and Grindley, N.D. (2008). Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* *47*, 6103-6116.
- Kazmierczak, K.M., Davydova, E.K., Mustaev, A.A., and Rothman-Denes, L.B. (2002). The phage N4 virion RNA polymerase catalytic domain is related to single-subunit RNA polymerases. *The EMBO journal* *21*, 5815-5823.
- Kennedy, W.P., Momand, J.R., and Yin, Y.W. (2007). Mechanism for de novo RNA synthesis and initiating nucleotide specificity by t7 RNA polymerase. *J Mol Biol* *370*, 256-268.
- Lassila, J.K., Zalatan, J.G., and Herschlag, D. (2011). Biological phosphoryl-transfer reactions: understanding mechanism and catalysis. *Annu Rev Biochem* *80*, 669-702.
- Lee, Y.S., Kennedy, W.D., and Yin, Y.W. (2009). Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* *139*, 312-324.
- Li, Y., Korolev, S., and Waksman, G. (1998). Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J* *17*, 7514-7525.
- Nakamura, T., Zhao, Y., Yamagata, Y., Hua, Y.J., and Yang, W. (2012). Watching DNA polymerase eta make a phosphodiester bond. *Nature* *487*, 196-201.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* *276*, 307-326.
- Rechinsky, V.O., Chernov, B.K., Dragan, S.M., Kostyuk, D.A., Tunitskaya, V.L., and Kochetkov, S.N. (1995). Targeted mutagenesis identifies Asp-569 as a catalytically critical residue in T7 RNA polymerase. *Mol Gen Genet* *247*, 110-113.

Schmidt, M. (2008). Structure Based Kinetics by Time-Resolved X-ray Crystallography. In *Ultrashort Laser Pulses in Biology and Medicine*, M. Braun, P. Gilch, and W. Zinth, eds. (Springer Berlin Heidelberg), pp. 201-241.

Schotte, F., Lim, M., Jackson, T.A., Smirnov, A.V., Soman, J., Olson, J.S., Phillips, G.N., Jr., Wulff, M., and Anfinrud, P.A. (2003). Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science* *300*, 1944-1947.

Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* *391*, 231-232.

Steitz, T.A., and Steitz, J.A. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A* *90*, 6498-6502.

Temiaikov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S., and Vassylyev, D.G. (2004). Structural basis for substrate selection by t7 RNA polymerase. *Cell* *116*, 381-391.

Tomita, K., Ishitani, R., Fukai, S., and Nureki, O. (2006). Complete crystallographic analysis of the dynamics of CCA sequence addition. *Nature* *443*, 956-960.

Vassylyev, D.G., Vassylyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* *448*, 163-168.

Westenhoff, S., Nazarenko, E., Malmerberg, E., Davidsson, J., Katona, G., and Neutze, R. (2010). Time-resolved structural studies of protein reaction dynamics: a smorgasbord of X-ray approaches. *Acta Crystallogr A* *66*, 207-219.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg²⁺-ion catalysis and substrate specificity. *Mol Cell* *22*, 5-13.

Yin, Y.W., and Steitz, T.A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell* *116*, 393-404.

Yuan, J., Cheng, K.C., Johnson, R.L., Huang, R., Pattaradilokrat, S., Liu, A., Guha, R., Fidock, D.A., Inglesse, J., Wellems, T.E., *et al.* (2011). Chemical genomic profiling for antimalarial therapies, response signatures, and molecular targets. *Science* 333, 724-729.

Chapter 3

Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme

Manuscript from the work in this chapter is in preparation:

Basu, R.S., Warner, B. A. and Murakami, K.S. Structural basis of *de novo* initiation by bacterial RNA polymerase holoenzyme.

3.1 Abstract

Bacterial RNA polymerase (RNAP) holoenzyme containing initiation factor σ recognizes and binds to the promoter DNA to form a transcription-ready open complex. The bacterial RNAP initiates RNA synthesis by *de novo* transcription, the only step during transcription where RNAP accepts two nucleotide substrates and performs primer-independent phosphodiester bond formation. I present the first structure of a *de novo* transcription complex of cellular RNAP that reveals multiple contacts between RNAP and the +1 nucleotide, which promotes +1 nucleotide binding. The structure shows the base stacking interaction between the DNA template at -1 and the initiating nucleotide at +1, which explains the sequence preference of many RNAP promoters having purine and pyrimidine bases at -1 and +1 template positions, respectively. I also present a structure of a holoenzyme transcription complex containing a 6-mer RNA prepared by *in crystallo* transcription. The structure revealed that the RNA 5'-triphosphate reaches the σ region 3.2 and causes it to become disordered, suggesting that 6-mer RNA transcription is a first step in σ ejection from RNAP to form a stable transcription elongation complex with the core enzyme.

3.2 Introduction

During transcription, RNA polymerase (RNAP) repeats a series of mechanical motions and chemical reactions including 1) loading of a single nucleotide substrate at the $i + 1$ site; 2) a nucleotidyl transfer reaction with the RNA 3'-end at the i site; 3) release of pyrophosphate; and 4) translocation of a newly formed RNA 3'-end from the $i + 1$ to the i sites to start the next round of transcription (Cheung and Cramer, 2012). X-ray crystal structures of cellular RNAPs, including bacterial RNAP (Vassylyev et al., 2007b) and eukaryotic RNAP II (Pol II) (Wang et al., 2006), have depicted transcription elongation in detail and revealed several coordinated conformational changes in the trigger loop (TL), trigger helix and bridging helix of RNAP, associated with the transcription cycle (described in **Section 1.7**).

Cellular and bacteriophage RNAPs initiate transcription using nucleotide substrates via *de novo* transcription by loading nucleotide substrates at both the i and $i + 1$ sites opposite to the +1 and +2 DNA template bases to form a first phosphodiester bond. In contrast to the subsequent hundreds or thousands of NTP incorporation steps during transcription of an entire gene, the *de novo* step accepts an NTP at the unconventional i site and relies on one-time only stabilizing interaction(s) to form the priming base for +2 NTP incorporation. Later in transcription, the i site is occupied by the 3'-end of the RNA base with a monophosphate; whereas, NTP binds only to the $i + 1$ site. High-resolution crystal structures of the *de novo* transcription complex of bacteriophage T7 (Kennedy et al., 2007) and N4 (Gleghorn et al., 2011b) identified several unique interactions between the RNAP–DNA complex with the incoming NTP at the i site, which is essential for *de novo* transcription. However, due to structural differences between the cellular and bacteriophage RNAPs, insight from bacteriophage RNAP *de novo* transcription structures cannot be used for cellular RNAPs and high-resolution structure of the *de novo* transcription complex has so far not been reported for any cellular RNAP.

RNAP core enzyme from most bacteria consists of five subunits (α_I , α_{II} , β , β' and ω subunits) and is responsible for transcription elongation. For promoter-dependent transcription, the transcription initiation σ factor binds to the core enzyme to form the holoenzyme, which recognizes the promoter DNA sequence and enables *de novo* transcription (Murakami and Darst, 2003). Several important crystal structures have shed light on the early stages of transcription initiation, including σ factor binding by the core RNAP (Murakami, 2013; Murakami et al., 2002b) and subsequent binding of the holo RNAP to the promoter, with specific recognition of conserved upstream promoter elements (Murakami et al., 2002a; Zhang et al., 2012) (described in **Section 1.7.2**). However, the structural basis of the special *de novo* transcription initiation event has not been adequately addressed. A part of the σ factor, region 3.2, was considered to be positioned near the active site of RNAP (Murakami et al., 2002b) in the holoenzyme and is involved in *de novo* transcription by direct contact with the initiating nucleotide (Kulbachinskiy and Mustaev, 2006; Severinov et al., 1994). These assumptions were based on binding affinity measurements of the initiating NTP and crosslinking experiments. However, a recent crystal structure of the bacterial RNAP – promoter DNA complex (Zhang et al., 2012) suggested that σ region 3.2 is located too far from the active site to directly support the initiating NTP binding; therefore, it is possible that σ region 3.2 may facilitate *de novo* transcription by an allosteric mechanism, or conformation changes in the RNAP – promoter DNA complex may trigger movement of σ region 3.2 towards the active site when NTPs bind. Thus a high-resolution structure of the transcription initiation complex bound to the two initiation NTPs is necessary to establish the validity of this hypothesis and explain the mechanism of *de novo* initiation.

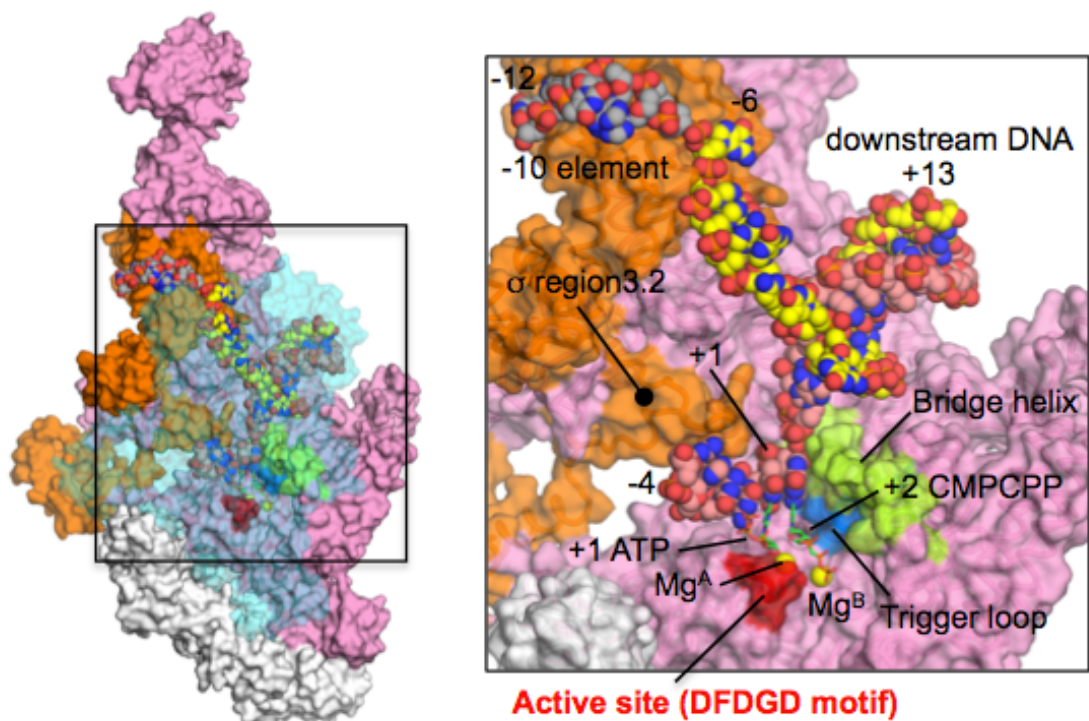


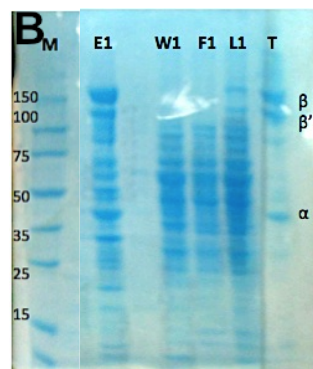
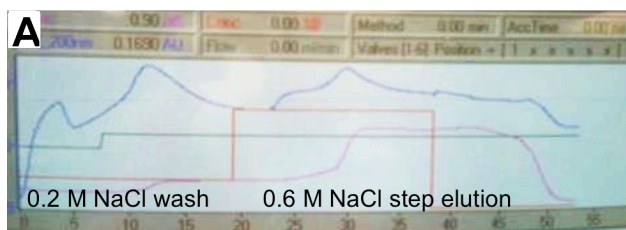
Figure 3.1 Overall structure of the *de novo* transcription complex. *T. thermophilus* RNAP is depicted as a molecular surface model. Each subunit of RNAP is depicted with a unique color (white, α and ω subunits; cyan, β subunit; pink, β' ; orange, σ^A). DNA is depicted as a sphere model with the same color scheme as in **a**. Right panel shows a magnified view of the boxed region in the left panel. For clarity, β subunit is transparent in the left panel and removed from the right panel. Several key motifs discussed in the text are highlighted.

Accordingly, in this chapter I report a crystal structure of the bacterial RNAP *de novo* transcription complex. This structure reveals key interactions among the core enzyme, σ factor, DNA, and initiating nucleotides. In addition, I also present the structure of a holoenzyme transcription complex with 6-mer RNA synthesized *in crystallo*. Several models of RNAP with abortive transcripts have been constructed to provide insight into abortive transcription or the transition from the initiation to elongation complex; however, the actual crystal structure of such a complex will provide more reliable insight into how σ release is initiated by the growing nascent transcript leading to establishment of the transcription elongation complex.

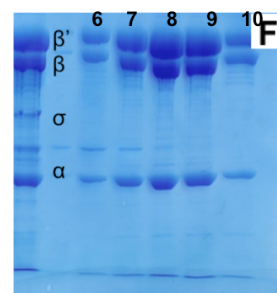
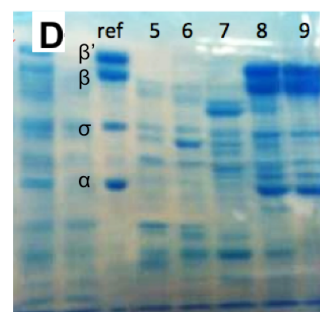
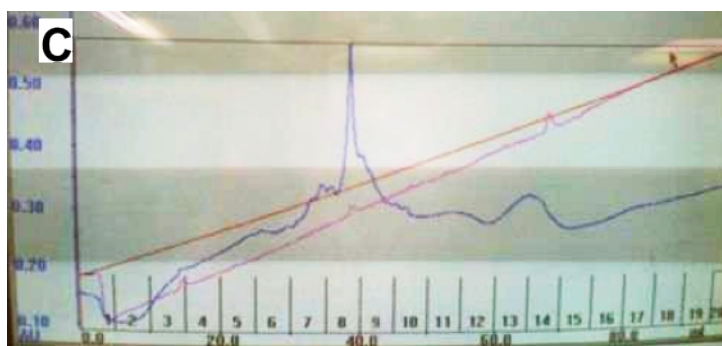
3.3 Materials and Methods

3.3.1 Purification of endogeneous *T. thermophilus* core RNAP and holoenzyme reconstitution

T. thermophilus HB27 cells were cultured by using a 300 Liter BioService fermentor at The Penn State fermentation facility, yielding roughly 1 kg of cell paste. About 100 g cell paste was suspended 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) and cells were lysed using an Emulsiflex C3 homogenizer (Avestin, Inc.) at 20,000 psi. After every 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 and glycerol was added to the supernatant to a concentration of 5% (v/v). RNAP in the soluble fraction was precipitated by adding 10 % polyethyleneimine (Polymin-P) solution (final 0.5 %) and the pellet was recovered by centrifugation. 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. This wash step was repeated with an additional 200 mL of wash buffer. Finally, 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 min at 4°C and 29,000 g. 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 to precipitate RNAP followed by centrifugation. The pellet was suspended in TGED buffer (10 mM Tris-HCl, pH 8 at 4°C, 5% glycerol, 0.1 mM EDTA and 1 mM DTT). RNAP core enzyme was purified by Heparin, ResourceQ (GE Healthcare) and SP sepharose column chromatography (**Figure 3.2 A-F**).



M - Novagen Perfect Protein Marker
T - Taq RNA polymerase reference
L1 - Lysate
F1 - Flow through
W1 - Wash
E1 - Elution



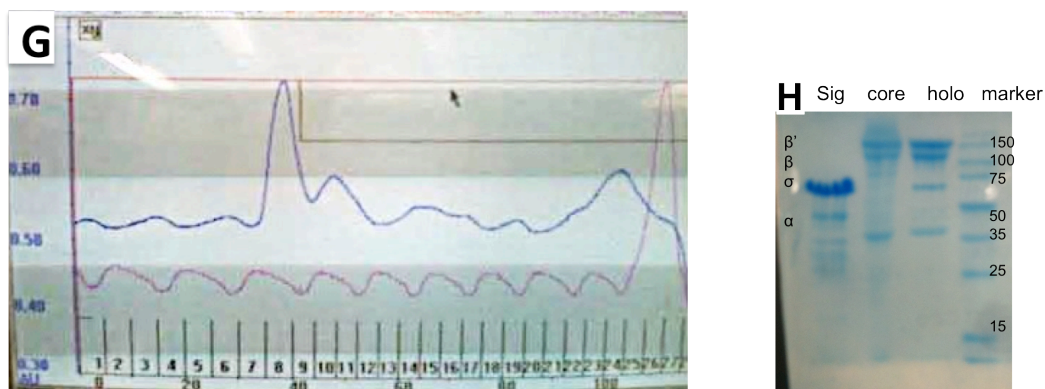


Figure 3.2 Purification of *T. thermophilus* RNAP core enzyme and preparation of holoenzyme by *in vitro* reconstitution. UV₂₈₀ chromatogram and Coomassie brilliant blue stained SDS polyacrylamide gels of successive column chromatography are shown. 50 ml Heparin column (A, B), 5 ml Resource Q (C, D) and 5 ml SP-Sepharose column (E, F) were used to purify endogeneous *T. thermophilus* RNAP core enzyme. After overnight incubation of the RNAP core enzyme with SigA, *T. thermophilus* holoenzyme was purified through a Superdex200 gel-filtration column (G, H). The blue lines on the chromatograms represent the UV (280 nm) absorbance. In SDS polyacrylamide gels, RNAP subunits (α , β , β' , σ), protein fractions from the UV chromatogram and molecular weight marker are labeled.

T. thermophilus SigA was expressed in BL21 (DE3) cells transformed with pET21a containing sigA. Cells were grown in 1.5 L LB containing 100 ug/ml ampicillin for 6 hr till OD=0.6, induced with 1 mM IPTG and grown for an additional 3 hr at 37°C. After cells were lysed by sonication, SigA was purified from the lysate by heat treatment in a 65°C water bath for 30 min. The suspension was centrifuged at 28,000 g at 4°C to remove the white precipitate. SigA was then purified from the supernatant by ResourceQ column (GE Healthcare) chromatography. SigA was stored in 50 mM Tris-HCl, pH 7.7, 400 mM NaCl and 15% glycerol at -20°C.

T. thermophilus RNAP holoenzyme was prepared by adding 3-fold excess amount of SigA to core RNAP and then purified by Superdex200 column chromatography (Figure 3.2 G-H).

3.3.2 Crystallization of the promoter DNA complex and preparation of the *de novo* transcription complex and transcription complex containing 6-mer RNA

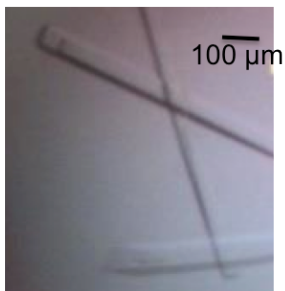


Figure 3.3 Crystals of RNAP-promoter DNA complex

The RNAP and promoter DNA complex was prepared by mixing 18 μM (24 μl) *T. thermophilus* holoenzyme (in 20 mM Tris-HCl, pH 7.7, 100 mM NaCl, and 1% glycerol) and 1 mM (0.65 μl) of the DNA scaffold (**Figure 3.4**) and incubated for 30 min at 22°C. Crystals were obtained by using hanging drop vapor diffusion by mixing equal volumes of the RNAP-DNA complex solution and crystallization solution (100 mM Tris-HCl, pH 8.7, 200 mM KCl, 50 mM MgCl₂, 10 mM Spermine tetra HCl and 10% PEG 4000) and incubating at 22°C over the same crystallization solution. The crystals were cryoprotected by soaking in the same constituents as the crystallization solution with stepwise increment of PEG 4000 and (2R, 3R)-(-)-2,3-butanediol (Sigma-Aldrich, Inc.) to final concentrations of 25% and 15%, respectively. The final cryoprotectant solution was also used for soaking nucleotides at room temperature to prepare the *de novo* transcription complex (5 mM ATP and 5 mM CMPCPP from Jena Bioscience, soaking for 30 min) and the transcription complex containing 6-mer RNA (5 mM each of ATP, CTP and UTP, soaking 5 hr).

3.3.3 X-ray data collections and structure determination

The native dataset was collected at the Macromolecular Diffraction at Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY). The crystals belong to the C centered monoclinic space group (**Table 3.1**) containing one *T. thermophilus* RNAP transcription complex per asymmetric unit. The *T. thermophilus* RNAP – promoter DNA complex (Zhang et al., 2012) was used as a search model for molecular replacement (McCoy et al., 2007). The data were processed by

HKL2000 (Otwinowski and Minor, 1997a). Rigid body refinements were performed and further adjustments to the model were performed manually. The resulting model phases allowed positioning the NTPs in the *de novo* transcription complex and the 6-mer RNA in the transcribing complex into their respective electron-density maps. Positional refinement and secondary structure restraints was performed by the program Phenix (Afonine et al., 2010).

Table 3.1: Data collection and refinement statistics

Complex PDB code	<i>De novo</i> N/A	6-mer RNA N/A
Data collection		
Space group	C2	C2
Cell dimensions		
<i>a</i> (Å)	184.49	187.09
<i>b</i> (Å)	102.16	102.08
<i>c</i> (Å)	294.72	297.26
<i>b</i> (°)	98.96	98.14
Resolution (Å)	50 – 2.85	50 – 3.00
Total reflections	260,182	273,453
Unique reflections	111,819	107,824
Redundancy	2.3 (1.4)*	2.5 (1.9)*
Completeness (%)	88.3 (67.7)*	94.4 (78.8)*
<i>I</i> / <i>s</i>	5.4 (1.4)*	6.6 (1.2)*
<i>R</i> _{sym}	0.144 (0.508)*	0.124 (0.719)*
Refinement		
Resolution (Å)	50 – 2.85	50 – 3.00
<i>R</i> _{work}	0.234	0.241
<i>R</i> _{free}	0.270	0.294
R.m.s deviations		
Bond length (Å)	0.006	0.002
Bond angles (°)	1.17	0.61

Data sets were collected at MacCHESS-F1, Ithaca, NY

*Highest resolution shells are shown in parenthesis

3.4 Results

3.4.1 Design of the X-ray crystallographic experiment to determine the *de novo* transcription complex

The bacterial RNAP holoenzyme – promoter DNA complex crystals were obtained by modifying a previously published method (Zhang et al., 2012). To obtain the RNAP *de novo* transcription complex crystals, the holoenzyme – promoter DNA crystals were soaked with the first two nucleotides of the transcription start site (**Figure 3.4**). To maintain the transcription bubble during crystallization of the holoenzyme – promoter DNA complex, Zhang *et al.* introduced noncomplementary DNA sequences around the transcription start site of the template DNA (Zhang et al., 2012). Based on the findings from the structural study of the N4 phage RNAP *de novo* transcription complex (Gleghorn et al., 2011b), we hypothesized that the template DNA around the transcription start site may participate in initiation NTP binding. Therefore, in this study, I modified the –1 template sequence to a purine (G), which mimics the natural sequence of most bacterial promoters recognized by the housekeeping σ factor (**Figure 3.4**) (Shultzaberger et al., 2007). By convention, in most elongation complexes the structural nomenclature for the incoming NTP register at the $i+1$ site is +1, which uniquely in *de novo* initiation is the +2 position on the template (genetic register). Consequently the i -site (referred to as the -1 structural register in elongation) is the +1 initiation position. Since the genetic registers are important to highlight the promoter position specific interactions in the initiation complex (at least till the point of promoter release), I will follow this unconventional structural nomenclature (equivalent to the genetic registers) in my studies. The sequence of the promoter template from the –1 to +3 positions is GTGA (transcription start site is underlined), this will form the *de novo* transcription complex if provided with ATP and a nonhydrolysable CTP, CMPCPP (Cytidine-5'-[(α,β)-methyleno]triphosphate). The structure was solved by molecular replacement at 2.85 Å resolution using the holoenzyme – promoter DNA complex structure as a reference model (Zhang et al., 2012). The structure showed strong unbiased $F_o - F_c$ electron densities at the active site corresponding to the +1 and +2 nucleotides (**Figure 3.5 A**).

The overall structure of RNAP was nearly identical with that of the reference model (PDB: 4G7H) (Zhang et al., 2012). The key changes include: conformational changes in the TL of the β' subunit (residues 1233–1255) and the amino acid side chains of the β and β' subunit around the active site (Figures 3.5 B and 3.6 A).

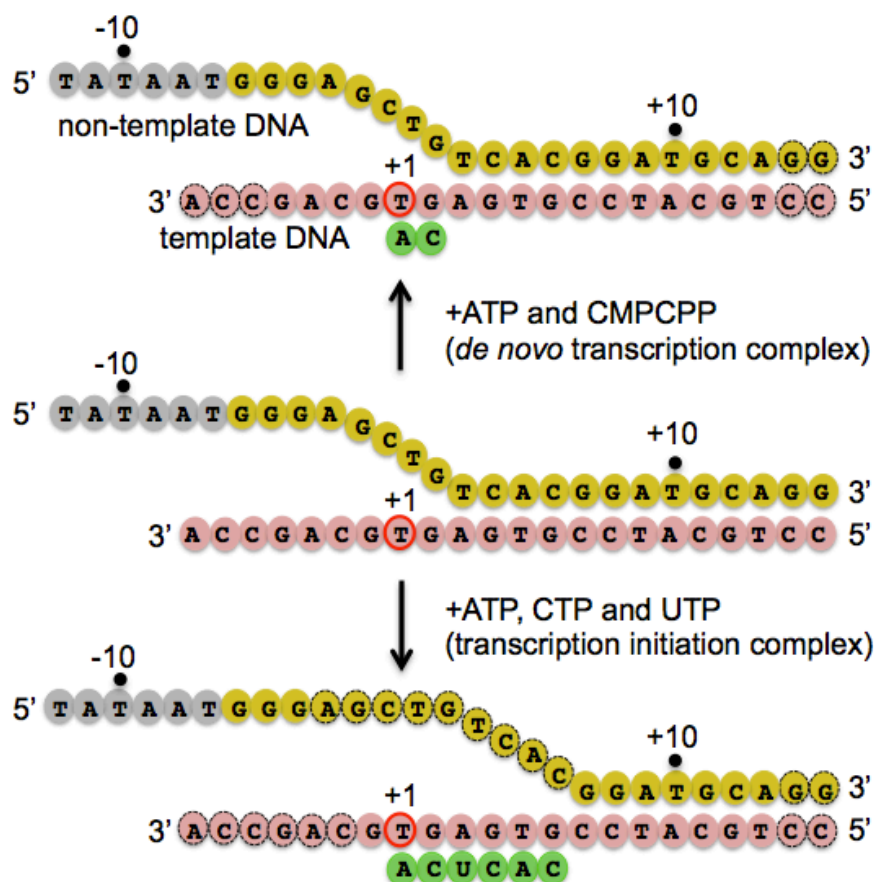


Figure 3.4 Schematic of the *de novo* transcription (top) and the holoenzyme transcription (bottom) scaffolds. Sequence of the nucleic acid scaffold used for holoenzyme – promoter DNA complex crystallization is shown (pink, template DNA; yellow, non-template DNA; gray, -10 element). Nucleotides and 6-mer RNA in the transcription complexes are depicted by green circles. Positions of DNA and transcription start site (red circle) are indicated. DNA bases outlined with black dashed lines are disordered in crystal structures.

3.4.2 The +1 NTP binding site in the *de novo* transcription complex

The *de novo* transcription complex binds two nucleotides at the active site: ATP at +1 and CMPCPP at +2. Binding of these NTPs engages residues around the active site

in interactions unique to *de novo* transcription and some that are universal in RNA elongation. However, the *de novo* transcription-specific interactions between RNAP and nucleotides appear to concentrate on the triphosphate of the +1 ATP, with multiple salt bridges with the β subunit Q β 567 (nonbridging oxygen of γ -phosphate), K β 838 (nonbridging oxygen of α -phosphate), K β 846 (nonbridging oxygen of α -phosphate), and H β 999 (nonbridging oxygens of β and γ -phosphates) (**Figure 3.5 B**). R704 of the β' subunit (R β' 704) also forms salt bridges with the 2'-OH and 3'-OH of the +1 ATP to facilitate its binding and to discriminate between rNTP and dNTP. Except for Q β 567, these residues are highly conserved in all cellular RNAPs. Therefore, these residues in all cellular RNAPs are likely used for binding +1 NTP during *de novo* transcription.

Based on a previous premise (Severinov et al., 1994), we investigated the possible interactions between σ region 3.2 and the +1 ATP binding site. The closest distance between σ region 3.2 (E324 side chain) and the γ -phosphate of +1 ATP in the *de novo* complex is approximately 10.5 Å. Instead, σ region 3.2 helps in positioning the upstream template DNA bases around -3/-4 (**Figure 3.5 A**). This rules out any direct interaction between σ region 3.2 and the +1 NTP during transcription initiation.

It is expected that template base pairing alone is not enough to stably bind an initiation NTP. We noticed a base stacking interaction between the -1 purine of the template DNA and the +1 ATP (**Figure 3.5 B**). This interaction likely serves as the additional DNA-ATP interaction that facilitates the stable binding of the +1 ATP during *de novo* transcription. The N-1 of the guanine base at the -1 template DNA position and β -phosphate of the +1 ATP are also connected by a water molecule, which provide additional hydrogen bonding interactions and further enhance +1 ATP binding.

The +1 ATP binding site overlaps with the 3'-end of the RNA binding site in the bacterial transcription elongation complex (**Figure 3.6 A**) (Vassylyev et al., 2007b) indicating that the +1 ATP binds to the *i* site for *de novo* transcription. In addition, in our structure, there is no *de novo* transcription specific nucleotide binding site that was proposed earlier in T7 RNAP transcription initiation complex (Kennedy et al., 2007).

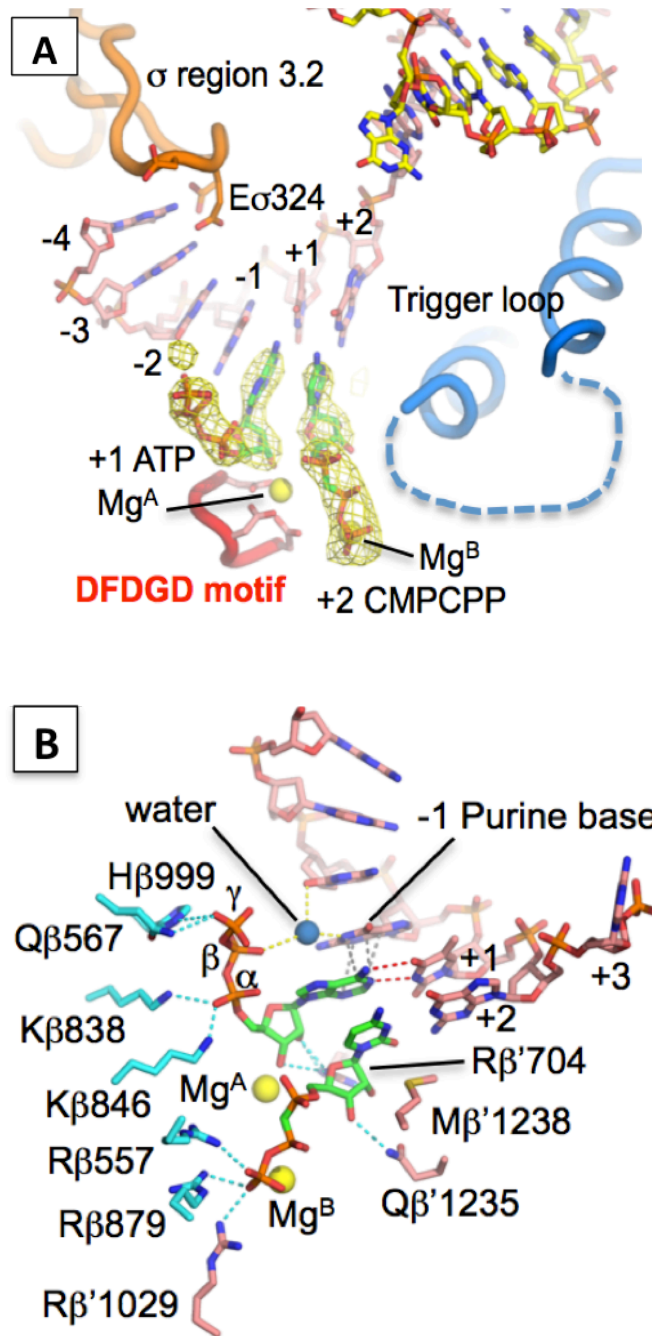


Figure 3.5 Nucleotide binding sites of the *de novo* transcription complex

(A) Active site structure of the *de novo* transcription complex. DNA, NTPs and amino acid side chains are shown as stick models, and the trigger loop, DFDGD motif and σ region 3.2 are shown as worm models and labeled. A disordered region of the trigger loop is shown as a dashed line. Fo – Fc electron density map (yellow mesh) showing +1 ATP and +2 CMPCPP is superposed on the final model.

(B) The +1 ATP bind at the active site through multiple interactions including base pairing with the +1 DNA base (red dashed lines), base stacking with the -1 purine base (gray dashed lines), water mediated interactions (yellow dashed lines) and salt bridges with side chains (cyan dashed lines). Amino acid side chains involved in the +2 CMPCPP binding are also indicated.

3.4.3 +2 NTP binding site in the *de novo* transcription complex

The +2 CMPCPP is positioned through template base pairing with the +2 DNA, and also interacts with side chains of the β' subunit TL and basic residues at the bottom of the secondary channel (**Figure 3.5 B**). In the holoenzyme – promoter DNA complex, the TL is in an open conformation and the tip of the loop (residues 1238–1251) is disordered (Zhang et al., 2012), whereas, in the transcription elongation complex with the nucleotide substrate at the $i + 1$ site, the TL is in a closed conformation and the trigger helix showed no disordered regions (Vassylyev et al., 2007b) (**Figure 3.6 A**). In the *de novo* transcription complex, even though the TL is in a more closed conformation compared to the holoenzyme – promoter DNA complex, the middle portion of the TL (residues 1239–1251) was still disordered. Residues M β' 1238 and Q β' 1235 from the N-terminal α -helix of the TL interact with the base and 3'-OH of CMPCPP, respectively. Since the trigger helix was not properly folded, it resulted in no direct interactions between the triphosphate group of +2 CMPCPP and R β' 1239 or H β' 1243 of the TL, causing CMPCPP to attain the pre-insertion conformation. The only interactions between RNAP and the CMPCPP triphosphate are those with a patch of positive residues, R β 557, R β 879 and R β' 1029, at the rim of secondary channel (**Figure 3.5 B**). In particular, the triphosphate in this pre-insertion conformation is far from the reacting distance of the 3'-OH of the +1 ATP (**Figure 3.6 B**). This position is similar to the pre-insertion site observed in the *T. thermophilus* transcription elongation complex with the inhibitor streptolydigin (Vassylyev et al., 2007) and in the Pol II transcription elongation complex (Kettenberger et al., 2004).

The catalytic metal (Mg^A) is bound stably at the active site through the aspartate triad (residues at 739, 741, and 743) of the DFDGD motif of the β' subunit. However, Mg^B is too far away to be coordinated by the aspartate triad (**Figure 3.6 B**) and instead is primarily coordinated by the β and γ -phosphates of the +2 CMPCPP. Consequently, the α -phosphate of CMPCPP is approximately 5.4 Å away from 3'-OH of the +1 ATP, thereby forming the nonreactive pre-insertion state. During *de novo* transcription, the TL may come to a closed conformation and fold into the trigger helix to push the +2 NTP further to form the first phosphodiester bond.

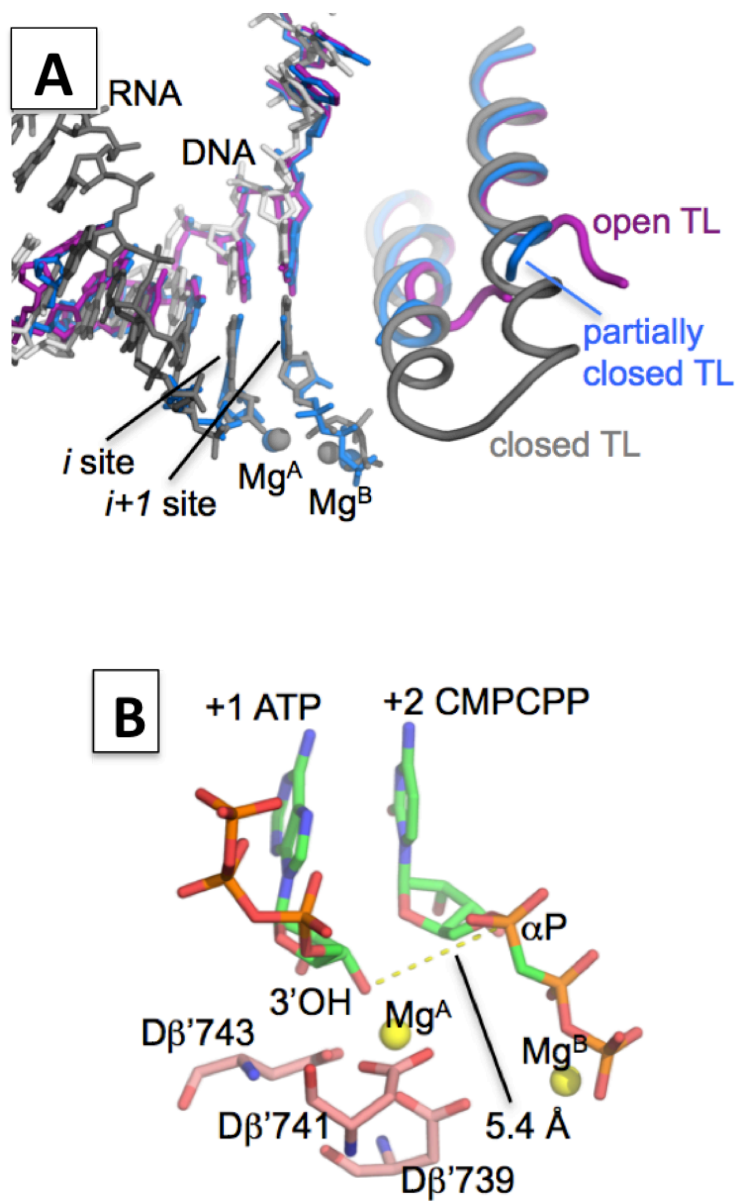


Figure 3.6 (A) Comparison of the *de novo* transcription complex (blue) to holoenzyme – promoter DNA (PDB: 4G7H, magenta), and transcription elongation complexes (PDB: 2O5J, gray and white). RNAP structures were superposed on their active site domains of β' subunit. The nucleotide binding *i* and *i* + 1 sites are indicated. Trigger loop (β' residues 1222 – 1265) of these structures are shown as tube models. **(B)** Conformation of two nucleotides, +1 ATP and +2 CMPCPP and coordinated metals in the *de novo* transcription complex active site is shown. The distance between the reacting groups 3'OH (+1) and 5' αP (+2) is shown.

3.4.4 Structure of a transcription initiation complex containing 6-mer RNA

Crystal structures of transcription complexes containing short RNAs of 2 to 7 nts have been determined for the Pol II system. However, these complexes were prepared by incubating the Pol II or Pol II/TFIIB with DNA template and synthetic RNA oligos (Cheung et al., 2011; Liu et al., 2011; Sainsbury et al., 2013). Therefore, these complexes are devoid of a triphosphate group at the RNA 5'-end, which may play important roles in the early stages of transcription such as DNA/RNA hybrid stabilization, separation of RNA from template DNA or removal of the transcription initiation factor from RNAP. To obtain the structural basis for the early stages of transcription containing σ factor and the natural form of RNA with a 5'-triphosphate group, we aimed to prepare the transcription complex by an *in crystallo* approach. For this purpose, we soaked the holoenzyme – promoter DNA complex crystals with nucleotides ATP, CTP, and UTP. The sequence of the template from –1 to +7 position is GTGAGTGC (transcription start site is underlined). If the RNAP was active in the crystal, soaking it with these

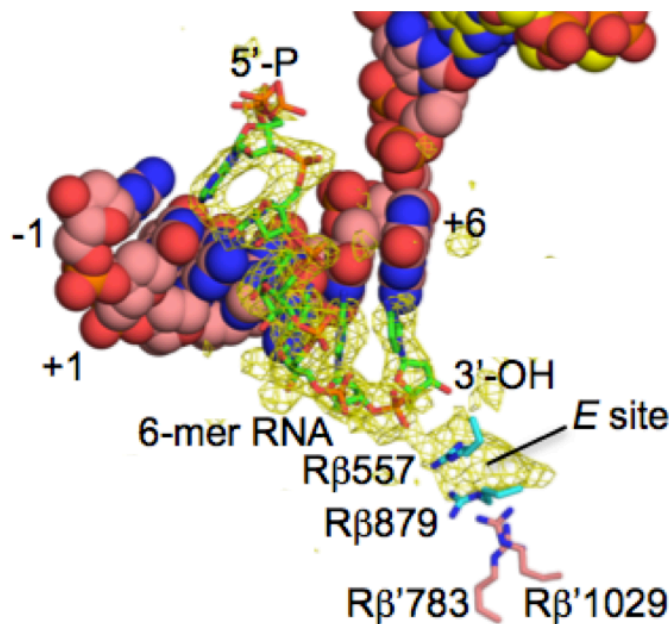


Figure 3.7 The holoenzyme transcription complex containing 6-mer RNA. Fo – Fc electron density map (yellow mesh) showing 6-mer RNA and a nucleotide bound at the E-site of the holoenzyme transcription complex. Amino acid side chains (stick) involved in the nucleotide binding at the E site are labeled.

nucleotides would generate a holoenzyme – DNA complex with transcribed RNA (**Figure 3.4**). The structure was solved at 3 Å resolution and it showed continuous electron density for RNA synthesized along the template DNA from +1 to +6 (**Figure 3.7**), indicating that *T. thermophilus* RNAP was active in this crystal form and capable of transcribing RNA up to 6 bases.

In comparison to the *de novo* transcription complex, the holoenzyme transcription complex with 6-mer RNA showed primary changes in σ region 3.2 and the upstream regions of nontemplate DNA. In the *de novo* transcription complex, the template DNA bases at the –4 and –3 positions were flipped out to make contacts with the acidic residues of σ region 3.2, whereas in the holoenzyme transcription complex, DNA bases near σ region 3.2 were rotated to base pair with the nascent 6-mer RNA (**Figure 3.8 A**). The 5'-triphosphate of the RNA transcript extended to σ region 3.2, causing the tip of σ region 3.2 (residues 321–327) to be disordered (**Figure 3.8 B**). This is likely due to the repulsion between the 5'-triphosphate of the RNA and the acidic cluster of σ region 3.2. When RNAP continues transcription, σ region 3.2 has to be pushed further to accommodate the longer nascent transcript extending toward σ region 4.1 (**Figure 3.8 A**). As the DNA/RNA hybrid extended in length, the initial contacts between RNAP and the triphosphate group of the +1 ATP (in *de novo* complex) were lost (**Figure 3.5 B**). The downstream duplex DNA was shifted 6 bases upstream due to transcription, while the upstream -10 DNA element and σ region 2.3 contacts were maintained, causing the nontemplate DNA bases from –3 to +6 to be disordered due to scrunching of the nontemplate DNA strand (**Figure 3.9**).

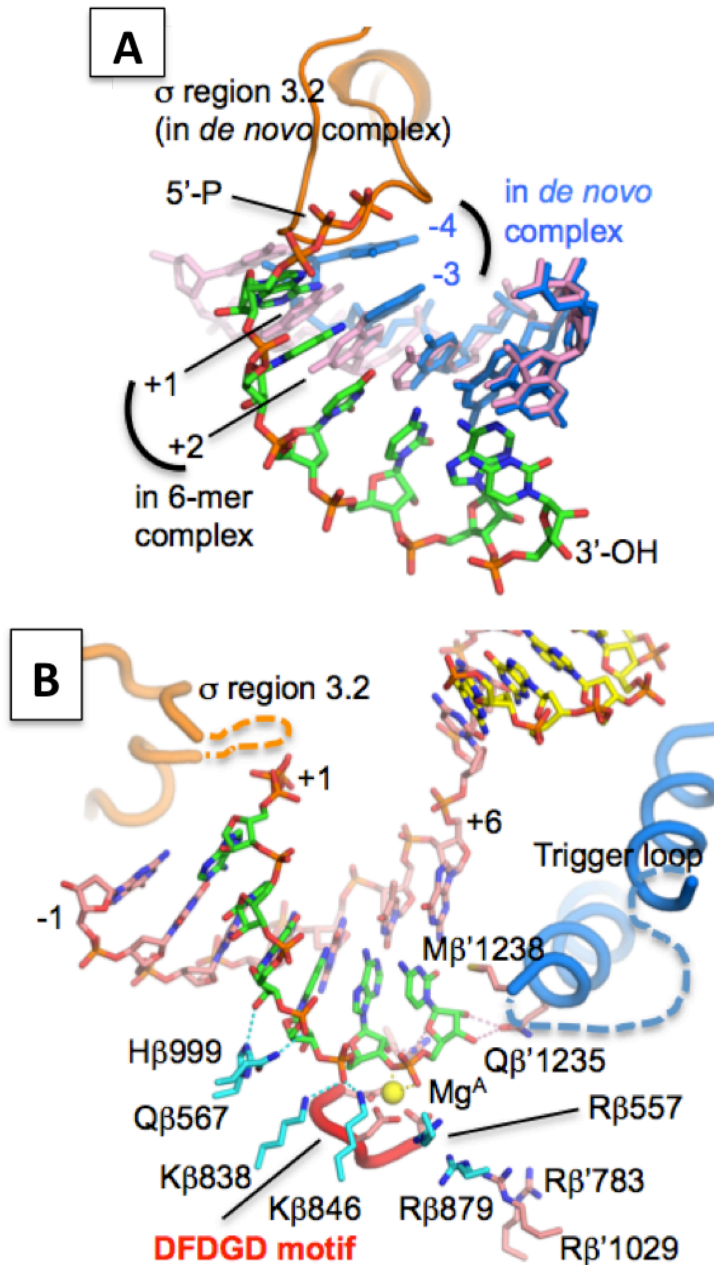


Figure 3.8 (A) Comparison of the *de novo* transcription (blue, template DNA; orange, σ region 3.2) and holoenzyme transcription complexes (pink, template DNA; green, 6-mer RNA). RNAP structures were superposed on their active site domains of β' subunit. **(B)** Active site structure. DNA, RNA and amino acid side chains contacting with the 6-mer RNA and nucleotide at the E site are shown as stick models. Disordered regions of the trigger loop (residues 1239 – 1253) and the σ region 3.2 (residues 321 – 327) are indicated by dashed lines.

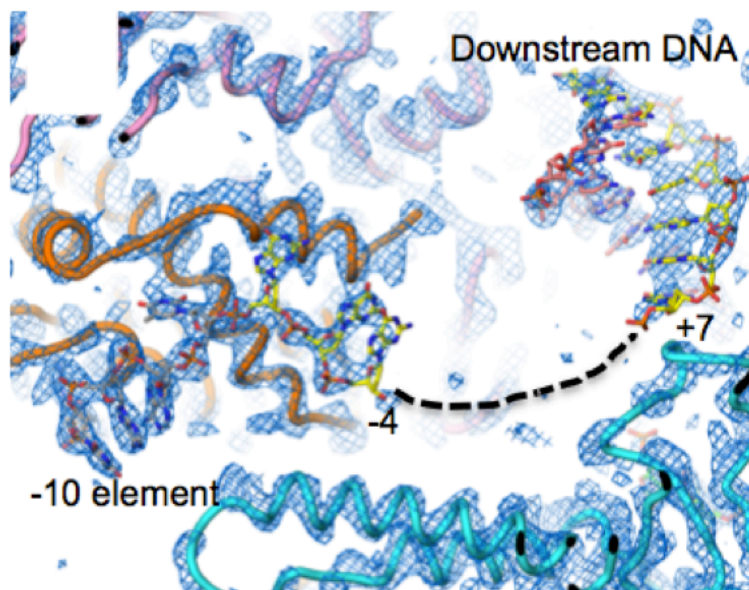


Figure 3.9 2Fo – Fc electron density map (blue mesh) of the holoenzyme transcription complex. RNAP (cyan, β ; pink, β' ; orange, σ^A) and DNA (pink, template DNA; yellow, non-template DNA; gray, -10 element) are depicted by worm and stick models, respectively. The non-template DNA between -3 and +6 is completely disordered, which is indicated by a black dashed line.

Overall, the DNA/RNA hybrid in this holoenzyme transcription complex is a typical A-form duplex with no intrinsic tilting as observed earlier in Pol II transcription initiation complexes (Cheung et al., 2011; Liu et al., 2011), clearly forming perpendicular canonical base pairs with respective template bases throughout the length of the nascent RNA (**Figure 3.8 A**). However, the 3'-RNA base is positioned at the $i + 1$ site indicating that this complex is in a pre-translocated state (**Figure 3.10**). Movement of the DNA/RNA hybrid to a post-translocated state may be constrained due to steric hindrance between 5'-end of RNA and σ region 3.2 and ejection of σ region 3.2 may be restricted due to crystal packing.

Although the transcription complex was still in the pre-translocated state, the pyrophosphate by-product was not visible and the middle portion of the TL was completely disordered (residues 1239–1253) (**Figure 3.8 A**). A bigger lobe of electron density appeared next to the RNA 3'-end (**Figure 3.7**). This density was close to the rim of the secondary channel and overlapped with the temporary nucleotide binding *E* site defined earlier in Pol II (Westover et al., 2004a).

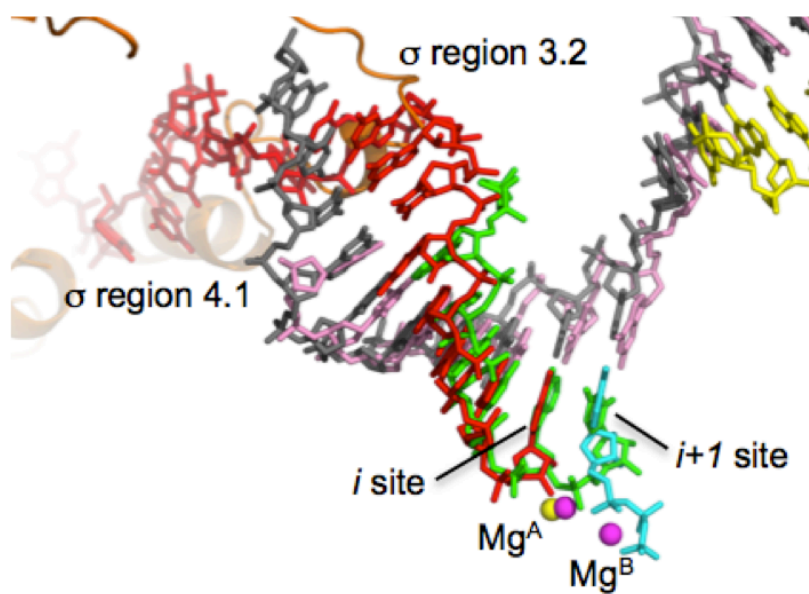


Figure 3.10. Model of longer RNA transcript into the RNA exit tunnel explains σ release. Structures of the holoenzyme transcription (pink, template DNA; green, 6-mer RNA; yellow, Mg^{2+} ; orange, σ factor) and the transcription elongation complexes (PDB: 2O5J, gray, template DNA; red, RNA; cyan, incoming NTP; magenta, Mg^{2+}) were superposed on their active site domains of β' subunit. The nucleotide binding i and $i+1$ sites are indicated.

3.5 Discussion

The +1 nucleotide binding is special in all RNAPs because it is able to nucleate primer-independent transcription. The *de novo* transcription complex structure reported in this study identifies the *de novo* transcription specific interactions for binding the first NTP (**Figure 3.5 B**). A network of salt-bridges and hydrogen-bonds between the +1 NTP triphosphate and RNAP side chains are critical for the NTP binding. The +1 NTP binding site in the *de novo* transcription complex is identical to the i site of the transcription elongation complex, indicating that requirements of phosphodiester bond formation are similar in both transcription initiation and elongation complexes. The structures that I report here uses a template DNA with sequence closest to the natural primary σ -dependent promoters, which reveal a base stacking interaction between the purine at -1 of the template DNA, while the purine at +1 (ATP) provides added stability to the +1 NTP binding (**Figure 3.5 B**). This purine–purine base stacking during the *de novo*

transcription explains the preference of purine and pyrimidine at the -1 and $+1$ template DNA sequences of most bacterial σ^{70} (Shultzaberger et al., 2007) and σ^E (Rhodius and Mutalik, 2010) promoters as well as eukaryotic Pol II promoters (Butler and Kadonaga, 2002). A similar base pair stacking interaction was noticed in the structure of the single-subunit N4 phage RNAP *de novo* transcription complex (Gleghorn et al., 2011b), indicating that this mechanism may be universal.

Although the role of σ factor in promoter recognition is well established, its role in *de novo* transcription remains to be elucidated. Our structure demonstrates the absence of any direct contact between σ region 3.2 and the $+1$ ATP (**Figure 3.5 A**) (Kulbachinskiy and Mustaev, 2006; Severinov et al., 1994). σ region 3.2 seems to act as a weight on the template DNA to ideally position the DNA bases at $+1$ and $+2$, thereby supporting the binding of the incoming nucleotides at the i and $i + 1$ sites. Therefore, deletion of σ region 3.2 can force the template DNA around the transcription start site to adopt a floppy conformation that hampers stable binding of NTPs. The eukaryotic counterpart of σ region 3.2, the B-reader loop of TFIIB, analogously inserts its loop deep into the active site to ensure proper template positioning (Sainsbury et al., 2013).

After transcript initiation, when the transcribed RNA reaches an optimal length of 6 nts, its 5'-triphosphate reaches σ region 3.2, thereby causing the tip of σ region 3.2 to become disordered (**Figure 3.8 B**), most likely due to the repulsion between the RNA 5'-triphosphate and the acidic cluster in σ region 3.2. Transcription of 6-mer RNA may be the first step of σ ejection and the subsequent transition from transcription initiation to stable transcription elongation.

Structures of unstable intermediates and the significance of transcription initiation spanning *de novo* transcription, abortive initiation and the transition to elongation, have not been described adequately for cellular RNAPs. Further investigations using time-resolved trigger-freeze crystallography, could provide answers to long-standing questions such as acid–base mechanism of the nucleotidyl transfer and the nucleotide addition cycle. By demonstrating that the holoenzyme transcription complex is active and capable

of synthesizing up to 6-mer RNA within the crystals, I could establish a system that can be used to study transcription by Raman crystallography and time-resolved trigger-freeze crystallography. Experiments to trace events in transcription and nucleotide addition cycle in cellular RNAPs, analogous to the studies performed on single-subunit RNAPs (Basu and Murakami, 2013; Chen et al., 2011), will be required to answer these questions for cellular RNAP.

3.6 Acknowledgments

We thank Irina Artsimovitch and Shun-ichi Sekine for providing information of culturing *T. thermophilus* cells and RNAP purification. I thank the staff at The Penn State Fermentation Facility for helping me culture the *T. thermophilus* cells and MacCHESS for supporting crystallographic data collection. This work was supported by NIH grant GM087350. Figures were prepared using PyMOL.

3.6 References

- Afonine, P.V., Mustyakimov, M., Grosse-Kunstleve, R.W., Moriarty, N.W., Langan, P., and Adams, P.D. (2010). Joint X-ray and neutron refinement with phenix.refine. *Acta Crystallogr D* 66, 1153-1163.
- Basu, R.S., and Murakami, K.S. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J Biol Chem* 288, 3305-3311.
- Butler, J.E., and Kadonaga, J.T. (2002). The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes Dev* 16, 2583-2592.
- Chen, Y., Basu, R., Gleghorn, M.L., Murakami, K.S., and Carey, P.R. (2011). Time-resolved events on the reaction pathway of transcript initiation by a single-subunit RNA polymerase: Raman crystallographic evidence. *J Am Chem Soc* 133, 12544-12555.

Cheung, A.C., and Cramer, P. (2012). A movie of RNA polymerase II transcription. *Cell* *149*, 1431-1437.

Cheung, A.C., Sainsbury, S., and Cramer, P. (2011). Structural basis of initial RNA polymerase II transcription. *Embo J* *30*, 4755-4763.

Gleghorn, M.L., Davydova, E.K., Basu, R., Rothman-Denes, L.B., and Murakami, K.S. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc Natl Acad Sci U S A* *108*, 3566-3571.

Kennedy, W.P., Momand, J.R., and Yin, Y.W. (2007). Mechanism for *de novo* RNA synthesis and initiating nucleotide specificity by σ^70 RNA polymerase. *J Mol Biol* *370*, 256-268.

Kettenberger, H., Armache, K.J., and Cramer, P. (2004). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* *16*, 955-965.

Kulbachinskiy, A., and Mustaev, A. (2006). Region 3.2 of the sigma subunit contributes to the binding of the 3'-initiating nucleotide in the RNA polymerase active center and facilitates promoter clearance during initiation. *J Biol Chem* *281*, 18273-18276.

Liu, X., Bushnell, D.A., Silva, D.A., Huang, X., and Kornberg, R.D. (2011). Initiation complex structure and promoter proofreading. *Science* *333*, 633-637.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* *40*, 658-674.

Murakami, K.S. (2013). X-ray crystal structure of Escherichia coli RNA polymerase sigma70 holoenzyme. *J Biol Chem* *288*, 9126-9134.

Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* *13*, 31-39.

- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O., and Darst, S.A. (2002a). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296, 1285-1290.
- Murakami, K.S., Masuda, S., and Darst, S.A. (2002b). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296, 1280-1284.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276, 307-326.
- Rhodium, V.A., and Mutalik, V.K. (2010). Predicting strength and function for promoters of the *Escherichia coli* alternative sigma factor, sigmaE. *Proc Natl Acad Sci U S A* 107, 2854-2859.
- Sainsbury, S., Niesser, J., and Cramer, P. (2013). Structure and function of the initially transcribing RNA polymerase II-TFIIB complex. *Nature* 493, 437-440.
- Severinov, K., Fenyo, D., Severinova, E., Mustaev, A., Chait, B.T., Goldfarb, A., and Darst, S.A. (1994). The sigma subunit conserved region 3 is part of "5'-face" of active center of *Escherichia coli* RNA polymerase. *J Biol Chem* 269, 20826-20828.
- Shultzaberger, R.K., Chen, Z., Lewis, K.A., and Schneider, T.D. (2007). Anatomy of *Escherichia coli* sigma70 promoters. *Nucleic Acids Res* 35, 771-788.
- Vassilyev, D.G., Vassilyeva, M.N., Zhang, J., Palangat, M., Artsimovitch, I., and Landick, R. (2007). Structural basis for substrate loading in bacterial RNA polymerase. *Nature* 448, 163-168.
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127, 941-954.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* 119, 481-489.

Zhang, Y., Feng, Y., Chatterjee, S., Tuske, S., Ho, M.X., Arnold, E., and Ebright, R.H. (2012). Structural basis of transcription initiation. *Science* 338, 1076-1080.

Chapter 4

Summary and Future directions

My doctoral studies contributed to the current knowledge of the nucleotide addition cycle of both single-subunit and multi-subunit RNA polymerases (RNAPs). With time-resolved crystal structures of bacteriophage RNAP transcription complexes, I constructed the atomic resolution movie of the nucleotidyl-transfer reaction in single-subunit RNAP. On the other hand, with structures of bacterial transcription initiation complexes, I defined the structural basis of *de novo* initiation in multi-subunit RNAPs. Nevertheless, further studies on both projects are required to fill in remaining gaps of the transcription mechanism of RNAP. This chapter aims to provide the guiding steps for this endeavor. Studies on the single-subunit N4 mini-vRNAP can be extended to the broader molecular mechanism of transcription elongation, as it is suspected to be somewhat unique owing to its factor-dependence during RNA separation from template DNA. In this chapter, I presented the significance and progress toward X-ray crystal structure determination of the N4 mini-vRNAP transcription elongation complex. Additionally, studies on the *T. thermophilus* initiation complex need to be continued to address the persisting ambiguities regarding the early stages of transcription and the nucleotide addition cycle. Here, I will summarize these gaps and outline the impending experiments to clarify these issues.

4.1 Studies of N4 mini-vRNAP transcription

The first topic of my dissertation stemmed from an important structural study from our lab on N4 mini-vRNAP at its early stage of transcription. With the crystallization of the N4 mini-vRNAP - promoter DNA binary complex, structures of several transcription initiation complexes were solved at high-resolution (~ 2.0 Å), including those with nucleotides, their analogues, and metal cofactors (Gleghorn et al., 2011; **Section 1.6.3**). Together, these complexes (two substrate complexes trapped at pre-catalytic stages and a post-catalytic product complex) were arranged as structural

snapshots in the order of their occurrence along a proposed reaction pathway. The proposed mechanism for the nucleotidyl-transfer reaction, including nucleotide uptake, active site conformational changes and reaction catalysis, were in agreement with most biochemical kinetic data coming from A-family polymerases such as DNA pol I (Bermek et al., 2011a; Joyce et al., 2008) and T7 phage RNAP. However, there was no direct evidence for the proposed temporal order of these intermediates. To answer this question we collaborated with the Carey laboratory at Case Western Reserve University to apply their specialization of Raman spectroscopy to crystals of transcription complexes. This study was able to monitor the nucleotidyl-transfer reaction *in crystallo* in real-time, which added a broad temporal resolution to the reaction (Chen et al., 2011). With my experiments I intended to trap natural kinetic intermediates in transcription without using any non-reactive analogue. To this end I used a simple soak-trigger-freeze strategy to determine their high-resolution crystal structures. The resulting structures revealed unprecedented temporal details of active site dynamics based on the shapes and intensities of electron density around the active site (Basu and Murakami, 2013). This direct atomic view of transient, elusive intermediates of the nucleotidyl-transfer reaction in single-subunit A-family polymerases proceeded many steps forward in understanding the reaction mechanism. This work verified hypotheses based on elegant biochemical and biophysical studies on bacterial DNAP I, phage RNAPs, as well as mitochondrial DNAP and RNAP.

Beyond the core nucleotidyl-transfer reaction, the molecular mechanism of transcription initiation and elongation in single-subunit RNAPs is based on the studies of the single T7 RNAP. However, there is substantial divergence among the members of this polymerase family, including protein sequence, promoter types, and involvement of transcription factors. N4 mini-vRNAP is a prime example of such divergence, which has obvious consequences on the mini-vRNAP transcription mechanism. The most striking difference being its factor dependence (described in **Section 1.6**). The unique features of initiation were adequately addressed by the structure of N4 mini-vRNAP – promoter hairpin binary complex (Gleghorn et al., 2008). However, a further analysis may be performed to find the roles of promoter sub-regions in N4 mini-vRNAP activation,

including advances presented in this dissertation. Similarly, the mechanism of transcription elongation was also expected to be different in N4 mini-vRNAP and also other single-subunit RNAPs. This difference may be because many of the single-subunit RNAPs utilize transcription factors and sequences of their NTD, which are major determinants in the transition from initiation to elongation complex and are highly variable. To explore these differences, I carried out structural studies on N4 mini-vRNAP transcription elongation and my progress in this endeavor is presented in this chapter.

4.1.1 N4 mini-vRNAP activation by hairpin-form promoter DNA

Before I started my project of N4 mini-vRNAP transcription, the mechanism of promoter recognition had been described in detail through the structure of the N4 mini-vRNAP - promoter DNA complex, along with biochemical data (Gleghorn et al., 2008); **Section 1.6.2**). The N4 vRNAP promoter DNA plays an important role in changing the vRNAP conformation. This unique feature of N4 mini-vRNAP activation could be further dissected to identify the minimal promoter region that may trigger the structural transitions necessary for transcription initiation. Prior studies using two short hairpin promoters, one consisting of the hairpin only (HP5), and the other containing the hairpin plus two single stranded bases (HP3; **Table 4.1**) resulted in the initial crystallization condition of the N4 mini-vRNAP binary complex with these promoters. With the help of an undergraduate student, Scott Pandya, I was able to prepare well-diffracting crystals (2.4 Å) (**Figure 4.1**) to solve their structures.

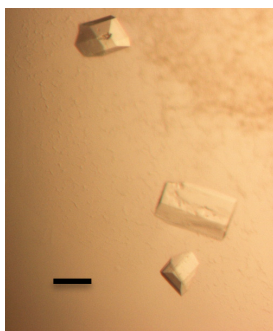


Figure 4.1: Crystals of mini vRNAP:HP3 binary complex. (The black line shows a scale of 100 μm)

Table 4.1 Crystallization condition for N4 mini-vRNAP short hairpin complex

Hairpin sequence	Crystallization condition	Crystal size	Final Cryoprotectant solution
$\begin{array}{c} \text{G}^{\text{CGAAGAA}^{5'}} \\ \text{A}^{\text{GCTTC}_3} \end{array}$ -4-3 HP3	25% PEG 3350 0.1 M Tris-HCl pH 9.0 0.2 M Ammonium dibasic citrate 0.5% w/v of n-Octyl- β -D-glucoside	200 μm x 80 μm x 40 μm	40% PEG 3350 0.1 M Tris-HCl pH 9.0 0.2 M Ammonium dibasic citrate
$\begin{array}{c} \text{G}^{\text{CGAAG}^{5'}} \\ \text{A}^{\text{GCTTC}_3} \end{array}$ -4 HP5	14% PEG 4000 0.1 M Na cacodylate pH 7.4 0.1 M MgSO_4	200 μm x 30 μm x 30 μm	Not determined

Table 4.2 Data collection and Refinement Statistics of N4 mini-vRNAP - short hairpin (HP3) binary complex

Crystal Parameters	
Space Group	P2 ₁ 2 ₁ 2 ₁
Cell Dimensions	
a, b, c (Å)	83.386, 113.697, 278.594
Data Collection	
Resolution (Å)	50-2.40 (2.43-2.40)*
R _{sym}	0.07 (0.78)*
I/ σ	11.34 (1.27)*
Completeness	85.6 (81.2)*
Redundancy	3.2 (2.6)*
Refinement	
Resolution (Å)	48.4-2.40
No. of reflections	88,544
R _{work} /R _{free}	22.12/28.48
R.M.S Deviations	
Bond Angle (°)	1.19
Bond Length (Å)	0.009

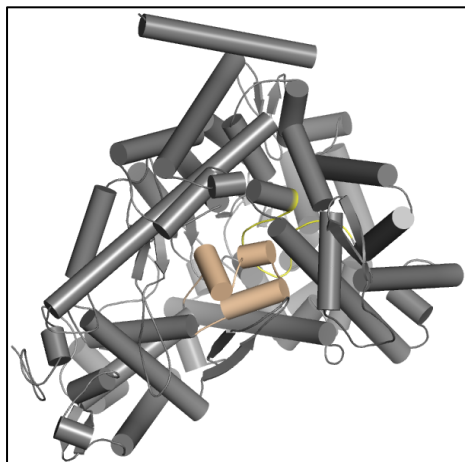
Data sets were collected remotely at Berkeley Centre for Structural Biology, Berkeley, CA.

*Highest resolution shell is shown in parenthesis

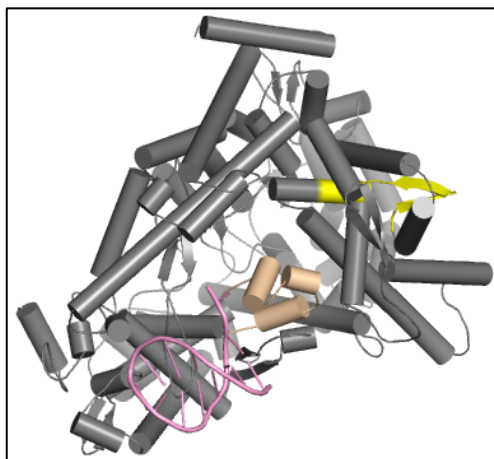
The structure of the N4 mini-vRNAP complex with the short hairpin was solved by molecular replacement with coordinates of the only mini-vRNAP structure as a search model. After rigid body and TLS refinement using Phenix (Adams et al., 2010a) the Fo-Fc electron density map showed ~14 bases of promoter hairpin DNA. The overall structure of both RNAP and DNA is almost identical to that of binary complex (BC) with regular length of promoter DNA (**Figure 4.2 B and C**). The BC with HP3 DNA (BC-HP3) shows the same trace for DNA and shares identical interactions of the hairpin head and stem with RNAP as in BC. Despite the lack of 10 bases of the single-stranded region including the +1 transcription start site, the plug module and the motif B loop that maintain the inactive-form vRNAP in the absence of promoter DNA are completely dislodged from the active site in BC-HP3 to generate an active-form RNAP (**Figure 4.2 A and B**). Only subtle differences were seen in the α -carbon traces of the plug module and the two 5' unpaired bases (**Figure 4.2 F**). Two amino acid side chains around the active site, R318 (NTD) and R421 (Thumb), are responsible for template DNA positioning in BC and are oriented differently in BC-HP3. (**Figure 4.2 G**). R318 which shares cation- π interactions with the -2 template DNA base and R421, which forms a salt bridge with the phosphate backbone between -2 and -1 of template DNA in the BC have displaced side chains due to the lack of the respective DNA bases. Thus this minimal promoter HP3 is enough to form the vRNAP-DNA complex and change vRNAP conformation to the active form, whereas the template positioning may only be done when bases up to -1 are present.

To further test whether the hairpin stem without the two base 5' overhang in the HP3 promoter was sufficient to form the active complex we tried crystallization of the minimum hairpin promoter HP5 using previously screened crystallization conditions by Gleghorn (Gleghorn, 2008). However the crystals were so thin that they showed no diffraction. We were also unable to improve the crystal size through optimization, maybe hinting at unstable binding of this hairpin to RNAP.

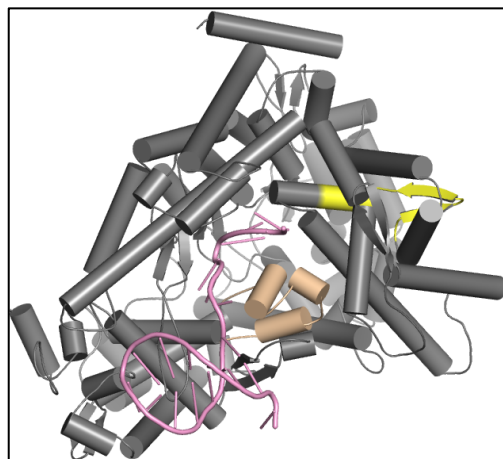
A



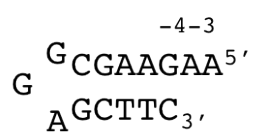
B



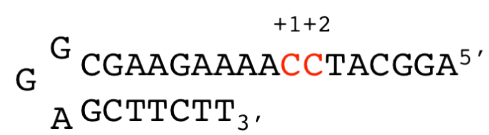
C



D



E



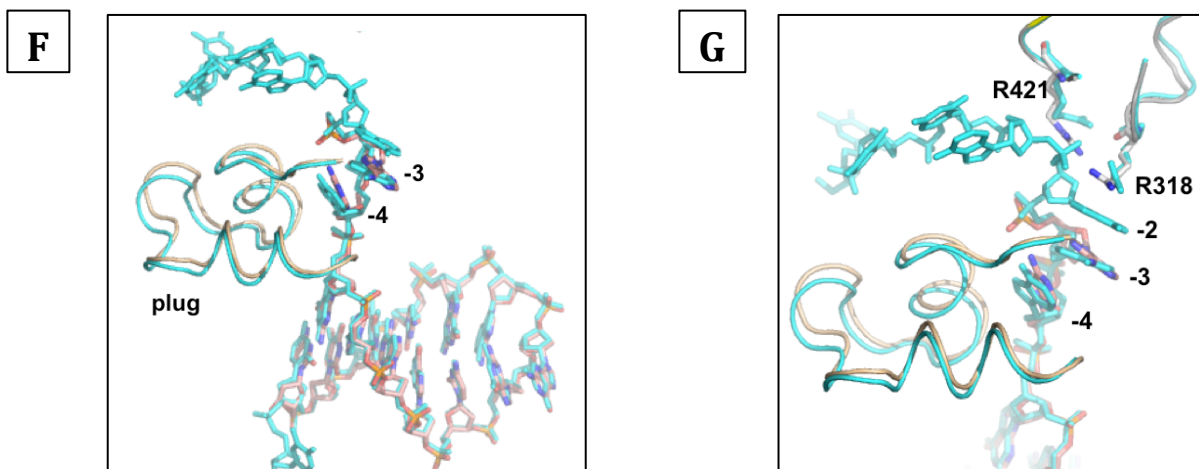


Figure 4.2: RNAP activation by short hairpin HP3. Structures of N4 mini-vRNAP apo-enzyme (A), complex with the short hairpin HP3 (B) and with hairpin promoter (C) are shown with cylindrical α -helices. RNAP is colored in grey, except for plug (wheat) and motif B loop (yellow). The promoter DNA is in pink. Sequences of the hairpin promoter (E) and the truncated hairpin, HP3 (D) are shown. (F) Only the promoter and the subtle changes of the plug and R318 and R421 are shown by superposing the HP3 complex onto the BC structure (cyan). The orientation is 180° rotated from the views in A-C. (G) is a closer view of (F).

4.1.2 Toward crystal structure determination of the N4 mini-vRNAP transcription elongation complex.

Transcription elongation in N4 mini-vRNAP is substantially different from T7 RNAP because the N4 polymerase requires external factor *EcoSSB* for displacing nascent transcript from template DNA (Davydova and Rothman-Denes, 2003). Moreover, the NTD that showed the largest conformational changes on the transition from initiation to elongation (Section 1.5) is highly divergent in sequence throughout the single-subunit RNAP family. Thus the N4-specific elongation requirements needed to be demonstrated by a structure.

Assembling the N4-mini-vRNAP elongation complex (EC) cannot rely on the straightforward spontaneous binding of RNAP with a DNA/RNA scaffold because of an active site occluded by the globular plug module and the motif B loop (Murakami et al., 2008). Thus the challenge for loading the DNA/RNA scaffold to form an EC is the displacement of these structural features. The plug module is stabilized in the active site

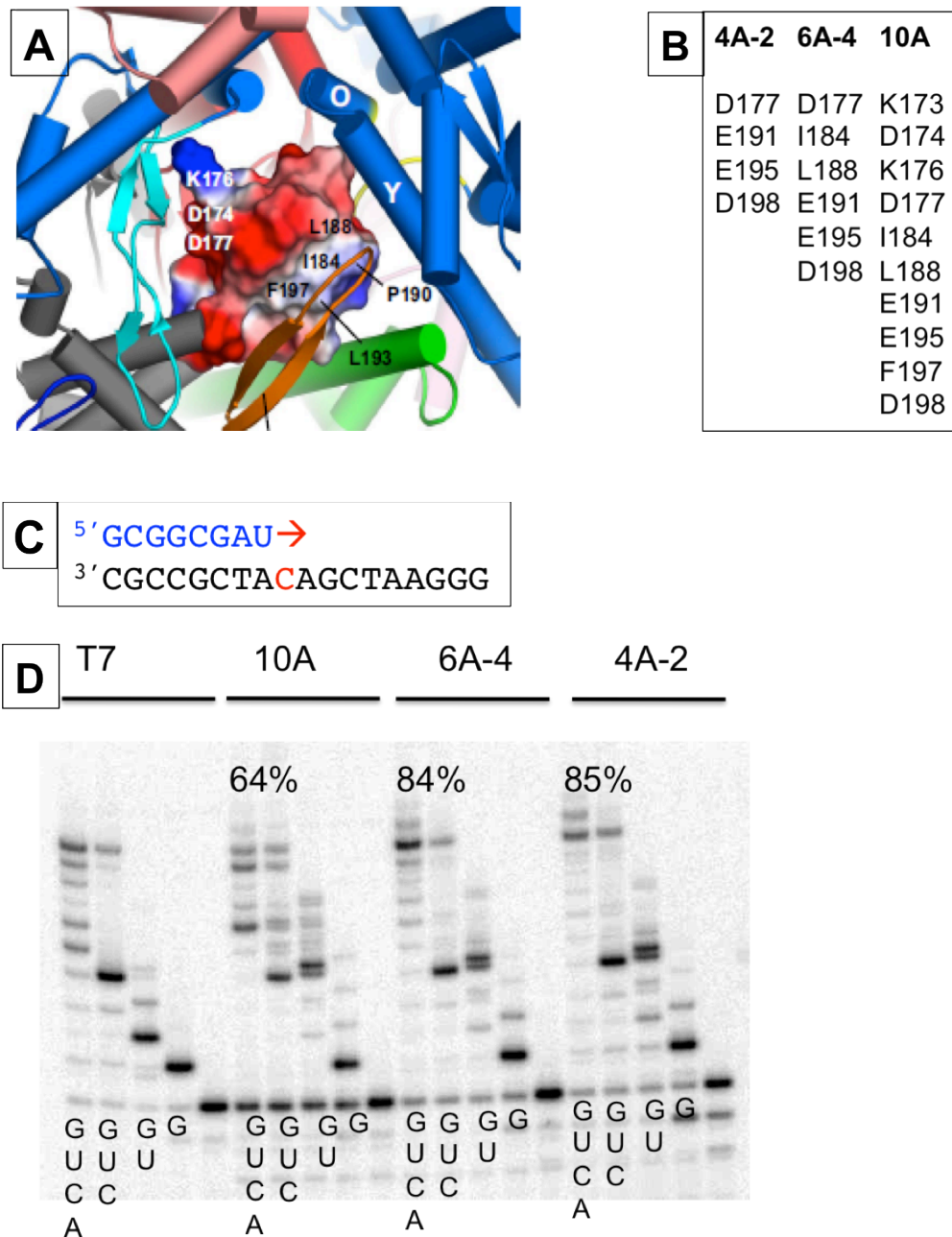


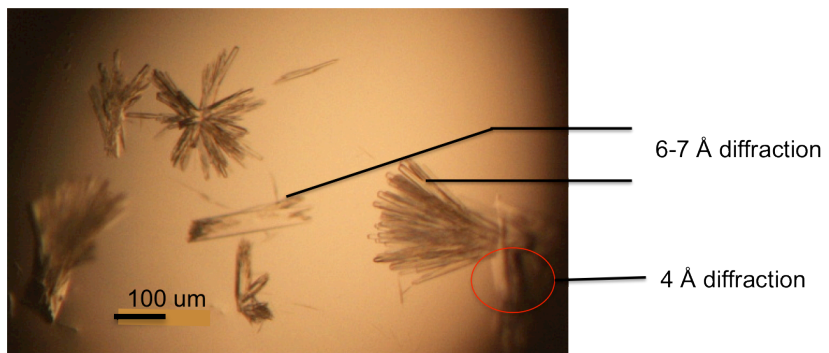
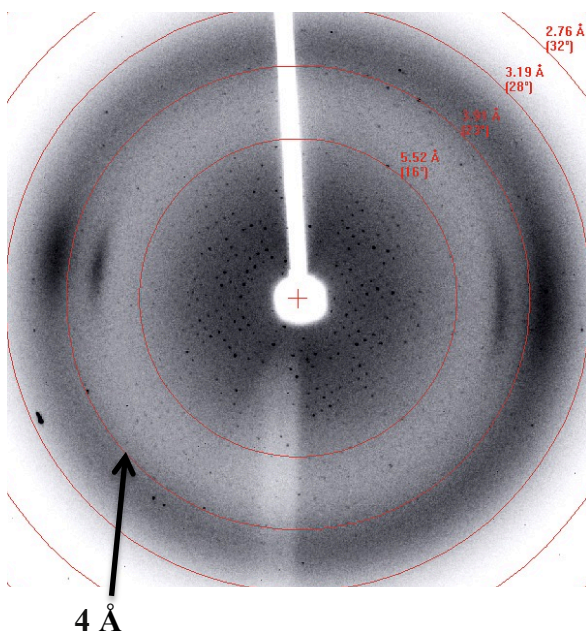
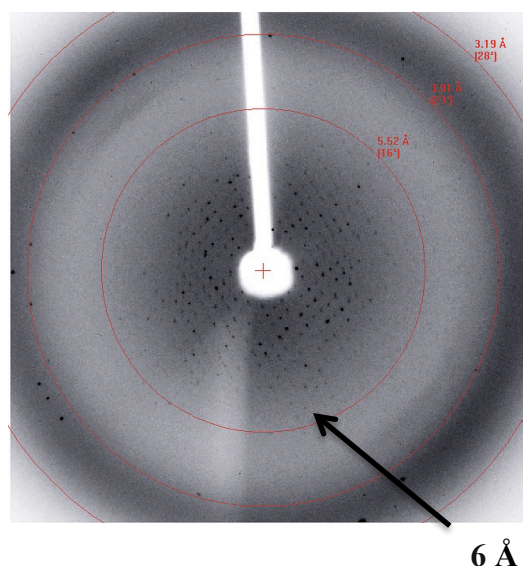
Figure 4.3 Assembly of N4 mini-vRNAP elongation complex. (A) Close-up view of the DNA binding channel, showing the plug as a surface representation with electrostatic distributions colored as red (-ve), white (neutral) and blue (+ve). The hydrophilic and hydrophobic surface residues interacting with the walls are labeled. (Figure taken from (Murakami et al., 2008)). Alanine mutants successful in scaffold loading and transcript elongation have been listed (B). The DNA (black)/RNA (blue) scaffold sequence (C) is the same as that used for T7 EC (Durniak et al., 2008), the first templating sequence (red) has been marked with an arrow. Transcription assay was performed on scaffold containing 5' P³²-labeled RNA (D). 1 μM scaffold was preincubated with 1 μM RNAP for 60 hrs at room temperature in buffer (20mM Tris pH 8.0, 15mM MgCl₂, 0.1% Tween 20, 5mM β-mercaptoethanol). Reaction was started with addition of 100 μM NTP mix at 37°C and stopped after 2 mins. Transcription assay was performed by Eda Koculi in the Rothman-Denes lab.

by means of both hydrophilic and hydrophobic surface interactions with the surrounding walls. (**Figure 4.3 A**; (Murakami et al., 2008). The Rothman-Denes group at The University of Chicago tried deleting the plug or parts of motif B loop, but that resulted in inactive RNAPs that were also unable to load the DNA/RNA scaffold. The same group also performed alanine scanning mutagenesis on the surface charged residues of the plug and succeeded in obtaining three multi-alanine mutants that could load an DNA/RNA scaffold and extend the RNA primer (**Figure 4.3 B and C**).

I used the same conditions as the transcription assay (**Table 4.3**) to assemble the EC and set up crystallization trials for each of the three mutant polymerases (4A-2, 6A-4 and 10A) with commercial screening kits. The most promising condition was obtained with the 6A-4 mutant protein using JCSG+ Suite (Qiagen), which upon optimizing gave me mostly needle shaped crystals, the best of which had an approximate size of 200 μm x 60 μm x 60 μm (**Figure 4.4 A**). I tested their diffraction at the home source X-ray facility and the best resolution I obtained was 4 Å. The crystals tested positive for the presence of RNA and the ability to extend the transcript using fluoro-labeled and radio-labeled UTP (**Figure 4.4 B and C**)

Table 4.3 Crystallization of N4 mini-vRNAP elongation complex

Scaffold sequence	EC assembly conditions	Crystallization condition	Other Comments
5'CGGCGAU→ 3'CGCCGCTACAGCTAAGGG	10 mg/ml N4 mini-vRNAP incubated with 1:1 molar ratio of DNA/RNA scaffold 60 hours at room temperature Buffer – 10 mM Tris pH 8.0, 20 mM NaCl	18% PEG 3350 0.15 M Mg Formate 3% v/v DMSO	Good quality crystals also obtained with detergent – decyl maltoside as additive. Best crystals were formed in microbatch under 1:1 paraffin oil and Al's oil.

A**B****C**

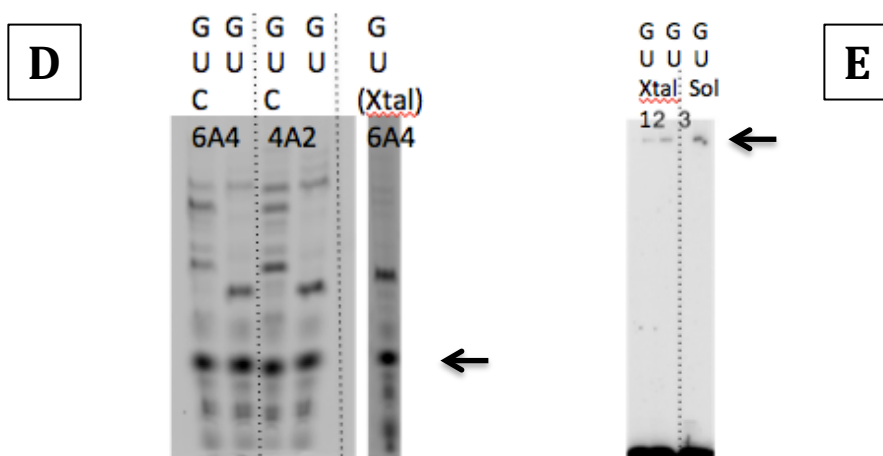


Figure 4.4. Crystals of N4 mini-vRNAP elongation complex (A). Preliminary crystals had frayed edges. The best diffraction obtained was from a central portion of the crystal cluster (C), but typical diffraction of crystals were ~ 6 Å (D). Labeled resolution arcs are marked in red on the diffraction pattern. 2-3 Crystals were washed in crystallization solution, dissolved in transcription buffer and tested for primer extension using Fluor-labeled UTP (D) and radiolabeled UTP (E). On comparing the gel bands (marked with arrows) with same reaction in solution testified its ability to extend RNA.

However, the diffraction quality and reproducibility of the crystals are sensitive to the age of RNA and therefore I could not collect any diffraction datasets with these crystals. Further optimization trials were performed, which included changing the His₆-tag from the C-terminus to N-terminus of RNAP and Seleno-methionine labeled RNAP expression, none of which could produce crystals. The hurdle for reproducing and improving these crystals is probably the long incubation time (60 hrs) for EC formation. Conditions capable of making high percentages of EC in smaller incubation time need to be optimized. Since RNA is sensitive to degradation it needs to be annealed to the elongation scaffold every time before use. Preliminary ITC results showed that the binding affinity of the scaffold is quite low (~ 10 μ M). Hence, there might be considerable heterogeneity in the EC solution. To ensure a homogeneous population of only scaffold loaded complex, the RNAP-scaffold mix could be passed through a heparin column post-incubation and the elution (which should contain the scaffold bound RNAP) may be used for crystallization. Another approach that could be tried is to extend the RNA by a single-base, to cause a more homogeneous population of the transcription-ready conformation of RNAP.

4.2 Studies on *T. thermophilus* RNAP initiation complex

My structures of the ternary bacterial initiation complex provided an explanation for the *de novo* substrate binding and initiation of transcription. Previously, this feature of RNAP had only been structurally explained in single-subunit RNAP. However, my structure likely provides a structural basis for the mechanism in all multi-subunit RNAPs, linking structure to the universal genetic preference of promoter sequence around the transcription start site. Another of my structures explains the structural basis for the transition of a bacterial initiation complex to the stable elongation complex. Although the role of participating structural features had been proposed using previous native structures of template-bound RNAP and modeling a certain length of RNA, my structure is the first direct evidence showing that ejection of the σ region 3.2 is the first step of transition between initiation and elongation.

Several steps of the transcription mechanism in multi-subunit RNAP are still based on biophysical studies and structures of complexes trapped in specific conditions. However, complete elucidation of some of the dynamics during the nucleotide addition cycle, such as the exact synchrony of the trigger loop and bridge helix motions with NTP binding and translocation could highly benefit from a time-resolved structural characterization. Our results of *in crystallo* transcription of the 6-mer transcript in the *T. thermophilus* initiation complex, promises to reveal direct evidence for some of these enzyme kinetics. Additionally, another aspect of multi-subunit RNAP transcription that is not well understood is the transition from abortive initiation to elongation. Since this is a complex multi-step process it may also be dissected with our *in crystallo* time-resolved approach.

4.3 Acknowledgements

I thank Scott Pandya for setting crystal trays for the N4 mini-vRNAP:HP3 complex, Hemant Yennawar and Neela Yennawar at the Penn State X-ray core facility to support crystal screening experiments. I thank Eda Koculi from Lucia Rothman-Denes' lab for

establishing the conditions for elongation complex formation in N4 mini-vRNAP, all the transcription assays and providing me with the mutant constructs for crystal work.

4.4 References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* *66*, 213-221.

Basu, R.S., and Murakami, K.S. (2013). Watching the bacteriophage N4 RNA polymerase transcription by time-dependent soak-trigger-freeze X-ray crystallography. *J Biol Chem* *288*, 3305-3311.

Bermek, O., Grindley, N.D., and Joyce, C.M. (2011). Distinct roles of the active-site Mg²⁺ ligands, Asp882 and Asp705, of DNA polymerase I (Klenow fragment) during the prechemistry conformational transitions. *J Biol Chem* *286*, 3755-3766.

Chen, Y., Basu, R., Gleghorn, M.L., Murakami, K.S., and Carey, P.R. (2011). Time-resolved events on the reaction pathway of transcript initiation by a single-subunit RNA polymerase: Raman crystallographic evidence. *J Am Chem Soc* *133*, 12544-12555.

Davydova, E.K., and Rothman-Denes, L.B. (2003). Escherichia coli single-stranded DNA-binding protein mediates template recycling during transcription by bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci U S A* *100*, 9250-9255.

Durniak, K.J., Bailey, S., and Steitz, T.A. (2008). The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* *322*, 553-557.

Gleghorn, M.L., Davydova, E.K., Basu, R., Rothman-Denes, L.B., and Murakami, K.S. (2011). X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides. *Proc Natl Acad Sci U S A* *108*, 3566-3571.

Gleghorn, M.L., Davydova, E.K., Rothman-Denes, L.B., and Murakami, K.S. (2008). Structural basis for DNA-hairpin promoter recognition by the bacteriophage N4 virion RNA polymerase. *Mol Cell* *32*, 707-717.

Gleghorn, M.L. (2008). X-ray crystal structures explain DNA hairpin promoter binding and *de novo* transcription initiation for the N4 virion RNA polymerase. (Doctoral dissertation)

Joyce, C.M., Potapova, O., Delucia, A.M., Huang, X., Basu, V.P., and Grindley, N.D. (2008). Fingers-closing and other rapid conformational changes in DNA polymerase I (Klenow fragment) and their role in nucleotide selectivity. *Biochemistry* *47*, 6103-6116.

Murakami, K.S., Davydova, E.K., and Rothman-Denes, L.B. (2008). X-ray crystal structure of the polymerase domain of the bacteriophage N4 virion RNA polymerase. *Proc Natl Acad Sci U S A* *105*, 5046-5051.

VITA
Ritwika Basu

EDUCATION

Ph.D. Biochemistry, Microbiology and Molecular Biology, The Pennsylvania State University, University Park, PA, USA (2014)

M.Sc. Biophysics, Molecular Biology and Genetics, Calcutta University, Kolkata, India (2008)

B.Sc. Chemistry, Jadavpur University, Kolkata, India (2006)

HONORS

1. Best Student Paper Prize from Cornell High Energy Synchrotron Source, Ithaca, NY (2012)
2. Best Student Poster award at the FASEB Summer Research Conference: Nucleic Acid Enzymes, Snowmass Village, Colorado, (2012)
3. Paul Berg Prize in Molecular Biology for outstanding graduate student research, Eberly College of Science, Pennsylvania State University (2011)
4. Homer F. Braddock, Nellie H. and Oscar L. Roberts Fellowships from Department of Biochemistry and Molecular Biology, Pennsylvania State University (2008)

PUBLICATIONS

1. **Basu, R. S.**, Warner, B. A. and Murakami, K. S. “Structural basis of transcription initiation by bacterial holoenzyme.” *Manuscript in preparation*.
2. **Basu, R. S.** and Murakami, K. S. (2013) Bacteriophage RNA polymerases (Chapter 10; pp 237 – 250) in *Nucleic Acid Polymerases* (eds. Murakami, K. S. and Trakselis, M. A.) Copyright © Springer-Verlag Berlin Heidelberg 2014
3. **Basu, R.**, Murakami, K. S. (2013). “Watching the bacteriophage N4 RNA polymerase transcription by Time-dependent soak-trigger-freeze X-ray Crystallography” *J. Biol. Chem.* 288, 3305-3311.
4. Chen, Y., **Basu, R.**, Gleghorn, M. L., Murakami, K. S., Carey, P. R. (2011). “Time-Resolved Events on the Reaction Pathway of Transcript Initiation by a Single-Subunit RNA Polymerase: Raman Crystallographic Evidence.” *J. Am. Chem. Soc.* 133, 12544–12555.
5. Gleghorn, M.L., Davydova, E.K., **Basu, R.**, Rothman-Denes, L.B., and Murakami, K.S. (2011). “X-ray crystal structures elucidate the nucleotidyl transfer reaction of transcript initiation using two nucleotides.” *Proc Natl Acad Sci U S A* 108, 3566-3571.

PRESENTATIONS

1. **Poster Presentation:** “X-ray structural snapshots of bacterial *de novo* transcription initiation”, FASEB Summer Research Conference: Mechanism and Regulation of Prokaryotic Transcription, Vermont Academy, Saxtons River, Vermont, June 2013
2. **Oral Presentation:** “Viewing enzyme motions in real-time by X-ray crystallography: Studies of nucleotidyl-transfer reaction in bacteriophage RNA polymerase”, FASEB Summer Research Conference: Nucleic Acid Enzymes, Snowmass Village, Colorado, June 2012
3. **Poster Presentation:** “Structure-based kinetics of the nucleotidyl transfer reaction by time-resolved trigger-freeze X-ray crystallography”, FASEB Summer Research: Nucleic Acid Enzymes, Snowmass Village, Colorado, June 2012
4. **Poster Presentation:** “Structure-based kinetics of the nucleotidyl transfer reaction by time-resolved trigger-freeze X-ray crystallography”, FASEB Summer Research Conference: Mechanism and Regulation of Prokaryotic Transcription, Vermont Academy, Saxtons River, Vermont, June 2011