

The Pennsylvania State University

The Graduate School

STRUCTURAL BASIS OF BACTERIAL TRANSCRIPTION
INVESTIGATION OF MECHANISMS OF REITERATIVE TRANSCRIPTION AND
RIBOSOMAL RNA TRANSCRIPTION BY X-RAY CRYSTALLOGRAPHY AND CRYO-
ELECTRON MICROSCOPY

A Dissertation in
Biochemistry, Microbiology and Molecular Biology

by
Yeonoh Shin

© 2020 Yeonoh Shin

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2020

The dissertation of Yeonoh Shin was reviewed and approved by the following:

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

Paul Babitzke
Professor of Biochemistry and Molecular Biology

David Gilmour
Professor of Biochemistry and Molecular Biology

Lu Bai
Associate Professor of Biochemistry and Molecular Biology
Associate Professor of Physics

Philip C. Bevilacqua
Professor of Chemistry and Biochemistry and Molecular Biology

Wendy Hanna-Rose
Department Head, Biochemistry and Molecular Biology
Professor of Biochemistry and Molecular Biology

ABSTRACT

Visualizing the three-dimensional structure of a macromolecule in a biological system provides many new insights into the function and mechanism of the biological process. This can be achieved by solving the atomic resolution structures of macromolecules by X-ray crystallography and cryo-electron microscopy (cryo-EM).

I investigated the mechanism of two different types of bacterial transcription systems by structural biology approaches. The first type is conditional reiterative transcription which regulates gene expression by nucleoside triphosphate substrate availability. Using time-dependent soak-trigger-freeze X-ray crystallography, I determined the transcript initiation and reiterative transcription complex structures from the *Bacillus subtilis* *pyrG* promoter that show 1) the unique pathway of RNA extension; and 2) a mismatch between template DNA and the incoming nucleotide followed by base sharing with an upstream template DNA during the reiterative RNA transcription. I also determined the structures of transcript initiation complexes from the *Escherichia coli* *pyrBI* promoter and *B. subtilis* *pyrG* promoter variants that show an alternative mechanism of RNA slippage and extension pathway compared with the *pyrG* promoter.

Ribosomal RNA (rRNA) is the highest expressed gene in rapidly growing bacteria; however, it is drastically downregulated by the DksA/ppGpp binding to RNAP during stationary phase and stressed conditions. I investigated the mechanism of extraordinary active rRNA transcription and its DksA/ppGpp-dependent inhibition by cryo-EM structure determination of RNAP and rRNA promoter complex and biochemical assays. The comparison of the apo-form of RNAP, closed and open complex structures, and DNA opening assay reveals the role of σ non-conserved region (σ NCR) rotation in facilitating DNA unwinding. The structures also show the structural basis of RNAP-UP element binding and the “open complex scrunching” by which the template DNA contacts with σ finger and β' lid and transcript start site selection at the rRNA

promoter. DksA/ppGpp binding to RNAP induces the movement of RNAP mobile domains that influence the stability of the open complex and hinder the DNA unwinding of promoters containing GC-rich discriminator sequence. I propose that promoter-dependent pleotropic effect of DksA/ppGpp is based on the duplex stability in discriminator sequences.

TABLE OF CONTENTS

LIST OF FIGURES	ix
LIST OF TABLES	xii
ABBREVIATIONS	xiii
ACKNOWLEDGEMENTS	xv
Chapter 1. Introduction to transcription by bacterial RNA polymerase	1
1.1 Abstract	2
1.2 Structure of bacterial RNA polymerase	2
1.2.1 Structural biology: X-ray crystallography and cryo-electron microscopy	2
1.2.2 Overall structure of bacterial RNA polymerase	4
1.2.3 DNA binding main channel, secondary channel and RNA exit channel of RNAP	9
1.2.4 Active site of bacterial RNA polymerase	10
1.2.5 Nucleotidyl transfer reaction	12
1.3 Bacterial transcription cycle	13
1.3.1 Initiation	13
1.3.2 Elongation	14
1.3.3 Termination	15
1.4 Dissertation format	17
1.5 References	18
Chapter 2. Structural basis of the productive reiterative transcription from the <i>pyrG</i> promoter	24
2.1 Abstract	25
2.2 Reiterative transcription	25
2.2.1 Productive reiterative transcription – <i>pyrG</i> operon of <i>B. subtilis</i>	26
2.2.2 Using <i>T. thermophilus</i> RNAP crystalline state to study the mechanism of transcription	28
2.3 Materials and Methods	30
2.3.1 Fermentation of <i>T. thermophilus</i> and Purification and Crystallization of <i>T.</i> <i>thermophilus</i> RNAP	30
2.3.2 X-ray Data collections and Structure Determinations	30
2.4 Results	33
2.4.1 The structure of the RNAP- <i>pyrG</i> promoter reiterative transcription complex at 30 min	33
2.4.2 Capturing transcription initiation complex from the <i>pyrG</i> promoter	39
2.4.3 Comparison of the distances between bases to characterize hydrogen bonds	42
2.4.4 Monitoring DNA translocation state during reiterative transcription in solution using 2-aminopurine fluorescence signal	43

2.4.5 Measurement of DNA translocation kinetics during reiterative transcription.....	47
2.5 Discussion.....	48
2.5 Acknowledgements.....	52
2.5 References.....	52
Chapter 3. Structural basis of non-productive reiterative transcription from the <i>pyrBI</i> and <i>pyrG</i> promoter variants	57
3.1 Abstract.....	58
3.2 Non-productive reiterative transcription – <i>pyrBI</i> operon of <i>E. coli</i>	58
3.3 Materials and Methods.....	61
3.3.1 Purification and crystallization of <i>T. thermophilus</i> RNAP	61
3.3.2 X-ray Data collections and Structure Determinations.....	61
3.3.3 <i>In vitro</i> transcription assay	64
3.4 Results.....	64
3.4.1 <i>In vitro</i> transcription assay using <i>pyrBI</i> promoter DNA	64
3.4.2 Structure of a transcript initiation complex from the <i>pyrBI</i> promoter.....	65
3.4.3 Investigation of the role of -1 base of template DNA in reiterative transcription.....	66
3.4.4 Structural analysis of the role of -1 base of template DNA from the <i>pyrG</i> promoter in reiterative transcription.....	69
3.5 Discussion.....	71
3.6 Acknowledgements.....	73
3.7 References.....	73
Chapter 4. Structural basis of ribosomal RNA transcription regulation	76
4.1 Abstract.....	77
4.2 Ribosomal RNA (rRNA) transcription in <i>E. coli</i>	78
4.3 Cryo-electron microscopy (Cryo-EM).....	82
4.4 Materials and Methods.....	84
4.4.1 Purification of <i>E. coli</i> RNA polymerase σ^{70} holoenzyme	84
4.4.2 Purification of <i>E. coli</i> RNA polymerase mutant	88
4.4.3 Preparation of <i>rrnBP1</i> dsDNA	88
4.4.4 Preparation of RNAP- <i>rrnBP1</i> DNA complex without or with DksA/ppGpp and vitrification of cryo-EM grids	88
4.4.5 Cryo-EM data collection	89
4.4.6 Cryo-EM image processing.....	91
4.4.7 Structure refinement.....	91
4.4.8 DNA opening assay by monitoring Cy3 fluorescence signal.....	92
4.4.9 DNA duplex free energy calculation.....	92
4.5 Results.....	93
4.5.1 Cryo-EM structure of the RNAP- <i>rrnBP1</i> promoter complex (closed and open complexes).....	93
4.5.1.1 The RNAP- <i>rrnBP1</i> promoter closed complex (RPc).....	93
4.5.1.2 The RNAP- <i>rrnBP1</i> promoter open complex (RPo).....	98
4.5.2 Cryo-EM structure of the RNAP- <i>rrnBP1</i> promoter complex in the presence of DksA/ppGpp (RP-DksA/ppGpp)	101

4.5.2.1 Binding modes of DksA/ppGpp to RNAP	101
4.5.2.2 RP-DksA/ppGpp (I) structure represents an early stage intermediate (RPi1) on the pathway to RPo	104
4.5.2.3 RP-DksA/ppGpp (II) structure represents a late stage intermediate (RPi2) on the pathway to RPo	107
4.5.3 Monitoring DNA opening in solution using Cy-3 fluorescence signal	109
4.5.4 DNA duplex free energy calculation.....	110
4.6 Discussion	112
4.7 Acknowledgements	117
4.8 References.....	117
Chapter 5. Summary and Future Directions.....	124
5.1 Abstract	124
5.2 Reiterative transcription	125
5.2.1 Current state of knowledge	125
5.2.2 Future directions.....	126
5.2.2.1 Visualizing the non-productive reiterative RNA by determining the cryo-EM structure of the reiterative transcription complex.....	126
5.2.2.2 How poly-G RNA from the <i>pyrG</i> promoter switches to canonical transcription?	127
5.2.2.3 How about other <i>pyrG</i> homopolymeric tract?	131
5.3 Ribosomal RNA transcription regulation.....	134
5.3.1 Current state of knowledge	134
5.3.2 Future directions.....	136
5.3.2.1 Biochemical assay using RNAP derivatives to test DNA unwinding efficiency	136
5.3.2.2 Cryo-EM structures of transcription initiation complex and elongation complex.....	137
5.3.2.3 A RNAP inhibitor ureidothiophene as a control compound in testing open complex formation	139
5.4 Cryo-EM structure of human DNA polymerase δ replication complex.....	141
5.4.1 Human DNA polymerase δ (Pol δ)	141
5.4.2 PCNA, RFC, and RPA	143
5.4.3 Cryo-EM structure of yeast DNA polymerase δ (Pol δ)	145
5.4.4 Materials and Methods	147
5.4.4.1 Purification of human DNA polymerase δ (Pol δ)	147
5.4.4.2 Vittrification of cryo-EM grids.....	151
5.4.4.3 Cryo-EM data collection	151
5.4.4.4 Processing cryo-EM data.....	153
5.4.5 Results	153
5.4.6 Future directions.....	155
5.5 Acknowledgements	156
5.6 References.....	156
Appendix A Fermentation of <i>T. thermophilus</i> HB27 cells	160
Appendix B Purification of <i>T. thermophilus</i> RNAP σ^A holoenzyme.....	161

Appendix C Crystallization of <i>T. thermophilus</i> RNAP σ^A holoenzyme – Promoter DNA complex.....	166
Appendix D Processing of cryo-EM data with Relion	169
Appendix E Processing of cryo-EM data with cryoSPARC	170

LIST OF FIGURES

Figure 1.1: Structural overview of <i>E. coli</i> RNAP- σ^{70} holoenzyme.....	6
Figure 1.2: Structural overview of <i>T. thermophilus</i> RNAP- σ^{70} holoenzyme.....	7
Figure 1.3: Scheme of RNAP core enzyme assembly..	8
Figure 1.4: The general organization of group 1 house keeping σ factor.	9
Figure 1.5: The overview of the active site of bacterial RNAP.	11
Figure 1.6: Mechanism of Two-Mg ²⁺ - mediated catalysis of nucleotidyl transfer reaction ...	13
Figure 1.7: The overview of bacterial transcription cycle	16
Figure 2.1: Model for CTP-mediated regulation of <i>pyrG</i> expression in <i>B. subtilis</i>	28
Figure 2.2: <i>T. thermophilus</i> RNAP crystalline state.	29
Figure 2.3: The fork-DNA used for crystallization.....	30
Figure 2.4: The structure of the reiterative transcription complex at 30 min (RTC-30') reveals the location of the reiterative transcript within RNAP.	35
Figure 2.5: Comparison of RNA paths within different transcription complexes.	36
Figure 2.6: Characterization of reiterative transcription.	37
Figure 2.7: Characterization of the <i>pyrG</i> homopolymeric tract that determines the maximum length of reiterative transcripts.	38
Figure 2.8: Experimental scheme of the reiterative transcription complex preparation by time-dependent soak-trigger-freeze X-ray crystallography.	39
Figure 2.9: The structures of the RTC-3' (A), RTC-5' (B) and RTC-7' (C) from the <i>pyrG</i> promoter.	41
Figure 2.10: Monitoring equilibrium DNA translocation state in solution using fluorophore 2-aminopurine.	45
Figure 2.11: Control experiment using 2-AP to monitor translocation state of DNA in solution.....	46
Figure 2.12: Kinetics analysis of DNA translocation rate.	48
Figure 2.13: Proposed model of RNA slippage during reiterative transcription from the <i>pyrG</i> promoter.....	51

Figure 3.1: Model for UTP-mediated regulation of <i>pyrBI</i> expression in <i>E. coli</i>	60
Figure 3.2: Characterization of reiterative transcription from the <i>pyrBI</i> promoter.....	65
Figure 3.3: The structures of the RNAP- <i>pyrBI</i> promoter binary complex (A) and transcription initiation complex (B).....	66
Figure 3.4: Characterization of the role of the -1 template DNA base of the <i>pyrG</i> promoter in reiterative transcription.....	68
Figure 3.5: Structural analysis of the role of the -1 template DNA base of the <i>pyrG</i> promoter in reiterative transcription.....	70
Figure 4.1: Sequences of the seven rRNA operons in <i>E. coli</i>	78
Figure 4.2: Sequences of the <i>rrnBP1</i> (A), <i>T7A1</i> (B) and <i>argI</i> (C) promoters.....	79
Figure 4.3: RNAP – promoter DNA interactions on the pathway to forming RPo	81
Figure 4.4: Overview of cryo-EM structure determination workflow and vitrified specimen.....	83
Figure 4.5: Purification of <i>E. coli</i> RNAP σ^{70} holoenzyme.....	87
Figure 4.6: Data collection strategy	90
Figure 4.7: The cryo-EM density map of <i>E.coli</i> RNAP and <i>rrnBP1</i> promoter DNA closed complex (RPC).....	94
Figure 4.8: The cryo-EM structure showing the interactions between the α CTDs of <i>E. coli</i> RNAP and UP element of <i>rrnBP1</i> promoter DNA.	95
Figure 4.9: The interactions between the -10 element and σ region 2.3	95
Figure 4.10: The interactions between the σ NCR and B'CT	97
Figure 4.11: The cryo-EM structure of the <i>E.coli</i> RNAP and <i>rrnBP1</i> promoter DNA open complex (RPo)	100
Figure 4.12: The structure of DksA	102
Figure 4.13: Cryo-EM maps of the (A) RP-DksA/ppGpp (I) and (B) RP-DksA/ppGpp (II) ..	103
Figure 4.14: Magnified views of RP-DksA/ppGpp (I) (A) and RP-DksA/ppGpp (II) (B)	106
Figure 4.15: The conformational changes of RNAP induced by the DksA/ppGpp binding....	108
Figure 4.16: DNA opening assay using Cy-3 labeled promoters.....	110
Figure 4.17: Promoter opening depends on the discriminator sequence.....	111

Figure 4.18: Cartoon model showing the dynamics of the σ NCR during the DNA binding and opening	115
Figure 4.19: Schematic representations of the promoter-specific transcription regulation by DksA/ppGpp	116
Figure 5.1: Cartoon model showing switching from the reiterative transcription complex to the elongation complex	128
Figure 5.2: Experimental scheme of the complex preparation to determine ntermediate structures switching from the reiterative transcription to the canonical transcription by time-dependent soak-trigger-freeze X-ray crystallography	128
Figure 5.3: The electron density maps (3.8Å) of the active sites of RNAPs	130
Figure 5.4: Experimental scheme of the complex preparation and the structure	130
Figure 5.5: The fork-DNAs used for <i>in crystallo</i> transcription containing various slippage-prone homopolymeric tract sequence found in the <i>pyrG</i> promoter	131
Figure 5.6: Electron density maps of the active sites of RNAPs with the <i>pyrG</i> homopolymeric tract variants.....	123
Figure 5.7: The <i>rrnBP1</i> promoter DNA with a substitution at +4 position	138
Figure 5.8: Cryo-EM structure of the active site of RNAP containing 6-mer RNA.....	138
Figure 5.9: Investigation of the binding mode of ureidothiophene (Urd).....	140
Figure 5.10: DNA polymerase structure and catalysis	142
Figure 5.11: PCNA, RFC, and RPA	144
Figure 5.12: The cryo-EM structure of DNA bound yeast Pol δ holoenzyme.....	146
Figure 5.13: Proposed model of yeast Pol δ -PCNA complex.....	147
Figure 5.14: Purification of human Pol δ	149
Figure 5.15: Representative images of human Pol δ holoenzyme on C-Flat Holey Carbon grids (Cu 2/1, 400 mesh)	152
Figure 5.16: The yeast Pol δ -DNA complex structure and human Pol δ density map at 10 Å resolution.....	154
Figure 5.17: Schematic of Pol δ replication complex preparation.....	155

LIST OF TABLES

Table 2.1: Data collection and refinement statistics of binary complex and RTC-30'	31
Table 2.2: Data collection and refinement statistics of RTC intermediates.....	32
Table 3.1: Data collection and refinement statistics of <i>pyrBI</i>	62
Table 3.2: Data collection and refinement statistics of -1 tDNA variants	63

ABBREVIATIONS

2-AP – 2-aminopurine

β 'CT - β ' clamp toe

σ NCR – σ non conserved region

BH – bridge helix

B-ME - mercaptoethanol

B. subtilis – Bacillus subtilis

cryo-EM – Cryo electron microscopy

dDNA – downstream DNA

DPBB – double psi beta-barrel

dsDNA – double stranded DNA

DTT - dithiothreitol

E. coli – Escherichia coli

ITR – initially transcribed region

kDa – kilo dalton

ntDNA – non template DNA

NTP – nucleoside triphosphate

PCNA – proliferating cell nuclear antigen

PMSF - phenylmethylsulfonyl fluoride

Pol δ – human DNA polymerase δ

ppGpp – guanosine tetraphosphate

RFC – replication factor C

RNAP – RNA polymerase

RPc – bacterial RNA polymerase closed complex

RPO – bacterial RNA polymerase open complex

γ -proteins – ribosomal proteins

rRNA – ribosomal RNA

RTC – reiterative transcription complex

tDNA – template DNA

TH – trigger helix

TIC – transcript initiation complex

TL – trigger loop

T. thermophilus – *Thermus thermophilus*

ACKNOWLEDGEMENTS

First of all, I would like to thank my dissertation advisor, Dr. Katsuhiko Murakami, for giving me the opportunity to work under his guidance and the support, motivation, enthusiasm and patience he has shared with me through this thesis project. His recommendations and valuable advice were vital in successful completion of this dissertation. I'd especially like to thank him for constant encouragement and support in learning new experimental approaches and pursuing future career.

I would also like to express gratitude to the committee members, Drs. Paul Babitzke, David Gilmour, Lu Bai and Philp C. Bevilacqua for their suggestions, recommendations, and advice throughout my project. I really appreciate them for finding time for me in their busy schedules for me, and giving me precious feedback.

Also, thank all of the current and past Murakami lab members at Penn State University for helpful discussion and friendships: Zuhaib Qayyum, Catherine Sutherland, Vadim Molodtsov, Shoko Murakami, Anoop Narayanan and Manju Narwal.

Next, I thank staffs at Penn State for helping me in accomplishing research projects in various ways: Mark Signs and Ashik Sathish at fermentation facility, Carol Bator, Joseph Sunghyun Cho and Missy Hazen at Cryo-EM facility, Hemant Yennawar and Neela Yennawar at X-ray crystallography facility.

I thank our collaborators – Dr. Charles L. Turnbough Jr at University of Alabama at Birmingham, Dr. Richard Pomerantz at Temple University, Dr. Nikolay Zenkin at New Castle University, UK, Dr. Andrey Kulbachinskiy at Russian Academy of Sciences and Dr. Mark Hedglin at Penn State University for various materials and comments.

I also thank staffs at Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline for support of crystallographic data collection and National

Cancer Institute's Cryo-EM Facility at Frederick for support of cryo-electron microscopy data collection.

Lastly, I dedicate this dissertation to my family - my parents Mr. Yongsig Shin and Mrs. Soonhee Kim, and my sister Ms. Aeyoung Shin for their unconditional love.

Chapter 1

Introduction to transcription by bacterial RNA polymerase

Publication in preparation for the work in this chapter is:

Qayyum, M.Z., **Shin, Y.**, and Murakami, K.S. (2020). RNA polymerase reaction in bacteria. The Encyclopedia of Systems Biology. 3rd edition, Springer, 2020.

1.1 Abstract

RNA polymerase (RNAP) is the central enzyme of gene expression that transcribes DNA to RNA. In prokaryotes, a single DNA-dependent RNAP is responsible for transcribing all RNAs including mRNA, tRNA, rRNA, and regulatory RNAs. With over sixty years of investigations since cellular RNAP's discovery in the late 1950s, a wealth of knowledge has accumulated elucidating the structures and functions of bacterial RNAPs. In this chapter, I will introduce a general structural feature of bacterial RNAP and a schematic of the bacterial transcription cycle for better understanding the following chapters.

1.2 Structure of bacterial RNA polymerase

1.2.1 Structural biology: X-ray crystallography and cryo-electron microscopy

Over the past century, X-ray crystallography was the primary method of choice for determining the high-resolution structures of bacterial RNAPs. Historically, endogenous and thermostable RNAPs isolated from *Thermus aquaticus* (Campbell et al., 2001; Murakami et al., 2002a; Murakami et al., 2002b; Zhang et al., 1999) and *Thermus thermophilus* (Vassylyev et al., 2002) have been used for determining the X-ray crystal structure. *Escherichia coli* RNAP expressed and purified by the plasmid-based over-expression system has been used since 2013 as well (Murakami, 2013). Endogenous bacterial RNAP can be purified by performing a RNAP-DNA co-precipitation method using polymin P followed by column chromatography (**Appendix A and B**), but it is tedious and time consuming. Also, the amount of the purified RNAP is often insufficient for structural and biochemical studies. The plasmid-based over-expression system

(Arsimovitch et al., 2003; Murakami, 2013) allowed researchers to purify wild-type and variant bacterial RNAPs more conveniently to crystallize for structural determination, and to biochemically characterize for structure-function comparisons. Protein crystals contain high solvent content with an average volume of > 50%, which allows for ligands to diffuse into the binding sites inside the crystals. For example, *E. coli* RNAP crystals have been used successfully in investigating the binding modes of RNAP inhibitors (Molodtsov et al., 2013; Molodtsov et al., 2017). Additionally, crystal structure analysis of *E. coli* RNAPs with RNAP binding factors such as DksA and TraR provided insights into mechanisms of transcriptional regulation (Molodtsov et al., 2018).

The method of choice for determining the structure of bacterial RNAPs during the last few years drastically changed from X-ray crystallography to cryo-electron microscopy (cryo-EM) due to its advancement of determining atomic resolution structure of macromolecules (Kuhlbrandt, 2014; Li et al., 2013; Lyumkis, 2019; Nogales, 2016; Punjani et al., 2017). Cryo-EM is able to determine structures without crystallization and requires a much lower amount of samples compared to conventional X-ray crystallography. Additionally, cryo-EM brought many opportunities to investigate structure and function of macromolecule in near physiological condition, which could not be addressed by X-ray crystallography due to crystal packing. Protein dynamics can also be addressed by cryo-EM by classifying individual macromolecules into distinct classes during heterogeneous refinement. For example, a recent cryo-EM structure of the *E. coli* RNAP and DNA complex revealed the interaction between an amino acid residue in the σ non-conserved region (NCR) and a DNA, which was not observed from the previously determined X-ray crystal structure (Narayanan et al., 2018). Additionally, recent cryo-EM structures of *M. tuberculosis* RNAP, *M. tuberculosis* RNAP-promoter DNA, and *M. tuberculosis* RNAP-fidaxomicin complexes

revealed the dynamics of the RNAP clamp and binding mode of the fidaxomicin to RNAP (Boyaci et al., 2018; Boyaci et al., 2019).

1.2.2 Overall structure of bacterial RNA polymerase

The bacterial RNAP catalytic core enzyme associates with a transcription stimulating factor called σ to form the RNAP holoenzyme (Murakami, 2013) (**Figures 1.1 and 1.2**). In *E. coli* and *T. thermophilus*, the catalytic core enzyme is composed of five subunits ($\alpha_2\beta\beta'\omega$, MW = ~380 kDa) and is responsible for an RNA synthesis by transcribing a template DNA sequence. The two largest subunits, β and β' (150 and 155 kDa in *E. coli*, 125 and 175 kDa in *T. thermophilus*, respectively) of core RNAP form a cleft for DNA binding in the main channel where template DNA enters and makes the overall shape of RNAP resemble a crab claw with the size of ~150 Å long (from the back to the tip of the claws), 115 Å tall, and 110 Å wide (Murakami and Darst, 2003). The active site of RNAP where RNA synthesis takes place is located at the bottom of the cleft. Two α subunits (α and α' , 36.5 kDa each in *E. coli*, 35 kDa each in *T. thermophilus*) of core RNAP act as a scaffold for assembling of β and β' of core RNAP subunits (Ishihama, 1992; Sutherland and Murakami, 2018; Murakami et al., 1997) (**Figure 1.3**). The α subunit consists of two structural domains including the N-terminal (α NTD, 1-235 residues in *E. coli*) and C-terminal domains (α CTD, 248-329 residues in *E. coli*) connected by a short flexible linker (236-247 residues in *E. coli*). The ω subunit (10.2 kDa in *E. coli*, 12 kDa in *T. thermophilus*) is the non-essential subunit for cell growth, and reconstituting RNAP without the ω subunit did not affect RNAP transcription activity (Gentry et al., 1991; Igarashi et al., 1989). The ω subunit interacts with the double psi beta-barrel (DPBB) domain of the β' subunit and has been

proposed to play a role in responding to the bacterial alarmone ppGpp, the protecting DPBB domain from damage and/or serving as a β ' chaperone (Sutherland and Murakami, 2018).

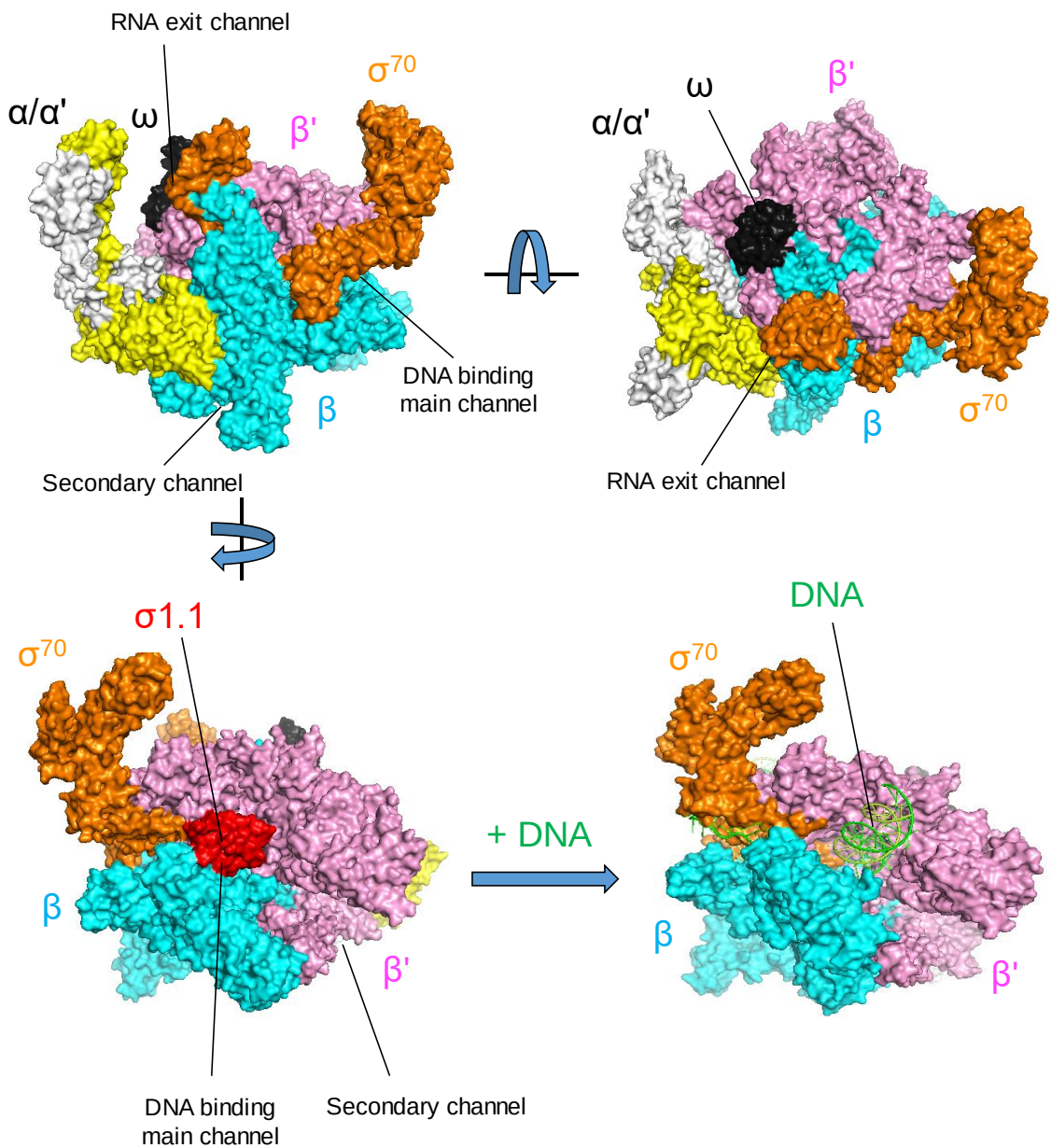


Figure 1.1: Structural overview of *E. coli* RNAP- σ^{70} holoenzyme. Each subunit (α : yellow, α' : white, β : cyan, β' : pink, ω : gray, σ : orange and red) is shown as a surface model. Locations of the DNA binding main channel, secondary channel, and RNA exit channel are indicated. σ region 1.1 (red) is located within the DNA binding main channel as an apo-form holoenzyme before RNAP binds DNA. After the RNAP binds to DNA (green, cartoon model), σ region 1.1 is ejected from the main channel to allow the DNA enter the active site located inside of the cleft formed by the β and β' subunits. The figure was generated from structures determined in this dissertation (Chapter 4).

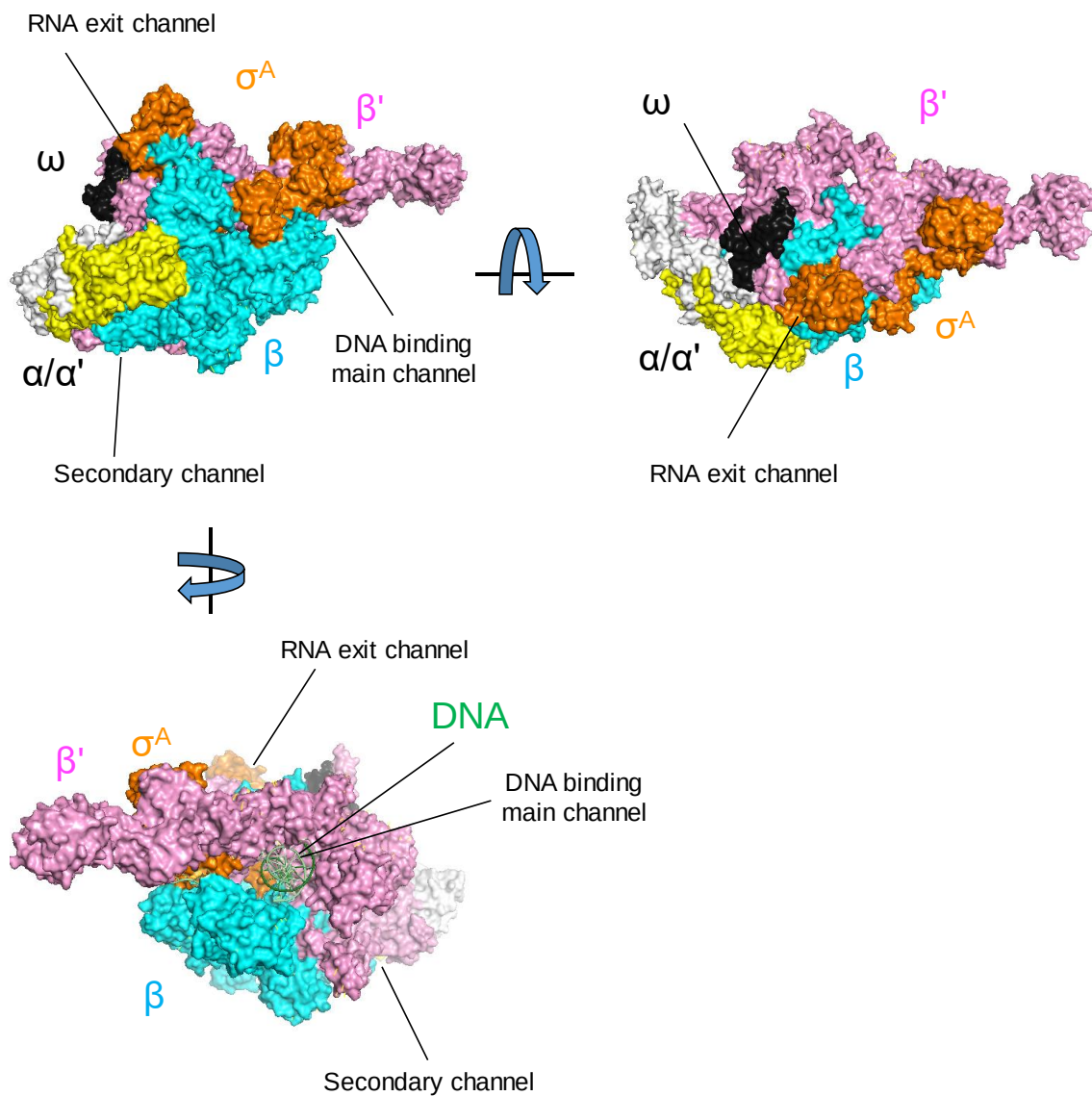


Figure 1.2: Structural overview of *T. thermophilus* RNAP- σ^A holoenzyme. Each subunit (α : yellow, α' : white, β : cyan, β' : pink, ω : gray, σ : orange) is shown as a surface model. Locations of the DNA binding main channel, secondary channel, and RNA exit channel are indicated. The fork-DNA bound at the DNA binding channel is shown as a cartoon model (green). The figure was generated from structures determined in this dissertation (**Chapters 2 and 3**).

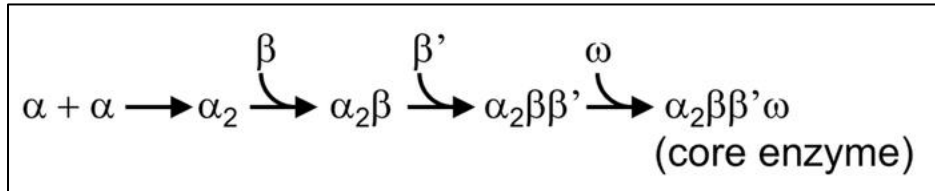


Figure 1.3: Scheme of RNAP core enzyme assembly. Two α subunits serve as a scaffold for assembling of the two largest subunits of RNAP, β and β' . The ω subunit interacts with the DPBB domain of β' subunit for a complete core enzyme. Figure from the reference (Sutherland and Murakami, 2018).

The overall architecture around the cleft and active site is highly conserved among all cellular RNAPs, indicating that the catalytic mechanism of RNA synthesis from bacteria to human is conserved. The other subunit (σ) recognizes the promoter sequence and is thus required for promoter-specific transcript initiation (Feklistov et al., 2014). σ factor was first jointly discovered unexpectedly in 1968 by researchers at Harvard and Rutgers universities, who studied the subunit composition of RNAP (Burgess et al., 1969). The core RNAP itself was inefficient and non-specific in transcribing DNA to RNA while adding another protein they identified from a different purification protocol greatly increased the efficiency and specificity of the transcription activity. They called the stimulating protein as a σ factor (sigma for stimulation).

In *E. coli* and *T. thermophilus*, σ^{70} (70 kDa) and σ^A (48 kDa) are the major housekeeping σ factor (group 1) that direct the majority of transcription during normal exponential-phase growth. Group 1 σ factor is composed of four domains $\sigma^{1.1}$, σ^2 , σ^3 and σ^4 , which are connected by flexible linkers (Murakami and Darst, 2003; Paget and Helmann, 2003) (**Figure 1.4**). σ non-conserved region (σ_{NCR} , N128 to R373 in *E. coli*) is between the conserved regions 1.2 and 2.1, and its length and sequence varies among primary σ factors (Narayanan et al., 2018). The major contacts that are important in forming the RNAP holoenzyme include σ^{70} regions 2 and 4 to the coiled-coil domain of the β' subunit (β' 262 – 309) and the flexible flap domain of the β

subunit, respectively. Domain 2 and domain 4 recognize the -10 and – 35 elements (Feklistov and Darst, 2011), respectively for promoter specific transcript initiation. Other *E. coli* σ factors such as σ^{32} (rpoH, heat shock), σ^{38} (rpoS, starvation/stationary phase), σ^{54} (rpoN, nitrogen-limitation sigma factor) and σ^E (rpoE, extracytoplasmic function) are utilized under certain growth and stress conditions and direct RNAP to transcribe the specific genes (Douceff et al., 2007; Feklistov et al., 2014; Helmann, 2002; Tanaka et al., 1993).

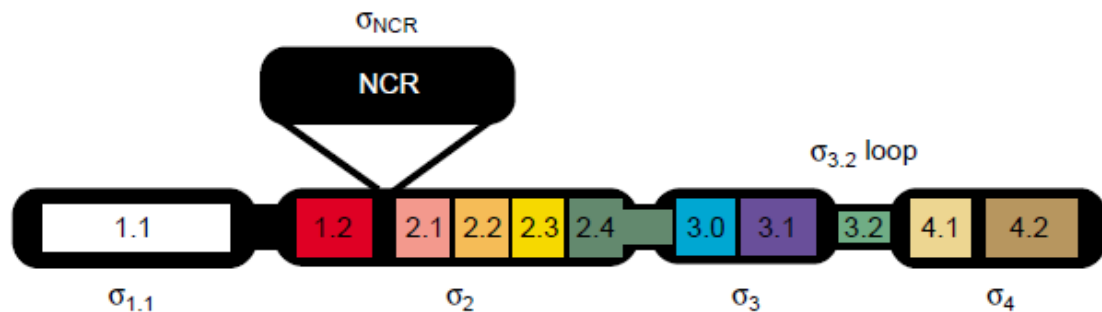


Figure 1.4: The general organization of group 1 house keeping σ factor (σ^{70} in *E. coli* and σ^A in *T. thermophilus*). Four domains are connected by flexible linkers. σ non-conserved region, which its length and sequence varies among primary σ factors, is located between the conserved regions 1.2 and 2.1. Figure from the reference (Murakami and Darst, 2003).

1.2.3 DNA binding main channel, secondary channel and RNA exit channel of RNAP

The RNAP active site, which is buried within a cleft between the β and β' subunits, is connected with the outside of RNAP through three channels (DNA binding main, secondary, and RNA exit) to synthesize RNA efficiently (**Figures 1.1 and 1.2**). Double-stranded DNA binds at the positively charged DNA binding main channel (also known as primary channel), followed by the DNA melting and entering of a single-stranded template DNA into the active site of RNAP (Murakami and Darst, 2003). It has been suggested that the opening of the RNAP clamp

facilitates the template DNA nucleation to the active site while the clamp closing stabilizes the transcription bubble and transcription ready open complex formation (Chakraborty et al., 2012; Feklistov et al., 2017). The secondary channel is a funnel-shaped pore separated from the main channel by the bridge helix. It has been proposed that small molecules such as ribonucleotides gain access to the active site through the secondary channel for RNA synthesis (Cramer et al., 2000; Nickels and Hochschild, 2004). The secondary channel also serves as binding sites for transcription elongation factor GreB (Opalka et al., 2003) and regulation factors DksA and TraR (Molodtsov et al., 2018). The RNA exit channel is tunneled from the active site to the outside of RNAP covered by σ region 4 under the flap domain of the β subunit. The σ finger (σ region 3.2) positions in the RNA exit channel in the active site of RNAP, and therefore must be dissociated from the core RNAP during the initiation to elongation transition (Basu et al., 2014) (**Figure 1.5**). It has been proposed that the σ finger facilitates initiating NTP binding in the active site of RNAP by contacting the template DNA strand (Pupov et al., 2018). The σ finger is also responsible for abortive transcription, the process in which RNAP repetitively releases short nascent RNAs during transcript initiation (Revyakin et al., 2006; Lee and Borukhov, 2016).

1.2.4 Active site of bacterial RNA polymerase

The active site of bacterial RNAP contains highly conserved structural motifs that participate in RNA synthesis (**Figure 1.5**). These motifs include the bridge helix (BH) which guides the template DNA strand into the active site of RNAP for RNA synthesis and the trigger loop (TL) which folds into two α -helical hairpin trigger helices (TH) for nucleotide addition. (Mishanina et al., 2017; Zhang et al., 2010). In the absence of NTPs or the presence of an incorrect NTP at the active site, the TL is located away from the insertion site by farther than 30

Å. When a correct NTP is present at the insertion site, the TL changes the conformation to TH and aligns the reactive group 3' OH of the RNA and the triphosphate group of the NTP for the catalysis (**Figure 1.6**). The BH also undergoes conformational change and coordinates catalysis by facilitating TH formation (Cheung and Cramer, 2012; Svetlov and Nudler, 2009).

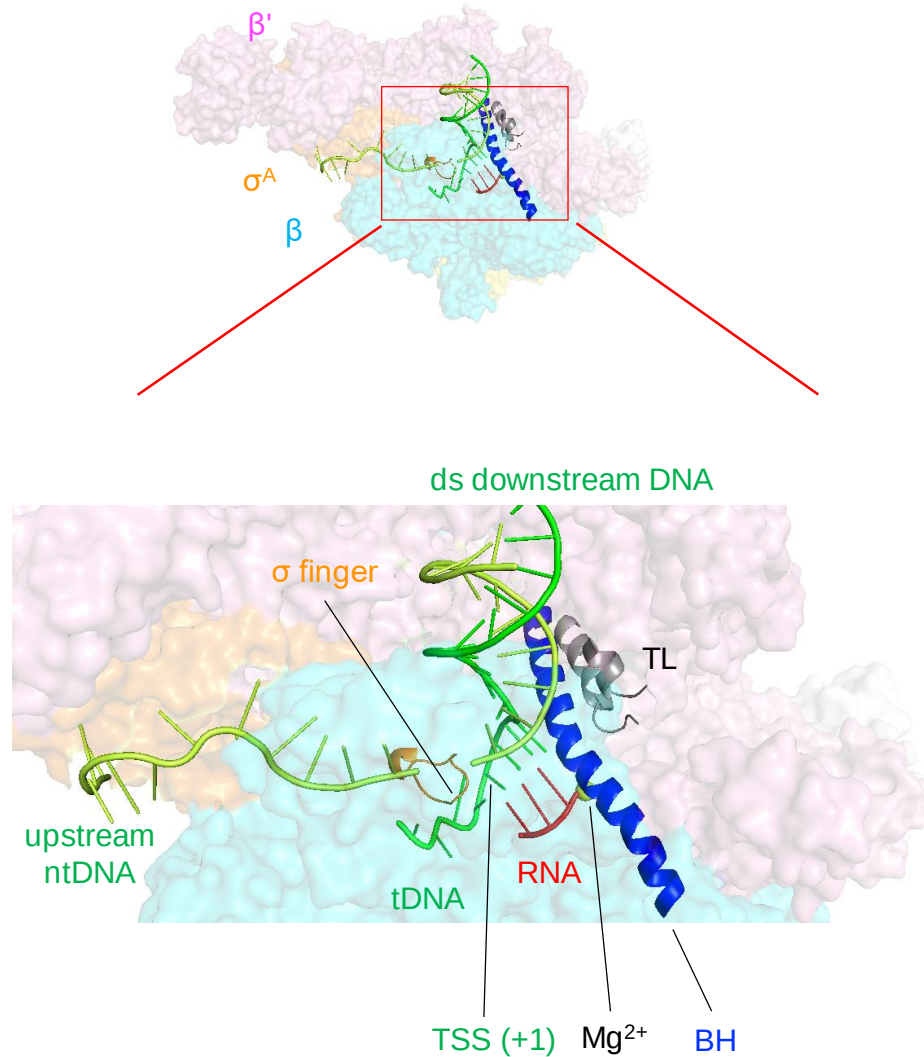


Figure 1.5: The overview of the active site of bacterial RNAP. *T. thermophilus* RNAP in Figure 2 is used as an example. β , β' , σ^A are shown as transparent surface model. The bridge helix (BH, blue), trigger loop (TL, gray), σ finger (orange), template DNA (green), non-template DNA (light green), RNA (red) and active site Mg^{2+} (yellow sphere) are shown as a cartoon model. Disordered TL indicates this domain is highly flexible. Figure modified from a structure determined in this dissertation (**Chapter 2, RTC-5'**)

1.2.5 Nucleotidyl transfer reaction

RNAP carries out the nucleotidyl transfer reaction via a two-metal ion catalytic mechanism. RNAP chelates two Mg^{2+} ions by three aspartic acid residues in the conserved motif **NADF $\underline{D}$ GD** found in DPBB domain of β' subunit, which is found in all cellular RNAPs (Basu et al., 2014; Qayyum et al., 2020) (**Figure 1.6**). The first active site Mg^{2+} (catalytic metal cation, Mg^A), is a Lewis acid that facilitates attacking of a phosphate group of an incoming nucleotide by the RNA 3'-OH group. The second Mg^{2+} (nucleotide-binding metal cation, Mg^B) is coordinated by the triphosphate of the incoming nucleotide and stabilizes a pentacovalent phosphate intermediate during the nucleotidyl transfer reaction. Mg^A has a higher affinity to the active site and binds even in the absence of DNA or NTP while Mg^B has a lower affinity and transiently binds with an incoming nucleotide.

RNAP positions the template DNA so that the 3'-hydroxyl of the RNA is located at the i site (also known as the primer or product site, p site), while the incoming nucleotide (NTP) bind at the $i + 1$ site (also known as the N-site). The RNA state right after phosphodiester bond formation is called the pre-translocated state. After forming a new DNA/RNA hybrid, the hybrid translocates one base upstream which positions the next template DNA at the $i + 1$ site (Brueckner et al., 2009). At this stage, the RNA is in the post-translocated state. Based on structures solved in this dissertation (**Chapter 2**), the byproduct of phosphodiester bond formation, a pyrophosphate (PPi) stays in place of the Mg^B at the $i + 1$ site until the cognate NTP is inserted, suggesting a possible role of PPi in selecting a correct NTP.

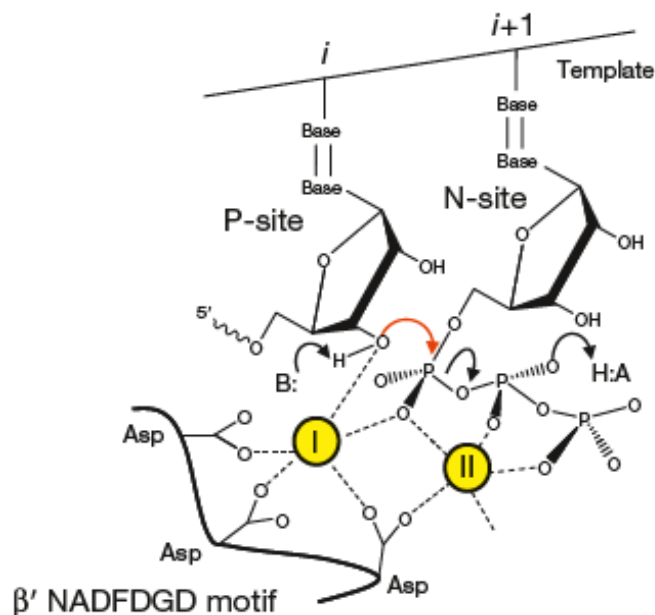


Figure 1.6 : Mechanism of Two-Mg²⁺- mediated catalysis of nucleotidyl transfer reaction. Two nucleotides at the *i* and *i* + 1 sites of the active site are shown. The red arrow illustrates the nucleophilic attack of an RNA 3' oxyanion on the NTP α -phosphate, and the reaction involves the stabilization of a trigonal-bipyramidal transition state by two Mg²⁺ ions (I = Mg^A, II = Mg^B). Three aspartic acid residues in the conserved motif NADFDGD of the β' subunit are shown. Figure from the reference (Qayyum et al., 2020).

1.3 Bacterial transcription cycle

1.3.1 Initiation

The bacterial transcription cycle can be divided into multiple steps (**Figure 1.7**). The first step is recognizing and specific binding in the promoter DNA sequence by the RNAP holoenzyme (**Figure 1.7A**) to form the closed RNAP promoter complex (RPC) (**Figure 1.7B**). The major interactions include σ^{70} conserved regions 2.4 and 4.2 to -10 (TATAAT) and -35

(TTGACA) promoter elements, respectively. Once the RPc is formed, a series of rapid and large conformational changes result in at least two intermediates followed by the formation of transcription competent RNAP open promoter complex (RPo) (**Figure 1.7C**). In the RPo, the double-stranded DNA is melted (between -11 to +2) to form a transcription bubble and the transcription start site (+1 of tDNA) is positioned at the *i* site within the active site of RNAP. The double-stranded downstream DNA from +5 to around +12 is clamped by the RNAP DNA-binding cleft. When RNAP takes initiating nucleotides into the active site through the secondary channel, RNAP initiates RNA synthesis *de novo*. After undergoing abortive transcript initiation several times, the RNA eventually displaces the σ finger and σ region 4 from the β flap domain, followed by destabilization of the interaction between σ region 4 and the -35 element. These changes allow RNAP to escape from the promoter and translocate downstream DNA transitioning into elongation phase (**Figure 1.7D**). Although it has been generally thought that the σ factor completely dissociates from core RNAP during elongation, σ factor may remain associated with core RNAP throughout transcription (Harden et al., 2016) or access the transcript elongation complex *in trans* (Goldman et al., 2015).

1.3.2 Elongation

Throughout transcript elongation, RNAP maintains a stable 9 -10 bp DNA/RNA hybrid with the transcription bubble size of 11-13 nt (Svetlov and Nudler, 2009; Vassilyev et al., 2007). The bacterial RNAP elongation complex is highly stable and processive. RNAP extends the transcript with high-fidelity at an estimated error rate of less than 10^{-5} and with the extension rate of 50-100 nucleotides s^{-1} . In *E. coli*, the transcription elongation factor NusA enhances intrinsic termination as well as transcriptional pausing by interacting with nascent RNA hairpin

(Toulokhonov et al., 2001) while NusG increases transcription processivity by suppressing RNAP pausing (Yakhnin et al., 2016) (**Figure 1.7E**). NusG also contacts the elongation factor NusE, which is also the ribosomal S10 subunit. This interaction has been suggested to play a role in the coupling of transcription and translation by a RNAP and a ribosome, respectively (**Figure 1.7F**).

1.3.3 Termination

There are two types of termination: ρ -dependent and ρ -independent (intrinsic). The ρ (Rho) protein is a hexameric ATP-dependent helicase, which loads onto nascent RNA strands at rho utilization sites (rut). Upon binding the RNA, ρ translocates in the 5'→3' direction along the RNA toward the EC, resulting in the release of the RNA from the RNAP elongation complex (**Figure 1.7G**). RNAP terminates transcription in the absence ρ , when the RNAP encounters an intrinsic termination signal that is 20 to 35 nt long self-complementary G/C rich RNA sequence that forms a hairpin immediately followed by a 7-9 nt long U-rich tract in the RNA. The RNA hairpin destabilizes the interaction between nucleic acid and RNAP, which eventually collapses the transcription bubble and dissociates the transcript from the RNAP.

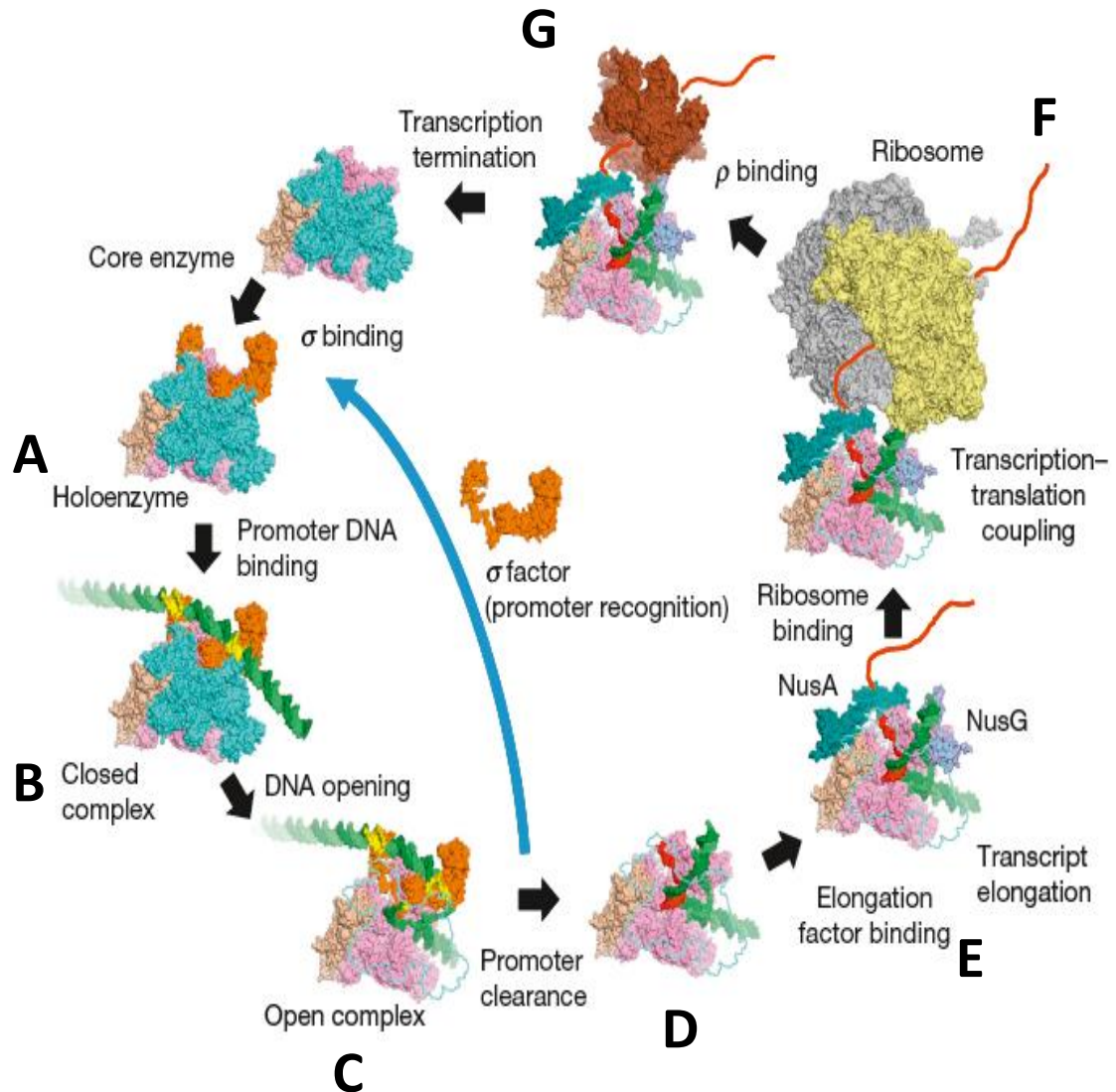


Figure 1.7 : Overview of the bacterial transcription cycle. Core RNAP subunits (α : light brown, β : cyan, β' : pink, ω : white), σ factor (orange), DNA (template: green, non-template: light green, -35 and -10 elements: yellow), RNA (red) are shown. NusA, NusG, ribosome (30 S: yellow, 50 S: gray), and ρ are also shown and labeled. (A) core RNAP and σ factor binds to form a holoenzyme. (B) The holoenzyme recognizes and binds the specific promoter sequences (-35 and -10 elements) to form a closed complex. (C) Large conformational changes lead to transcription competent open complex. (D) After undergoing abortive transcript initiation several times, the RNAP transits to the elongation phase. (E) During elongation, transcript elongation factors such as NusA and NusG interact with the RNAP to pause or facilitate transcript elongation. (F) A ribosome begins to translate RNA to protein by coupling with the RNAP. (G) binding of ρ terminates transcription. Figure from the reference (Qayyum et al., 2020).

1.4 Dissertation format

Chapter 2 starts with the description of non-canonical form of RNA synthesis called “reiterative transcription”, which has been shown to play a regulatory role in bacteria. I studied the mechanism of reiterative transcription from the *pyrG* promoter of *Bacillus subtilis* by structural and biochemical approaches. Using time-dependent soak-trigger-freeze X-ray crystallography, I was able to determine a series of reiterative transcription complex structures that revealed the detailed mechanism of reiterative transcription. I also performed *in vitro* transcription and fluorescence assays to correlate the structural observations in solution. These structures and structure-based biochemical assay revealed an alternative RNA extension pathway for the first time, and provided insight into the base-sharing mechanism during reiterative transcription. I also established the fermentation, purification and crystallization protocols that researchers can follow in the future for X-ray crystallographic studies.

In **Chapter 3**, I continued investigating the mechanism of another type of reiterative transcription from the *pyrBI* promoter of *E. coli*. The fate of reiterative transcripts is known to be fundamentally different from the transcripts of the *pyrG* promoter. I solved the transcription initiation complex structures with DNAs containing the *pyrBI* promoter and *pyrG* promoter variant sequences and characterized reiterative transcription from these promoters by *in vitro* transcription assays. The structures and biochemical assays showed an alternative RNA extension pathway compared to the RNA path from the *pyrG* promoter and revealed the role of -1 base of template DNA in determining the RNA path.

In **Chapter 4**, I investigated the mechanism of extraordinary active ribosomal RNA (rRNA) transcription and its DksA/ppGpp-dependent inhibition by cryo-EM structure determination of RNAP and rRNA promoter complex and biochemical assays. In *E. coli*, rRNA are the most transcribed genes in rapidly growing bacteria during vegetative growth; however, it

is drastically downregulated by the DksA/ppGpp binding to RNAP during stationary phase and stressed conditions. The open complex structure shows the open complex scrunching by which the template DNA contacts with σ finger and β' lid and transcript start site selection at the rRNA promoter. DksA/ppGpp binding to RNAP induces the movement of RNAP mobile domains that influence the stability of the open complex and hinder the DNA unwinding of promoters containing GC-rich discriminator sequence. I explain the pleotropic effect of DksA/ppGpp in bacterial RNAP transcription based on the difference in DNA duplex stability of discriminator sequences.

Lastly, In **Chapter 5**, I summarize the findings from my studies and remaining questions that still need to be answered in the future. I present the preliminary data that I am still investigating and suggest the experiments to be performed in the future. I conclude the chapter with my new research project that I will be continuing in the future.

1.5 References

- Artsimovitch, I., Svetlov, V., Murakami, K.S., and Landick, R. (2003). Co-overexpression of Escherichia coli RNA polymerase subunits allows isolation and analysis of mutant enzymes lacking lineage-specific sequence insertions. *J Biol Chem* 278, 12344-12355.
- Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem* 289, 24549-24559.
- Boyaci, H., Chen, J., Jansen, R., Darst, S.A., and Campbell, E.A. (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* 565, 382-385.

Boyaci, H., Chen, J., Lilic, M., Palka, M., Mooney, R.A., Landick, R., Darst, S.A., and Campbell, E.A. (2018). Fidaxomicin jams Mycobacterium tuberculosis RNA polymerase motions needed for initiation via RbpA contacts. *Elife* 7.

Brueckner, F., Ortiz, J., and Cramer, P. (2009). A movie of the RNA polymerase nucleotide addition cycle. *Curr Opin Struct Biol* 19, 294-299.

Burgess, R.R., Travers, A.A., Dunn, J.J., and Bautz, E.K. (1969). Factor stimulating transcription by RNA polymerase. *Nature* 221, 43-46.

Campbell, E.A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., and Darst, S.A. (2001). Structural mechanism for rifampicin inhibition of bacterial rna polymerase. *Cell* 104, 901-912.

Chakraborty, A., Wang, D., Ebright, Y.W., Korlann, Y., Kortkhonjia, E., Kim, T., Chowdhury, S., Wigneshweraraj, S., Irschik, H., Jansen, R., *et al.* (2012). Opening and closing of the bacterial RNA polymerase clamp. *Science* 337, 591-595.

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

Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* 288, 640-649.

Doucleff, M., Pelton, J.G., Lee, P.S., Nixon, B.T., and Wemmer, D.E. (2007). Structural basis of DNA recognition by the alternative sigma-factor, sigma54. *J Mol Biol* 369, 1070-1078.

Feklistov, A., Bae, B., Hauver, J., Lass-Napiorkowska, A., Kalesse, M., Glaus, F., Altmann, K.H., Heyduk, T., Landick, R., and Darst, S.A. (2017). RNA polymerase motions during promoter melting. *Science* 356, 863-866.

Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* 147, 1257-1269.

Feklistov, A., Sharon, B.D., Darst, S.A., and Gross, C.A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* 68, 357-376.

Gentry, D., Xiao, H., Burgess, R., and Cashel, M. (1991). The omega subunit of Escherichia coli K-12 RNA polymerase is not required for stringent RNA control in vivo. *J Bacteriol* 173, 3901-3903.

Goldman, S.R., Nair, N.U., Wells, C.D., Nickels, B.E., and Hochschild, A. (2015). The primary sigma factor in Escherichia coli can access the transcription elongation complex from solution in vivo. *Elife* 4.

Harden, T.T., Wells, C.D., Friedman, L.J., Landick, R., Hochschild, A., Kondev, J., and Gelles, J. (2016). Bacterial RNA polymerase can retain sigma70 throughout transcription. *Proc Natl Acad Sci U S A* 113, 602-607.

Helmann, J.D. (2002). The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46, 47-110.

Igarashi, K., Fujita, N., and Ishihama, A. (1989). Promoter selectivity of Escherichia coli RNA polymerase: omega factor is responsible for the ppGpp sensitivity. *Nucleic Acids Res* 17, 8755-8765.

Ishihama, A. (1992). Role of the RNA polymerase alpha subunit in transcription activation. *Mol Microbiol* 6, 3283-3288.

Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* 343, 1443-1444.

Lee, J., and Borukhov, S. (2016). Bacterial RNA Polymerase-DNA Interaction-The Driving Force of Gene Expression and the Target for Drug Action. *Front Mol Biosci* 3, 73.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10, 584-590.

Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* 294, 5181-5197.

Mishanina, T.V., Palo, M.Z., Nayak, D., Mooney, R.A., and Landick, R. (2017). Trigger loop of RNA polymerase is a positional, not acid-base, catalyst for both transcription and proofreading. *Proc Natl Acad Sci U S A* 114, E5103-E5112.

Molodtsov, V., Nawarathne, I.N., Scharf, N.T., Kirchhoff, P.D., Showalter, H.D., Garcia, G.A., and Murakami, K.S. (2013). X-ray crystal structures of the Escherichia coli RNA polymerase in complex with benzoxazinorifamycins. *J Med Chem* 56, 4758-4763.

Molodtsov, V., Scharf, N.T., Stefan, M.A., Garcia, G.A., and Murakami, K.S. (2017). Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in Mycobacterium tuberculosis. *Mol Microbiol* 103, 1034-1045.

Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S.E., and Murakami, K.S. (2018). Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol Cell* 69, 828-839 e825.

Murakami, K., Kimura, M., Owens, J.T., Meares, C.F., and Ishihama, A. (1997). The two alpha subunits of Escherichia coli RNA polymerase are asymmetrically arranged and contact different halves of the DNA upstream element. *Proc Natl Acad Sci U S A* 94, 1709-1714.

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.

Narayanan, A., Vago, F.S., Li, K., Qayyum, M.Z., Yernool, D., Jiang, W., and Murakami, K.S. (2018). Cryo-EM structure of *Escherichia coli* sigma(70) RNA polymerase and promoter DNA complex revealed a role of sigma non-conserved region during the open complex formation. *J Biol Chem* 293, 7367-7375.

Nickels, B.E., and Hochschild, A. (2004). Regulation of RNA polymerase through the secondary channel. *Cell* 118, 281-284.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nat Methods* 13, 24-27.

Opalka, N., Chlenov, M., Chacon, P., Rice, W.J., Wriggers, W., and Darst, S.A. (2003). Structure and function of the transcription elongation factor GreB bound to bacterial RNA polymerase. *Cell* 114, 335-345.

Paget, M.S., and Helmann, J.D. (2003). The sigma70 family of sigma factors. *Genome Biol* 4, 203.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14, 290-296.

Pupov, D., Petushkov, I., Esyunina, D., Murakami, K.S., and Kulbachinskiy, A. (2018). Region 3.2 of the sigma factor controls the stability of rRNA promoter complexes and potentiates their repression by DksA. *Nucleic Acids Res* 46, 11477-11487.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314, 1139-1143.

Sutherland, C., and Murakami, K.S. (2018). An Introduction to the Structure and Function of the Catalytic Core Enzyme of *Escherichia coli* RNA Polymerase. *EcoSal Plus* 8.

- Svetlov, V., and Nudler, E. (2009). Macromolecular micromovements: how RNA polymerase translocates. *Curr Opin Struct Biol* 19, 701-707.
- Tanaka, K., Takayanagi, Y., Fujita, N., Ishihama, A., and Takahashi, H. (1993). Heterogeneity of the principal sigma factor in *Escherichia coli*: the *rpoS* gene product, sigma 38, is a second principal sigma factor of RNA polymerase in stationary-phase *Escherichia coli*. *Proc Natl Acad Sci U S A* 90, 8303.
- Touloukhonov, I., Artsimovitch, I., and Landick, R. (2001). Allosteric control of RNA polymerase by a site that contacts nascent RNA hairpins. *Science* 292, 730-733.
- Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, 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.
- Vassylyev, D.G., Vassylyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448, 157-162.
- Yakhnin, A.V., Murakami, K.S., and Babitzke, P. (2016). NusG Is a Sequence-specific RNA Polymerase Pause Factor That Binds to the Non-template DNA within the Paused Transcription Bubble. *J Biol Chem* 291, 5299-5308.
- 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, J., Palangat, M., and Landick, R. (2010). Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struct Mol Biol* 17, 99-104.

Chapter 2

Structural basis of the productive reiterative transcription from the *pyrG* promoter

Publications from the work in this chapter are:

1) **Shin, Y.**, Hedglin, M., and Murakami, K.S. (2020). Structural basis of reiterative transcription from the *pyrG* and *pyrBI* promoters by bacterial RNA polymerase. *Nucleic Acids Research*, **48**, 2144-2155.

2) Murakami, K.S*, **Shin, Y***, Turnbough, C.L., and Molodtsov, V. (2017). X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway. *Proc Natl Acad Sci U S A* **114**, 8211-8216.

* co-first author

2.1 Abstract

Reiterative transcription is a non-canonical form of RNA synthesis in which a nucleotide specified by a single base in the template DNA is repetitively added to the 3' end of the nascent transcript. This process plays a regulatory role in bacteria. In this chapter, I present a series of reiterative transcription and transcript initiation complex structures from the *pyrG* promoter of *B. subtilis* determined by soak trigger freeze X-ray crystallography. The structures reveal that 1) 5' end of nascent RNA initially extends toward the main channel of RNA polymerase (RNAP) without clashing with the σ finger; 2) extra G bases are added at the RNA 3' end by the G-G mismatch between template DNA and incoming GTP followed by base sharing; and 3) RNAP maintains stable 3 base pair DNA:RNA hybrid.

2.2 Reiterative transcription

During canonical transcription, RNAP transcribes the sequence of a template DNA to RNA one base at a time based on Watson-Crick complementary base pairing. However, during reiterative transcription (also known as transcript slippage, RNA polymerase stuttering, and pseudo-templated transcription), a single base in a template DNA produce multiple bases in the RNA product. This is due to the repetitive addition of the same base at the 3' end of RNA without translocation of template DNA (Anikin et al., 2010; Jacques and Kolakofsky, 1991; Turnbough, 2011; Turnbough and Switzer, 2008). Interestingly, this phenomenon has been observed for the first time in the early 1960s by Prof. Paul Berg (Chamberlin and Berg, 1962), Penn State alumnus and Nobel Laureate at Stanford University. This observation was considered as a simple transcription error for decades; however, extensive research has demonstrated that reiterative

transcription plays key regulatory roles in all domain of life and also in virus (Barr and Wertz, 2001; Cheng et al., 2001; Guo and Roberts, 1990; Liu et al., 1994; Meng et al., 2004; Molodtsov et al., 2014; Raabe et al., 1998; van Leeuwen et al., 1998). During transcript initiation, reiterative transcription primarily occurs within homopolymeric tracts in template DNA as short as three residues, but it has been shown that it can also occur within dinucleotide (Gilman et al., 2009; Pal and Luse, 2002) and trinucleotide (Fabre et al., 2002) repeat sequences. Longer homolymeric tracts are generally required during elongation and termination (Wagner et al., 1990; Xiong and Reznikoff 1993; Zhou et al., 2013). The fates of transcripts produced by reiterative transcription can be either productive if they can be elongated by canonical transcription after the addition of extra nucleotides or non-productive if the transcripts are released from the RNAP-DNA transcription complex after the addition of extra nucleotides (Liu et al., 1994; Meng et al., 2014).

This dissertation describes two distinct examples of gene regulation in bacteria by conditional reiterative transcription. **Chapter 2** focuses on regulation of the *pyrG* operon of *B. subtilis*, which involves productive reiterative transcription. **Chapter 3** focuses on regulation of the *pyrBI* operon of *E. coli*, which involves non-productive reiterative transcription.

2.2.1 Productive reiterative transcription – *pyrG* operon of *B. subtilis*

A well characterized example involving reiterative transcription in regulating gene expression is CTP-mediated control of *pyrG* expression in *B. subtilis* (Meng et al., 2004). *pyrG* encodes CTP synthetase, which aminates UTP to form CTP when the concentration of CTP in the cell is low. Early efforts in finding a transacting regulatory protein or RNA-binding protein that downregulates gene expression at low CTP levels was not successful. However, in 2004, Dr. Charles Turnbough at The University of Alabama showed that the *pyrG* gene expression is

conditionally regulated by intracellular CTP level due to its programmed template nucleotide sequence; 1) slippage prone homopolymeric sequence GGG followed by CTC in initially transcribed region (ITR); and 2) a long pyrimidine-rich attenuator sequence (**Figure 2.1**). In the presence of high CTP concentrations, RNAP transcribes the DNA sequence including ITR via canonical transcription until read through the upstream segment of the attenuator sequence, which forms the transcription termination hairpin that eliminates *pyrG* expression (**Figure 2.1, left**). On the other hand, when the CTP concentration is low, RNAP pauses after synthesizing 5'-GGG-3' RNA, allowing RNAP to insert extra G bases instead of C residues with the reiterative transcription process (**Figure 2.1, right**). The extra G tract then forms an antiterminator hairpin with a downstream segment of the nascent transcript that precludes intrinsic transcription termination at an attenuator preceding the *pyrG*. As reiterative transcription at the *pyrG* promoter eventually switches to canonical transcription to express the gene, it represents a classical example of productive reiterative transcription (non-productive reiterative transcription will be described in **Chapter 3**).

Although reiterative transcription clearly plays important roles in controlling gene expression, the molecular mechanism was unknown as there was no structural information showing how it differs from canonical transcription. In this chapter, I describe how I started this project by biochemically supporting the first reiterative transcription complex, followed by how I continued to investigate more detailed mechanism of reiterative transcription at the *pyrG* promoter using time-dependent soak-trigger-freeze X-ray crystallography (Basu and Murakami, 2013)

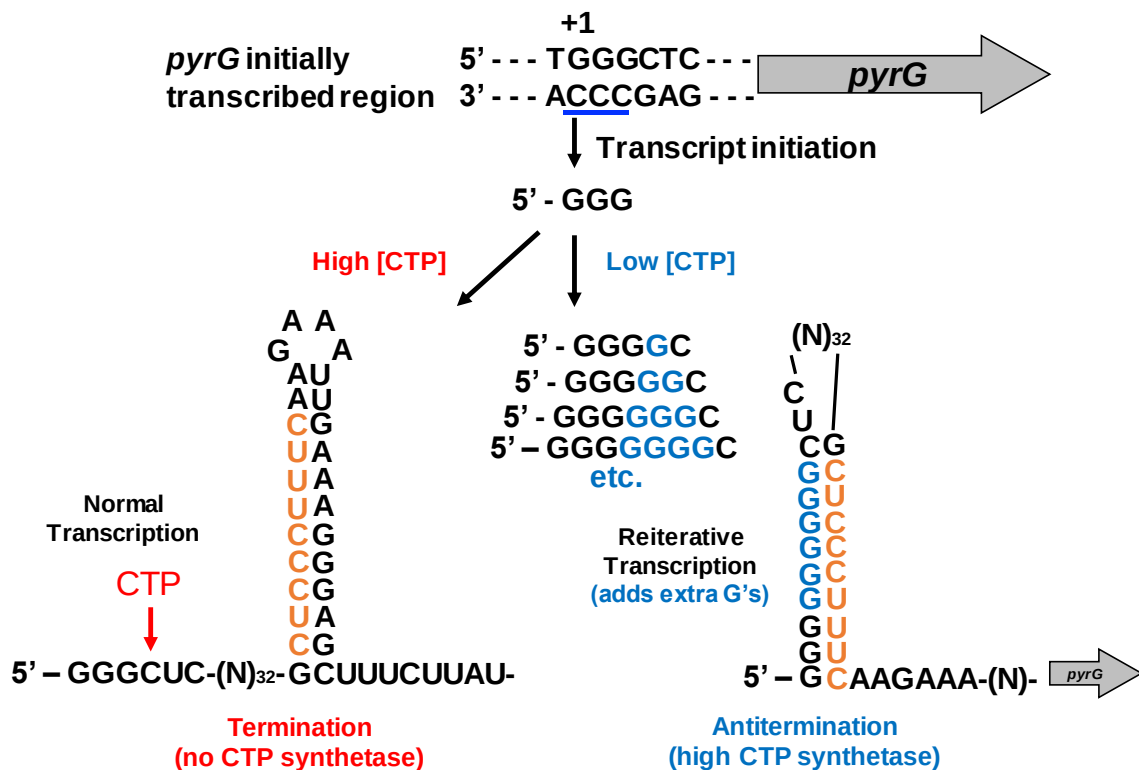


Figure 2.1: Model for CTP-mediated regulation of *pyrG* expression in *B. subtilis*. High CTP concentration allows RNAP to transcribe by canonical transcription by inserting CTP after the first three bases of the ITR (5'-GGG-3'), resulting in the formation of the transcription termination hairpin within the attenuator sequence (orange). *pyrG* expression is eliminated (left). Low CTP concentration allows reiterative transcription that adds extra G bases to the nascent RNA. The resulting poly-G forms the anti-termination hairpin by base pairing with the first part of attenuator sequence. *pyrG* is expressed (right). Figure from the references (Meng et al., 2004; Shin et al., 2020).

2.2.2 Using *T. thermophilus* RNAP crystalline state to study the mechanism of transcription

To study the structural basis of the mechanism of reiterative transcription, I used *T. thermophilus* RNAP crystalline state (**Figure 2.2**) as 1) *B. subtilis* RNAP crystallization system is not available to date, 2) *T. thermophilus* RNAP crystals diffract to higher resolution than *E. coli*

RNAP crystals, which is important for determining positions of DNA and RNA bases within the RNAP-DNA complex, and 3) published studies show *T. thermophilus* RNAP remains active in crystals (Basu et al., 2014; Zhang et al., 2012), and therefore is suitable for NTP soaking to trigger transcription *in crystallo*. I also confirmed *in vitro* that reiterative transcription was essentially the same when the *pyrG* promoter was transcribed by *B. subtilis*, *T. thermophilus*, and *E. coli* RNAPs (See **section 2.4.1**).

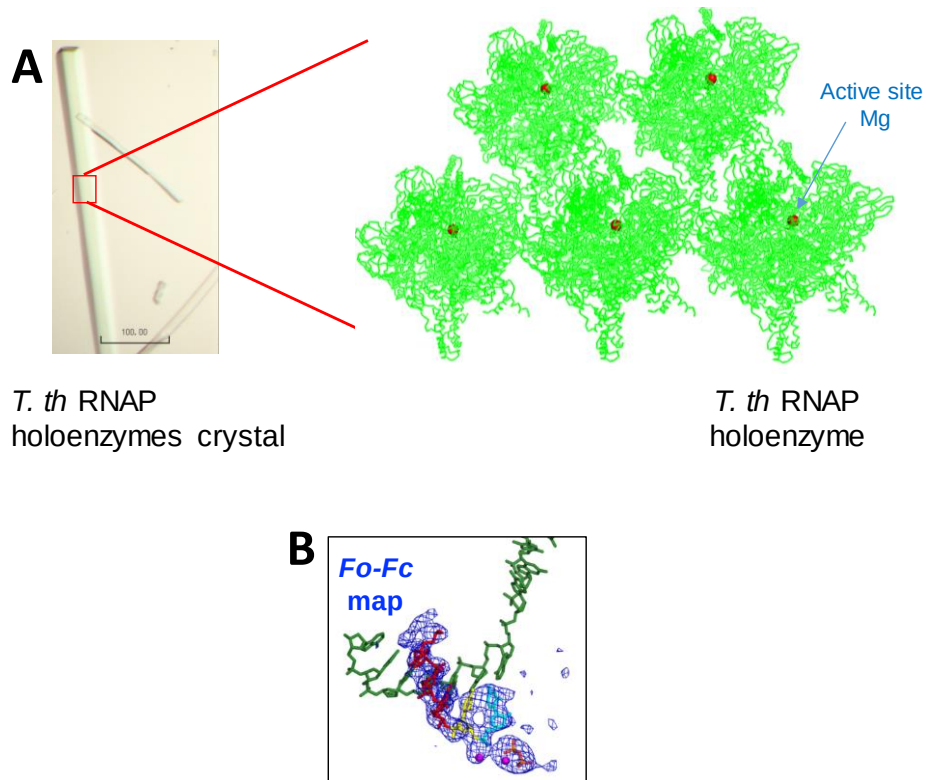


Figure 2.2: *T. thermophilus* RNAP crystalline state (A) Full-sized *T. thermophilus* RNAP holoenzyme crystal showing ordered state of enzymes. About 60% of crystal content is solvent, which allows delivering nucleotides to the active site of RNAP. (B) *Fo-Fc* map generated from X-ray diffraction of the crystal, which shows the precise location of DNA and RNA bases.

2.3 Materials and Methods

2.3.1 Fermentation of *T. thermophilus* and Purification and Crystallization of *T. thermophilus* RNAP

Endogenous *T. thermophilus* RNAP was purified as described in **Appendix A** and **B**. *T. thermophilus* RNAP – DNA complex containing the *pyrG* promoter sequence (**Figure 2.3**) was crystallized as described in **Appendix C**. Binary complex crystals were sequentially transferred to cryoprotectant and soaked into a final cryo-solution containing 2 mM GTP to trigger RNA synthesis *in crystallo*. The reaction was stopped by freezing crystals in liquid nitrogen.

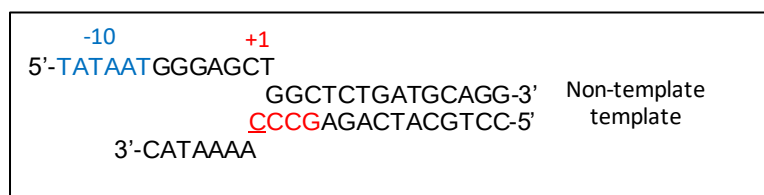


Figure 2.3: The fork-DNA used for crystallization. DNA contains slippage-prone homopolymeric sequence found in *pyrG* (red). Transcription start site is underlined and -10 element is indicated in blue.

2.3.2 X-ray Data collections and Structure Determinations

The X-ray datasets were collected at the Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY) and structures were determined using crystallographic software: HKL2000, Phenix and Coot (Afonine et al., 2010; Basu et al., 2014; Emsley and Cowtan, 2004; Otwinowski and Minor 1997;). Final structures were deposited in the Protein Data Bank (PDB) database with their PDB ID codes listed in **Tables 2.1 and 2.2**.

Table 2.1: Data collection and refinement statistics of binary complex and RTC-30'

Complex PDB code	RP_{pyrG-o} 5VOI	RTC-30' 5VO8
Data collection		
Space group	C2	C2
Cell dimensions		
<i>a</i> (Å)	183.18	186.54
<i>b</i> (Å)	103.47	101.98
<i>c</i> (Å)	294.44	296.11
β (°)	99.14	98.56
Resolution (Å)	30 – 2.80	30 – 3.30
Total reflections	440,744	253,251
Unique reflections	132,496	78332
Redundancy	3.3 (2.3)	3.2 (2.7)
Completeness (%)	98.2 (89.5)	94.2 (84.1)
<i>I</i> / σ	8.6 (1.67)	9.9 (2.11)
Wilson B factor	47.8	81.8
<i>R</i> _{sym} (%)	10.3 (66.5)	8.7 (47.1)
CC ^{1/2}	(0.474)	(0.743)
Refinement		
Resolution (Å)	30 – 2.80	30 – 3.30
<i>R</i> _{work}	0.219	0.197
<i>R</i> _{free}	0.245	0.249
No. of atoms	28,432	28,563
<i>B</i> factors	64.0	121.0
R.m.s deviations		
Bond length (Å)	0.021	0.015
Bond angles (°)	1.49	1.66
Clashscore	10	13
Ramachandran favored, %	98.0	96.0
Ramachandran outliers, %	0.1	1.1

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

*Highest resolution shells are shown in parentheses

Table 2.2: Data collection and refinement statistics of RTC intermediates

Complex	RTC-3'	RTC-5'	RTC-7'
PDB code	6OY5	6OY6	6OY7
Data collection			
Space group	C2	C2	C2
Cell dimensions			
<i>a</i> (Å)	186.09	186.88	185.68
<i>b</i> (Å)	102.41	102.14	101.61
<i>c</i> (Å)	297.23	297.36	295.80
β (°)	99.00	98.81	98.76
Resolution (Å)	50 – 3.0	50 – 3.1	50 – 3.04
Total reflections	295,628	261,015	316,859
Unique reflections	99,590	88,633	100,023
Redundancy	3.0 (2.5)	2.9 (2.3)	3.2 (2.7)
Completeness (%)	90.0 (64.4)	88.0 (56.0)	95.1 (72.7)
<i>I</i> / σ	11.9 (0.85)	9.64 (0.64)	9.54 (0.91)
CC ^{1/2}	(0.544)	(0.488)	(0.520)
Refinement			
Resolution (Å)	44 – 3.1	44 – 3.1	48 – 3.04
<i>R</i> _{work}	0.207	0.212	0.207
<i>R</i> _{free}	0.259	0.264	0.248
No. of atoms	28,534	28,578	28,534
R.m.s deviations			
Bond length (Å)	0.011	0.012	0.012
Bond angles (°)	1.546	1.551	1.612
Clashscore	10.6	9.39	12.22
Ramachandran favored, %	95.96	95.96	95.45
Ramachandran outliers, %	0.78	0.75	1.10

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

*Highest resolution shells are shown in parentheses

2.4 Results

2.4.1 The structure of the RNAP-*pyrG* promoter reiterative transcription complex at 30 min

The first structure of the reiterative transcription complex (RTC) was determined by soaking the binary complex to a cryo-solution containing GTP for 30 min (RTC-30') at room temperature (done by Prof. Katsuhiko Murakami). The structure reveals previously unknown RNA extension strategy 1) the 5' end of the reiterative transcript (8-mer poly-G RNA) detours from the canonical RNA exit channel and extends toward the main channel of the RNAP without clashing with the σ finger (**Figure 2.4**); and 2) a stable maintenance of a 3-bp DNA/RNA hybrid (**Figure 2.5**). The RNA is in the post-translocated state (i.e. the 3' end of the RNA is base-pairing with the +3C template DNA at i site) and the +4G template DNA was positioned at $i + 1$ site (**Figure 2.5**).

My earliest graduate research started with characterizing the reiterative transcription from the *pyrG* promoter in solution to validate the RTC-30' structure. I showed that the reiterative transcription results were similar when the synthetic *pyrG* promoter was transcribed by *E. coli*, *T. thermophilus* or *B. subtilis* RNAPs (**Figure 2.6A**), indicating the universal nature of the reiterative transcription reaction in bacteria. Additionally, I showed the reiterative transcription was essentially the same with the synthetic DNA scaffold and the natural *pyrG* promoter under the same reaction conditions (**Figure 2.6B**). Soaking GTP directly into the crystal (*in crystallo* transcription) also verified that *T. thermophilus* RNAP crystals remained transcriptionally active in the crystalline state by producing and retaining poly-G transcripts up to 8 nt in length, similar to the transcripts produced in solution (**Figure 2.6C**).

Lastly, I tested the barrier hypothesis (**Figure 2.4B**) by performing *in vitro* transcription using as template DNA two variants of the *pyrG*-like DNA scaffold in which the initially

transcribed sequence was changed from C₃ to either T₃ or G₃ (**Figure 2.7**). In the presence of only the corresponding initiating nucleotides ATP or CTP, *T. thermophilus* RNAP efficiently synthesized reiterative transcripts with both promoter variants. With the T₃ promoter, a ladder of poly-A transcripts up to 8 residues in length was produced, the same maximum length observed with the standard template and poly-G transcripts. In contrast, with the G₃ promoter, a ladder of poly-C transcripts with lengths easily exceeding 40 residues was produced. These results and the fact that pyrimidines are smaller than purines, suggest that the maximum length of reiterative transcripts was defined by steric restriction imposed by the narrow opening into the main channel (**Figure 2.4**).

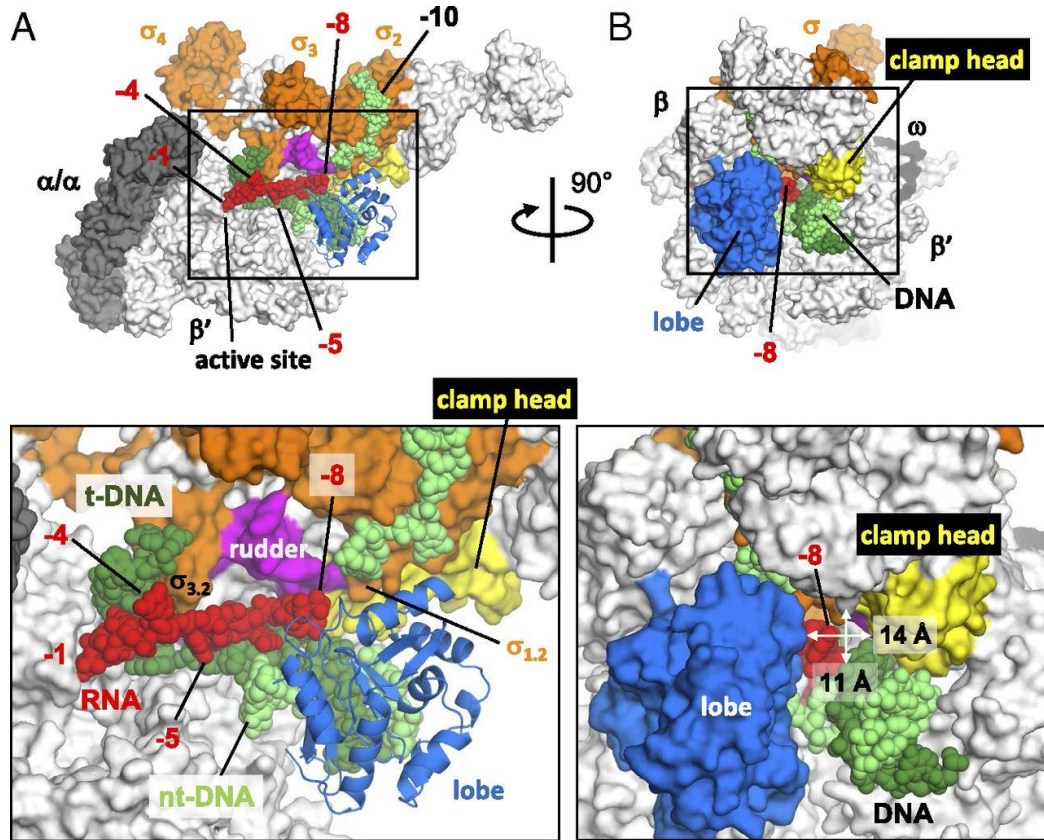


Figure 2.4: The structure of the reiterative transcription complex at 30 min (RTC-30') reveals the location of the reiterative transcript within RNAP. (A) 5' end of the 8-mer poly-G transcript approaches the entrance of the RNAP main channel. (B) The opening into the main channel that would allow the RNA to escape from RNAP is only 14 Å wide and 11 Å high, which could present a barrier to further RNA extension beyond the 8-mer RNA. Figure from the reference (Murakami et al., 2017)

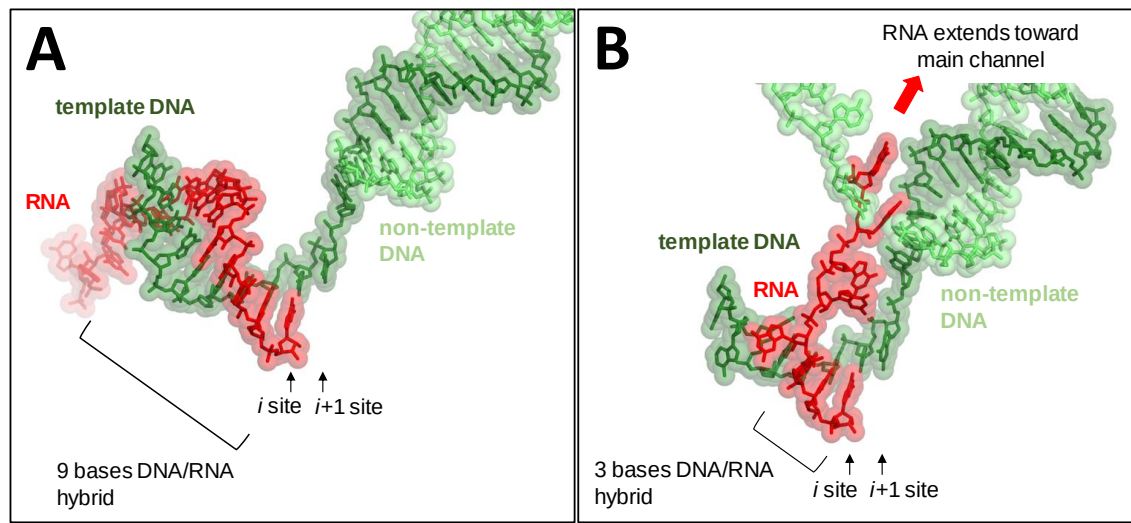


Figure 2.5: Comparison of RNA paths within different transcription complexes. (A) Elongation complex (PDB:2O5J) shows 9 bp DNA/RNA hybrid with 5' end of RNA extending toward the RNA exit channel (Vassilyev et al., 2007). (B) Reiterative transcription complex (PDB:5VO8) shows 3 bp DNA/RNA hybrid with 5' end of RNA extending toward the main channel. Figure modified from the reference (Murakami et al., 2017).

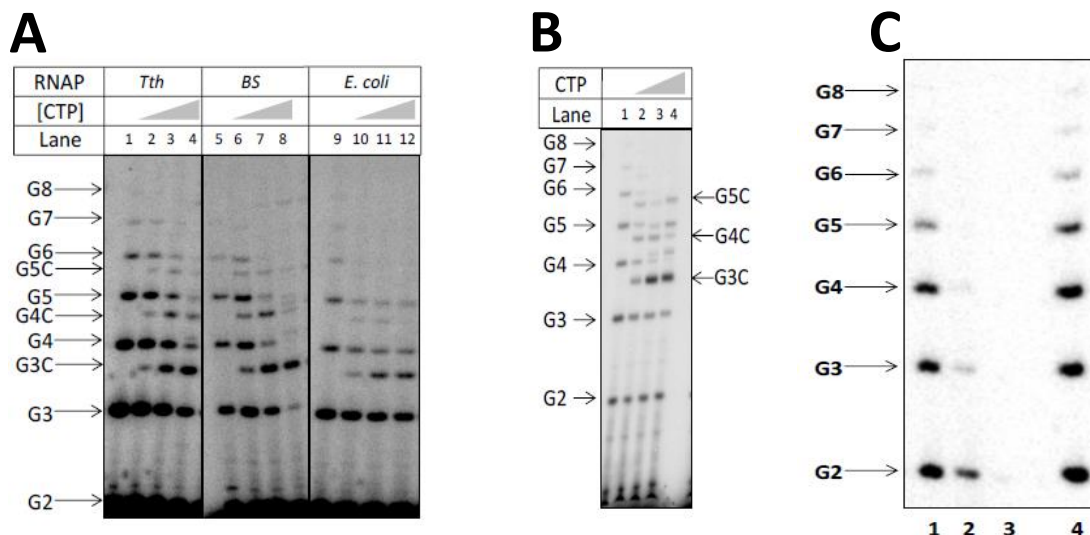


Figure 2.6: Characterization of reiterative transcription (A) *in vitro* transcription with fork-DNA used for crystallization (Figure 1.3) by different RNAPs. The reactions were performed in the presence of 100 μ M GTP and 1 μ Ci [γ - 32 P] GTP and different concentrations of CTP (0, 1, 10, and 100 μ M). (B) *in vitro* transcription with double-stranded DNA containing the *pyrG* promoter sequence by *T. th* RNAP. The reactions were performed in the presence of 100 μ M GTP and 1 μ Ci [γ - 32 P] GTP and different concentrations of CTP (0, 1, 10, and 100 μ M). (C) *in crystallo* transcription with a RTC crystal. Reaction was performed by directly soaking 400 μ M GTP and 1 μ Ci [γ - 32 P] GTP to the RTC crystal. Lane 1 shows RNAs released from the crystal, lane 2 and 3 are wash, and lane 4 shows RNAs retained in the RTC crystal. Figure from the reference (Murakami et al., 2017).

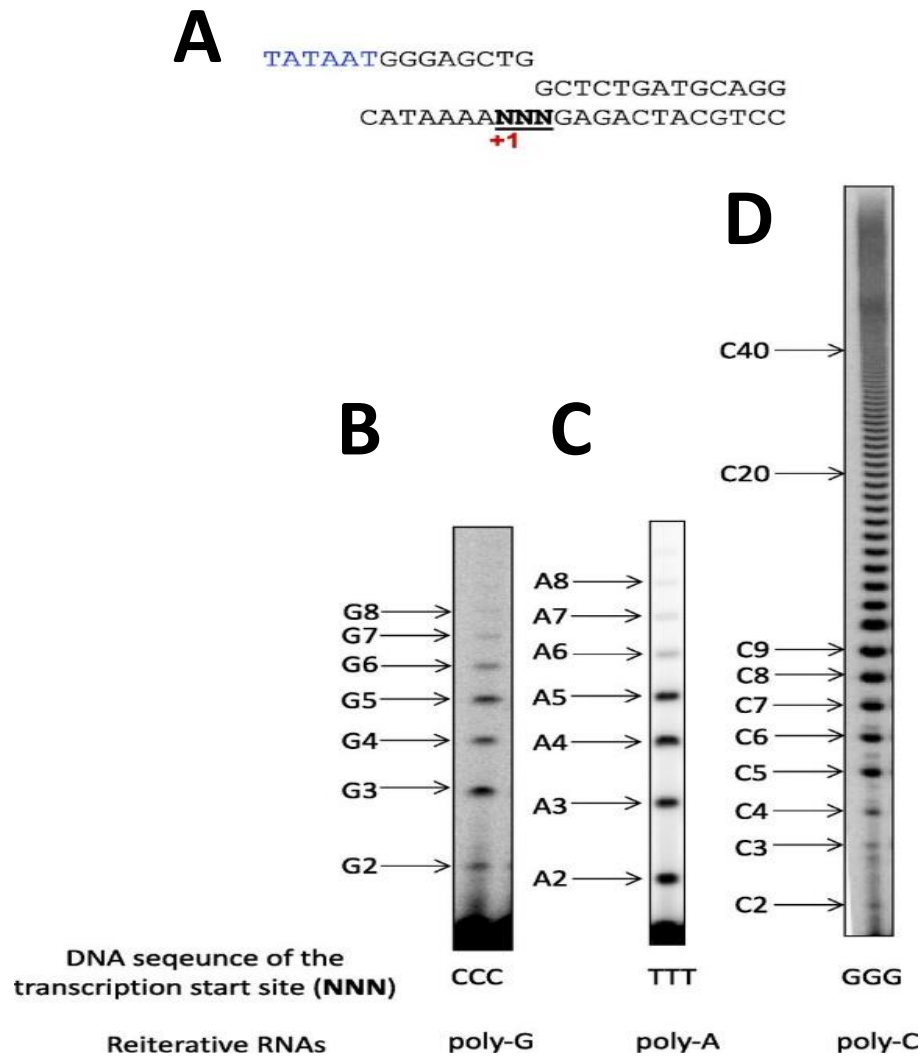


Figure 2.7: Characterization of the *pyrG* homopolymeric tract that determines the maximum length of reiterative transcripts. (A) The DNA scaffold containing variants *pyrG* homopolymeric tract (NNN) used for transcription assay. *In vitro* transcription assays were performed in the presence of 100 μ M GTP and 1 μ Ci [γ - 32 P] GTP (B), 100 μ M ATP and 1 μ Ci [γ - 32 P] ATP (C), or 100 μ M CTP and 1 μ Ci [α - 32 P] CTP (D). Figure from the reference (Murakami et al., 2017).

2.4.2 Capturing transcription initiation complex from the *pyrG* promoter

The structure of RTC-30' revealed an unexpected RNA extension pathway during reiterative transcription; however, several questions remain to be answered such as how RNA slips on tDNA and how the 5' end of RNA is guided toward the main channel of RNAP. Therefore, I further studied the mechanism of reiterative transcription from the *pyrG* promoter by determining a series of X-ray crystal structures of the RNAP and *pyrG* promoter complex containing 2-, 3- and 4-mer RNAs. To do this, I applied time-dependent soak-trigger-freeze X-ray crystallography method used by a past lab member with N4 RNA polymerase. After reproducing multiple crystals of RNAP and *pyrG* promoter complex as described in **Appendix C**, I soaked these crystals to a cryo-solution containing GTP to trigger reiterative transcription and froze at multiple time points to capture intermediates (**Figure 2.8**).

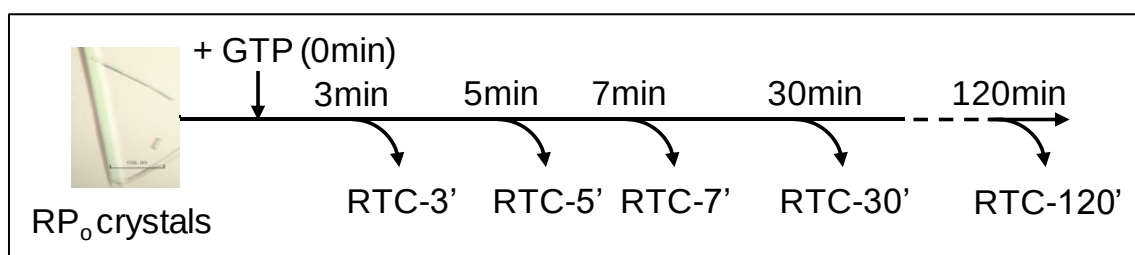


Figure 2.8: Experimental scheme of the reiterative transcription complex preparation by time-dependent soak-trigger-freeze X-ray crystallography. Figure from the reference (Shin et al., 2020)

After 3 min of GTP soaking (RTC-3', **Figure 2.9A**), RNAP synthesizes 2-mer RNA and the 3' end of the RNA is in a post-translocated state, forming a base pair with +2C template DNA (tDNA). Hereafter, RNA residues are counted -1, -2, -3 from the 3' end. The +3C tDNA is positioned at the $i+1$ site and forms a base pair with an incoming GTP. The α -phosphate of GTP is 5.9 Å away from the 3'-OH of RNA and the trigger loop is in the open conformation, indicating that the GTP is positioned at the pre-insertion site. The structure of RTC after 4 min of GTP soaking (RTC-4', structure not shown) was the same as the structure of the RTC-3'.

After 5 min of GTP soaking (RTC-5', **Figure 2.9B**), RNA extends to 4-mer and the 3' end of the RNA is in a post-translocated state, forming a base pair with +3C tDNA, while +4G tDNA is positioned at the $i+1$ site. The first three bases of RNA are synthesized by canonical transcription whereas the 4th base is added by reiterative transcription. The 5' end of 4-mer RNA (-4G) is inserted into the RIF-binding pocket, as observed in the previously published RTC structure (RTC-30') containing 8-mer RNA, indicating that the 5'-end of RNA base fits in the RIF-binding pocket followed by extension toward the main channel of RNAP. In the RTC-5', incoming GTP bound at the $i+1$ site forms a mismatch with the +4G tDNA and the trigger loop is in the open conformation

At 7 min (RTC-7', **Figure 2.9C**), RNA is extended to 6-mer and the 5'-end RNA base positions at fork loop 2 of the β subunit and the downstream edge of the transcription bubble. The 3' end of the RNA is in a post-translocated state, forming a base pair with +3C tDNA, while +4G tDNA positions at the $i+1$ site as observed in the RTC-5'. The NTP binding $i+1$ site is empty but traps pyrophosphate. The +4G tDNA is waiting for GTP for further extension of the poly-G transcript. I also prepared the RTC by soaking GTP for 2 hours, but there was no further RNA extension beyond 8-mer RNA (structure not shown) as observed in the RTC-30', indicating that the 8-mer RNA is the longest RNA produced by *in crystallo* transcription. A series of structures of RTC show that 1) extra G bases are added at the RNA 3' end by the G-G mismatch between the

+4G tDNA and incoming GTP; and 2) the fitting of -4G RNA in the RIF-binding pocket is an obligatory step before RNA extension beyond 4-mer RNA.

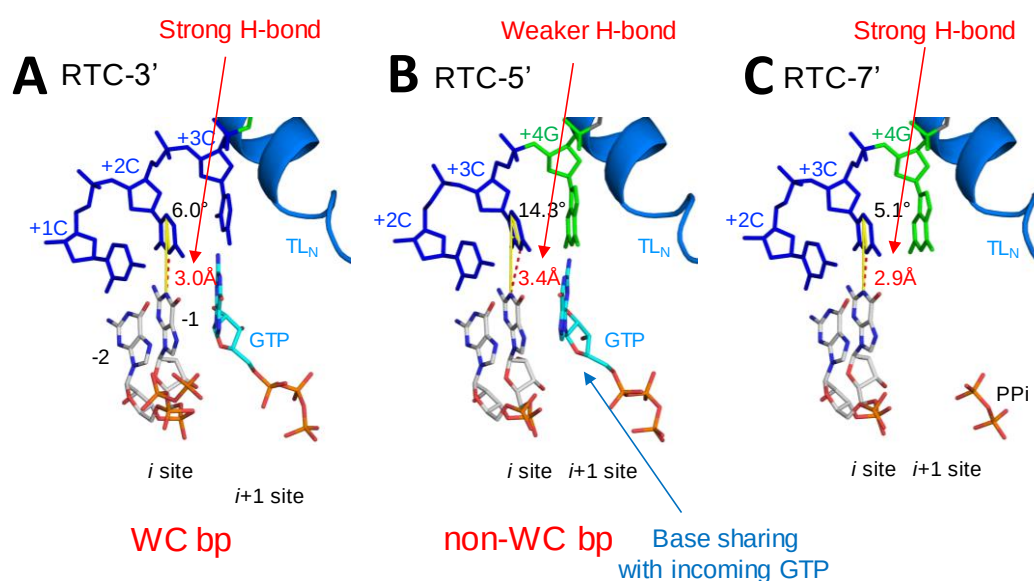


Figure 2.9: The structures of the RTC-3' (A), RTC-5' (B) and RTC-7' (C) from the *pyrG* promoter. RNA (white carbon), tDNA (blue and green) and incoming GTP at pre-insertion site (cyan carbon) are shown as stick models. The RNAP trigger loop (TL) is shown as a blue-ribbon model. Only partial portion of RNA and tDNA are shown for the clarity. The distance between N1 atom (RNA G base) and N3 atom (tDNA C base) is shown as a red dashed line. The angle of the DNA and RNA base pairing (From N1(RNA G base) to C5(tDNA C base) to N3 (tDNA C base)) is shown as yellow lines. Figure modified from the reference (Shin et al., 2020).

2.4.3 Comparison of the distances between bases to characterize hydrogen bonds

I captured the progression of RNA synthesis, from 2-mer to 8-mer, within the RTC crystals. In case of transcription from the *pyrG* promoter, the first three bases of RNA are synthesized by canonical transcription (*e.g.* RTC-3') and then RNA synthesis is continued by reiterative transcription (*e.g.* RTC-5', RTC-7' and RTC-30'). To gain insight into the RNA slippage mechanism, I analyzed the DNA and RNA base pairing at the *i* site of these structures. I assessed the distance between bases and the planarity of the base pair, which could be affected by GTP binding at the *i*+1 site during canonical and reiterative transcription. The distance between N1 of the G base of the RNA and N3 of the C base of the tDNA accommodated at the *i* site of RTC-3' is 3.0 Å whereas, it is extended to 3.4 Å in case of the RTC-5' containing the G-G mismatch at the *i*+1 site (**Figures 2.9 A and B**). An average distance between C (DNA) and G (RNA) bases in atomic resolution DNA/RNA hybrid crystal structures is 2.9 Å with a minimum 2.75 Å and maximum 3.15 Å (Saenger, 1984; Saenger and Suck, 1973; Voet and Rich, 1970). I therefore concluded that the base pair between tDNA (+3C) and 3' end of RNA is wobbled when RNAP switches the mode of RNA synthesis from canonical to reiterative transcription. The distance between C-G bases at the *i* site returns to 2.9 Å without GTP bound at the active site (RTC-7', **Figure 2.9C**) or when the RTC has completed poly-G RNA extension (RTC-30'). Not only a distance between bases, but also planarity of base pair at the *i* site, is impaired during reiterative transcription. In the RTC-5', +3C base of the tDNA is tilted about 15° toward the incoming GTP bound at the *i*+1 site (**Figure 2.9B**). A major difference found in the RTC-5' compared with other RTC structures is the presence of an incoming GTP at the *i*+1 site, forming a G-G mismatch with the tDNA +3G base, which may wobble base pairing at the *i* site and initiate RNA slippage.

2.4.4 Monitoring DNA translocation state during reiterative transcription in solution using 2-aminopurine fluorescence signal

The RTC-5', RTC-7' and RTC-30' structures show that the +4G DNA base is translocated into the $i+1$ site when extra G residues are added to RNA during reiterative transcription (**Figure 2.9B**). To validate this observation in solution, I monitored tDNA position during transcription by fluorescence signal of 2-aminopurine (2-AP) incorporated in the tDNA. 2-AP is an adenine analog (**Figure 2.10A**) (Sowers et al., 1986) and its fluorescence intensity is affected by its environment (Guest et al., 1991; Raney et al., 1994). 2-AP displays weak and strong fluorescence signals when it stacked and unstacked with neighboring bases, respectively (**Figures 2.10B and 2.11A**) (Malinen et al., 2012; Sullivan et al., 1997)

I prepared two DNA scaffolds containing a guanine DNA base as a quencher of 2-AP fluorescence with 2-AP at the +4 and +5 positions of the tDNA, respectively. One scaffold contains the tDNA sequence 3'-CATGX-5' (transcription start site is underlined, X=2-AP) for canonical transcription (**Figure 2.11B**) and another scaffold contains a RNA slippage prone tDNA sequence, 3'-CCCGX-5', for reiterative transcription (**Figure 2.10C**). For this assay, I used the *E. coli* RNAP σ^{70} holoenzyme since it is proficient at reiterative transcription from the *pyrG* promoter (Murakami et al., 2017).

In the case of the canonical transcription DNA template, 2-AP displays low fluorescence in the RNAP-DNA complex (2-AP positioned at +5 site remains stacked with a quencher G base at +4 site) whereas it shows increased fluorescence after adding GTP, UTP and ATP in the RNAP-DNA complex (**Figure 2.11C**). The control experiments show that in the absence of any of these NTPs, there was no change in fluorescence. The result indicates that after 3-mer RNA synthesis, the +4G DNA base moves to the active site of RNAP ($i+1$ site), leaving the 2-AP unstacked with +4G (**Figure 2.11A**).

In the case of the reiterative transcription DNA template, 2-AP fluorescence increases upon addition of GTP (but not ATP, CTP or UTP) to the RNAP-DNA complex for synthesizing 3-mer or longer RNA (**Figure 2.10D**), demonstrating that the +4G tDNA is translocated to the $i+1$ site of the RNAP active site, which is consistent with the observation from the structural analysis of RTC (**Figure 2.9**).

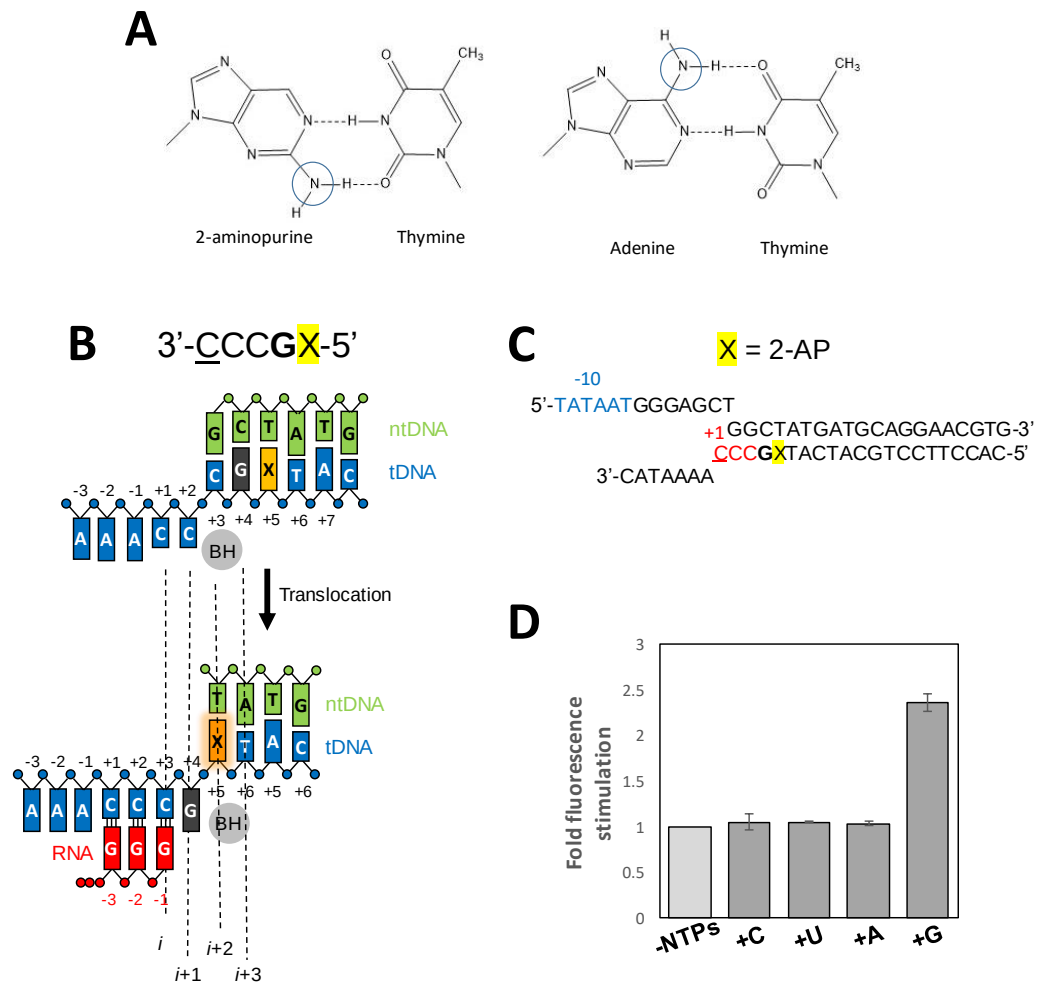


Figure 2.10: Monitoring equilibrium DNA translocation state in solution using fluorophore 2-aminopurine. (A) Chemical composition and arrangement of 2-AP resembles the adenine base and 2-AP base-pair with thymine. (B) Reaction scheme of reiterative transcription at the *pyrG* promoter. Before the addition of GTP (top), 2-AP (X, yellow box) is located further from the edge of bubble and quenched by neighboring base G (low fluorescence intensity). After the addition of GTP (bottom), G base quencher translocates to active site and leaves 2-AP unstacked (high fluorescence intensity). (C) The DNA containing transcription bubble, *pyrG* ITR, G base quencher at +4 position and 2-AP at +5 position. (D) Increase in fluorescence signal when GTP (but not ATP, CTP, UTP) was added to RNAP-DNA complex for synthesizing 3-mer or longer RNA. Figure modified from the reference (Shin et al., 2020).

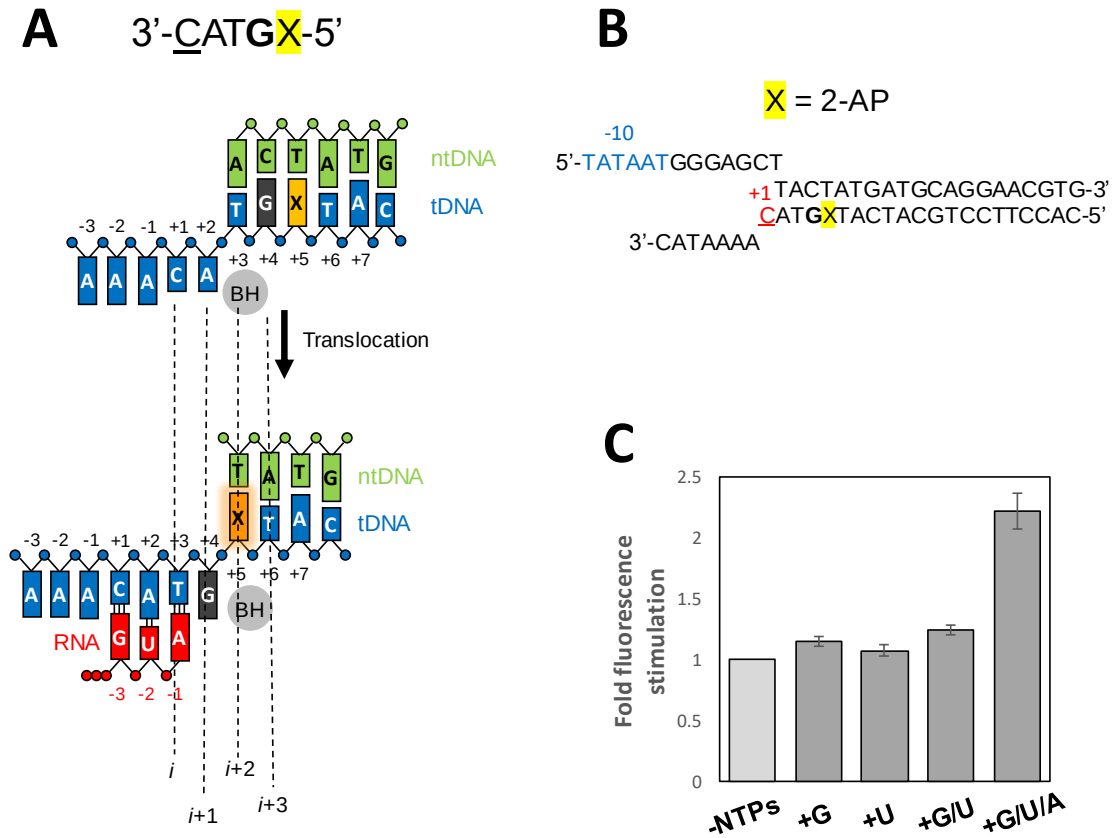


Figure 2.11: Control experiment using 2-AP to monitor translocation state of DNA in solution. (A) Reaction scheme of transcription at the control promoter. Before the addition of NTPs (top), 2-AP (X, yellow box) is located further from the edge of bubble and quenched by neighboring base G (low fluorescence intensity). After the addition of GTP/UTP/ATP (bottom), G base quencher translocates to active site and leaves 2-AP unstacked (high fluorescence intensity). (B) The DNA containing transcription bubble, G base quencher at +4 position and 2-AP at +5 position. This DNA does not contain the *pyrG* homopolymeric tract at ITR. (C) Increase in fluorescence signal when GTP/UTP/ATP was added to RNAP-DNA complex for synthesizing 3-mer RNA. Figure modified from the reference (Shin et al., 2020).

2.4.5 Measurement of DNA translocation kinetics during reiterative transcription

A series of the RTC structures determined in this study revealed that RNAP requires 4~5 min to synthesize 5-mer RNA from the *pyrG* promoter by *in crystallo* transcription, which is substantially slower than RNA synthesis from a promoter without a slippage prone sequence. For example, the structure of the initially transcribing complex containing 6-mer RNA (PDB: 4Q5S) (Basu et al., 2014), which contains the tDNA sequence 3'-TGAGTGC-5', requires only 20 sec to produce 6-mer RNA *in crystallo* in the presence of ATP, CTP and UTP.

To measure the speed of RNA synthesis from slippage prone DNA in solution, I monitored the fluorescence of 2-AP embedded in tDNA at the +5 position with stopped-flow technique. I used the same DNA for determining the DNA translocation state during reiterative and non-reiterative transcription. These DNA templates require at least 3-mer RNA synthesis for enhancing the 2-AP fluorescence signal. The data were best fit to a single exponential equation and kinetic values of the fluorescence signal from 2-AP are shown in **Figure 2.12**. The rates of DNA translation of the RNA slippage prone and the canonical transcription DNA templates are $\sim 0.401\text{ s}^{-1}$ and $\sim 1.00\text{ s}^{-1}$, respectively, indicating that RNA synthesis from the slippage prone DNA is substantially slower than from the DNA for canonical transcription. I showed that DNA containing homopolymeric tract sequence have the general trend of slower kinetics both *in crystallo* and *in vitro*. However, it should be noted that different RNAPs may possess different initiation kinetics, and therefore, should not be directly correlated in terms of RNAPs. More comprehensive experiments using complete dsDNA containing various slippage prone DNA sequences need to be performed to address the difference in kinetics.

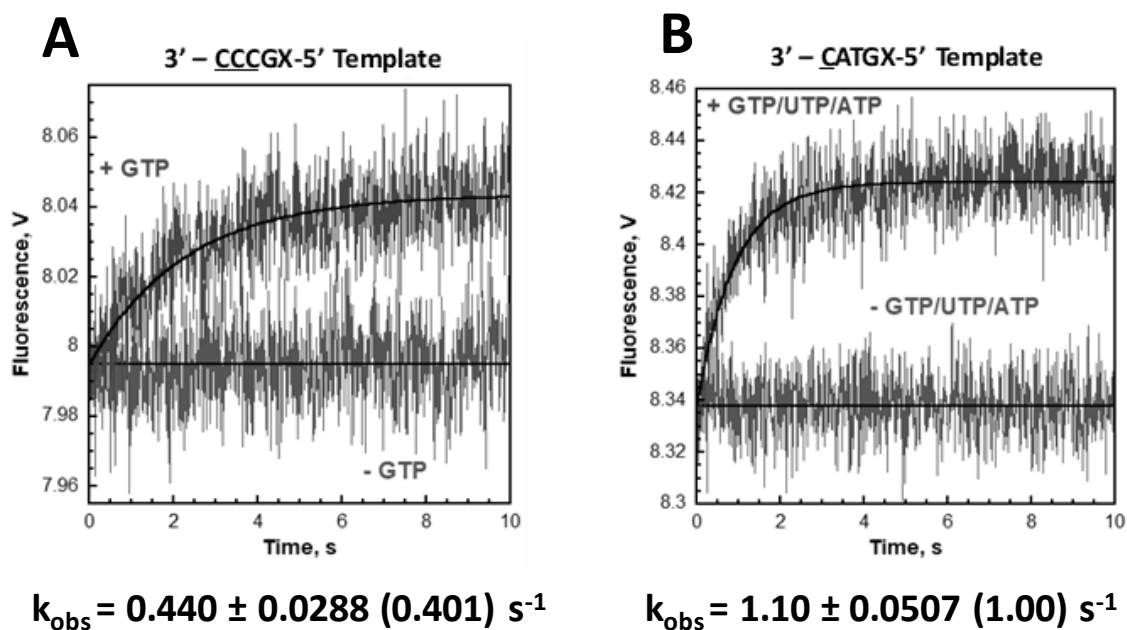


Figure 2.12: Kinetics analysis of DNA translocation rate. In reiterative transcription DNA template (A) and the canonical transcription DNA template (B) by measuring change in 2-AP signal upon the addition of nucleotides with stopped-flow. Figure from the reference (Shin et al., 2020)

2.5 Discussion

In this study, I determined the structure of a bacterial RNAP in the process of reiterative transcription, revealing a unique mechanism for extending RNA within the transcription initiation complex. This process includes an atypical RNA extension pathway, a stably maintained 3-bp DNA/RNA hybrid, and extension of the nascent RNA without displacement of σ 3.2 (**Figures 2.4-2.5**). I also analyzed how RNAP synthesizes RNA from the *pyrG* promoter by reiterative transcription. The first three RNA bases are synthesized without RNA slippage, which is a simple copy of DNA sequence to RNA. In the case of a CTP limited condition, RNAP extends RNA with GTP by reiterative transcription (**Figure 2.1**). The structure of RTC-5' containing 4-mer RNA,

which represents the transcription complex right after the mode of RNA synthesis has switched, revealed that +4G tDNA is positioned at the active site of RNAP ($i+1$ site) and forms a mismatch pair with an incoming GTP using an *anti-anti* conformation (**Figure 2.9B**). The structure also revealed that a base pair between tDNA (+3C tDNA) and the 3' end of RNA positioned at the i site is wobbled (N3 of C and N1 of G, 3.4 Å) and the +3C tDNA base is tilted toward an incoming GTP, likely due to a mismatch between +4G tDNA and incoming GTP (**Figure 2.9**). The opening of the trigger loop causes the α -phosphate of GTP to be 6.1 Å away from the 3'-OH of RNA and maintain GTP in a nonreactive state for the nucleotidyl transfer reaction. I speculate that the trigger loop would then adapt the completely closed conformation, forming the trigger helix, to load the GTP in the reactive state followed by the nucleotidyl transfer reaction (**Figure 2.13A**). RNA extension may trigger base sharing of the +3C tDNA with guanine bases of RNA at the i and $i+1$ sites and would promote shifting of the DNA/RNA hybrid in a stepwise manner (**Figure 2.13B**). This hypothesis is supported by the observation that the base pairs upstream from the i site are not wobbled. After RNA is extended, only RNA translocates in the upstream direction to prepare for the next cycle of reiterative transcription (**Figure 2.13C**). This prediction is supported by the results of 2-AP fluorescence-based DNA translocation assay (**Figures 2.10 – 2.12**). Extending RNA from 3-mer to 4-mer determines the fate of RNA synthesis from the *pyrG* promoter. Translocating +4G tDNA at the active site ($i+1$ site) after 3-mer RNA synthesis is also observed in the DNA translocation assay by monitoring the 2-AP fluorescence signal from tDNA. The +4G tDNA moving at the active site is essential for regulating *pyrG* expression depending on CTP availability; when the CTP concentration is high enough, CTP is loaded at the active site, forms a base pair with +4G tDNA and RNAP undergoes canonical transcription (**Figure 2.1**). Previous study by Meng et al observed the similar reiterative transcription pattern when the +4G tDNA was replaced with either A or T (Meng et al., 2014), further suggesting that the mismatch between the +4 tDNA and incoming GTP favors transcript slippage at *pyrG* promoter.

At some point during further RNA extension up to approximately eight bases, there can be a switch to nonreiterative transcription. This switch is accompanied by an increase in the length of the DNA/RNA hybrid, which would displace σ finger to expose the dedicated RNA exit channel of the core enzyme. Possibly, the 5' end of the nascent RNA transcript is then redirected toward the RNA exit channel. I speculate that reiterative transcription pauses when the 5'-end RNA reaches the narrow opening of the main channel of RNAP, providing time for CTP to be incorporated at the 3' end of RNA and switch from reiterative to canonical transcriptions. Increasing the length of the DNA/RNA hybrid triggers the eventual displacement of σ from core enzyme. Without σ -factor, the main and RNA exit channels are separated only by a weak interaction between the lid region (β' -subunit) and the inner surface of the flap domain (β -subunit), which can be disrupted by opening of the RNAP clamp and/or flap. Therefore, disruption of this interaction during the transition from reiterative to nonreiterative transcription might be involved in redirecting the RNA toward the dedicated RNA exit channel.

Interestingly, in all known examples of reiterative transcription leading to release of transcripts during initiation, the promoters involved differ from the *pyrG* promoter in a similar way. All of these promoters include an initially transcribed region in which the homopolymeric tract at which reiterative transcription occurs is located one or two bases downstream from the transcription start site (Turnbough and Switzer, 2008). Thus, at the start of reiterative transcription, the DNA/RNA hybrid will be one or two base pairs longer than that observed at the *pyrG* promoter. This small difference might be sufficient to prevent a switch from reiterative to canonical transcription. Although the RNAP structural features controlling these alternative fates are not known, it is possible that the presence of the 5' end base(s) preceding the homopolymeric tract causes nascent transcripts to be directed into a pathway within RNAP that precludes transcribing downstream DNA sequence. Confirming such a model will require structural comparisons of reiterative transcription complexes that differ in transcript fate.

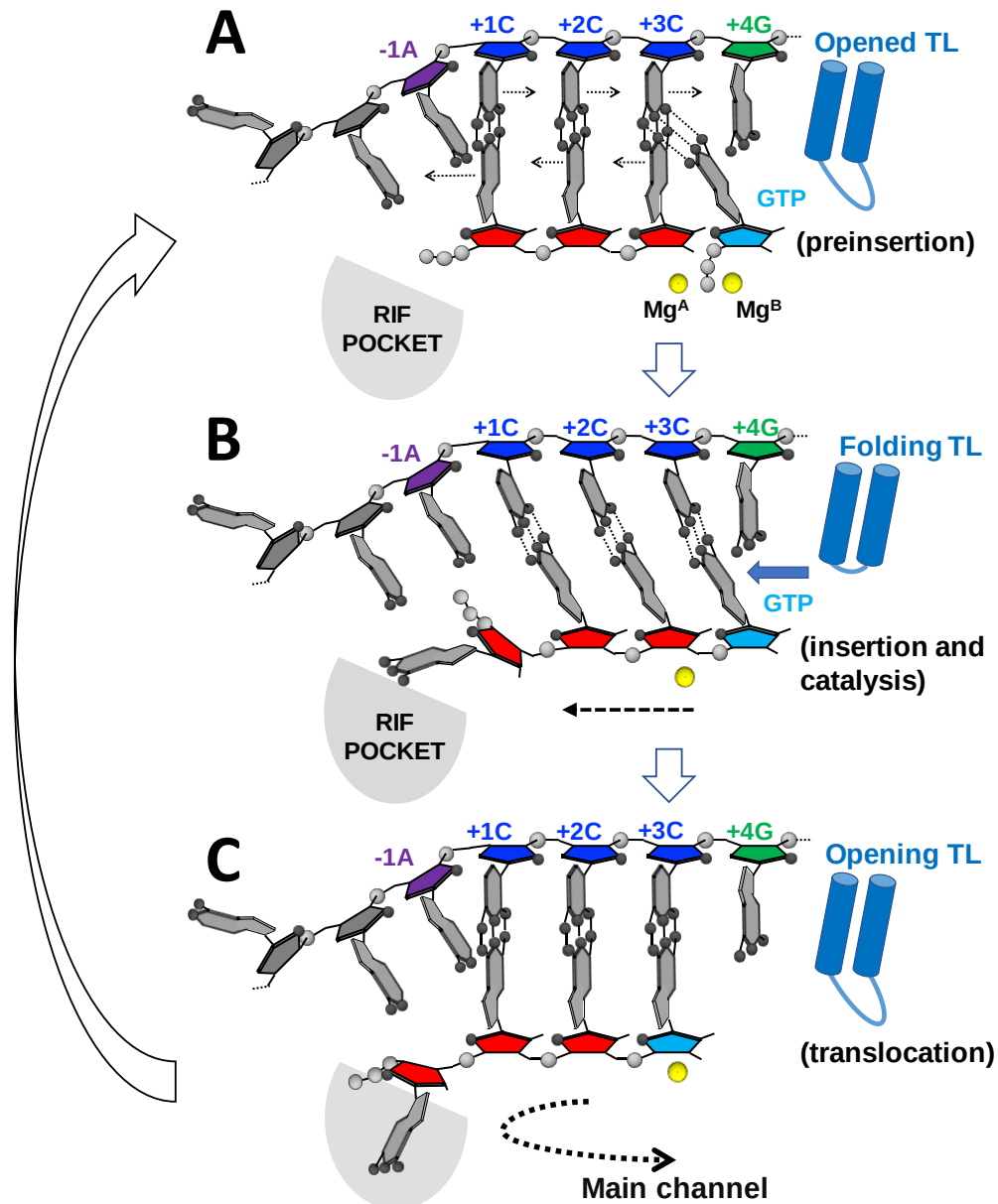


Figure 2.13: Proposed model of RNA slippage during reiterative transcription from the *pyrG* promoter. Cartoon model of the active site of RNAP. *pyrG* initially transcribed region sequence (template DNA, -1A: purple, +1C, +2C, +3C: blue, +4G: green), nascent RNA (5'-GGG-3', red), incoming GTP (cyan), active site Mg²⁺ (yellow sphere), trigger loop (blue cartoon), and rifampin-binding pocket (RIF POCKET, gray) are shown. Figure from the reference (Shin et al., 2020).

2.6 Acknowledgements

I thank the staffs at the MacCHESS for support of crystallographic data collection. I thank Dr. Charles L Turnbough Jr. at University of Alabama at Birmingham and Dr. Vadim Molodtsov, a past member of Murakami Lab for the discussion and comments. I thank Mark Signs at the fermentation facility at Penn State for training with fermentor and Dr. Mark Hedglin at Penn State for supporting with stopped-flow kinetics experiment. This work was supported by NIH grants GM94466 to Charles L Turnbough Jr. and GM087350 and GM131860 to Katsuhiko Murakami.

2.7 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 crystallographica Section D, Biological crystallography* 66, 1153-1163.
- Anikin, M., Molodtsov, V., Temiakov, D., and McAllister, W.T. (2010). Transcript Slippage and Recoding. In *Nucleic Acids and Molecular Biology* (Springer New York), pp. 409-432.
- Barr, J.N., and Wertz, G.W. (2001). Polymerase slippage at vesicular stomatitis virus gene junctions to generate poly(A) is regulated by the upstream 3'-AUAC-5' tetranucleotide: implications for the mechanism of transcription termination. *J Virol* 75, 6901-6913.
- 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.

- Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *J Biol Chem* 289, 24549-24559.
- Chamberlin, M., and Berg, P. (1962). Deoxyribonucleic acid-directed synthesis of ribonucleic acid by an enzyme from *Escherichia coli*. *Proc Natl Acad Sci U S A* 48, 81-94.
- Cheng, Y., Dylla, S.M., and Turnbough, C.L., Jr. (2001). A long T•A tract in the *upp* initially transcribed region is required for regulation of *upp* expression by UTP-dependent reiterative transcription in *Escherichia coli*. *Journal of bacteriology* 183, 221-228.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta crystallographica Section D, Biological crystallography* 60, 2126-2132.
- Fabre, E., Dujon, B., and Richard, G.F. (2002). Transcription and nuclear transport of CAG/CTG trinucleotide repeats in yeast. *Nucleic Acids Res* 30, 3540-3547.
- Gilman, B., Drullinger, L.F., Kugel, J.F., and Goodrich, J.A. (2009). TATA-binding protein and transcription factor IIB induce transcript slipping during early transcription by RNA polymerase II. *J Biol Chem* 284, 9093-9098.
- Guest, C.R., Hochstrasser, R.A., Sowers, L.C., and Millar, D.P. (1991). Dynamics of mismatched base pairs in DNA. *Biochemistry* 30, 3271-3279.
- Guo, H.C., and Roberts, J.W. (1990). Heterogeneous initiation due to slippage at the bacteriophage 82 late gene promoter *in vitro*. *Biochemistry* 29, 10702-10709.
- Jacques, J.P., and Kolakofsky, D. (1991). Pseudo-templated transcription in prokaryotic and eukaryotic organisms. *Genes & Development* 5, 707-713.

- Liu, C., Heath, L.S., and Turnbough, C.L., Jr. (1994). Regulation of *pyrBI* operon expression in *Escherichia coli* by UTP-sensitive reiterative RNA synthesis during transcriptional initiation. *Genes Dev* 8, 2904-2912.
- 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.
- Meng, Q., Turnbough, C.L., Jr., and Switzer, R.L. (2004). Attenuation control of *pyrG* expression in *Bacillus subtilis* is mediated by CTP-sensitive reiterative transcription. *Proc Natl Acad Sci U S A* 101, 10943-10948.
- Molodtsov, V., Anikin, M., and McAllister, W.T. (2014). The presence of an RNA:DNA hybrid that is prone to slippage promotes termination by T7 RNA polymerase. *J Mol Biol* 426, 3095-3107.
- Murakami, K.S., Shin, Y., Turnbough, C.L., Jr., and Molodtsov, V. (2017). X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway. *Proc Natl Acad Sci U S A* 114, 8211-8216.
- Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* 276, 307-326.
- Pal, M., and Luse, D.S. (2002). Strong natural pausing by RNA polymerase II within 10 bases of transcription start may result in repeated slippage and reextension of the nascent RNA. *Mol Cell Biol* 22, 30-40.

Raabe, M., Linton, M.F., and Young, S.G. (1998). Long runs of adenines and human mutations. *American journal of medical genetics* 76, 101-102.

Raney, K.D., Sowers, L.C., Millar, D.P., and Benkovic, S.J. (1994). A fluorescence-based assay for monitoring helicase activity. *Proc Natl Acad Sci U S A* 91, 6644-6648.

Saenger, W. (1984). *Principles of nucleic acid structure* (New York: Springer-Verlag).

Saenger, W., and Suck, D. (1973). The relationship between hydrogen bonding and base stacking in crystalline 4-thiouridine derivatives. *European journal of biochemistry* 32, 473-478.

Sowers, L.C., Fazakerley, G.V., Eritja, R., Kaplan, B.E., and Goodman, M.F. (1986). Base pairing and mutagenesis: observation of a protonated base pair between 2-aminopurine and cytosine in an oligonucleotide by proton NMR. *Proc Natl Acad Sci U S A* 83, 5434-5438.

Sullivan, J.J., Bjornson, K.P., Sowers, L.C., and deHaseth, P.L. (1997). Spectroscopic determination of open complex formation at promoters for *Escherichia coli* RNA polymerase. *Biochemistry* 36, 8005-8012.

Turnbough, C.L. (2011). Regulation of gene expression by reiterative transcription. *Current Opinion in Microbiology* 14, 142-147.

Turnbough, C.L., Jr., and Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* 72, 266-300.

van Leeuwen, F.W., de Kleijn, D.P., van den Hurk, H.H., Neubauer, A., Sonnemans, M.A., Sluijs, J.A., Koycu, S., Ramdjielal, R.D., Salehi, A., Martens, G.J., *et al.* (1998). Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. *Science (New York, NY)* 279, 242-247.

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

Voet, D., and Rich, A. (1970). The crystal structures of purines, pyrimidines and their intermolecular complexes. *Progress in nucleic acid research and molecular biology* 10, 183-265.

Wagner, L.A., Weiss, R.B., Driscoll, R., Dunn, D.S., and Gesteland, R.F. (1990). Transcriptional slippage occurs during elongation at runs of adenine or thymine in *Escherichia coli*. *Nucleic acids research* 18, 3529-3535.

Xiong, X.F., and Reznikoff, W.S. (1993). Transcriptional slippage during the transcription initiation process at a mutant *lac* promoter *in vivo*. *J Mol Biol* 231, 569-580.

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.

Zhou, Y.N., Lubkowska, L., Hui, M., Court, C., Chen, S., Court, D.L., Strathern, J., Jin, D.J., and Kashlev, M. (2013). Isolation and characterization of RNA polymerase *rpoB* mutations that alter transcription slippage during elongation in *Escherichia coli*. *J Biol Chem* 288, 2700-2710.

Chapter 3

Structural basis of non-productive reiterative transcription from the *pyrBI* and *pyrG* promoter variants

Publication from the work in this chapter is:

Shin, Y., Hedglin, M., and Murakami, K.S. (2020). Structural basis of reiterative transcription from the *pyrG* and *pyrBI* promoters by bacterial RNA polymerase. *Nucleic Acids Research*, 48, 2144-2155.

3.1 Abstract

In this chapter, I present a transcription initiation complex structure from the *pyrBI* promoter of *E. coli* determined by soak trigger freeze X-ray crystallography. The structure reveals that the 5' end of the nascent RNA extends toward the canonical RNA exit channel of RNAP and clash with the σ finger. The reiterative transcription complex structures from the *pyrG* promoter variant of *B. subtilis* that has base substitution at -1 template DNA position show an extra base pair within DNA/RNA hybrid and 5' end of RNA extending toward the RNA exit channel. The structures and biochemical assay suggest an alternative mechanism of RNA slippage from the *pyrBI* promoter and the extension of non-productive RNA requires the σ dissociation from the core enzyme.

3.2 Non-productive reiterative transcription – *pyrBI* operon of *E. coli*

In **Chapter 2**, I described a classical example of the productive reiterative transcription – the *pyrG* operon of *B. subtilis*. Another well-known example is UTP-sensitive regulation of transcription initiation of the *pyrBI* operon of *E. coli*, which encodes two subunits of the pyrimidine nucleotide biosynthetic enzyme aspartate transcarbamylase (Cheng et al., 2001; Liu et al., 1994; Tu and Turnbough, 1997; Turnbough 2011; Turnbough and Switzer, 2008). In this case, reiterative transcription occurs within a T₃ tract located in the initially transcribed region 5'-AATTTG (nontemplate strand sequence) of the *pyrBI* promoter (Liu et al., 1994). Reiterative transcription produces transcripts with the sequence 5'-AAUUU_n (where n = 1 to >100), especially all of which are released from the transcription initiation complex. The extent of this nonproductive reiterative transcription is directly proportional to the intracellular level of UTP, so that *pyrBI* gene expression is reduced when UTP levels are high (**Figure 3.1**).

In addition to fundamentally different fate of the reiterative transcripts, reiterative transcription at the *pyrBI* promoter is different compare to reiterative transcription at the *pyrG* promoter in 1) there is no attenuator involved in controlling the gene regulation at the *pyrBI* promoter; 2) reiterative transcription starts after the 5 bp DNA/RNA hybrid in the *pyrBI* promoter, which is two base pairs longer compared to reiterative transcription at the *pyrG* promoter; and 3) the presence of the non-homopolymeric tract at the 5' end of ITR preceding the homopolymeric tract.

To study similarity and difference of reiterative transcription from the *pyrBI* and *pyrG* promoters, I determined the crystal structures of the RNAP and DNA complex containing the *pyrBI* promoter and its TIC by soaking crystal into a solution containing ATP and UTP. Additionally, I investigated the role of adenine bases of tDNA of the *pyrG* promoter by determining the crystal structures of the RNAP - *pyrG* promoter variant complexes.

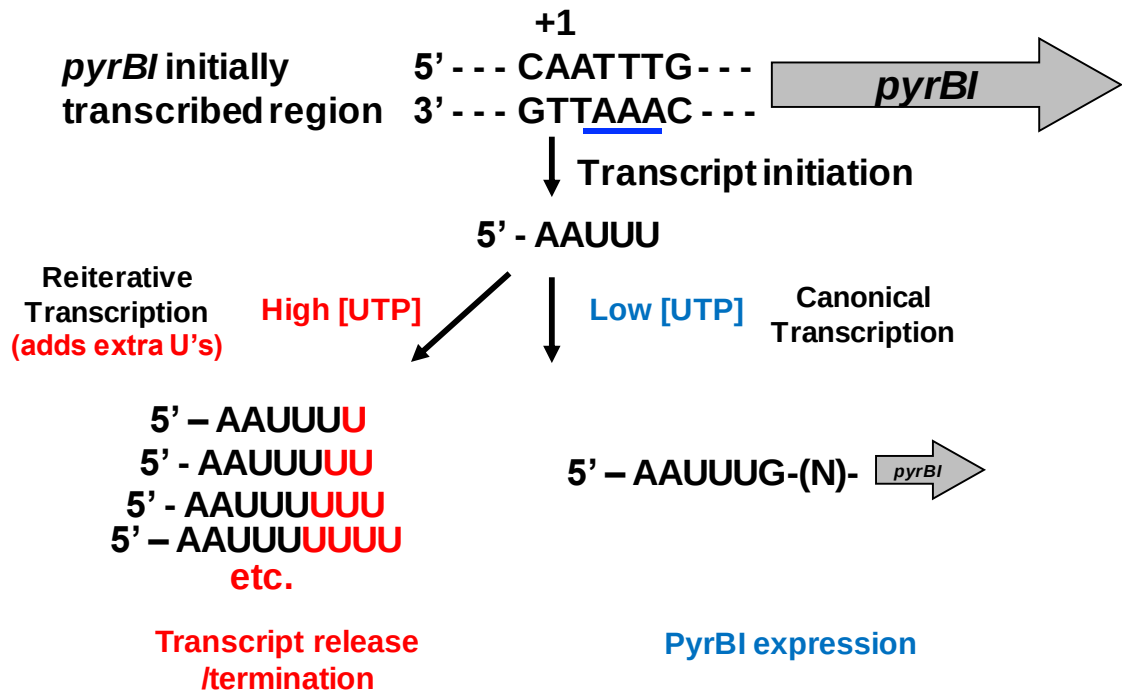


Figure 3.1: Model for UTP-mediated regulation of *pyrBI* expression in *E. coli*. High UTP concentration allows reiterative transcription that adds extra U bases to the nascent RNA (5'-AAUUU-3'). These transcripts are released from the transcript initiation complex, thereby *pyrBI* expression is reduced (left). Low UTP concentration allows canonical transcription by inserting GTP after the first five bases (5'-AAUUU-3') are synthesized, thereby *pyrBI* expression is induced (right). Figure from the reference (Shin et al., 2020).

3.3 Materials and Methods

3.3.1 Purification and crystallization of *T. thermophilus* RNAP

Endogenous *T. thermophilus* RNAP was purified as described in **Appendix A and B**. The *T. thermophilus* RNAP – DNA complex containing the *pyrBI* promoter sequence (**Figure 3.2A**) was crystallized as described in **Appendix C**. The binary complex crystal was sequentially transferred to cryoprotectants and soaked into a final cryo-solution containing 5 mM ATP and 5 mM UTP for 2 hr to trigger RNA synthesis *in crystallo*. For the *T. thermophilus* RNAP – DNA complex containing the *pyrG* variant sequences, the binary complex crystals were soaked into a final cryo-solution containing 2 mM GTP for 1 hr. The reactions were stopped by freezing the crystals in liquid nitrogen.

3.3.2 X-ray Data collections and Structure Determinations

The X-ray datasets were collected at the Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS) F1 beamline (Cornell University, Ithaca, NY) and structures were determined using crystallographic software: HKL2000, Phenix and Coot (Afonine et al., 2010; Basu et al., 2014; Emsley and Cowtan, 2004; Otwinowski and Minor 1997;). Final structures were deposited in the Protein Data Bank (PDB) database with their PDB ID codes listed in **Tables 3.1 and 3.2**.

Table 3.1: Data collection and refinement statistics of *pyrBI*

Complex	<i>pyrBI</i> -R _{Po}	<i>pyrBI</i> -TIC
PDB code	6P70	6P71
Data collection		
Space group	C2	C2
Cell dimensions		
<i>a</i> (Å)	185.18	187.51
<i>b</i> (Å)	100.85	100.66
<i>c</i> (Å)	294.87	296.30
β (°)	98.81	98.32
Resolution (Å)	50 – 3.05	50 – 2.85
Total reflections	342,091	435,858
Unique reflections	101,835	122,765
Redundancy	3.4 (2.9)	3.6 (2.5)
Completeness (%)	99.7 (97.0)	96.2 (77.7)
<i>I</i> / σ	10.9 (1.21)	9.86 (1.0)
CC ^{1/2}	(0.730)	(0.647)
Refinement		
Resolution (Å)	46 – 3.05	49 – 2.92
<i>R</i> _{work}	0.207	0.211
<i>R</i> _{free}	0.253	0.249
No. of atoms	28,359	28,417
R.m.s deviations		
Bond length (Å)	0.011	0.012
Bond angles (°)	1.451	1.513
Clashscore	10.14	6.83
Ramachandran favored, %	97.1	97.2
Ramachandran outliers, %	0.32	0.37

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

*Highest resolution shells are shown in parentheses

Table 3.2: Data collection and refinement statistics of -1 tDNA variants

Complex	-1G	-1C	-1T
PDB code	6OVR	6OVY	6OW3
Data collection			
Space group	C2	C2	C2
Cell dimensions			
<i>a</i> (Å)	185.90	184.83	186.19
<i>b</i> (Å)	101.42	102.48	101.83
<i>c</i> (Å)	295.20	295.60	295.76
β (°)	98.64	98.80	98.64
Resolution (Å)	50 – 2.83	50 – 3.0	50 – 2.77
Total reflections	471,538	398,603	518,675
Unique reflections	123,691	106,213	136,486
Redundancy	3.8 (3.3)	3.8 (3.0)	3.8 (3.1)
Completeness (%)	97.1 (76.6)	97.6 (81.6)	97.7 (78.2)
<i>I</i> / σ	16.3 (1.27)	9.15 (1.21)	16.24 (1.29)
CC ^{1/2}	(0.740)	(0.644)	(0.744)
Refinement			
Resolution (Å)	42 – 2.84	46 – 3.0	37 – 2.77
<i>R</i> _{work}	0.221	0.213	0.220
<i>R</i> _{free}	0.267	0.237	0.257
No. of atoms	28,535	28,273	28,430
R.m.s deviations			
Bond length (Å)	0.013	0.012	0.012
Bond angles (°)	1.087	1.573	1.643
Clashscore	24.08	12.47	12.46
Ramachandran favored, %	93.9	92.9	92.5
Ramachandran outliers, %	0.87	0.9	1.1

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

*Highest resolution shells are shown in parentheses

3.3.3 *In vitro* transcription assay

Transcription assay on the *pyrBI* promoter was performed in 10 µl transcription buffer (40 mM Tris-HCl (pH 8 at 25 °C), 30 mM KCl, 10 mM MgCl₂, 15 µM acetylated BSA, 1 mM DTT) containing 250 nM RNAP holoenzyme, 250 nM DNA, 100 µM ATP, 100 µM GTP, 1 µCi [γ-³²P] ATP and different concentrations of UTP (5 mM to 10 µM). Transcription assays on the *pyrG* promoter variants were performed in 10 µl containing 250 nM RNAP holoenzyme, 250 nM DNA, ³²P-labelled GpGpG primer and 100 µM GTP. The reactions were incubated at 37 °C for 10 mins and stopped by adding 10 µl stop buffer containing 90% formamide, 50 mM EDTA, xylene cyanol and bromophenol blue. RNA products were electrophoretically separated on a denaturing 24 % polyacrylamide / 7 M urea gel and visualized with a phosphorimager (Typhoon 9410; GE Healthcare).

3.4 Results

3.4.1 *In vitro* transcription assay using *pyrBI* promoter DNA

First, I tested RNAP activity with a fork-DNA containing *pyrBI* homopolymeric sequence (**Figure 3.2A**) by performing *in vitro* transcription assay. ATP, GTP and a different concentration of UTP was added to the RNAP-DNA binary complex. The results showed that reiterative transcription from the fork-DNA with the *pyrBI* promoter is depending on the concentration of UTP (i.e. produces much longer RNA products in the presence of the high concentration of UTP) (**Figure 3.2B**)

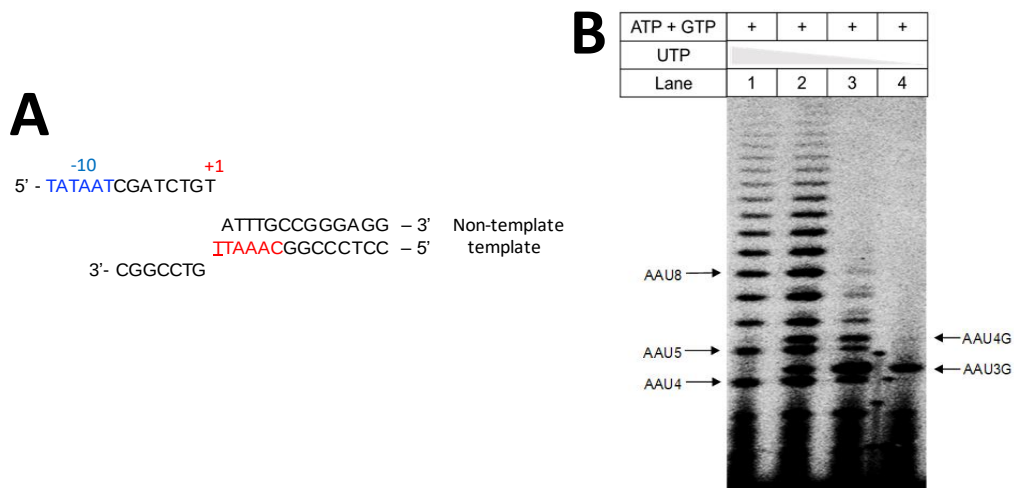


Figure 3.2: Characterization of reiterative transcription from the *pyrBI* promoter. (A) The fork DNA containing -10 element (blue) and the *pyrBI* initially transcribed region (red). The transcription start site is underlined. (B) *In vitro* transcription assay from the *pyrBI* promoter. RNAP produces the longer transcripts in the presence of high UTP concentrations. UTP concentrations: lane 1, 5 mM; lane 2, 1 mM; lane 3, 100 μ M; lane 4, 10 μ M. Figure modified from the reference (Shin et al., 2020).

3.4.2 Structure of a transcript initiation complex from the *pyrBI* promoter

The structures before (open complex) and after (transcription initiation complex) the addition of nucleotides were determined at 3.0 Å and 2.9 Å resolution, respectively (**Figure 3.3**). In the RNAP and DNA open complex structure (**Figure 3.3A**), thymine bases of tDNA position at the *i* and *i*+1 sites of the RNAP active site, showing that the RNA synthesis starts from +1 position. The TIC structure (**Figure 3.3B**) shows that RNAP active site containing 4-mer RNA (5'-AAUU) with a UTP positioned at the pre-insertion site. RNA extends directly toward the σ finger and steric clash between the RNA triphosphate and σ finger prohibits RNA extension beyond 4-mer RNA.

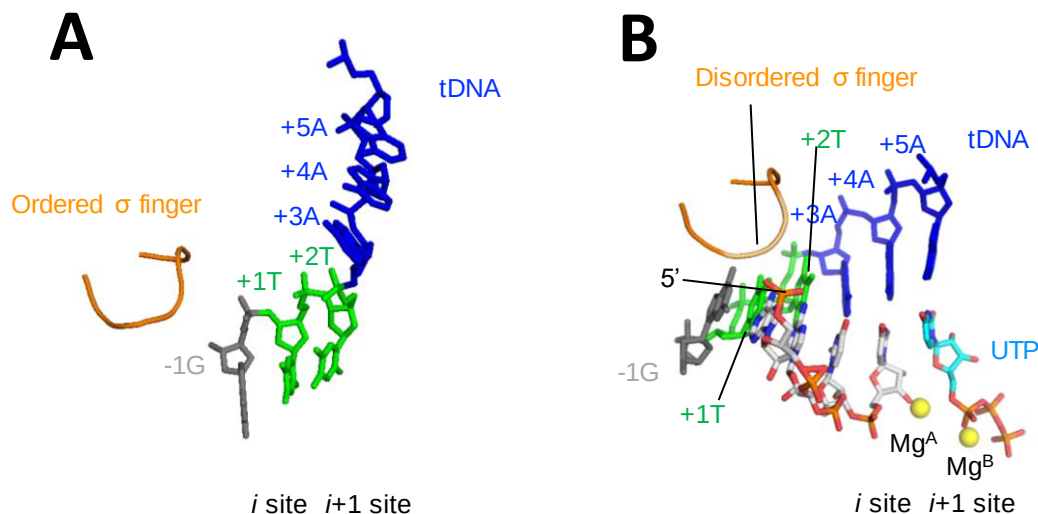


Figure 3.3: The structures of the RNAP-pyrBI promoter binary complex (A) and transcription initiation complex (B). tDNA (-1G: gray, +1T, +2T: green, +3A, +4A, +5A: blue) RNA (white carbon) and incoming UTP (cyan carbon) are shown as stick models. σ finger is shown as a ribbon model and the Mg^{2+} ions bound at the active site of RNAP are shown yellow spheres. Figure from the reference (Shin et al., 2020).

3.4.3 Investigation of the role of -1 base of template DNA in reiterative transcription

The RTC-5', RTC-7' and RTC-30' structures at the *pyrG* promoter (**Chapter 2**) showed the -1A base of tDNA partially overlapping with the -3G base of RNA (**Figure 3.4A**). Such base stacking is only possible in the presence of a purine (tDNA) and purine (RNA) combination (adenine in tDNA and guanine in RNA in the case of *pyrG* promoter transcription) (**Figure 3.4B**). Consistent with this fact, the upstream sequence of the transcription start site of the *pyrG* promoter is highly conserved in other closely related bacteria, and particularly, tDNA bases at positions -1 and -2 are adenine in a majority of promoters (**Figure 3.4C**). I therefore hypothesized that the -1A tDNA may block RNA extension toward the RNA exit channel, thereby the -4G

RNA base is pushed into the RIF-binding pocket when RNA slips on the tDNA. I investigated the role of adenine bases of tDNA by *in vitro* transcription. The *pyrG* promoter from *B. subtilis* contains four adenine bases from the -4 to -1 positions in the tDNA. I prepared a series of *pyrG* promoter variants with substitutions from the -4 to -1 positions (**Figure 3.4D**) and tested their abilities to produce reiterative transcripts. The reaction was performed in the presence of a pGpGpG primer complementary to the tDNA positions from +1 to +3 and GTP for efficient RNA extension. The wild-type promoter produces poly-G RNA around eight bases in length, and adenine bases at -4, -3 and -2 tDNA can be substituted with thymine without changing the activity of reiterative transcription (**Figure 3.4E, lanes 1, 2 and 4**). Guanine substitutions at -2 and -1 positions of tDNA did not influence the transcription (**Figure 3.4E, lane 3**). In contrast, *pyrG* derivatives containing thymine or cytosine substitutions at the -1 position increased the length of poly-G RNA transcript substantially (20-mer or longer) (**Figure 3.4E, lanes 5-9**). These RNA products produced from lanes 5-9 are non-productive, which cannot be switched back to canonical transcription (personal communication with Dr. Charles Turnbough). These results indicate that the tDNA base at the -1 position determines the length of reiterative RNA product from the *pyrG* promoter; short RNAs are synthesized in the presence of adenine and guanine (purine bases) and longer RNA are produced in the presence of thymine and cytosine (pyrimidine bases).

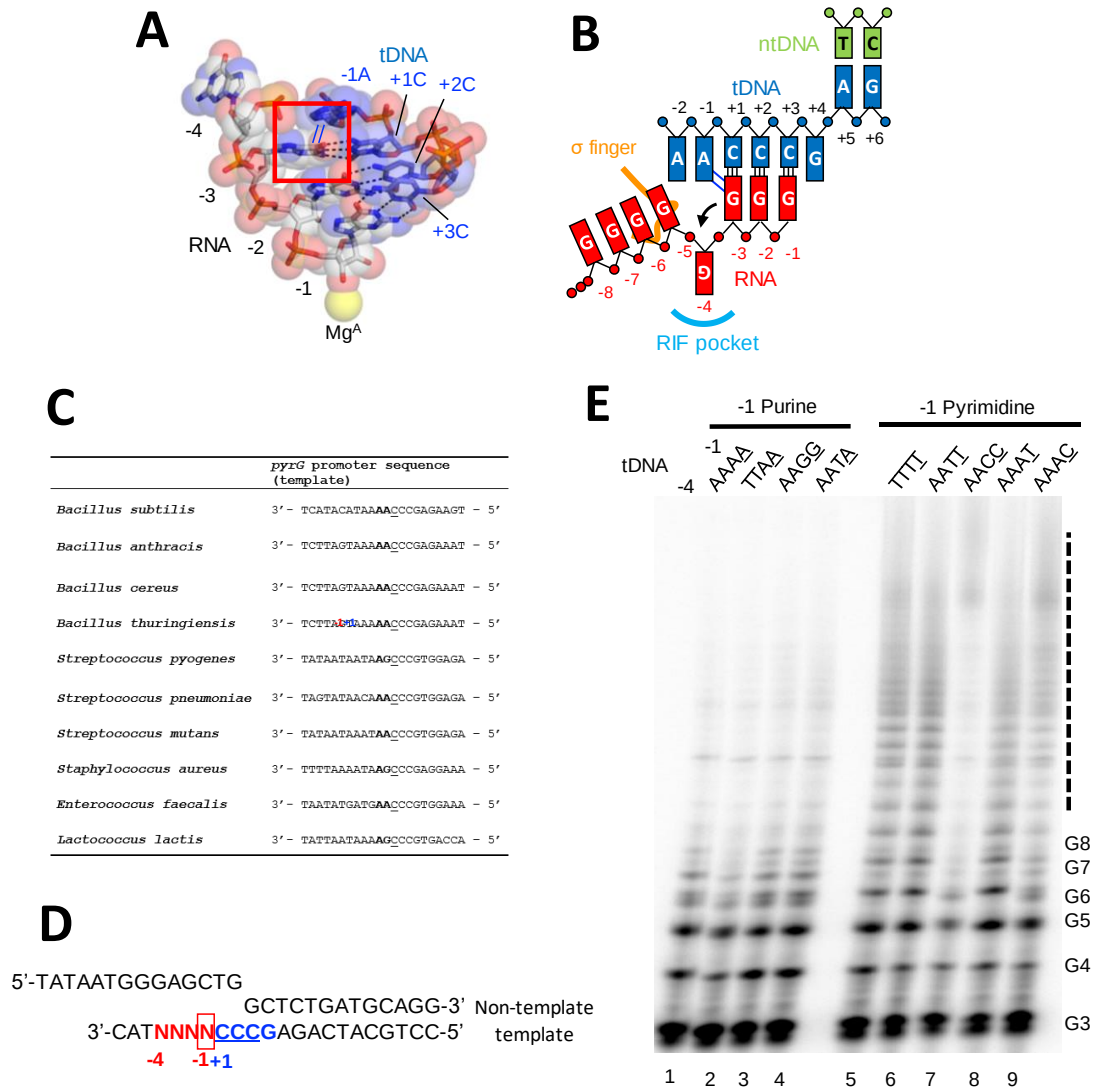
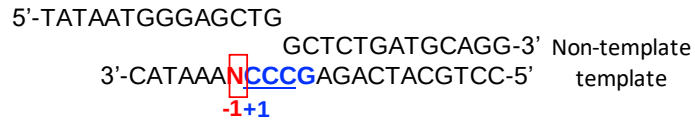


Figure 3.4: Characterization of the role of the -1 template DNA base of the *pyrG* promoter in reiterative transcription. (A) Stick models showing base pairings (black dashed lines) and base stacking interactions (blue solid lines) from the DNA/RNA hybrid of RTC-7' structure. (B) schematic cartoon representation of reiterative transcription from the *pyrG* promoter. Base stacking interaction between adenine base of tDNA at -1 position and guanine base of RNA is shown as blue lines. (C) Template DNA sequences of the *pyrG* promoter from the closely related bacteria. Transcription start sites are underlined and purines tracks from -2 to -1 positions are in bold. (D) The fork DNA containing -10 element and *pyrG* initially transcribed region (blue). Different -4 to -1 sequences are used (red). Transcription start site is indicated as +1. (E) *In vitro* transcription assays with *pyrG* promoter variants comparing poly-G RNA production. The sequences of tDNA from -4 to -1 positions are indicated. The sequences of -1 position is underlined. Poly-G RNAs longer than 8 bases are indicated as a dashed line. Figure adapted from the reference (Shin et al., 2020).

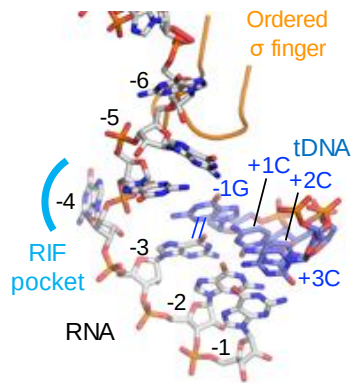
3.4.4 Structural analysis of the role of -1 base of template DNA from the *pyrG* promoter in reiterative transcription

To understand how -1 base of tDNA determines the length of reiterative transcription products, I determined the crystal structures of RTC containing *pyrG* variants having tDNA bases of -1G, -1C and -1T (**Figure 3.5A**). All RTC structures show electron densities corresponding to poly-G RNA products starting from the RNAP active site; however, the position of the 5' end of RNA is different depending on the sequence of tDNA. In the case of the RTC containing -1G tDNA (RTC-1G) (**Figure 3.5B**), the RNA is accommodated as found in the wild-type RTC; the RNA forms a 3-bp hybrid with tDNA, the -3G RNA base partially overlaps with -1G tDNA base, the -4G RNA base fits into the RIF-binding pocket and the 5' end of RNA extends toward the main channel. In sharp contrast, the RTCs with -1C and -1T tDNA bases (RTC-1C and RTC-1T) (**Figures 3.5 C and D**) contain a 4-bp DNA/RNA hybrid; -4G RNA base forms a Watson-Crick base pair (C-G) with tDNA base (-1C) in the RTC-1C and a wobble base pair (T-G) with tDNA base (-1T) in the RTC-1T. In both structures, the triphosphate of RNA contacts the tip of the σ finger. The path of RNA in the RTC-1C and RTC-1T is the same as the initially transcribing complex containing a 6-bp DNA/RNA hybrid (PDB:4G7H) (Basu et al., 2014).

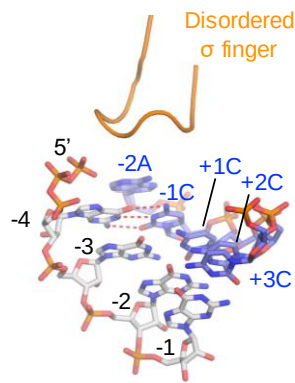
A RPo + GTP, 1hr



B RTC-1G



C RTC-1C



D RTC-1T

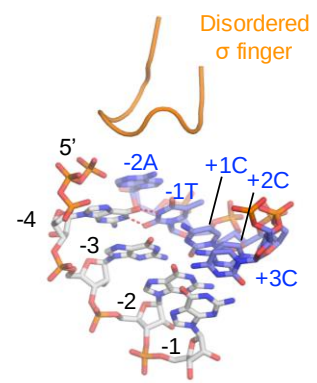


Figure 3.5: Structural analysis of the role of the -1 template DNA base of the *pyrG* promoter in reiterative transcription. (A) The fork DNA containing -10 element and *pyrG* initially transcribed region (blue). The position of -1 tDNA is red-boxed. Transcription start site is indicated as +1. (B-D) The structures of the RTC with *pyrG* promoter variants having -1G tDNA (B), -1C tDNA (C) and -1T tDNA (D). RNAs (white carbon) and tDNAs (blue carbon) are shown as stick models and σ finger is depicted as a ribbon model. Base stacking interaction in RTC-1G is shown as blue lines (B) and base pairing interactions in RTC-1C and RTC-1T are shown as red dashed lines (C and D).

3.5 Discussion

In this chapter, I structurally and biochemically analyzed the mechanism of the non-productive reiterative transcription from the *pyrBI* and *pyrG* promoter variants. Although the reiterative RNAs produced from the natural *pyrG* promoter containing the adenine base at -1 tDNA position can be switched back to normal elongation (**Chapter 2**), the substitution of this base to a pyrimidine (cytosine or thymine) makes the reiterative RNAs non-productive, which can not be switched to canonical elongation (**Chapter 3**). Similarly, non-productive reiterative transcription controls the gene expression from the *pyrBI* promoter (**Figure 3.1**). Based on the structures determined in this study, the reiterative RNAs produced from these promoters extend toward the canonical RNA exit channel and collide with the σ finger, which would eventually trigger the release of the σ factor from the core RNAP for longer RNA extension (**Figures 3.3 and 3.5**). *In vitro* transcription assay shows RNAP produces much longer reiterative RNA from these promoters relative to the wild-type *pyrG* promoter (**Figures 3.2 and 3.4**).

It has been demonstrated that two short segments of the 5' untranslated region of *pyrG* are required for CTP concentration-dependent regulation, including the ITR of the *pyrG* promoter (5'-GGGC; transcription start site is underlined) and the pyrimidine-rich sequence of the attenuator. In addition to these DNA *cis*-elements, in this chapter, I shed light on the function of the purine-rich sequence of tDNA just upstream of the transcription start site. These bases, particularly a purine base at -1 tDNA position, maintain the length of DNA/RNA hybrid to three bases and guide the 5' end of RNA toward the RIF-binding pocket of RNAP, which is an obligatory step for RNA extension toward the main channel of RNAP. Reiterative RNA synthesis pauses around 8-10 bases when the 5'-end RNA reaches the narrow opening of the main channel of RNAP, may providing time for CTP to be incorporated at the 3'-end of RNA and for RNAP to switch the mode of RNA synthesis from the reiterative to canonical transcription. Maintaining the three-base

DNA/RNA hybrid length in the *pyrG* RTC is critical to the regulation of *pyrG* gene expression depending on the CTP availability. In the ITR, not only a run of three but also a run of four or five G residues permits reiterative transcription; however, these extra Gs in the ITR result in less than optimum regulation of *pyrG* expression under CTP limited conditions. A run of five or more G residues suppresses RNA slippage due to suppression of the DNA/RNA hybrid melting (Elsholz et al., 2007).

The mechanism of RNA extension observed for the *pyrG* promoter will not hold for other promoters engaged in reiterative transcription, such as *pyrBI* (5'-A*ATTTG*; transcription start site is underlined and slippage-prone sequence is italicized) (Liu et al., 1994), *codBA* (5'-A*TTTTTTG*) (Qi and Turnbough, 1995) and *upp-uraA* (5'-G*ATTTTTTTG*) (Tu and Turnbough) in *E. coli*, because these promoters produce 5-10 bases of RNA before slipping and synthesize much longer stretches of reiterative RNAs (30 nucleotides or longer). Furthermore, once reiterative transcription starts from these promoters, RNA synthesis does not switch to canonical transcription. the mechanism of reiterative transcription from these promoters might be similar to *pyrG* promoter variants containing -1T or -1C as investigated in this chapter; these variants produce much longer reiterative RBA relative to the wild-type *pyrG* promoter. This hypothesis is consistent with the structural observation that the crystal containing the *pyrBI* promoter in the presence of ATP and UTP substrates formed 4-mer RNA (5'-AAUU-3') and its 5' end collides with the σ finger. Lastly, I also speculate that different propeller twist angles between the bases in A:T and G:C base pairs (Yoon et al., 1988) may be observed in an alternative mechanism of reiterative transcription.

The direct visualization of RNAP accommodating a full-length and long stretch of reiterative RNA from these *pyrBI*-type promoters cannot be addressed by X-ray crystallography because σ factor release is not permitted in the RNAP crystals. Elucidating such structures will be

achieved in the future by cryo-electron microscopy (cryo-EM), a powerful method to determine atomic resolution structure in solution (**Chapters 4 and 5**) (Kuhlbrandt, 2014).

3.6 Acknowledgements

I thank the staffs at the MacCHESS for support of crystallographic data collection. I thank Dr. Charles L Turnbough Jr. at University of Alabama at Birmingham and Dr. Vadim Molodtsov, past member of Murakami Lab for the discussion and comments. I thank Mark Signs at the Penn State fermentation facility for training with fermentor. This work was supported by NIH grants GM087350 and GM131860 to Katsuhiko Murakami.

3.7 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 crystallographica Section D, Biological crystallography* 66, 1153-1163.

Basu, R.S., Warner, B.A., Molodtsov, V., Pupov, D., Esyunina, D., Fernandez-Tornero, C., Kulbachinskiy, A., and Murakami, K.S. (2014). Structural basis of transcription initiation by bacterial RNA polymerase holoenzyme. *The Journal of biological chemistry* 289, 24549-24559.

Cheng, Y., Dylla, S.M., and Turnbough, C.L., Jr. (2001). A long T. A tract in the upp initially transcribed region is required for regulation of upp expression by UTP-dependent reiterative transcription in *Escherichia coli*. *Journal of bacteriology* 183, 221-228.

Elsholz, A.K.W., Jorgensen, C.M., and Switzer, R.L. (2007). The number of G residues in the *Bacillus subtilis* *pyrG* initially transcribed region governs reiterative transcription-mediated regulation. *J Bacteriol* 189, 2176-2180.

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

Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science (New York, NY)* 343, 1443-1444.

Liu, C., Heath, L.S., and Turnbough, C.L., Jr. (1994). Regulation of *pyrBI* operon expression in *Escherichia coli* by UTP-sensitive reiterative RNA synthesis during transcriptional initiation. *Genes & development* 8, 2904-2912.

Meng, Q., Turnbough, C.L., and Switzer, R.L. (2004). Attenuation control of *pyrG* expression in *Bacillus subtilis* is mediated by CTP-sensitive reiterative transcription. *Proc Natl Acad Sci U S A* 101, 10943-10948.

Murakami, K.S., Shin, Y., Turnbough, C.L., Jr., and Molodtsov, V. (2017). X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway. *Proc Natl Acad Sci U S A* 114, 8211-8216.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in enzymology* 276, 307-326.

Qi, F., and Turnbough, C.L., Jr. (1995). Regulation of *codBA* operon expression in *Escherichia coli* by UTP-dependent reiterative transcription and UTP-sensitive transcriptional start site switching. *J Mol Biol* 254, 552-565.

Shin, Y., Hedglin, M., and Murakami, K.S. (2020). Structural basis of reiterative transcription from the pyrG and pyrBI promoters by bacterial RNA polymerase. *Nucleic Acids Res* 48, 2144-2155.

Tu, A.H., and Turnbough, C.L., Jr. (1997). Regulation of upp expression in *Escherichia coli* by UTP-sensitive selection of transcriptional start sites coupled with UTP-dependent reiterative transcription. *Journal of bacteriology* 179, 6665-6673.

Turnbough, C.L. (2011). Regulation of gene expression by reiterative transcription. *Current Opinion in Microbiology* 14, 142-147.

Turnbough, C.L., Jr., and Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* 72, 266-300.

Yoon, C., Prive, G.G., Goodsell, D.S., Dickerson, R.E. (1988). Structure of an alternating-B DNA helix and its relationship to A-tract DNA. *Proc Natl Acad Sci U S A* 85, 6332-6.

Chapter 4

Structural basis of ribosomal RNA transcription regulation

Publication in preparation from the work in this chapter is:

Shin, Y*, Qayyum, M.Z*, Pupov, D., Esyunina, D., Kulbachinskiy, A., and Murakami, K.S.

(2020). Structural basis of ribosomal RNA transcription regulation. *Manuscript in preparation*

*** co-first author**

4.1 Abstract

Ribosomal RNA (rRNA) is the highest expressed gene in bacteria during vegetative growth whereas it is drastically downregulated by the DksA/ppGpp binding to RNAP during the stationary phase and stressed conditions. *rrnBP1* is the most well characterized rRNA promoter in *E. coli*. It contains several unique characters compared with others such as: 1) containing AT-rich UP element; and 2) GC-rich discriminator sequence downstream from the -10 element; and 3) initiating transcription 9 bases downstream from the -10 element. To investigate the mechanism of extraordinary regulation of rRNA transcription, I determined the cryo-EM structures of *E. coli* RNAP σ^{70} holoenzyme bound at the *rrnBP1* promoter in the closed complex (RPc), open complex (RPo) and in the presence of DksA and ppGpp (RP-DksA/ppGpp). The RPc structure reveals that the UP element bound form of the carboxyl-terminal domain of α subunit (α CTD) is a dimer, the σ factor does not recognize the DNA sequence of -10 element in duplex form and movement of the σ non-conserved region (σ NCR) to contact with the β' clamp toe domain (β' CT). The RPo structure shows the “open complex scrunching” by which the template DNA contacts with σ finger and β' lid. DksA/ppGpp binding triggers conformational changes of RNAP mobile domains that influence the stability of promoter complex. Results from the structural analysis and Cy-3 fluorophore labeled DNA opening assay suggest roles of the interaction between σ NCR and β' CT in accommodating duplex form -10 element DNA on the σ region 2 and its unwinding during the RPo formation. Although DksA/ppGpp binding to RNAP inhibits transcription of rRNA genes, many genes are unaffected while some genes are activated. I discuss the promoter-dependent pleotropic effect of DksA/ppGpp based on the duplex stability in discriminator sequences.

4.2 Ribosomal RNA (rRNA) transcription in *E. coli*

Rapidly growing *E. coli* contains higher number of ribosomes compared with slow growing cells (Bremer and Dennis, 2008). The rate of ribosome production is primarily determined by the ribosomal RNA (rRNA) transcription (Nomura, 1999; Zengel and Lindahl, 1994). In *E. coli*, the amount of rRNA transcription reaches around 70 % of total transcription during the exponential growth while it decreases drastically during the stationary phase and stressed conditions (aka stringent response) to save cell resources (Bremer and Dennis, 2008). The synthesis of ribosomal proteins (γ -proteins) is also inhibited by feedback mechanisms. When the γ -proteins accumulate in excess of the rRNA, the γ -proteins act as translational repressors by binding to their own mRNAs and inhibit the synthesis of γ -proteins (Burgos et al., 2017). The growth rate dependent rRNA transcription is primarily achieved at transcript initiation step (Cashel and Gallant, 1969; Condon et al., 1995; Gralla, 2005; Paul et al., 2004b).

The *rrnBP1* promoter has been the best studied among seven rRNA operons (*rrnA-E* and *rrnG-H*, the DNA sequences are highly conserved) in *E. coli* genome (Condon et al., 1992)

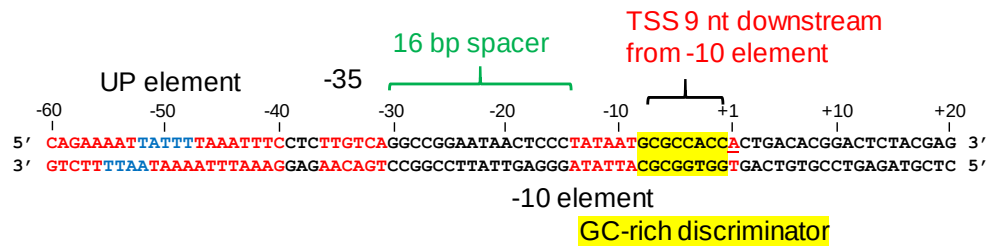
(Figure 4.1).

<i>E. coli</i>		-10	-7	+1	
<i>rrnB</i> P1	5'	CTCCC	TATAAT	GCGCCACCA	ACTGA-3' non-template
	3'	GAGGG	ATATTAC	GCGGTGGT	TGACT-5'
<i>rrnA</i> P1	3'	GAGGG	ATATTAC	GCGGTGGT	TGACT-5'
<i>rrnC</i> P1	3'	GAGGG	ATATTAC	GCGGTGGT	TGACT-5'
<i>rrnG</i> P1	3'	GAGGG	ATATTAC	GCGGTGGT	TGACT-5'
<i>rrnE</i> P1	3'	GAGGG	ATATTAC	GCGGAGGT	TAGCT-5'
<i>rrnH</i> P1	3'	GAGGG	ATATTAC	GCGGAGGT	TAGCT-5'
<i>rrnD</i> P1	3'	TAGGG	ATATTAC	GCGGAGGT	CAACT-5'

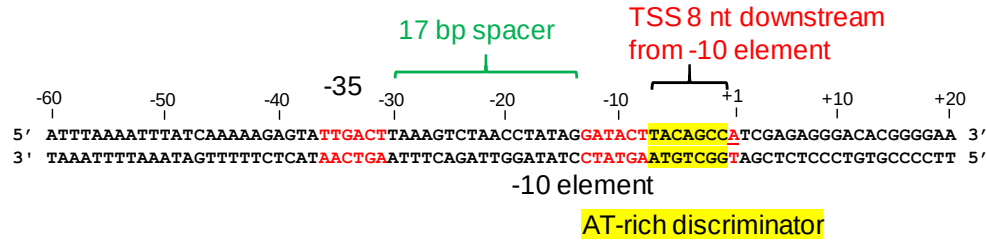
Figure 4.1: Sequences of the seven rRNA operons in *E. coli*. -10 element and transcription start site (+1, underlined) are indicated. The conservation of sequence at -7 position, which is shown to be important for open complex scrunching in this study is indicated in red.

rRNA promoters contain several unusual features compared to other promoters (e.g., *T7A1* and *argI*) including 1) the A + T rich UP element located upstream of the -35 element, which enhances the transcription activity as much as 30-fold; and 2) G+C rich discriminator sequence with the position of the transcription start site located 9 bp downstream of the -10 hexamer, and 3) the suboptimal 16 bp spacer between the -35 and -10 elements (Gaal et al., 1996; Ginsburg and Steitz, 1975; Haugen et al., 2008; Murakami et al., 1996; Pupov et al., 2018; Ross et al., 1993; Winkelman et al., 2016) (**Figure 4.2**).

A *rrnBP1* (inhibited by DksA)



B *T7A1* (insensitive to DksA)



C *argI* (activated by DksA)

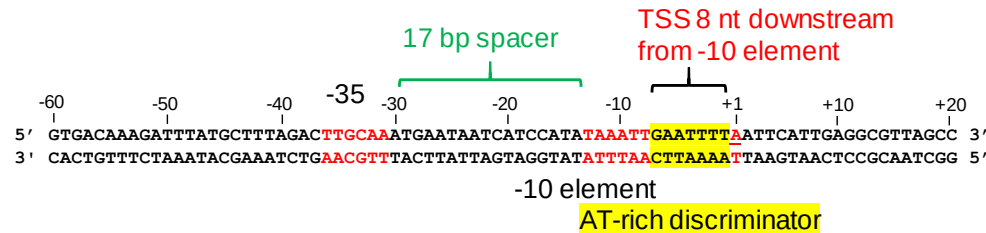
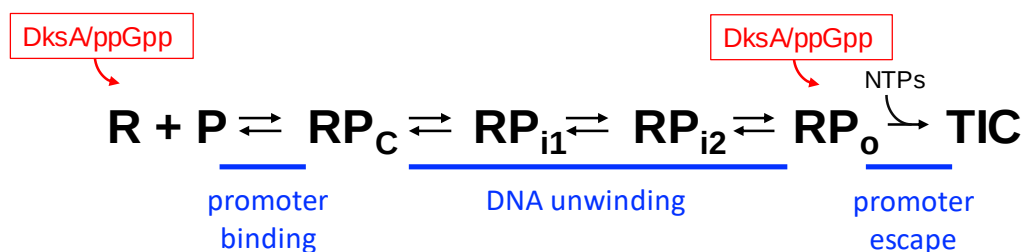


Figure 4.2: Sequences of the *rrnBP1* (A), *T7A1* (B) and *argI* (C) promoters. UP element, -35 element, -10 element and transcription start sites (+1) are indicated in red. Difference in discriminator sequences are highlighted in yellow.

The activity of rRNA transcription is primarily regulated by two small molecules including: 1) initiating nucleotide (iNTP) (ATP in all but one rRNA operon) and the bacterial alarmone ppGpp (guanosine tetraphosphate, also known as “magic spot”) (Cashel and Gallant, 1969; Hauryliuk et al., 2015; Potrykus and Cashel, 2008). Because of the intrinsic instability of rRNA promoter open complex (RPo), RNAP initiates RNA synthesis effectively only in the presence of the high concentration of the iNTP while RNAP predominantly forms the closed complex (RPc) and eventually dissociate from the rRNA promoter in the iNTP limited condition (Haugen et al., 2006; Haugen et al., 2008). ppGpp accumulates to high concentration during the stringent response to downregulate the rRNA transcription rate (Brown et al., 2002; Molodtsov et al., 2018; Paul et al., 2004a; Ross et al., 2016; Ross et al., 2013) by enhancing the action a RNAP binding factor DksA. DksA is a 17.5 kDa, α -helical protein, which interacts with the secondary channel of RNAP. The cellular concentration of DksA remains constant throughout different growth conditions (Brown et al., 2002; Molodtsov et al., 2018) because the transcription of *dksA* is negatively autoregulated via a feedback loop and translationally activated by CsrA (carbon storage regulator A), an RNA binding protein. (Edwards et al., 2011). Consequently, DksA can bind either apo-form or DNA bound form of RNAP (**Figure 4.3**). Although DksA/ppGpp inhibit the transcription from the rRNA promoter, it activates the transcription from the amino acid biosynthesis operons such as *argI* and has no effect in transcription from the *T7A1* promoter (Barker and Gourse., 2001) (**Figure 4.2**).



(R: RNAP, P: promoter DNA)

Figure 4.3: RNAP – promoter DNA interactions on the pathway to forming RPo. RNAP (R) interact with promoter DNA (P) to form the closed complex (RP_c). DksA/ppGpp can bind free RNAP. The closed complex unwinds DNA duplex through transitions of at least two known intermediates to form transcription competent open complex (RP_o). In the presence of NTPs, RPo initiate transcription (TIC). DksA/ppGpp can prevent RPo formation and/or facilitate RP_c formation from RPo.

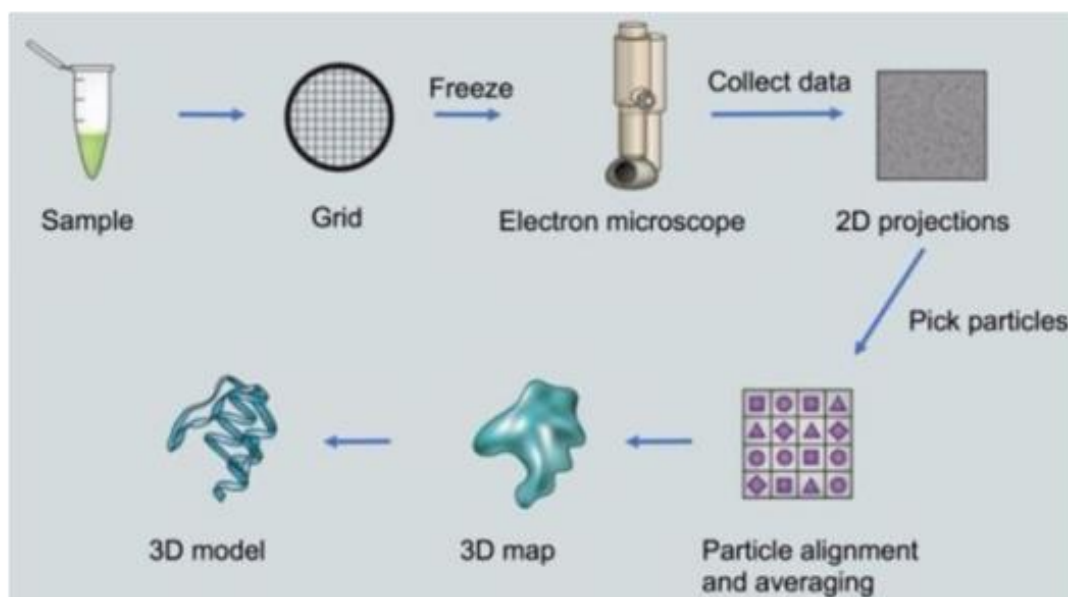
Our laboratory previously determined the X-ray crystal structure of *E. coli* RNAP σ^{70} holoenzyme with DksA/ppGpp to study the inhibitory mechanism by DksA/ppGpp (Molodtsov et al., 2018). The study showed that 1) DksA is capable of inhibiting the catalysis and influences the DNA binding of RNAP; 2) ppGpp acts as an allosteric effector of DksA function; and 3) ppGpp stabilizes DksA in more functionally important binding mode. However, the crystal structures lack promoter DNAs and therefore could not explain the pleiotropic effects of DksA/ppGpp in different types of promoters.

In this chapter, I present the cryo-EM structures of *E. coli* RNAP σ^{70} – *rrnBP1* promoter in the closed complex (RP_c), open complex (RP_o) and in the presence of the stringent response regulators DksA and ppGpp. The structures, structure-based DNA opening assay and DNA duplex free energy calculation reveal the role of the non-conserved region of σ^{70} (σ NCR) in facilitating DNA opening and explain the pleiotropic effect of DksA/ppGpp based on the duplex stability in discriminator sequences.

4.3 Cryo-electron microscopy (Cryo-EM)

The methodology of determining a macromolecular structure during last two years of my PhD training switched from X-ray crystallography to cryo-electron microscopy (cryo-EM) due to its advancement of determining the atomic resolution structure the macromolecule (Doerr, 2016; Lyumkis, 2019; Kuhlbrandt, 2014; Li et al., 2013; Nogales, 2016; Punjani et al., 2017) (**Figure 4.4**). Cryo-EM is able to determine the macromolecular structure without crystallization and requires much less amount samples compared to conventional X-ray crystallography. Additionally, cryo-EM brought many opportunities to investigate the structure and function of macromolecule in near physiological condition, which often could not be addressed by X-ray crystallography due to the crystal packing. Particularly, cryo-EM allows us to prepare a RNAP-DNA complex without a synthetic bubble promoter and address the protein dynamics by classifying individual macromolecule into distinct classes during heterogeneous refinement. Nonetheless, the preparation of homogeneous and stable complex as well as the optimization of vitrification conditions remain one of the obstacles in determining high-resolution structures by cryo-EM.

A



B

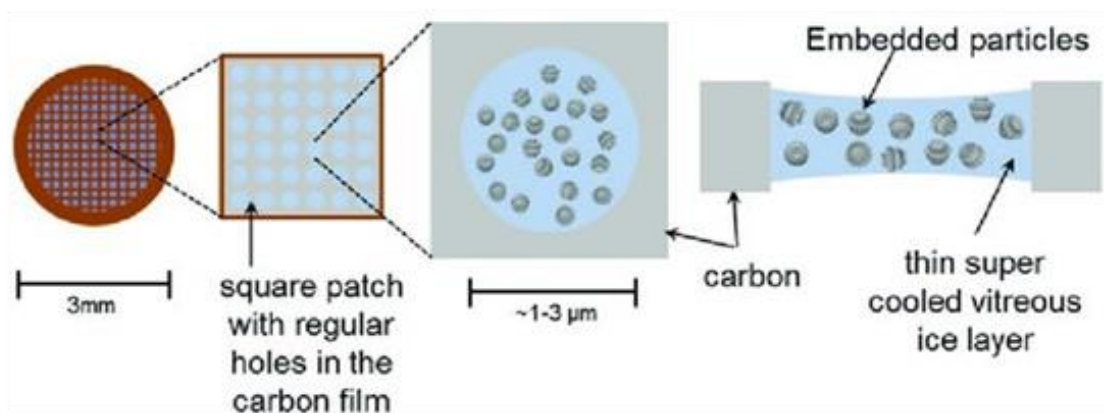


Figure 4.4: Overview of cryo-EM structure determination workflow and vitrified specimen.

(A) A sample is applied on a cryo-EM grid followed by freezing. After data collection with a microscope equipped with a direct electron detector, the images are processed using softwares such as Relion and cryoSPARC for motion and CTF corrections, particle picking, 2D and 3D classifications. Resulting 3D map is analyzed to build a 3D protein model (Doerr, 2016). (B) A 3mm copper mesh grid covered with a holey carbon film. A sample is applied on a cryo-EM grid to promote its absorption followed by blotting with a filter paper to remove the excess liquid. To create a protein suspension in a layer of ice (vitrification), the grid is then rapidly frozen in liquid ethane. Figure from the reference (Costa et al., 2017).

4.4 Materials and Methods

4.4.1 Purification of *E. coli* RNA polymerase σ^{70} holoenzyme

All *E. coli* core RNAP subunits were expressed in BL21 (DE3) cells transformed with pIA900 plasmid encoding *rpoA*, *rpoB*, *rpoC* and *rpoZ* genes. Transformed cells were grown in 1.5 L LB media in 4 L flask containing 100 μ g/ml ampicillin till OD600 = 0.8, protein expressions were induced by adding 1 mM IPTG and were continued for 4 hr at 30 °C. About 20 g of cell paste was suspended in 50 ml of lysis buffer (20 mM Tris-HCl, pH 8 at 4 °C, 10% glycerol, 1 mM EDTA, 2 mM PMSF, 10 mM B-ME) and cells were lysed by sonication (3sec on, 15sec off for total of 5 min on time). The suspension was centrifuged for 30 min/4 °C/17,000 rpm and the polymin P was added to supernatant to final 0.5%. The tubes were shaken on a shaking plate in the cold room for 30 min and centrifuged for 15 min/4 °C/12,000 rpm. The supernatant was discarded and 30 ml of extraction buffer (20 mM Tris-HCl, pH8 at 4 °C, 1M NaCl, 5% glycerol, 0.5 mM EDTA, 1mM PMSF, 5mM B-ME) was added to the pellet. The pellet was resuspended until there is no visible flakes and centrifuged for 15 min/4 °C/15,000 rpm. Supernatant was transferred to new tubes and 1.3 volumes of prechilled saturated ammonium sulfate was added. The tube was shaken in the cold room for at least 30 min, and centrifuged 15 min/4 °C/17,000 rpm. Supernatant was discarded. The pellet was resuspended in 25 ml of resuspension buffer (20 mM Tris-HCl, pH8 at 4 °C, 50 mM NaCl, 5% glycerol, 0.1 mM EDTA, 1 mM PMSF, 5 mM B-ME) and dialyzed against 2L dialysis buffer (20 mM Tris-HCl, pH8 at 4 °C, 50 mM NaCl, 5% glycerol, 0.1 mM EDTA, 5mM B-ME).

The dialyzed sample was applied to BioRex 70 column equilibrated with TGE buffer (10 mM Tris-HCl, pH8 at 4 °C, 50 mM NaCl, 5% glycerol, 0.1 mM EDTA, 1 mM PMSF, 5 mM B-ME) and eluted over 0.1M – 0.6M salt gradient. **(Figures 4.5A and B)** Main peak fractions were combined and concentrated to 5 ml and diluted to 50 ml with TGE buffer without NaCl. The sample was applied to Heparin column (CV = 5 ml) equilibrated with TGE buffer with 0.1 M NaCl. The protein was eluted with stepwise elution at 0.6 M NaCl **(Figures 4.5C and D)**. The fraction containing core RNAP was saved at -80 °C.

E. coli σ^{70} was expressed in BL21 (DE3) cells transformed with pET15b- σ^{70} encoding *rpoD* gene. Transformed cells were grown in 1.5 L LB media in 4 L flask containing 100 µg/ml ampicillin till OD600 = 0.5, protein expressions were induced by adding 0.5 mM IPTG and were continued for 16 hr at 16 °C. About 20g of cell paste was suspended in 30 ml of lysis buffer (10 mM Tris-HCl, pH8 at 4 °C, 0.1 M NaCl, 10 mM imidazole, 5% glycerol, 1 mM B-ME, 1mM PMSF), cells were lysed by sonication (3sec on, 15sec off for total of 5 min on time). The suspension was centrifuged for 30 min/4 °C/17,000 rpm. Supernatant was applied to Ni column equilibrated lysis buffer. The protein was eluted with buffer containing 200 mM imidazole. The eluted protein was applied to Source 15Q column equilibrated with TGE buffer with 0.1 M NaCl followed by elution with salt gradient. The fractions containing σ^{70} (15 and 16) were combined and concentrated **(Figures 4.5E and F)**.

RNAP σ^{70} holoenzyme was reconstituted by adding 3-fold molar excess of σ^{70} to core RNAP and incubated at 30 °C for 30 min, followed by purification by Superdex200 size-exclusion column chromatography equilibrated with 10 mM HEPES-NaOH, (pH8 at

4°C), 50 mM NaCl, 0.1 mM EDTA and 5 mM DTT. The fractions 8-10 (RNAP σ^{70} holoenzyme) were combined and concentrated to 10 mg/ml (20 μ M). The fractions 12-14 (σ^{70}) were combined, concentrated and stored at -80 °C for later use (**Figures 4.5G and H**).

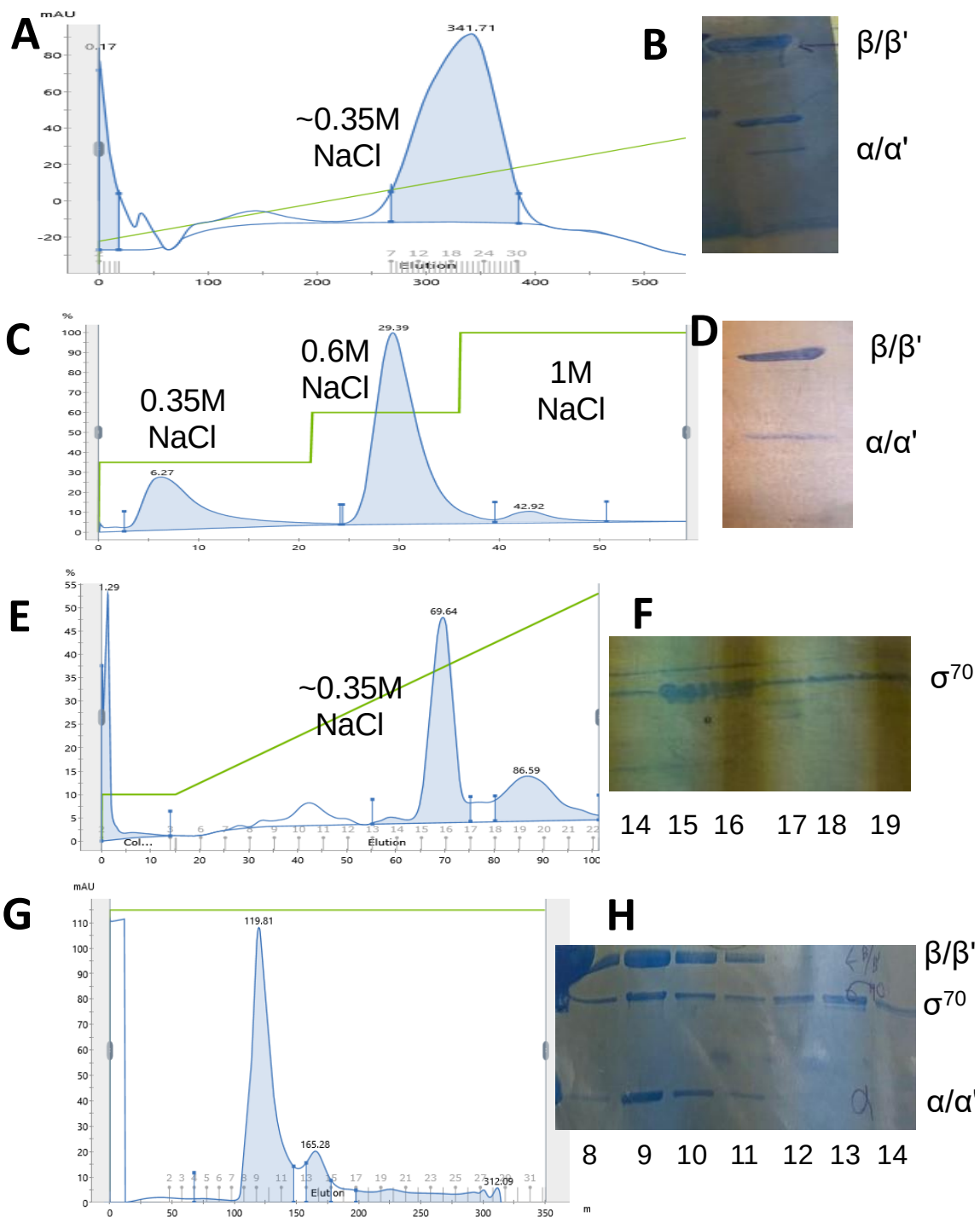


Figure 4.5: Purification of *E. coli* RNAP σ^{70} holoenzyme. (A) 50 ml BioRex70 column chromatography, (B) SDS-PAGE gel of main peak fractions from biorex70, (C) 5 ml Heparin column chromatography, (D) SDS-PAGE gel 0.6M elution from heparin, (E) 3 ml Source 15Q column chromatography, (F) SDS-PAGE gel of fractions from Source 15Q, (G) 320 ml SD 200 gel filtration column chromatography and (H) SDS-PAGE gel of main peak fractions from SD 200.

4.4.2 Purification of *E. coli* RNA polymerase derivative

The RNAP derivative, E175K β' core – WT σ^{70} holoenzyme was prepared as the WT RNAP

4.4.3 Preparation of *rrnBP1* dsDNA

The *rrnBP1* promoter DNA was synthesized (IDT) according to the native *rrnBP1* sequence (**Figure 4.2A**) and annealed in 40 μ L containing 10 mM Tris-HCl (pH 8.0), 50 mM NaCl, and 1 mM EDTA to the final concentration of 0.5 mM. The solution was heated at 95 °C for 10 min and then temperature was gradually decreased to 22 °C.

4.4.4 Preparation of RNAP-*rrnBP1* DNA complex without or with DksA/ppGpp and vitrification of cryo-EM grids

E. coli σ^{70} holoenzyme (20 μ M) and *rrnBP1* promoter DNA (40 μ M) (**Figure 3.2A**) were preincubated for 5 mins at 37 °C in the buffer (10 mM HEPES-NaOH, pH 8.0, 50 mM NaCl, 0.1 mM EDTA, 5mM DTT and 5 mM MgCl₂) for the RPo formation followed by adding ATP and CMPCPP (Cytidine-5'-[(α,β)-methyleno]triphosphate, Jena Bioscience) (2 mM each) for stabilizing the open complex with initiating nucleotides (iNTPs) bound at the active site.

CMPCPP was purchased from Jena Bioscience. In case of forming the RNAP-*rrnBP1*-DksA/ppGpp complex, DksA (100 μ M) and ppGpp (2 mM) were added to the RNAP followed by adding *rrnBP1* promoter DNA. To compare the effect of the order of DksA addition, another grid was prepared by adding DksA/ppGpp to the RNAP-*rrnBP1* promoter DNA complex. After mixing, the complexes were further incubated for 5 mins at 37 °C. Immediately before freezing, 8 mM CHAPSO (Hampton research) was added to the complexes and incubated for 10 mins at 25 °C followed by centrifugation for 1 min. 3.5 μ L sample was applied to glow-discharged C-Flat Holey Carbon grids (Cu 2/1, 400 mesh), blotted and plunge-frozen in liquid ethane using Vitrobot Mark IV (FEI, USA) at 95% humidity and 4°C.

4.4.5 Cryo-EM data collection

Adsorption of macromolecules at the air/water interface during cryo-EM grid preparation causes preferred particles orientation and increases chance of denaturing particles. I also have encountered a dissociation of σ factor from RNAP core and DNA from RNAP-holoenzyme after vitrification. CHAPSO is a zwitterionic surfactant which prevents adsorption of the macromolecules to air/water interface at the critical micelle concentration (8 mM) and broadens the distribution of particle orientations (Chen et al., 2019). However, the addition of CHAPSO causes aggregation and also accumulation of RNAP toward the edge of holes. Therefore, high concentration of RNAP (20 μ M) was required for cryo-EM grid preparation and also cryo-EM images were collected near the edges of holes (**Figure 4.6**).

Cryo-EM data were collected using the Titan Krios (Thermo Fisher) microscope equipped with a K3 direct electron detector (Gatan) at the National Cancer Institute's Cryo-EM Facility at Frederick with the software Latitude S (Gatan, Inc.) or with a Falcon II direct electron

detector (Gatan) at the Penn State Cryo-EM Facility with the software EPU (ThermoFisher Scientific). Each sample grid was imaged at 300 kV, with a defocus range of 1.0 – 3.0 μm . The magnification was 81,000X or 75,000X in electron counting mode (pixel size = 1.08 $\text{\AA}/\text{pixel}$ or 0.88 $\text{\AA}/\text{pixel}$, respectively) with a total dose of 45 electrons/ $\text{\AA}^2$ that is 1.02 electrons/frame/ $\text{\AA}^2$.

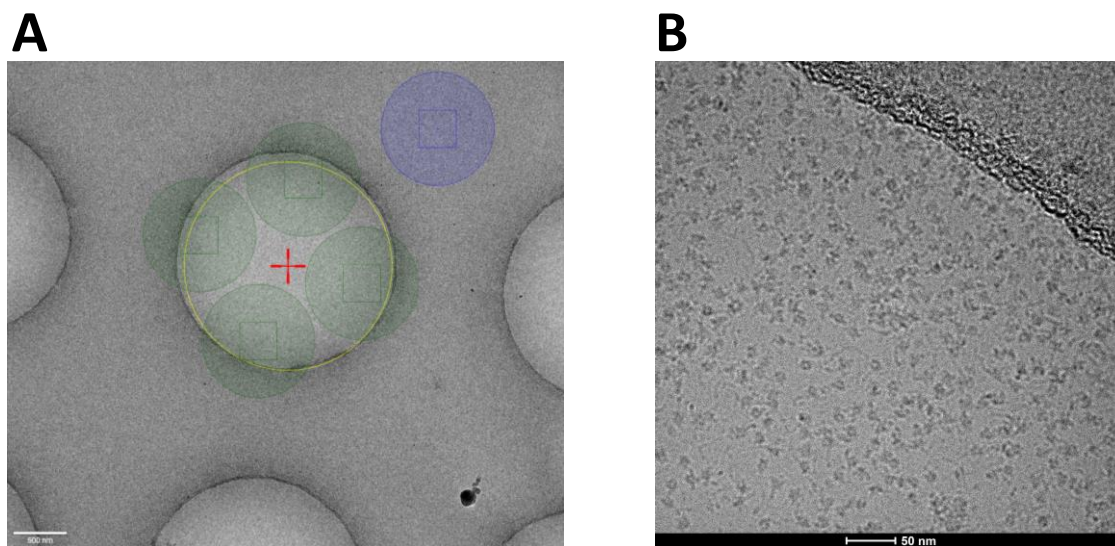


Figure 4.6: Data collection strategy. (A) a hole on the carbon grid (yellow circle), beam focusing (+ red sign) and auto focusing on the carbon area of the grid (blue circle) are shown. As CHAPSO accumulate RNAP toward the edge of holes, relatively high concentration of RNAP (20 μM) was used and data was collected from near the edges of holes of grids (green circles). (B) A representative cryo-EM image of the *E. coli* RNAP σ^{70} holoenzyme on C-Flat Holey Carbon grid (Cu 2/1, 400 mesh).

4.4.6 Cryo-EM image processing

Cryo-EM data was processed using either Relion 3.0.7 (**Appendix D**) or cryoSPARC V2.9.0 (**Appendix E**). From the collected movies, usable movies were aligned, dose weighted, and CTF fitted using MotionCor2 and Gctf (Relion) or Patch-Motion correction and Patch CTF (cryoSPARC). Initially, about 1000 particles were manually picked to generate a template followed by autopicking. The autopicked particles were 2D classified to 10 classes and good 2D classes were selected for 3D classification. After initial 3D classification, heterogeneous and homogeneous refinements, post-processing and particle polishing were performed to generate final cryo-EM density maps.

4.4.7 Structure refinement

To refine the closed and open complex structures, *E. coli* RNAP holoenzyme crystal structure (PDB: 4YG2) was manually fit into the cryo-EM density map using Chimera (Pettersen et al., 2004) and real-space refined using Phenix (Afonine et al., 2010). In the real-space refinement, domains of RNAP were rigid-body refined, then subsequently refined with secondary structure, Ramachandran, rotamer and reference model restraints. To refine the structure of RNAP-DksA/ppGpp, *E. coli* RNAP and DksA/ppGpp complex crystal structure (PDB: 5VSW) was manually fit into the cryo-EM density map using Chimera. DNA was manually built by using Coot (Emsley and Cowtan, 2004). The structure was refined the same as in the closed and open complex structures.

4.4.8 DNA opening assay by monitoring Cy3 fluorescence signal

DNAs containing *rrnBP1* and *T7A1* promoter sequences and Cy3 fluorophore attached to non-template T at position +3 was synthesized from IDT. dsDNA was prepared as described in **Section 4.4.3**. 100 nM *E. coli* RNAP (WT- σ^{70} holoenzyme or mutants) was preincubated with 500 nM DksA and 100 μ M ppGpp where indicated in binding buffer (10 mM Tris-HCl, pH 7.8, 20 mM NaCl, 10 mM MgCl₂, 1 mM DTT) at 37 °C for 5 min on the microplate (Corning). Cy3-labeled dsDNA (2 nM, either *rrnBP1* or *T7A1*) was rapidly mixed and measured Cy3 fluorescence immediately at 37 °C using microplate reader (Infinite M1000 pro, TECAN) with excitation at 550 and emission at 570 nm.

4.4.9 DNA duplex free energy calculation

DNA duplex free energies were analyzed based on the nearest-neighbor thermodynamics (Dienemann et al., 2019; Santa Lucia, 1998). Briefly, the python script was wrote so that, the script reads a sequence from a text file, calculates the DNA duplex free energy of dinucleotides, sums these values over a 8 bases window and report these sums for the first base of the central nucleotide of the window (e.g. sum for the first window base 1-8 will be reported for base 4).

4.5 Results

4.5.1 Cryo-EM structures of the RNAP-*rrnBP1* promoter complexes (closed and open complexes)

4.5.1.1 The RNAP-*rrnBP1* promoter closed complex (RPc)

To prepare the RNAP and *rrnBP1* promoter open complex, I preincubated *E. coli* RNAP with the *rrnBP1* promoter DNA at 37 °C for 5 min. After the addition of an iNTP (ATP) and a non-hydrolyzable nucleotide CMPCPP, the reaction was further incubated at 37 °C for 5 min to stabilize the open complex. In the course of cryo-EM structure determination of the RNAP-*rrnBP1* complex, 3D classification of particles based on the difference in DNA conformation resulted in two distinct 3D classes representing closed complex (RPc, about 13% of particles used for final reconstruction) and open complex (RPo, about 87% of particles used for final reconstruction) indicating spontaneous DNA closing of RPo occurred during the cryo-EM grid preparation due to naturally unstable RNAP-*rrnBP1* promoter open complex. It is also possible that the DNA forming the RPc never opened to form the open complex.

The RPc structure shows the RNAP bind the fully double stranded DNA from -60 to +2 including transcription bubble region of DNA (-10 element to transcription start site) (**Figure 4.7**). The downstream DNA from +3 position is disordered. The structure shows the α CTDs forms a dimer and bind minor groove of middle of UP element (-40 to -60) (**Figure 4.8**), which is in agreement with previous DNA footprinting result (Newlands et al., 1991). Alanine substitutions of the α CTDs residues R265, N294, K291 and K298 have shown to reduce the activity of *rrnBP1*

transcription, suggesting these residues play a role in the α CTDs – UP element interaction (Gaal et al., 1996; Murakami et al., 1996; Ozoline et al., 2001).

The R_Pc structure also shows σ region 2.3 (D417 to W434) contacts with -10 element by fitting into a DNA major groove without any specific amino acid–DNA base interaction (**Figure 4.9A**), suggesting the σ region 2.3 recognizes the shape formed around the -10 element. It is in agreement with previous proposal that the σ factor does not recognize the DNA sequence of -10 element in duplex form (Feklistov and Darst, 2011). Downstream part of DNA duplex bends about 13 degree centered at -14 position, which avoid steric clash between the -10 element and σ region 2.3 (**Figure 4.7**).

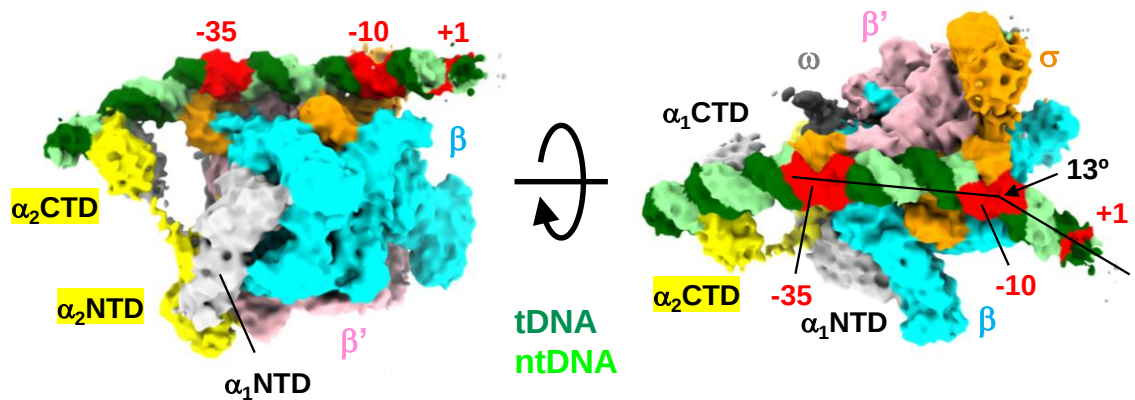


Figure 4.7:The cryo-EM density map of *E. coli* RNAP and *rrnBP1* promoter DNA closed complex (RP_c). β (cyan), β' (pink), α_1 (white), α_2 (yellow), ω (gray), σ (orange), tDNA (dark green), ntDNA (light green) are shown. -35 element, -10 element and +1 (tDNA) are highlighted in red.

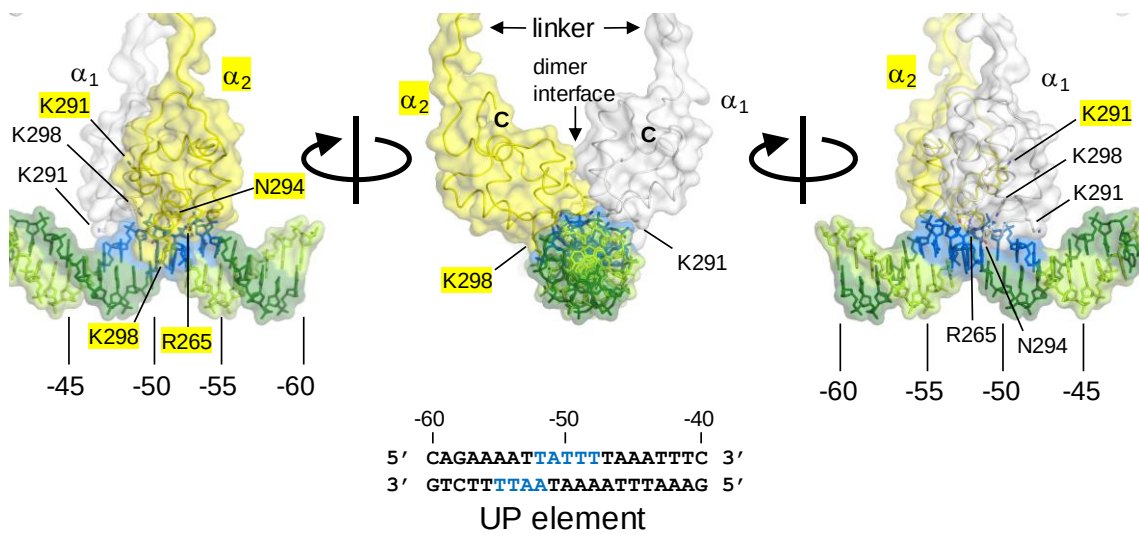


Figure 4.8: The cryo-EM structure showing the interactions between the α CTDs of *E. coli* RNAP and UP element of *rrnBP1* promoter DNA. α CTDs (white and yellow) are depicted as ribbon models with transparent surface. DNA (green, light green, blue) is shown as stick model. The amino acid residues making contacts with UP element are indicated. β

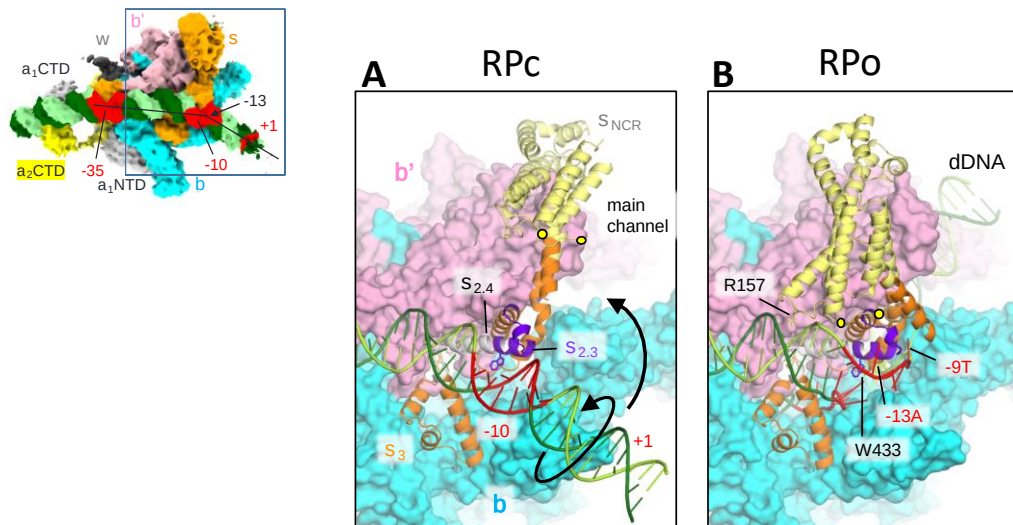


Figure 4.9: The interactions between the -10 element and σ region 2.3. (A) The structure of RPo showing fitting of σ 2.3 into the major groove of -10 element. (B) The structure of RPo showing flipped $-13A$ base into the pocket of σ 2.3. Downstream DNA (dDNA) buried in DNA binding channel is shown. W433 of σ 2.3, the amino acid residue that plays a key role in the DNA unwinding is shown.

To gain insight into the mechanisms of DNA binding and DNA opening, I compared RPc structure with previously published structures of apo-form RNAP and RPo (Narayanan et al., 2018) (**Figure 4.10**). The notable difference observed in RPc is the conformational change of the σ non-conserved region (σ NCR, N128 to R372 in *E. coli*). Upon RNAP binds dsDNA, the σ NCR undergoes a rigid rotation toward the DNA binding channel (formed by two pincers including the β' clamp and the β lobe/SI1) to establish interaction between the basic surface of σ NCR and the acidic surface of the clamp toe domain of β' subunit (β' CT, residues from 143 to 180 in *E. coli*) (**Figure 4.10A**), which is not observed in the apo-form RNAP structure (PDB: 6C9Y) and the RPo structure (PDB: 6CA0) (Narayanan et al., 2018).

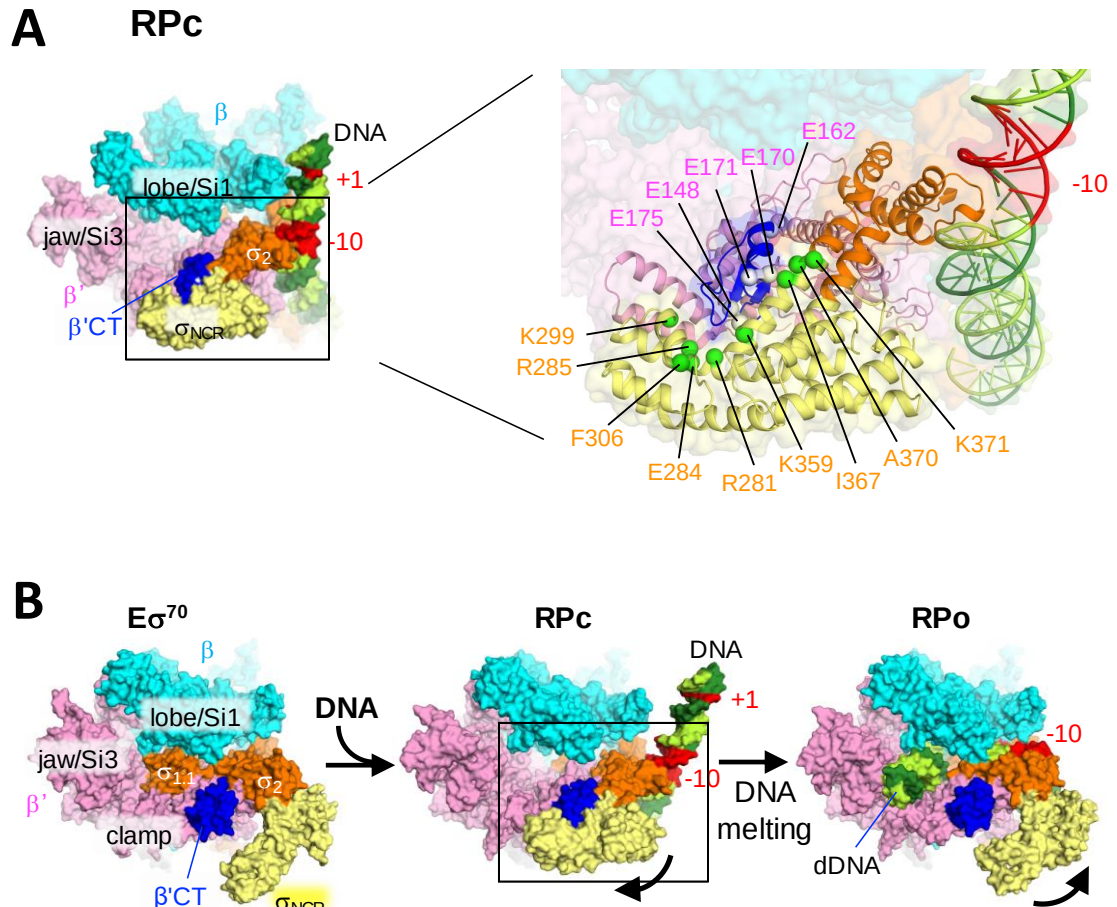


Figure 4.10: The interactions between the σ NCR and β 'CT. (A) The structure of RPc showing the contact between σ NCR and β 'CT. Amino acid residues involved in interactions are indicated. (B) The conformational change of σ NCR from apo-form RNAP to RPc to RPo. Upon σ NCR contacts β 'CT, DNA unwinds to form RPo, suggesting the role of σ NCR and β 'CT interaction in initial DNA unwinding.

4.5.1.2 The RNAP-*rrnBP1* promoter open complex (RPo)

Second class of RNAP and *rrnBP1* complex represents the open complex (RPo) containing *rrnBP1* dsDNA from -44 to +20 including single-stranded template DNA bubble (-13 to +2) in the active site of RNAP (**Figure 4.11**). Unexpectedly, RNAP synthesized 2-mer RNA using CMPCPP and ATP for the first and second iNTPs, respectively, by starting the transcription at -1 position (**Figure 4.11C**), which has been observed *in vitro* in all NTP concentrations (Winkelman et al., 2016). The 2-mer RNA and DNA template hybrid is in post-translocation state, resulting in the +1T and +2G of tDNA positioned at *i* and *i* + 1 sites, respectively. Physiologically relevant transcription start site (+1T of template DNA) is located at *i* site, therefore, the structure represents the functional RPo.

Compared to the RPc structure, the region around α CTDs and the UP element (**Figure 4.8**) is disordered in RPo structure. Additionally, -13A base of the ntDNA flips out and fits into the binding pocket of the σ 2.3, which may be linked to the rotation and fitting of the dsDNA into the DNA binding channel of RNAP (**Figure 4.9B**). The conformation of σ NCR and β' CT are the same as previously published RPo structure (PDB: 6CA0) (Narayanan et al., 2018).

The transcription start site of the *rrnBP1* promoter is located at 9 nt downstream from the -10 element compared to 8 nt in *T7A1* and *ArgI* promoters (**Figure 4.2**). Because of the extra base, it has been proposed that the template scrunches in the active site of RNAP (Winkelman et al., 2016). The RPo structure determined in this study revealed the open complex scrunching; the single-stranded template DNA is scrunched around -7 G base because it fits in a pocket surrounded by the lid of β' subunit, σ finger (σ region 3.2) and -6C template DNA base (**Figure 4.11A**). The RNAP residues making interactions with -7G base, D256 of β' lid and F522 of σ finger (-7G binding pocket, **Figure 4.11B**) are highly conserved in bacteria and substitution or

deletion of these residues shifts the transcription start site to 6 nt downstream from the -10 element (-3A position) (Pupov et al., 2018). The -7G and -6C bases of tDNA are also conserved in all seven rRNA promoters in *E. coli* (**Figure 4.1**) and in other gamma-proteobacteria. The substitution of -7G to -7C also shifts the transcription start site to 6 nt downstream from the -10 element (Winkelman et al., 2016), indicating that the binding -7G of tDNA to the -7G binding pocket is essential for the open complex scrunching and the transcription start site selection of the *rrnBP1* promoter. It is important to note that “the open complex scrunching” revealed in this study is different from “DNA scrunching” in which RNAP remains stationary and unwinds and pulls downstream DNA into itself during transcript initiation (Revyakin et al., 2006).

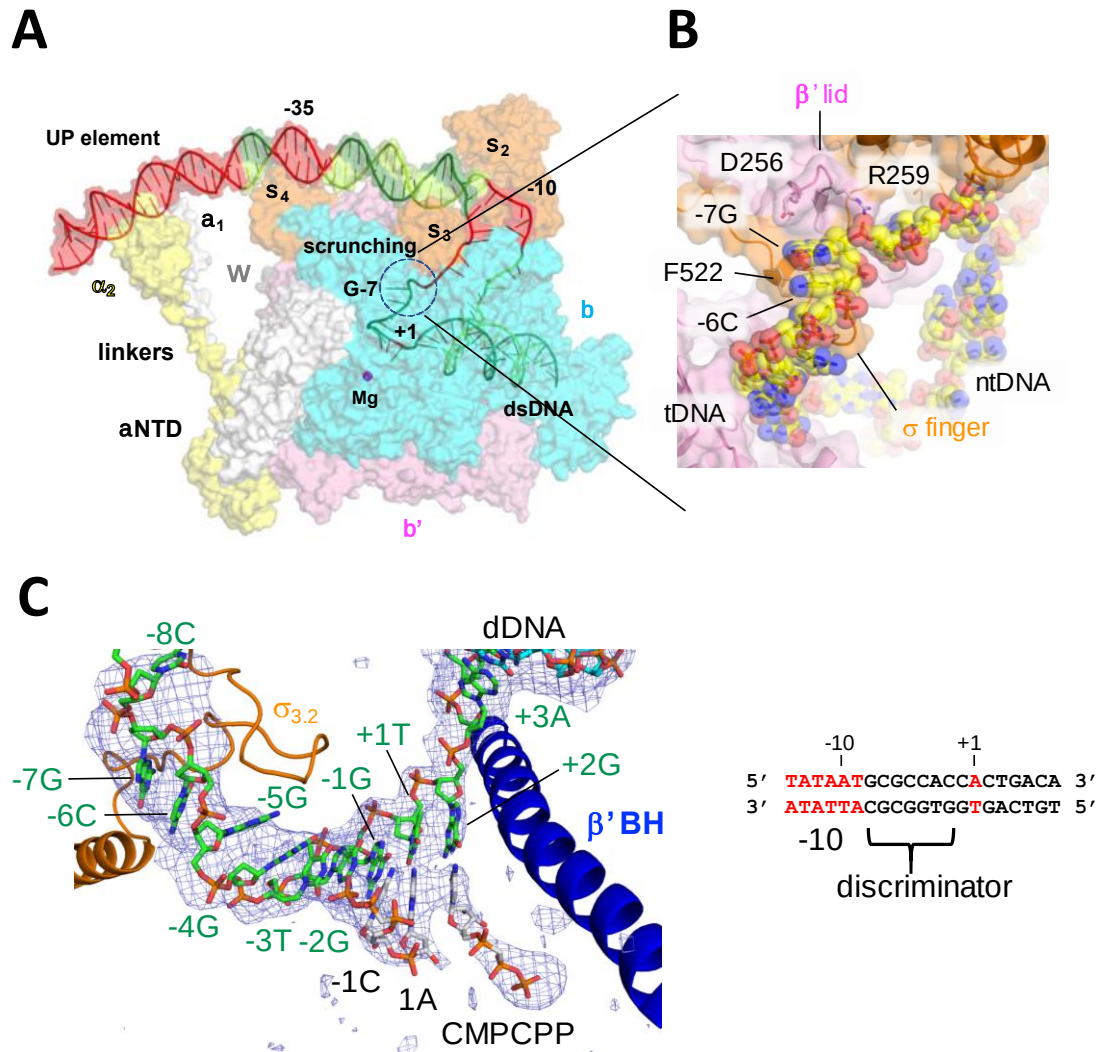


Figure 4.11: The cryo-EM structure of the *E. coli* RNAP and *rrnBP1* promoter DNA open complex (RPO). (A) The structure of RPO. RNAP is shown as a transparent surface and DNA is depicted with ribbon model. (B) Close-up view of the scrunching of the G-7 base of tDNA. Lid of RNAP β' subunit and σ finger are depicted as ribbon models with transparent surface. DNA is shown as stick model with transparent CPK representation. The residues making interactions with G-7 base are indicated. (C) The density map of active site of RNAP. DNA and RNA are showing as stick model. σ finger and bridge helix (BH) of β' subunit are shown as cartoon model.

4.5.2 Cryo-EM structure of the RNAP-*rrnBP1* promoter complex in the presence of DksA/ppGpp (RP-DksA/ppGpp)

4.5.2.1 Binding modes of DksA/ppGpp to RNAP

To study the mechanism of *rrnBP1* specific transcription inhibition by DksA/ppGpp, I determined the cryo-EM structure of RNAP and *rrnBP1* promoter complex in the presence of DksA/ppGpp. I prepared the RNAP complex in two alternative ways: 1) mixing DksA/ppGpp first to RNAP to form the RNAP-DksA/ppGpp complex followed by adding *rrnBP1* promoter DNA; and 2) mixing RNAP and *rrnBP1* promoter DNA first to form RPo followed by adding DksA/ppGpp (**Figure 4.3**). I prepared two sets of cryo-EM grid with these complexes, determined the structures independently, and found that both structures are essentially identical in terms of RNAP conformation as well as RNAP interactions with DksA/ppGpp and *rrnBP1* promoter. Therefore, I will present the structures prepared by mixing DksA/ppGpp first to free RNAP followed by the addition of *rrnBP1* promoter DNA, which is more physiologically relevant. The 3D classification of RNAP-*rrnBP1*-DksA/ppGpp complex (RP-DksA/ppGpp) particles resulted in two distinct 3D structures having either σ region 1.1 (class I, RP-DksA/ppGpp (I)) or downstream DNA (class II, RP-DksA/ppGpp (II)) within the DNA binding channel of RNAP (**Appendix E**). These structures also show two distinct position of the β β lobe/SI1 as described later. Both structures show the DksA binding at the RNAP secondary channel and the ppGpp binding at the sites I (β ' and ω binding interface) and site II (DksA and β ' binding interface) as observed in previously determined X-ray crystal structure (Molodtsov et al., 2018). The DksA interacts with RNAP via 1) the globular domain (G domain, residues 7-33 and 110-

134), which contacts the β' rim helix of RNAP, 2) the central domain, two long α -helices (coiled-coil (CC), residues 35-109), which contact the bridge helix (BH), the trigger loop (TL) and linkers connecting to the β' SI3, 3) the coiled-coil tip (CC tip, the linker part of CC, residues 69-75 which contacts the active site, and 4) the C-terminal α helix (CT-helix, residues 135-151), which contacts the β lobe/SI1 domain. (Perederina et al., 2004) (**Figures 4.12 and 4.13**). The CC of DksA blocks NTPs entering from the secondary channel to the RNAP active site and prevents the conformational change of TL during the RNA synthesis. Therefore, DksA must be dissociated from RNAP to initiate transcription.

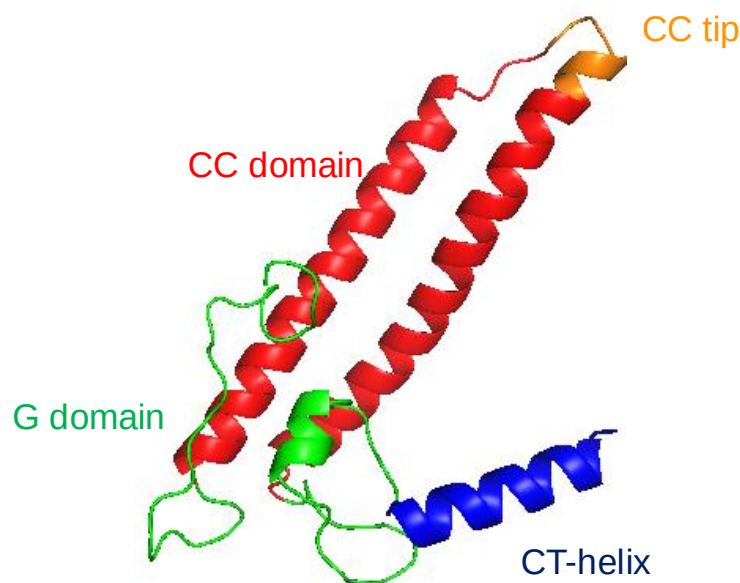


Figure 4.12: The structure of DksA. G domain (globular domain, residues 7-33 and 110-134, green), CC domain (central coiled coil domain, residues 35-109, red), CC tip (coiled-coil tip, residues 69-75, orange) and CT-helix (C-terminal α helix, residues 135-151, blue). The structure was generated from RP-DksA/ppGpp (I) structure after removing other parts.

A RP-DksA/ppGpp (I)

B RP-DksA/ppGpp (II)

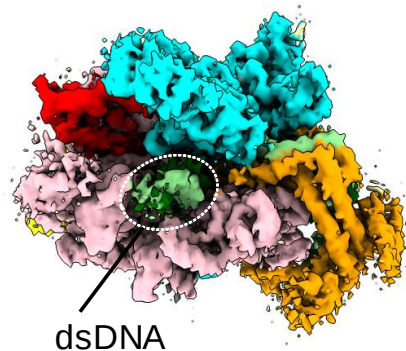
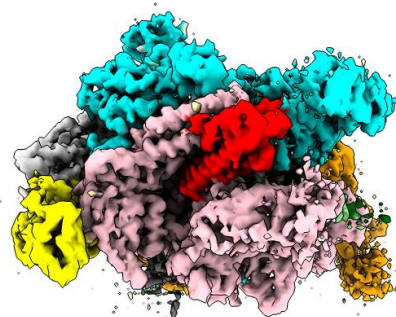
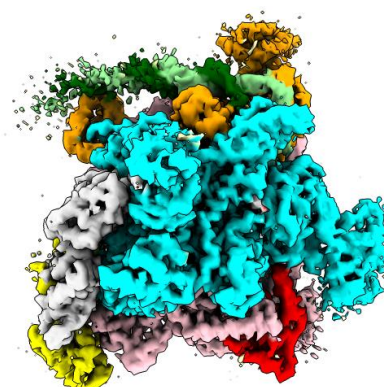
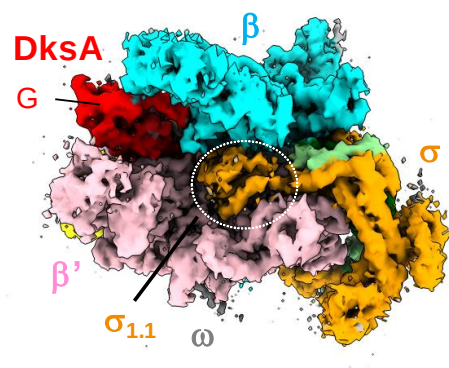
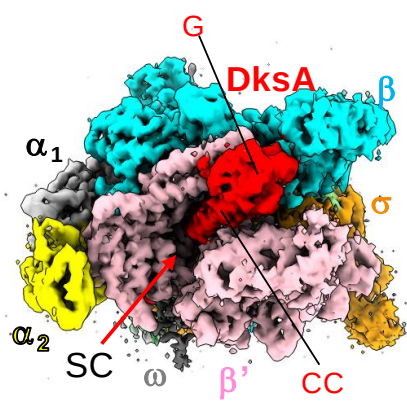
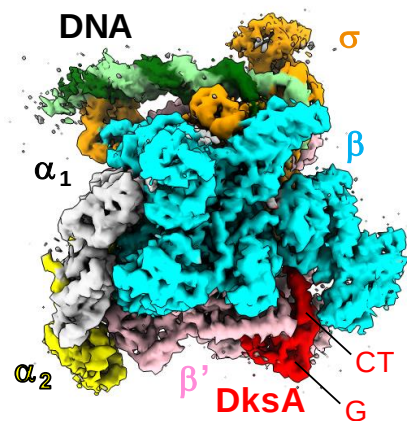


Figure 4.13: Cryo-EM maps of the (A) RP-DksA/ppGpp (I) and (B) RP-DksA/ppGpp (II). (A) cryo-EM density map of the RP-DksA/ppGpp(I). DNA, RNAP subunits and DksA domains (G, G-domain; CC, CC domain; CT, CT-helix) are indicated and colored (green, tDNA; light green, ntDNA; red, DksA). Relative orientations of proteins in the panels are indicated. σ 1.1 is highlighted by white dash oval. (B) cryo-EM density map of the RP-DksA/ppGpp(II). DNA in the complex is shown as cryo-EM difference density with transparent RNAP. Downstream DNA (dDNA) is highlighted by white dash oval.

4.5.2.2 RP-DksA/ppGpp (I) structure represents an early stage intermediate (RPi1) on the pathway to RPo

The DNA density of the class I complex (RP-DksA/ppGpp(I)) shows duplex DNA from -41 to -14 and single-stranded DNA of -10 element, but DNA from -5 to +20 of non-template DNA and from -14 to +20 of template DNA including transcription bubble and downstream duplex DNA are not traceable. Instead, the density corresponding the σ 1.1 occupies the DNA binding channel with closed RNAP clamp. During open complex formation, downstream duplex DNA displaces σ 1.1 from the DNA binding channel (Murakami and Darst, 2003). Therefore, the class I complex structure represents early stage intermediate (RPi1) from the closed to open complex formation (**Figures 4.3 and 4.13A**).

Compares to the RPe and RPo structures determined in this study, binding DksA/ppGpp to RNAP induces the movements of RNAP mobile domains around the DNA binding channel such as the 1) β lobe/SI1, 2) β' jaw/SI3 and 3) DNA binding clamp (**Figures 4.14A and 4.15A**). Movements of these domain may affect the promoter DNA interaction and reduce the open complex stability. The β lobe/SI1 domain moves toward DksA to establish the contact between the SI domain and DksA, which was not observed in a previously determined X-ray structure of RNAP and DksA/ppGpp complex (Molodtsov et al., 2018). Deletion of the β lobe/SI1 domain reduces the DksA affinity to RNAP (Parshin et al., 2015), indicating that the

movement of β lobe/SI1 domain plays a role in the enhancement of DksA affinity to RNAP
(Parshin et al., 2015).

4.5.2.3 RP-DksA/ppGpp (II) structure represents a late stage intermediate (RPi2) on the pathway to RPo

In the second class of RNAP-*rrnBP1*-DksA/ppGpp complex (RP-DksA/ppGpp(II)), downstream duplex DNA (from +3 to +20) occupies in the DNA binding channel instead of $\sigma 1.1$ found in the class I complex, but the DNA density corresponding transcription bubble (from -8 to +2) is not traceable. The DNA occupancy found in the class II complex is similar to a previously determined cryo-EM structure of the MTB RNA-promoter DNA complex (Boyaci et al., 2019), suggesting that the class II complex represents a late stage intermediate of open complex (**Figure 4.13B**). The RP-DksA/ppGpp (II) structure shows that the β lobe/SI1 and β' clamp move back to the positions as observed in the RPc and RPo structures determined in this study (**Figure 4.14B**), and the DksA is no longer contacting with the β lobe/SI1 domain, suggesting the transition from the RPi1 to RPi2 reduces the DksA affinity to RNAP, which may eventually dissociate DksA from RNAP to allow the iNTP binding at the RNAP active site and initiating transcription (**Figure 4.15B**).

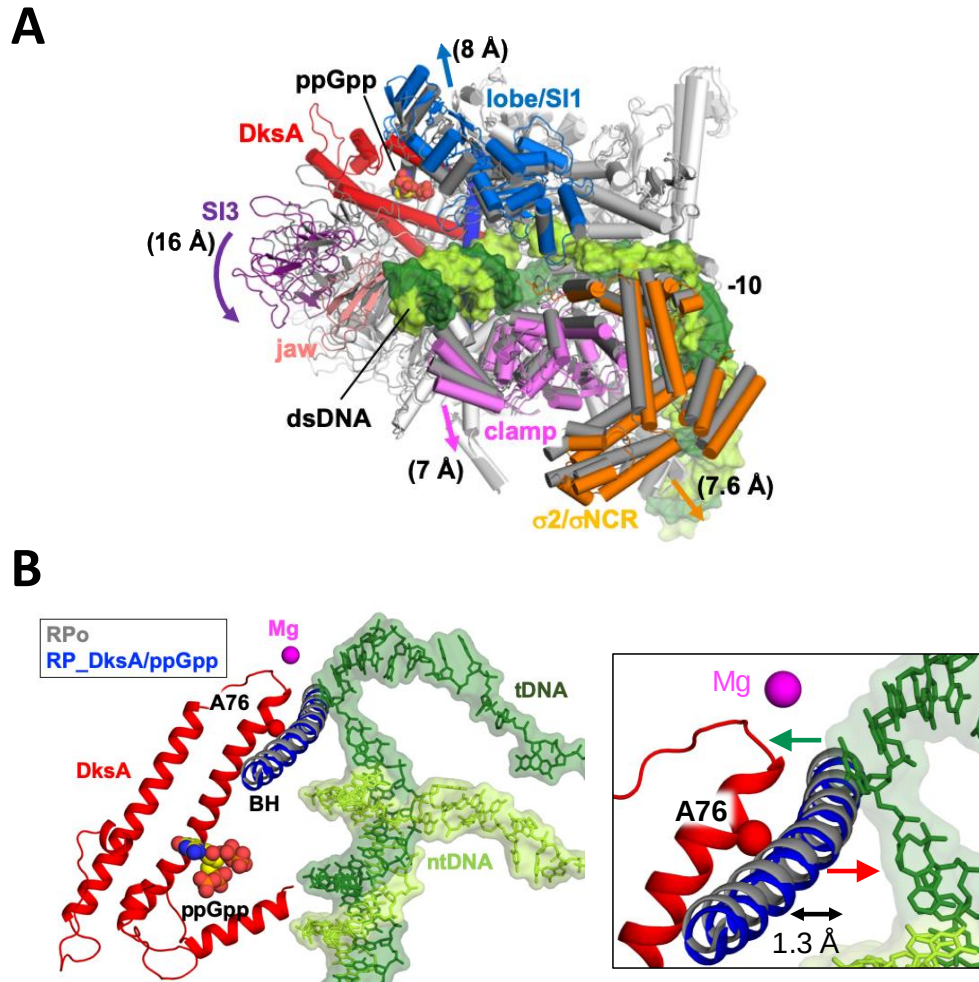


Figure 4.15: The conformational changes of RNAP induced by the DksA/ppGpp binding. (A) DksA/ppGpp binding to RNAP and DNA complex widens the DNA binding channel. RNAPs in the RP-DksA/ppGpp and RPo are superimposed with their α subunits and DPBB domains. RNAP (white with colors in the RP-DksA/ppGpp, gray in the RPo) and DksA (red) are depicted as cartoon representation and DNA is shown as surface model. Mobile domains surrounding the DNA binding channel are indicated (lobe/SI1, light blue; clamp, violet; jaw, salmon; SI3, deep purple; $\sigma 2/\sigma \text{NCR}$, orange) domains and their movements (directions and distances) upon the DksA/ppGpp binding are indicated. (B) Bridge helix (BH) interactions with DksA in RP-DksA/ppGpp and tDNA in RPo. A magnified view (right) of the CC tip of DksA and tDNA (+2 base) interacting with the BH. Directions of BH movements upon the DksA and tDNA interactions are shown as red and green arrows, respectively.

4.5.3 Monitoring DNA opening in solution using Cy-3 fluorescence signal

The structural analysis suggests a role of the interaction between σ NCR and β' CT in accommodating the -10 element DNA on the σ region 2 and its unwinding during the RPO formation. To investigate a role of the σ NCR movement during the open complex formation in solution, I compared DNA opening activity of the *E. coli* RNAP WT and derivative containing the β' E175K substitution, which weakens the σ NCR and β' CT interaction (Leibman and Hochschild, 2007). To test the DNA opening, I measured fluorescence signal of Cy3 attached at +3 positions in both *rrnBP1* or *T7A1* dsDNA promoters. Similar to 2-aminopurine (**Chapter 2**), Cy3 signal is sensitive to change in its environment (Ko and Heyduk, 2014). When Cy3-labeled dsDNA was mixed with the β' E175K variant, the signal was reduced to 25% of WT in both *rrnBP1* and *T7A1* promoters, indicating σ NCR and β' CT interaction plays an important role in DNA unwinding in solution (**Figure 4.16, 1 vs 4**).

I also tested the effect of DksA/ppGpp binding to RNAP in DNA opening in solution by measuring the Cy3 signal in RNAP-DksA/ppGpp complex. Compared with RNAP in the absence of DksA/ppGpp, the Cy3 signals were reduced to 40% and 75% in *rrnBP1* and *T7A1* promoters, respectively, indicating DksA/ppGpp action is more sensitive to *rrnBP1* promoter (**Figure 4.16, 1 vs 2 and 3**).

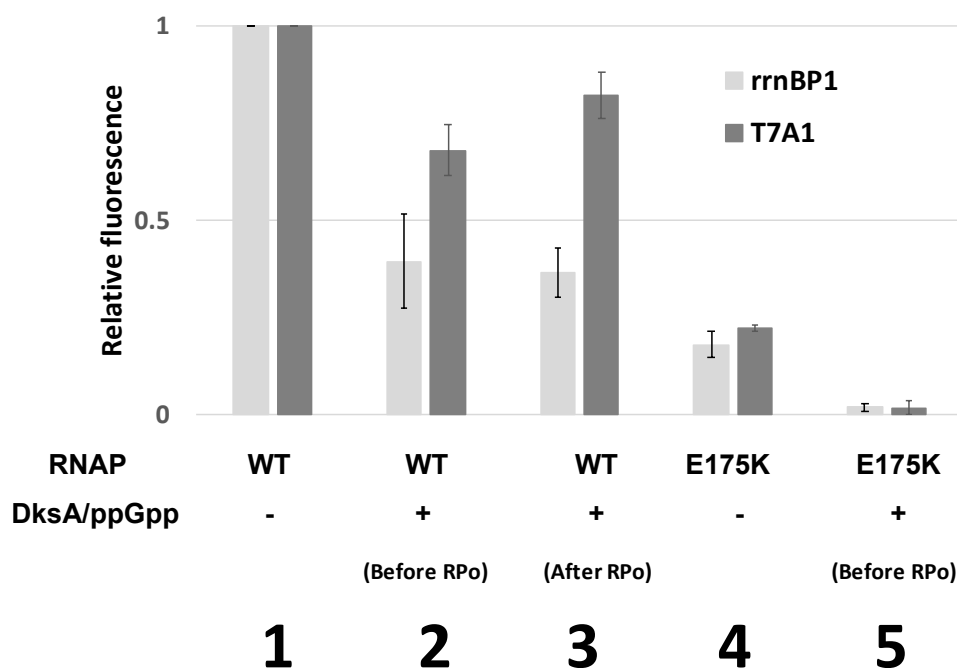


Figure 4.16: DNA opening assay using Cy-3 labeled promoters. (A) WT-RNAP in the absence of DksA/ppGpp. (B) WT-RNAP in the presence of DksA/ppGpp (added to free RNAP before the addition of the DNA). (C) WT-RNAP in the presence of DksA/ppGpp (added to RNAP-DNA open complex) (D) E175K RNAP in the absence of DksA/ppGpp. (E) E175K RNAP in the presence of DksA/ppGpp. Light gray = *rrnBP1* promoter, Dark gray = *T7A1* promoter.

4.5.4 DNA duplex free energy calculation

DNA opening assay shows that DksA/ppGpp inhibit the *rrnBP1* promoter unwinding more severely while *T7A1* promoter is still able to open as close to the condition without DksA/ppGpp (**Figure 4.16**). The RP-DksA/ppGpp structure shows that DksA/ppGpp opens the clamp (the lobe and jaw domains) of RNAP which facilitates the dsDNA loading to the main channel of RNAP to form RPi (**Figures 4.14 and 4.15**), but prevents σ NCR rotation that may facilitate DNA opening. I calculated DNA-duplex free energy of each promoter that show

different sensitivity to DksA/ppGpp action (**Figure 4.2**) by the nearest-neighbor method (SantaLucia, 1998) (**Figure 4.17**). As the DNA sequences suggest, the *rrnBP1* promoter with G-C rich discriminator has the lowest DNA duplex free energy (stable DNA duplex, less spontaneous DNA opening) while the *T7A1* and *argI* promoters with A-T rich discriminators have higher DNA duplex free energies (less stable DNA duplex, more spontaneous DNA opening) compared to *rrnBP1* promoter. The activated promoter (*ArgI*) has the highest free energy of the three promoters. More spontaneous opening of *T7A1* promoter around the TSS more likely to form the RPo by triggering the conformational change of the β lobe, disrupting the contact between the CT-helix of DksA and RNAP, and eventually dissociating the DksA from RNAP (**Figure 4.15B**). Less spontaneous opening of the *rrnBP1* promoter around the TSS less likely to form the RPo, resulting in compromise of the transcript initiation. Consistent with this idea, the genome-wide investigation of promoters found in *E. coli* showed that the promoters inhibited by DksA/ppGpp contain G-C rich discriminators while the promoters activated by DksA/ppGpp contain A-T rich discriminators (Sanchez-Vazquez et al., 2019).

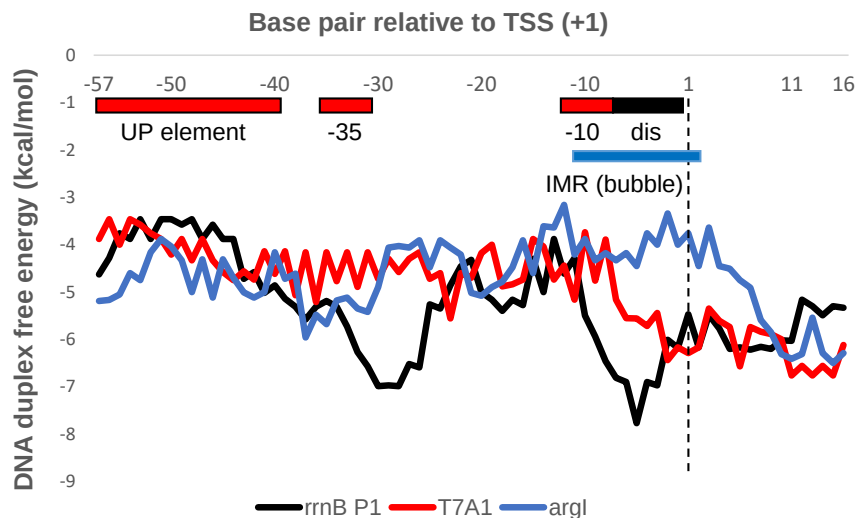


Figure 4.17: Promoter opening depends on the discriminator sequence. The DNA-duplex free energies of *rrnBP1* (black), *T7A1* (red) and *argI* (blue) are shown. UP element, -35 element and initially melted region (IMR), which is a part of transcription bubble are indicated.

4.6 Discussion

While it is still challenging, cryo-EM has become an indispensable tool in structural biology because of advantages such as 1) less amount of samples required compare to other conventional techniques like X-ray crystallography and NMR spectroscopy, 2) the samples are in near native state in solution, and 3) the 3D classification of particles may reveal dynamics of macromolecules.

In this chapter, I successfully prepared the RNAP-*rrnBP1* promoter complex and determined cryo-EM structures of the complex including the closed complex and open complex as well as two distinct complexes in the presence of DksA/ppGpp. In the presence of initiating nucleotides, the structures of RNAP-DNA binary complex could be classified into two native structures; closed complex and open complex. The RPc structure reveals that the UP element bound form of the carboxyl-terminal domain of α subunit (α CTD) is a dimer (**Figure 4.8**), the σ factor does not recognize the DNA sequence of -10 element in duplex form (**Figure 4.9**) and movement of the σ non-conserved region (σ NCR) to contact with the β' clamp toe domain (β' CT) (**Figure 4.10**). The RPo structure shows the “open complex scrunching” by which the template DNA contacts with σ finger and β' lid (**Figure 4.11**). DksA/ppGpp binding triggers conformational changes of RNAP mobile domains that influence the stability of promoter complex (**Figure 4.14 – 4.15**).

The comparison of apo-form RNAP, RPc and RPo structures reveal the dynamics of σ NCR during the DNA binding and opening (**Figures 4.18 and 4.19A-C**). Compares with an apo-holoenzyme, the RPc structure show that σ non-conserved region (σ NCR) swings toward the β' clamp domain (β' CT) upon binding DNA. The σ NCR- β' CT interaction unmask the σ region 2.3 from the acidic loop of σ NCR, which may provide time for DNA binding to form closed

complex (**Figure 4.18A and B**). σ region 2.3 interact with double-stranded form of -10 element without any specific amino acids-DNA base interaction in the RPc structure while -13A base of ntDNA flips out and fits into the binding pocket of σ region 2.3 in the RPo structure. The -13A base of ntDNA is highly conserved in -10 element of bacterial promoter, and the interaction is suspected to be important in initiating DNA unwinding.

The RPo structure shows the σ NCR rotates back to the same position as observed in the apo-holoenzyme. Based on the structure, the acidic loop of σ the NCR interact with the phosphate backbone of the DNA. Therefore, it is suspected that the charge-charge interactions or mechanical stress initiate the DNA unwinding (**Figure 4.18C**). The DNA opening assay using the Cy-3 fluorophore labeled DNA shows the RNAP derivative with a substitution in the β 'CT has reduced DNA opening activity, further suggesting σ NCR and β 'CT contacts plays an role in facilitating DNA unwinding. I can not rule out the possibility that the RNAP derivative may have reduced DNA binding affinity compared to WT RNAP. In the future, I am planning to compare the DNA binding affinity by the fluorescence anisotropy.

The σ NCR and β 'CT interaction was also previously proposed to be important for RNAP promoter escape and to hinder early elongation pausing based on the biochemical study (Leibman and Hochschild, 2007). The structural basis for the σ NCR and β 'CT interaction during the promoter escape will be studied in the future.

The RPo structure also reveals the structural basis of transcription start site selection of the *rrnBP1* promoter, which initiates transcription from 9 bases downstream from the -10 element. The tDNA scrunches inside the active site of RNAP and -7G base of tDNA fits into pocket to locate physiologically relevant transcription start site to *i* site, which allows RNAP to change transcription activity depending on ATP concentration. Also, the tDNA scrunching

destabilizes open complex and facilitates promoter escape. It will be interesting to test if changing the transcription start site to other base positions affect the rate or efficiency of rRNA synthesis.

The RP-DksA/ppGpp structures also can be classified into two native intermediate structures having either $\sigma 1.1$ (RP-DksA/ppGpp (I)) or DNA (RP-DksA/ppGpp (II)) in the DNA binding channel (**Figure 4.19E-F**). The DksA binds secondary channel of RNAP and $\sigma 1.1$ positions deeper inside the DNA binding channel in RP-DksA/ppGpp(I) structure compare to apo-form RNAP. Consequently, RNAP β' clamp rotate counter clockwise around DNA binding channel to avoid steric clash with $\sigma 1.1$, which would prevents σ NCR rotation toward β' CT domain and hinder DNA opening. The DksA-RNAP interaction also restricts the opening of the main channel which is required for the tDNA nucleation to the active site of RNAP for RPo formation.

Based on the structures determined in this study and DNA opening assay, I propose the following model to explain pleiotropic effect of DksA/ppGpp on different types of promoters. RNAP-DNA complex with only pre-melted or DNA with higher meltability (spontaneous opening) such as *T7A1* or *argI* can load the single stranded tDNA to the active site of RNAP through narrower DNA binding channel, which changes the BH conformation and trigger the DksA/ppGpp dissociation from the RNAP to allow RNAP to initiate transcription. In contrast, the G-C rich discriminator of *rrnBP1* promoter prevents spontaneous opening of dsDNA, results in inefficient RPo formation and inefficient transcript initiation. Confirming such a model will require more comprehensive biochemical and DNA opening assays using RNAP derivatives (**Chapter 5**).

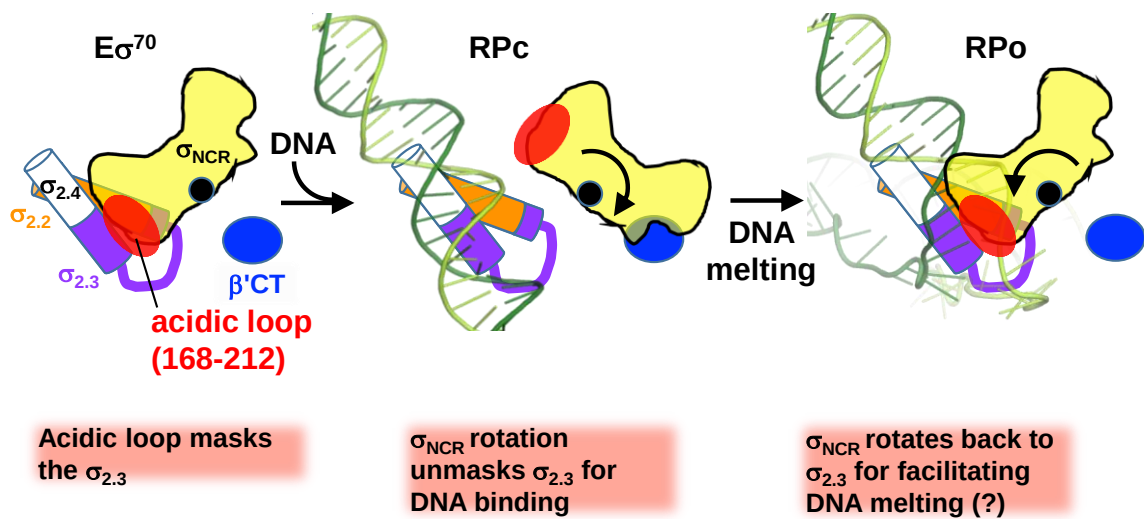


Figure 4.18: Cartoon model showing the dynamics of the σ_{NCR} during the DNA binding and opening. σ region 2.3 is masked by the acidic loop of σ_{NCR} in the apo-form RNAP ($E\sigma^{70}$). The interaction between σ_{NCR} and β' CT during closed complex formation expose the $\sigma_{2.3}$ which binds -10 element of DNA. The back rotation of σ_{NCR} may facilitates DNA opening by charge-charge interactions or mechanical stress between phosphate backbone of DNA and the acidic loop of σ_{NCR} .

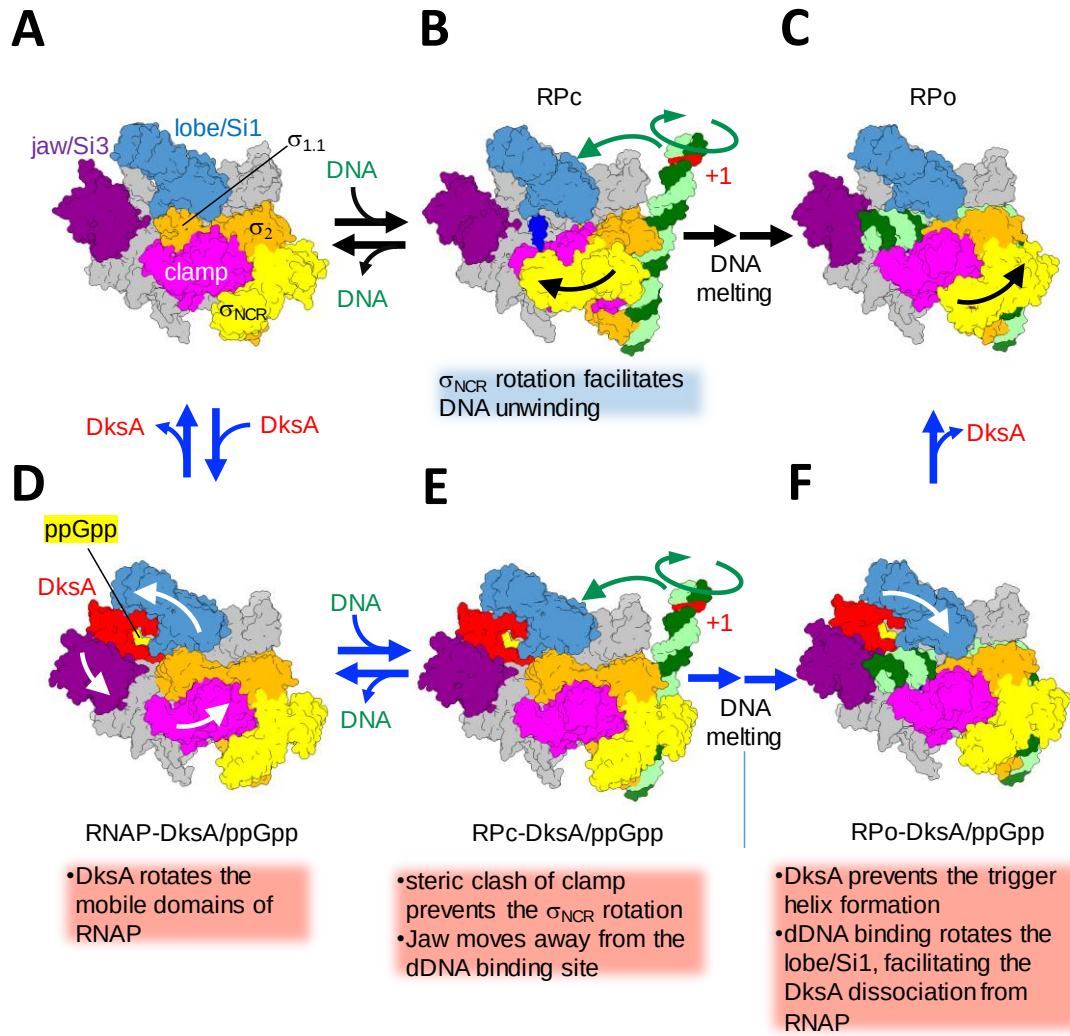


Figure 4.19: Schematic representations of the promoter-specific transcription regulation by DksA/ppGpp. Cartoon models of the RNAP. Apo-form RNAP (A) binds DNA to form closed complex (B). σ_{NCR} rotation toward β' clamp toe domain facilitates DNA unwinding and open complex formation (C). DksA/ppGpp binds secondary channel of RNAP and rotates the mobile domains (clamp, β lobe/Si1, and jaw/Si3) (D), which prevents σ_{NCR} rotation toward β' clamp toe domain (E). downstream DNA ejects $\sigma_{1.1}$, binds DNA binding channel and rotates the β lobe/Si1. DksA dissociates from RNAP by DNA with higher meltability to allow RNAP initiate transcription.

4.7 Acknowledgements

I thank the staffs at the National Cancer Institute's National Cryo-EM facility at the Frederick and Carol Bator at the Penn State Huck Institute Cryo-EM facility for support of the cryo-EM data collections. I thank M. Zuhaib Qayyum, a current postdoc in Murakami lab for helping with cryo-EM data processing, members of Andrey Kulbachinskiy lab at Russian Academy of Sciences for collaboration and providing some of plasmids for RNAP derivatives expressions, Christian Dienemann at Max Planck Institute for providing the method to calculate DNA duplex free energy, Tim Miyashiro at Penn State for TECAN microplate reader and Ann Hochschild at Harvard Medical School for providing some of plasmids for RNAP derivatives expressions. This work was supported by NIH grants GM087350 and GM131860 to Katsuhiko Murakami.

4.8 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 crystallographica. Section D, Biological crystallography* 66, 1153-1163.

Barker, M.M., and Gourse, R.L. (2001). Regulation of rRNA transcription correlates with nucleoside triphosphate sensing. *J Bacteriol* 183, 6315-6323.

Boyaci, H., Chen, J., Jansen, R., Darst, S.A., and Campbell, E.A. (2019). Structures of an RNA polymerase promoter melting intermediate elucidate DNA unwinding. *Nature* 565, 382-385.

Bremer, H., and Dennis, P.P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* 3.

Brown, L., Gentry, D., Elliott, T., and Cashel, M. (2002). DksA affects ppGpp induction of RpoS at a translational level. *J Bacteriol* 184, 4455-4465.

Burgos, H.L., O'Connor, K., Sanchez-Vazquez, P., Gourse, R.L. (2017). Roles of transcriptional and translational control mechanisms in regulation of ribosomal protein synthesis in *Escherichia coli*. *J Bacteriol* 199, e00407-17.

Cashel, M., and Gallant, J. (1969). Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature* 221, 838-841.

Chen, J., Noble, A.J., Kang, J.Y., and Darst, S.A. (2018). Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J Struct Biol X*, 100005.

Condon, C., Philips, J., Fu, Z.Y., Squires, C., and Squires, C.L. (1992). Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. *EMBO J* 11, 4175-4185.

Condon, C., Squires, C., and Squires, C.L. (1995). Control of rRNA transcription in *Escherichia coli*. *Microbiol Rev* 59, 623-645.

Costa, T.R.D., Ignatiou, A., Orlova, E.V. (2017). Structural analysis of protein complexes by cryo electron microscopy. *Methods Mol Biol* 1615, 377-413.

Dienemann, C., Schwalb, B., Schilbach, S., and Cramer, P. (2019). Promoter Distortion and Opening in the RNA Polymerase II Cleft. *Mol Cell* 73, 97-106 e104.

- Doerr, A. (2016). Single-particle cryo-electron microscopy. *Nat Methods* 13, 23.
- Edwards, A.N., Patterson-Fortin, L.M., Vakulskas, C.A., Mercante, J.W., Potrykus, K., Vinella, D., Camacho, M.I., Fields, J.A., Thompson, S.A., Georgellis, D., Cashel, M., Babitzke, P., and Romeo, T. (2011). Circuitry linking the Csr and stringent response global regulatory systems. *Mol Microbiol* 80, 1561-1580.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphic. *Acta Crystallogr D Biol Crystallogr* 60, 2126-2132.
- Feklistov, A., and Darst, S.A. (2011). Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit. *Cell* 147, 1257-1269.
- Gaal, T., Ross, W., Blatter, E.E., Tang, H., Jia, X., Krishnan, V.V., Assa-Munt, N., Ebright, R.H., and Gourse, R.L. (1996). DNA-binding determinants of the alpha subunit of RNA polymerase: novel DNA-binding domain architecture. *Genes Dev* 10, 16-26.
- Ginsburg, D., and Steitz, J.A. (1975). The 30 S ribosomal precursor RNA from *Escherichia coli*. A primary transcript containing 23 S, 16 S, and 5 S sequences. *J Biol Chem* 250, 5647-5654.
- Gralla, J.D. (2005). *Escherichia coli* ribosomal RNA transcription: regulatory roles for ppGpp, NTPs, architectural proteins and a polymerase-binding protein. *Mol Microbiol* 55, 973-977.
- Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* 125, 1069-1082.
- Haugen, S.P., Ross, W., and Gourse, R.L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* 6, 507-519.

- Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* 13, 298-309.
- Ko, J., and Heyduk, T. (2014). Kinetics of promoter escape by bacterial RNA polymerase: effects of promoter contacts and transcription bubble collapse. *Biochem J* 463, 135-144.
- Kuhlbrandt, W. (2014). Biochemistry. The resolution revolution. *Science* 343, 1443-1444.
- Leibman, M., and Hochschild, A. (2007). A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J* 26, 1579-1590.
- Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods* 10, 584-590.
- Lyumkis, D. (2019). Challenges and opportunities in cryo-EM single-particle analysis. *J Biol Chem* 294, 5181-5197.
- Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S.E., and Murakami, K.S. (2018). Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol Cell* 69, 828-839 e825.
- Murakami, K., Fujita, N., and Ishihama, A. (1996). Transcription factor recognition surface on the RNA polymerase alpha subunit is involved in contact with the DNA enhancer element. *EMBO J* 15, 4358-4367.
- Murakami, K.S., and Darst, S.A. (2003). Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* 13, 31-39.

Narayanan, A., Vago, F.S., Li, K., Qayyum, M.Z., Yernool, D., Jiang, W., and Murakami, K.S. (2018). Cryo-EM structure of Escherichia coli sigma(70) RNA polymerase and promoter DNA complex revealed a role of sigma non-conserved region during the open complex formation. *J Biol Chem* 293, 7367-7375.

Nogales, E. (2016). The development of cryo-EM into a mainstream structural biology technique. *Nat Methods* 13, 24-27.

Nomura, M. (1999). Regulation of ribosome biosynthesis in Escherichia coli and Saccharomyces cerevisiae: diversity and common principles. *J Bacteriol* 181, 6857-6864.

Parshin, A., Shiver, A.L., Lee, J., Ozerova, M., Schneidman-Duhovny, D., Gross, C.A., and Borukhov, S. (2015). DksA regulates RNA polymerase in Escherichia coli through a network of interactions in the secondary channel that includes Sequence Insertion 1. *Proc Natl Acad Sci U S A* 112, E6862-6871.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W., and Gourse, R.L. (2004a). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118, 311-322.

Paul, B.J., Ross, W., Gaal, T., and Gourse, R.L. (2004b). rRNA transcription in Escherichia coli. *Annu Rev Genet* 38, 749-770.

Perederina, A., Svetlov, V., Vassilyeva, M.N., Tahirov, T.H., Yokoyama, S., Artsimovitch, I., and Vassilyev, D.G. (2004). Regulation through the secondary channel--structural framework for ppGpp-DksA synergism during transcription. *Cell* 118, 297-309.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.

Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu Rev Microbiol* 62, 35-51.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14, 290-296.

Pupov, D., Petushkov, I., Esyunina, D., Murakami, K.S., and Kulbachinskiy, A. (2018). Region 3.2 of the sigma factor controls the stability of rRNA promoter complexes and potentiates their repression by DksA. *Nucleic Acids Res* 46, 11477-11487.

Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* 314, 1139-1143.

Ross, W., Gosink, K.K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., and Gourse, R.L. (1993). A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* 262, 1407-1413.

Ross, W., Sanchez-Vazquez, P., Chen, A.Y., Lee, J.H., Burgos, H.L., and Gourse, R.L. (2016). ppGpp Binding to a Site at the RNAP-DksA Interface Accounts for Its Dramatic Effects on Transcription Initiation during the Stringent Response. *Mol Cell* 62, 811-823.

Ross, W., Vrentas, C.E., Sanchez-Vazquez, P., Gaal, T., and Gourse, R.L. (2013). The magic spot: a ppGpp binding site on E. coli RNA polymerase responsible for regulation of transcription initiation. *Mol Cell* 50, 420-429.

Sanchez-Vazquez, P., Dewey, C.N., Kitten, N., Ross, W., and Gourse, R.L. (2019). Genome-wide effects on *Escherichia coli* transcription from ppGpp binding to its two sites on RNA polymerase. *Proc Natl Acad Sci U S A* 116, 8310-8319.

SantaLucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* 95, 1460-1465.

Winkelman, J.T., Chandrangsu, P., Ross, W., and Gourse, R.L. (2016). Open complex scrunching before nucleotide addition accounts for the unusual transcription start site of *E. coli* ribosomal RNA promoters. *Proc Natl Acad Sci U S A* 113, E1787-1795.

Zengel, J.M., and Lindahl, L. (1994). Diverse mechanisms for regulating ribosomal protein synthesis in *Escherichia coli*. *Prog Nucleic Acid Res Mol Biol* 47, 331-370.

Chapter 5

Summary and Future Directions

5.1 Abstract

Visualizing the three-dimensional structure of a macromolecule in a biological system provides many new insights into the function and mechanism of the biological process. This can be achieved by solving the atomic resolution structures of macromolecules by X-ray crystallography and cryo-EM. During my PhD training at Penn State, I studied the mechanisms of bacterial transcriptions by structural approaches.

First, I determined the various structures of bacterial RNA polymerase (RNAP) in the process of reiterative transcription from the *pyrG* and *pyrBI* promoters by using time-dependent soak-trigger-freeze X-ray crystallography (**Chapters 2 and 3**). I optimized the procedures of *T. thermophilus* RNAP purification and crystallization followed by the soaking experiments to reveal various transcription mechanisms and investigate binding modes of various ligands. Additionally, I prepared and determined the cryo-EM structures of *E. coli* RNAP – ribosomal RNA promoter DNA complexes in the closed complex, open complex and in the presence of DksA/ppGpp by cryo-EM (**Chapter 4**). I optimized the procedures of *E. coli* RNAP purification, complex preparation, vitrification and data collection strategy to elucidate the dynamics and flexibility of transcription by bacterial RNAP. In this chapter, I will summarize the significance of my studies and remaining questions that still need to be answered in the future. I will conclude with a new project that I will be working on, determining the structure of human DNA polymerase delta (Pol δ) complex in the process of DNA replication by cryo-EM.

5.2 Reiterative transcription

5.2.1 Current state of knowledge

For the first topic of my dissertation, I investigated the mechanism of reiterative transcription by bacterial RNAP from the most-well characterized promoters, *pyrG* of *B. subtilis* and *pyrBI* of *E. coli* (Liu et al., 1994; Meng et al., 2014; Turnbough 2011; Turnbough and Switzer, 2008). Although reiterative transcription plays a key role in controlling gene expression, much remained to be learned about the mechanism of this reaction and how it differs from canonical transcription. This is particularly true for the positioning of the nascent RNA inside RNAP. The initial study (**Chapter 2**) revealed a unique RNA extension pathway that detours from the canonical RNA exit channel and extends toward the main channel of RNAP without displacement of the σ finger (**Figure 2.4**) (Murakami et al., 2017). The structure also revealed a previously unobserved stable 3-bp DNA/RNA hybrid during the reiterative transcription (**Figure 2.5**). I continued to investigate more detailed mechanism of reiterative transcription by time-dependent soak-trigger-freeze X-ray crystallography. One of the biggest challenges in X-ray crystallography is preparing large amount of crystallizable bacterial RNAP. It was particularly more challenging to prepare endogenous *T. thermophilus* RNAP. However, my skill has been drastically improved over the past years and I was able to crystallize the *T. thermophilus* RNAP and promoter DNA complexes, prepare the transcript initiation complexes by soaking the NTP into the RNAP-DNA complex crystals and determine the high-resolution structures by X-ray crystallography. The structures of the transcript initiation complex revealed how the RNA slips on the template DNA and how the 5' end of the RNA is guided toward the main channel of RNAP (**Chapters 2 and 3**) (Shin et al., 2020). Using the structural and biochemical approaches, I

showed that extra G bases are added at the 3' end of the RNA by the G-G mismatch between the +4G template DNA (tDNA) and an incoming GTP, followed by the base sharing with the upstream +3C tDNA (**Figure 2.9**). From the analysis of the *pyrG* promoter variants, I showed that the base of tDNA at -1 position plays a base-stacking role with the 5' end of the RNA and thus determines the fate and the maximum length of the reiterative RNAs (**Figure 3.5**). Lastly, I showed that the non-productive reiterative RNAs synthesized from the *pyrBI* promoter extend toward the canonical RNA exit channel and clash with the σ finger, indicating the σ factor needs to be released from the RNAP core to observe the long-stretch of RNAs (**Figure 3.3**).

5.2.2 Future directions

5.2.2.1 Visualizing the non-productive reiterative RNA by determining the cryo-EM structure of the reiterative transcription complex

The direct visualization of RNAP accommodating a full-length and long stretch of reiterative RNA from these *pyrBI*-type promoters cannot be addressed by X-ray crystallography because the σ factor release is not permitted in the RNAP crystals. Elucidating such structures will be achieved in the future by cryo-EM (**Chapter 4**). Determining the structures by cryo-EM is more suitable to observe the conformational change of the σ finger, the dissociation of the σ factor and the extension of the RNA from the RNA exit channel. If this particular reiterative RNAs from the *pyrBI* promoter (AAUUUn) are too rapidly released from RNAP to observe, the alternative non-productive promoters such as the *pyrG* promoter variants (**Chapter 3**) may be used.

5.2.2.2 How poly-G RNA produced from the *pyrG* promoter switches to canonical transcription?

At any point during reiterative transcription from the *pyrG* promoter, there can be a switch to canonical transcription by incorporating a CTP (**Figure 2.1**). This switch increases the DNA/RNA hybrid length to 4-bp, followed by a UTP incorporation that makes the DNA/RNA hybrid length to 6-bp. I speculate that the increase in the DNA/RNA hybrid length changes the direction of the RNA extension from the main channel to the canonical RNA exit channel, resulting in clashing of 5' end of the RNA to the σ finger to release the σ factor, or the 6-bp DNA/RNA hybrid clash with the σ finger first, followed by changing the direction of the 5' end of the RNA (**Figure 5.1**). I investigated the switching mechanism by soaking 1 mM GTP for 1 hour followed by soaking either 4 mM CTP or 4 mM CTP and 4 mM UTP together into the *T.thermophilus* RNAP – *pyrG* promoter complex crystals (**Figure 5.2**).

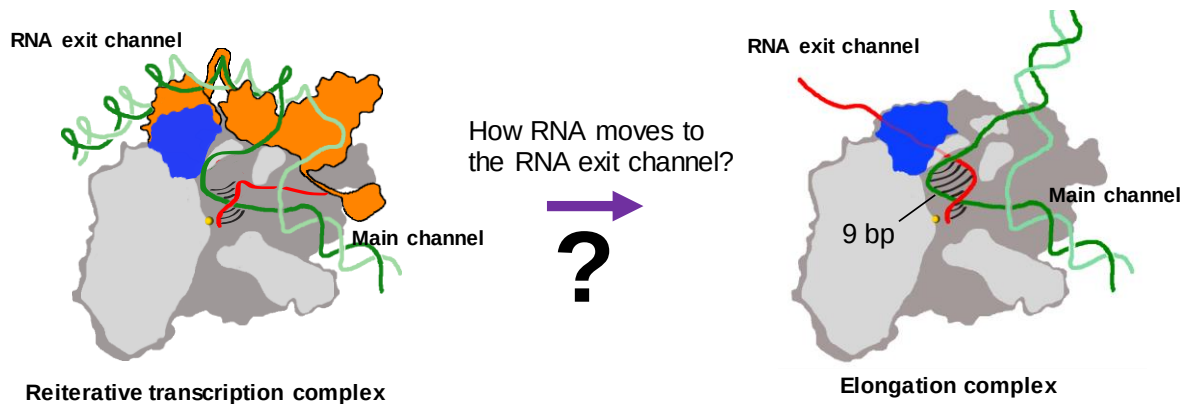


Figure 5.1: Cartoon model showing switching from the reiterative transcription complex to the elongation complex. Cross-sectional views of the RNAP holoenzyme β -flap, blue; σ , orange; rest of RNAP, gray; catalytic Mg^{2+} , yellow sphere), promoter DNA (template strand, dark green; nontemplate strand, light green) and the RNA transcript (red) are shown. Base pairs between the DNA and RNA are depicted as black lines. Figure modified from the reference (Murakami et al., 2017).

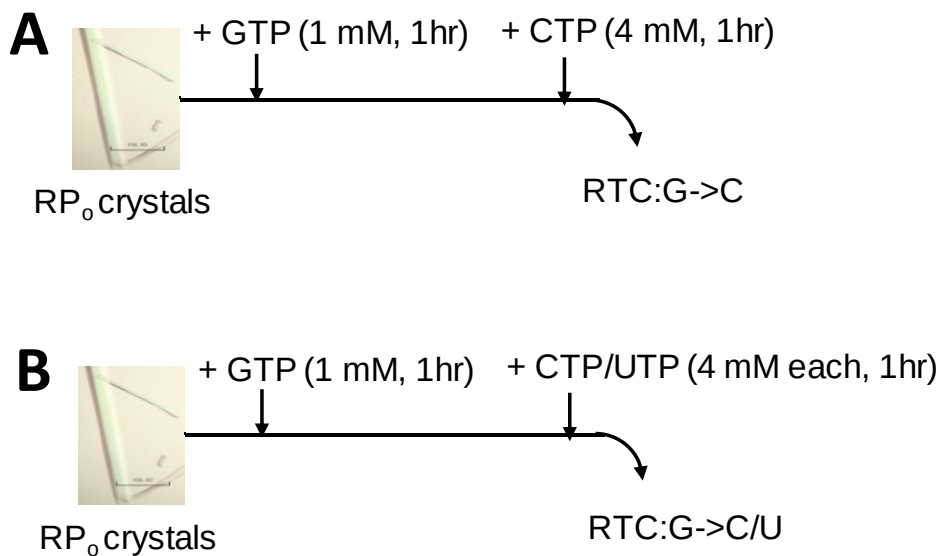
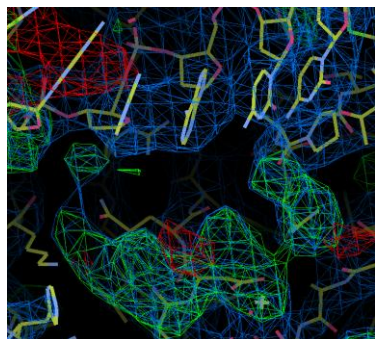


Figure 5.2: Experimental scheme of the complex preparation to determine intermediate structures switching from the reiterative transcription to the canonical transcription by time-dependent soak-trigger-freeze X-ray crystallography. (A) Soaking GTP followed by CTP that make 4-bp DNA/RNA hybrid. (B) Soaking GTP followed by CTP and UTP that make 6-bp DNA/RNA hybrid.

Although the active site of RNAP shows an electron density corresponding to the tDNA and RNA (**Figure 5.3**), the resolutions of these structures ($3.8\text{\AA}$) are not high enough to determine the exact locations and the bases of the tDNA and RNA. I speculate that this is due to the heterogeneous population of poly-G RNAs with various lengths and poly-G + C bases in the 3' end of RNA in the RTC:G-> C structure (e.g. GGGn and GGGnC) and poly-G + C/U bases in the 3' end of RNA in the RTC:G->C/U structure (e.g. GGGn, GGGnC, and GGGnCU). When I soaked GTP (1mM) and CTP (2mM) for 30 min into the *T.thermophilus* RNAP – *pyrG* promoter DNA binary complex, the active site shows the clear electron density of the RNA corresponding to 5'-GGGC-3' (Resolution = $3.0\text{\AA}$) produced *in crystallo* (**Figure 5.4**), which supports the heterogeneity hypothesis. Therefore, more controlled experiments to synthesize the same length of poly-G RNA (such as using a primer) *in crystallo* will be performed in the future.

A RTC:G->C



B RTC:G->C/U

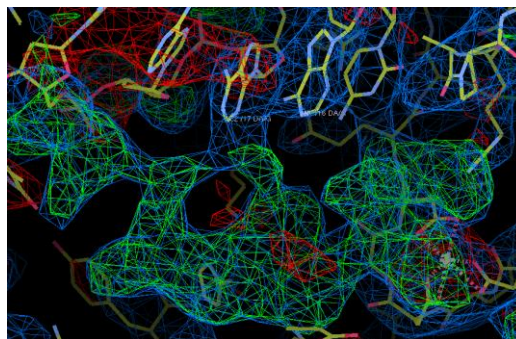
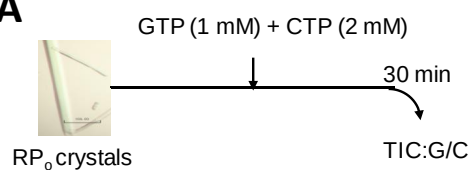


Figure 5.3: The electron density maps ($3.8\text{\AA}$) of the active sites of RNAPs (A) RTC:G->C structure. (B) RTC:G->C/U structure. Green mesh indicates the electron density of *in crystallo* synthesized RNA. Blue mesh and stick model indicates the template DNA.

A



B

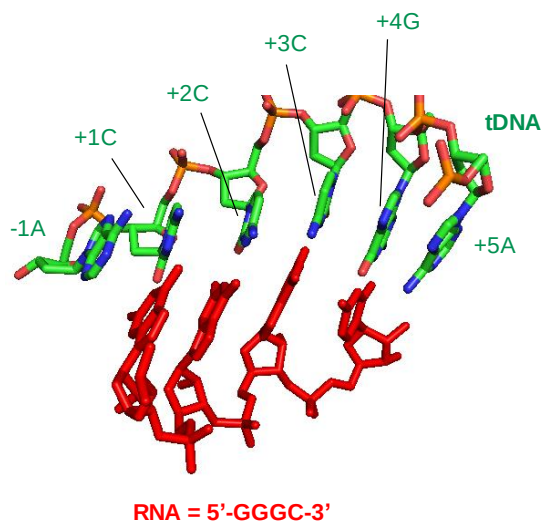


Figure 5.4: Experimental scheme of the complex preparation and the structure. (A) Soaking GTP and CTP together into the RNAP-*pyrG* promoter binary complex to make 4-mer RNA. **(B)** The active site of RNAP shows *in crystallo* synthesized 5'-GGGC-3' RNA.

5.2.2.3 How about other *pyrG* homopolymeric tracts?

The RTC-30' structure with 8-mer poly-G RNA shows that narrow opening of the main channel could provide a barrier to further RNA extension (**Figure 2.4**) (Murakami et al., 2017). Interestingly, changing the homopolymeric tract of the tDNA from natural *pyrG* tDNA, C₃ (RNA = GGGn) to T₃ (RNA = AAAn) did not change the reiterative transcription pattern while changing it to G₃ (RNA = CCCn) changed a ladder of transcripts which easily exceeded 40-mer in length (**Figure 2.7**). The DNA/RNA hybrid length for all of these transcripts is 3-bp. Poly-G and poly-A transcripts (purines) would have the base stacking interaction with -1A base of the tDNA while poly-C transcripts (pyrimidine) would not have. I crystallized *T.thermophilus* RNAP – *pyrG* promoter variants complexes containing different homopolymeric tract sequences and soaked corresponding NTPs to trigger *in crystallo* reiterative transcription (**Figure 5.5**).

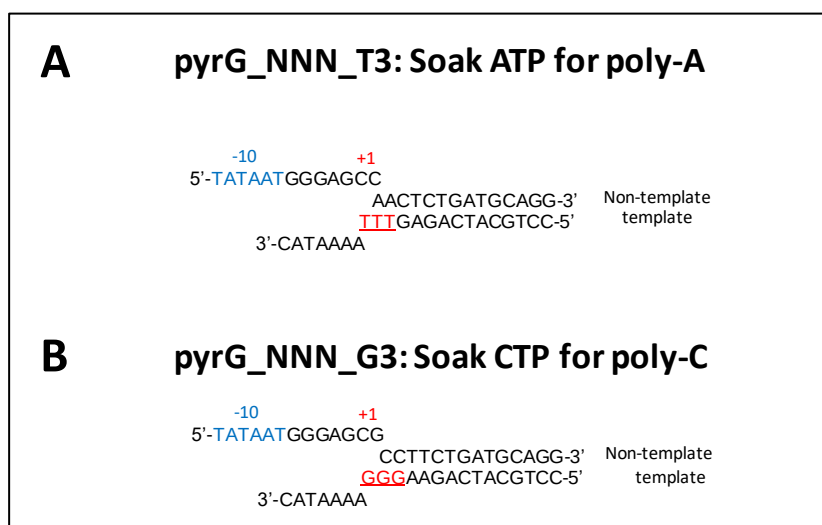


Figure 5.5: The fork-DNAs used for *in crystallo* transcription containing various slippage-prone homopolymeric tract sequences (NNN, underlined) found in the *pyrG* promoter (red). Transcription start site and -10 element are indicated in red and blue, respectively.

Both structures were solved at 3.2 Å resolution. The density map of poly-A transcripts synthesized from the *pyrG*_NNN_T3 DNA shows a 3-bp DNA/RNA hybrid with an incoming ATP at $i + 1$ site (**Figure 5.6A**). Similar to the poly-G transcripts synthesized from the *pyrG* promoter, I observed the electron density around the main channel of RNAP that may be corresponding to 5' end of the poly-A transcripts (**Figure 5.6B**). The density map of poly-C transcripts synthesized from the *pyrG*_NNN_G3 DNA was not as clear as the poly-A transcripts. I observed the electron density at the active site corresponding to 3-mer poly-C making a 3-bp DNA/RNA hybrid with the tDNA GGG tract, but the exact location of the 5' end of the poly-C transcript was not identifiable (**Figure 5.6C**). This may be due to the presence of a heterogeneous population of poly-C transcripts being synthesized and released from RNAP simultaneously.

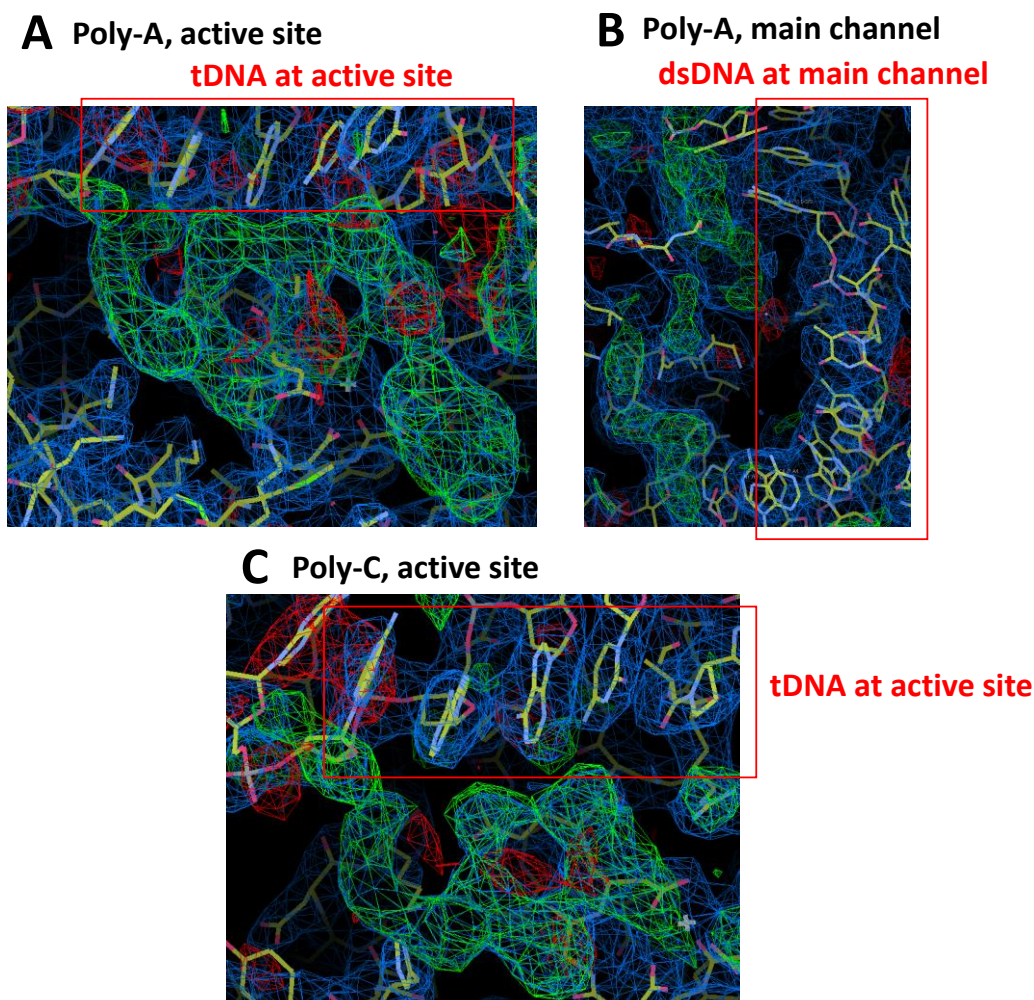


Figure 5.6: Electron density maps of the active sites of RNAPs with the *pyrG* homopolymeric tract variants (A) Electron density map of the active site of RNAP from the *pyrG_NNN_T3* DNA. Green mesh corresponds to *in crystallo* synthesized poly-A transcript. Blue mesh with a stick model indicates the template DNA. (B) Electron density map around the main channel of RNAP from *pyrG_NNN_T3* DNA. Green mesh corresponds to de novo synthesized poly-A transcript. Blue mesh with a stick model indicates downstream dsDNA at the main channel. (C) Electron density map of the active site of RNAP from the *pyrG_NNN_G3* DNA. Green mesh corresponds to *in crystallo* synthesized poly-C transcript. Blue mesh with stick model indicates template DNA.

5.3 Ribosomal RNA transcription regulation

5.3.1 Current state of knowledge

For the second topic of my dissertation, I investigated the mechanism of rRNA transcription and the rRNA transcription regulation by the RNAP secondary binding factor DksA. rRNA (*rrnBP1* in *E. coli* for example) is the highest expressed gene in bacteria during vegetative growth whereas it is drastically down-regulated by the DksA/ppGpp binding to RNAP during the stationary phase and stressed conditions (Bremer and Dennis, 2008; Nomura, 1999). The down-regulation of the rRNA transcription is due to the low NTP concentrations coupled with the inhibition action of DksA. The DksA action is enhanced by the alarmone “ppGpp”, the molecule which the concentration drastically increases during the stressed condition (Cashel and Gallant, 1969; Haugen et al., 2006; Haugen et al., 2008; Haurlyuk et al., 2015; Molodtsov et al., 2018; Paul et al., 2004a; Paul et al., 2004b; Potrykus and Cashel, 2008). However, the DksA/ppGpp has no effect on transcription from the *T7A1* promoter and activates transcription from the *argI* promoter. The sequence analysis of the *rrnBP1* promoter shows the unique features such as a longer G-C rich discriminator between the -10 element and the transcription start site, the presence of the UP element that enhances the rate of the transcription and a shorter spacer between -35 and -10 elements compared to the *T7A1* and the *argI* promoters (Barker and Gourse, 2001) (**Figure 4.2**)

Previously determined the X-ray crystal structure of *E. coli* RNAP σ^{70} holoenzyme with DksA/ppGpp revealed the binding mode of the DksA/ppGpp to RNAP, but the study could not adequately address the pleiotropic effect of the DksA/ppGpp due to the absence of a DNA in the structures (Molodtsov et al., 2018). Additionally, the crystal packing effect did not fully reveal

the conformational changes of RNAP that provide insights into the mechanism of the transcription regulation.

During my last 2-3 years of Ph.D. training, I investigated the mechanisms of the extraordinary regulation of the rRNA transcription and the rRNA transcription inhibition by DksA/ppGpp using cryo-EM. I determined the cryo-EM structures of *E. coli* RNAP σ^{70} holoenzyme bound at the *rrnBP1* promoter in the closed complex (RPc), open complex (RPo) and in the presence of DksA and ppGpp (RP-DksA/ppGpp). The RPc structure reveals that the UP element bound form of the α CTD is a dimer (**Figure 4.9**). The comparison of RPc structure to the apo-form RNAP and the RPo structure (**Figure 4.10**) reveals the σ NCR contacts the β 'CT to expose the σ region 2.3 to the double-stranded DNA during the closed complex formation. The DNA opening assay suggests that the interaction is important in DNA unwinding (**Figure 4.16**). The RPo structure also reveals the structural basis for scrunching of the -7G base of the tDNA into the pocket formed by the lid of RNAP β ' subunit and the σ finger (**Figure 4.11**).

I prepared the RNAP-*rrnBP1* promoter complex in the presence of DksA/ppGpp. 3D classification of the particles results in two distinct 3D classes – class I with σ region 1.1 and class II with downstream DNA at the DNA binding channel, respectively. Both classes show the DksA binds at the RNAP secondary channel and interacts with the bridge helix (**Figure 4.13**). The class I structure represents an earlier intermediate on the pathway to forming the RPo. The DksA binding induces the movement of the RNAP mobile domains including the β lobe/SI1, β 'jaw/SI3, and clamp (**Figures 4.14 and 4.15**). Based on the structure, it is speculated that the rotation of β ' clamp would sterically clash with the σ NCR rotation toward the β 'CT domain and hinder the DNA opening (**Figure 4.14B**). The class II structure represents a late intermediate on the pathway to forming the RPo. The structure shows that the conformations of the β lobe/SI1 and β ' clamp are back to the same positions as observed in the RPo, due to the ejection of the

$\sigma^{1.1}$ from the DNA binding channel followed by the entrance of the DNA (**Figure 4.14B**). The conformational change may dissociate the DksA from RNAP to allow transcript initiation (**Figure 4.15B**). The DNA opening assay suggests the DksA/ppGpp binding reduces the DNA unwinding activity of RNAP, with more effect from the *rrnBP1* promoter compared to the *T7A1* promoter (**Figure 4.16**).

The pleiotropic effect of the DksA can be explained by the difference in DNA duplex stability of the discriminator sequences. The DNA meltability around the G-C rich discriminator from the *rrnBP1* promoter is lower and less likely spontaneously open to dissociate the DksA from RNAP while the A-T rich discriminator from the *T7A1* and the *argI* promoters more likely spontaneously open to dissociate the DksA from RNAP to allow transcript initiation (**Figures 4.2, 4.15B and 4.17**).

5.3.2 Future directions

5.3.2.1 Biochemical assay using RNAP derivatives to test DNA unwinding efficiency

The structures reveal many intriguing conformational changes such as the contacts between the σ^{NCR} and β' CT during the closed and open complexes formations. To validate the structural observation *in vitro*, I obtained the plasmids and purified the RNAP derivatives including 1) deletion of the β' CT, 2) deletion of the acidic region (168-212) of the σ^{NCR} , 3) E148K, E171K, E175K (the β' CT residues contacting the σ^{NCR}) and 4) K299E (the σ^{NCR} residue contacting the β' CT). I will compare the RNAP derivatives with wild-type RNAP for the DNA opening activity using the Cy-3 labeled DNA.

5.3.2.2 Cryo-EM structures of transcription initiation complex and elongation complex

Hochschild group previously proposed the role of the contacts between the σ NCR and β 'CT in promoter escape (Leibman and Hochschild, 2007). To observe the σ NCR motion and other important interactions during the transitions from the early transcript initiation (RPo in this study) to a late transcript initiation (TIC), and an early elongation (EC), I prepared the RNAP-*rrnBP1* promoter complex with the *rrnBP1* promoter DNA having a substitution at +4 position (**Figure 5.7**) and incubated ATP, CTP and UTP (2mM each) for 10 min at 37 °C to solve the structure containing a 8-mer RNA. However, the structure unexpectedly shows that the transcription paused after synthesizing a 6-mer RNA (**Figure 5.8**). It is possible that a higher NTP concentration and a longer incubation time may need to observe the beginning of the σ finger release. Additionally, the base of the +6 position of the DNA may need to be substituted to another base, as the +6C may have a stronger pausing effect than other bases (research talk from the symposium – celebrating 50th anniversary of discovery of σ factor). Another type of promoter that shows less abortive transcription rate or less pausing may be used instead. The analysis of the cryo-EM density map shows two different conformations of the σ NCR, suggesting the σ NCR swings during the transition from the early TIC to the late TIC.

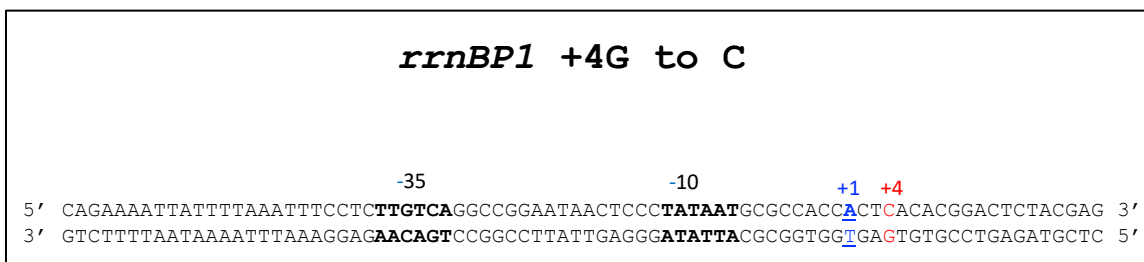


Figure 5.7: The *rrnBP1* promoter DNA with a substitution at +4 position. The transcription start site is underlined and indicated in blue. +4 position is indicated in red. The expected length of RNA synthesized is 8-mer with ATP, CTP, and UTP incubation.

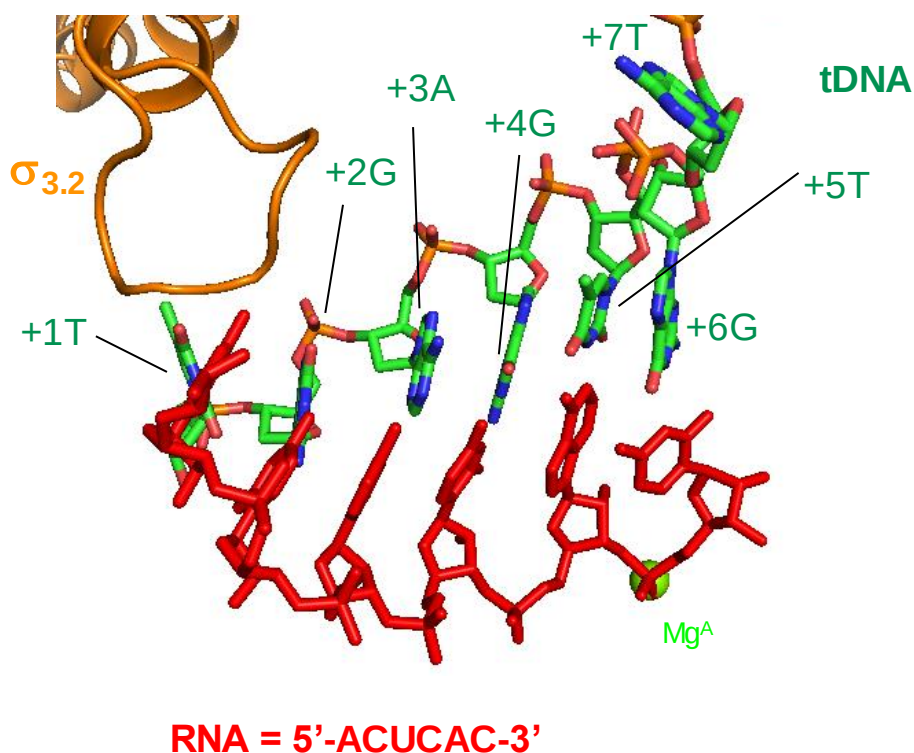


Figure 5.8: Cryo-EM structure of the active site of RNAP containing 6-mer RNA. The tDNA (green) and RNA (red) are shown as a stick model, the σ 3.2 (orange) is shown as a cartoon model and the active site Mg^{2+} is shown as a yellow sphere. The DNA used is shown in Figure 5.5.

5.3.2.3 A RNAP inhibitor ureidothiophene as a control compound in testing open complex formation

Personal communication with Dr. Murakami indicates that Dr. Nikolay Zenkin at Newcastle University, UK identified a compound ureidothiophene (Urd) (**Figure 5. 9A**), which binds the σ region 1.2 and inhibit an open complex formation by allosterically controls the recognition of -10 element by the σ region 2. Therefore, the Urd also may inhibit the σ NCR rotation and can be used in the future as a control compound in testing the σ NCR motion.

I previously soaked the Urd to the *T.thermophilus* RNAP – *pyrG* promoter binary complex crystal to determine the binding mode of the Urd, but I could not find the electron density corresponding to the Urd (**Figure 5. 9B**). I believe the compound is not suitable for the soaking experiment to this particular crystal packing or the Urd does not bind to the RNAP-DNA binary complex crystal that already formed an open complex. Therefore, the Urd (500 μ M in dimethylsulfoxide (DMSO)) was incubated with *E. coli* RNAP (15 μ M) at 37 °C for 30 min followed by the addition of 8 mM CHAPSO and vitrified for the cryo-EM structure determination. The cryo-EM data will be collected and processed in the future.

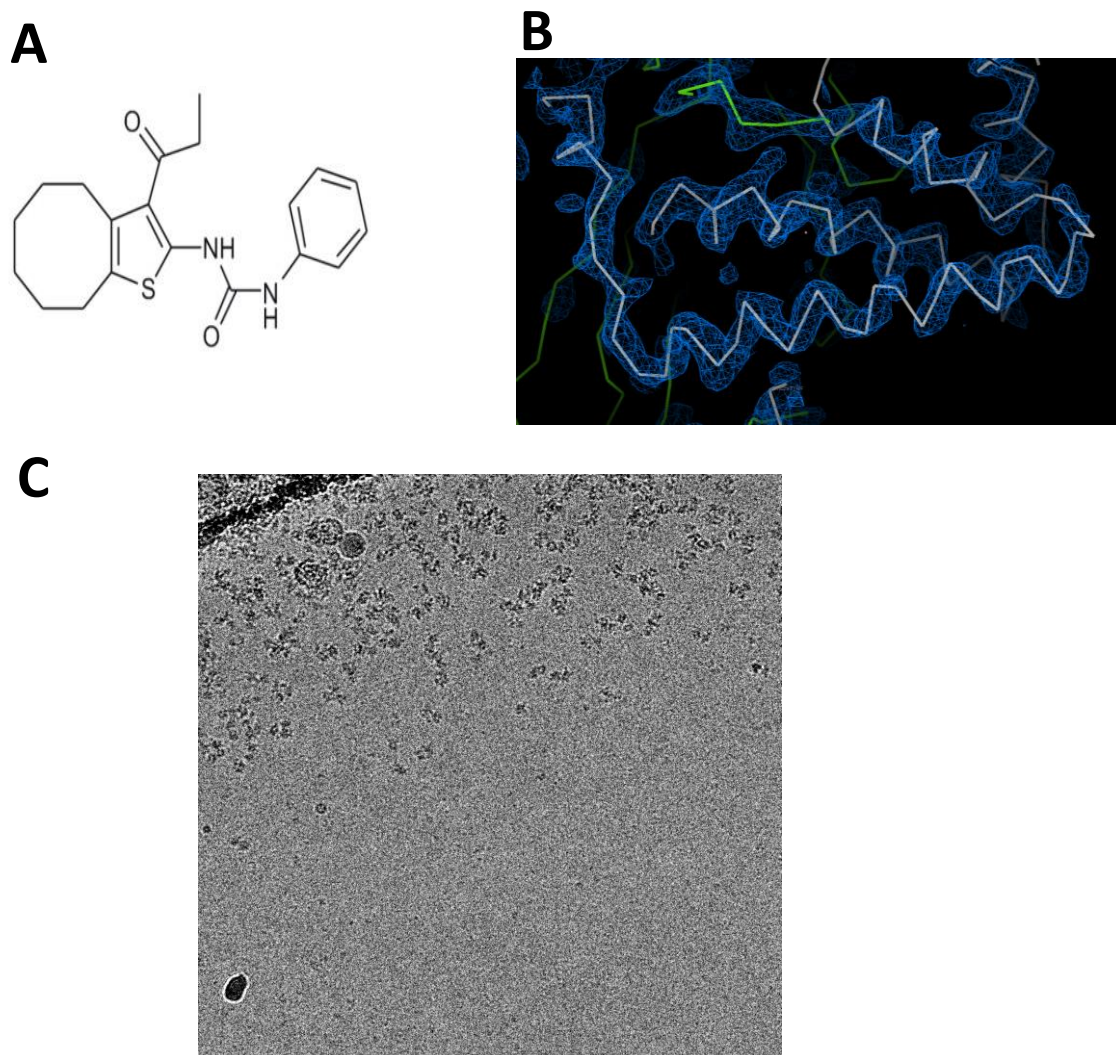


Figure 5.9: Investigation of the binding mode of ureidothiophene (Urd). (A) The structural formula of the Urd. (B) The *T. thermophilus* σ region 1.2 is shown as a white ribbon model. Blue mesh indicates the density of the σ region 1.2. green mesh which indicates the Urd compound binding site was not observed. (C) A representative cryo-EM image of the *E. coli* RNAP σ^{70} holoenzyme – Urd complex on C-Flat Holey Carbon grid (Cu 2/1, 400 mesh).

5.4 Cryo-EM structure of human DNA polymerase δ replication complex.

5.4.1 Human DNA polymerase δ (Pol δ)

The human genome encodes at least 15 DNA-dependent DNA-polymerases. A lagging strand DNA replication in eukaryotic cells is carried out by DNA polymerase δ (Pol δ) with coordinated actions of other proteins such as replication factor C (RFC), proliferating cell nuclear antigen (PCNA) and replication protein A (RPA). In this section, I briefly describe the experimental plan to determine the structure and dynamics of human apo-Pol δ , Pol δ -template/primer complex in the act of DNA synthesis and the holoenzyme replication complex Pol δ – template/primer – PCNA - RPA by cryo-EM.

Pol δ is one of β -family polymerase which is an essential enzyme for a lagging strand DNA replication and repair. Pol δ is highly conserved from yeast to humans with 3 subunits in yeast where 4 subunits in most other organisms including human. Pol δ in *Saccharomyces cerevisiae* is a heterotrimeric complex consists of the Pol3 catalytic subunit (125 kDa) with both polymerase (**Figure 5.10**) and 3'->5' exonuclease proofreading activities (Boulet et al., 1989; Hartwell, 1976; Simon et al., 1991), and Pol31 (55 kDa) and Pol32 (40 kDa) regulatory subunits (Jain et al., 2019). Pol δ in human is a heterotetrameric complex, consists of POLD1(p125) catalytic subunit, and POLD2(p66), POLD3(p50) and POLD(p12) regulatory subunits that interact with proliferating cell nuclear antigen (PCNA), the accessory DNA clamp that enhances the Pol δ processivity (Gulbis et al., 1996)

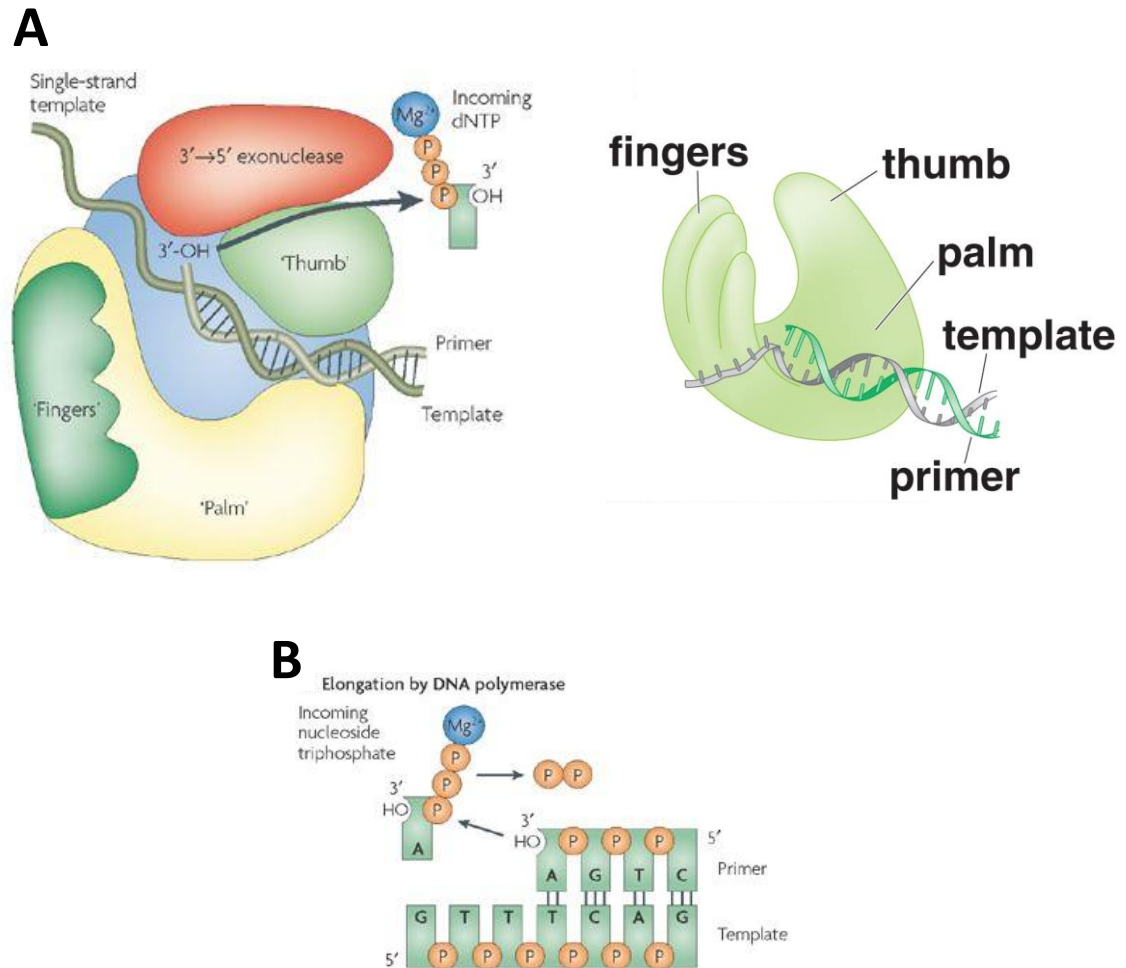


Figure 5.10: DNA polymerase structure and catalysis (A) The overall structure of polymerase domain resembles that of a human right hand forming palm (active site – catalyze the phosphoryl-transfer reaction), finger (interacts with the incoming dNTP and paired template base, and closes at each nucleotide addition step) and thumb (interacts with duplex DNA). (B) the elongation of the primer terminus by DNA polymerase proceeds by nucleophilic attack of the 3'-OH primer terminus on the α -phosphate of incoming dNTP. Figure from the reference (Loeb and Monnat, 2008).

5.4.2 PCNA, RFC, and RPA

PCNA encircles a dsDNA with its homotrimeric ring-shaped structure (molecular mass of each monomer = 29 kDa) and known to stabilize the Pol δ -DNA interaction and thereby enhances the processivity of Pol δ (Gulbis et al., 1996; Mondol et al., 2019) (**Figure 5.11A**). Replication factor C (RFC) is consists of one large subunit (145 kDa) and four small subunits (40, 38, 37 and 36 kDa) and plays a role of loading the PCNA onto a primer/template (P/T) junction with its DNA-dependent ATPase activity (Hedglin et al., 2013) (**Figure 5.11B**). Replication protein A (RPA) is a heterotrimeric protein (RPA70, RPA32, and RPA14) that binds a single-stranded DNA (ssDNA) with the high affinity (< pM in physiological condition) (**Figure 5.11C**). The RPA protects the ssDNA from nuclease degradations and prevents a hairpin formation. The RPA covers a range of 22-30 nt of the ssDNA (Bochkareva et al., 2002). The RPA also coordinates lagging strand polymerase switch from Pol α -primase to Pol δ (Yuzhakov et al., 1999), the sequential action of the DNA and FEN1 flap endonuclease that remove the RNA primer during Okazaki fragment maturation (Bae et al., 2001) and prohibits diffusion of the PCNA sliding clamp along the ssDNA (Hedglin and Benkovic, 2017).

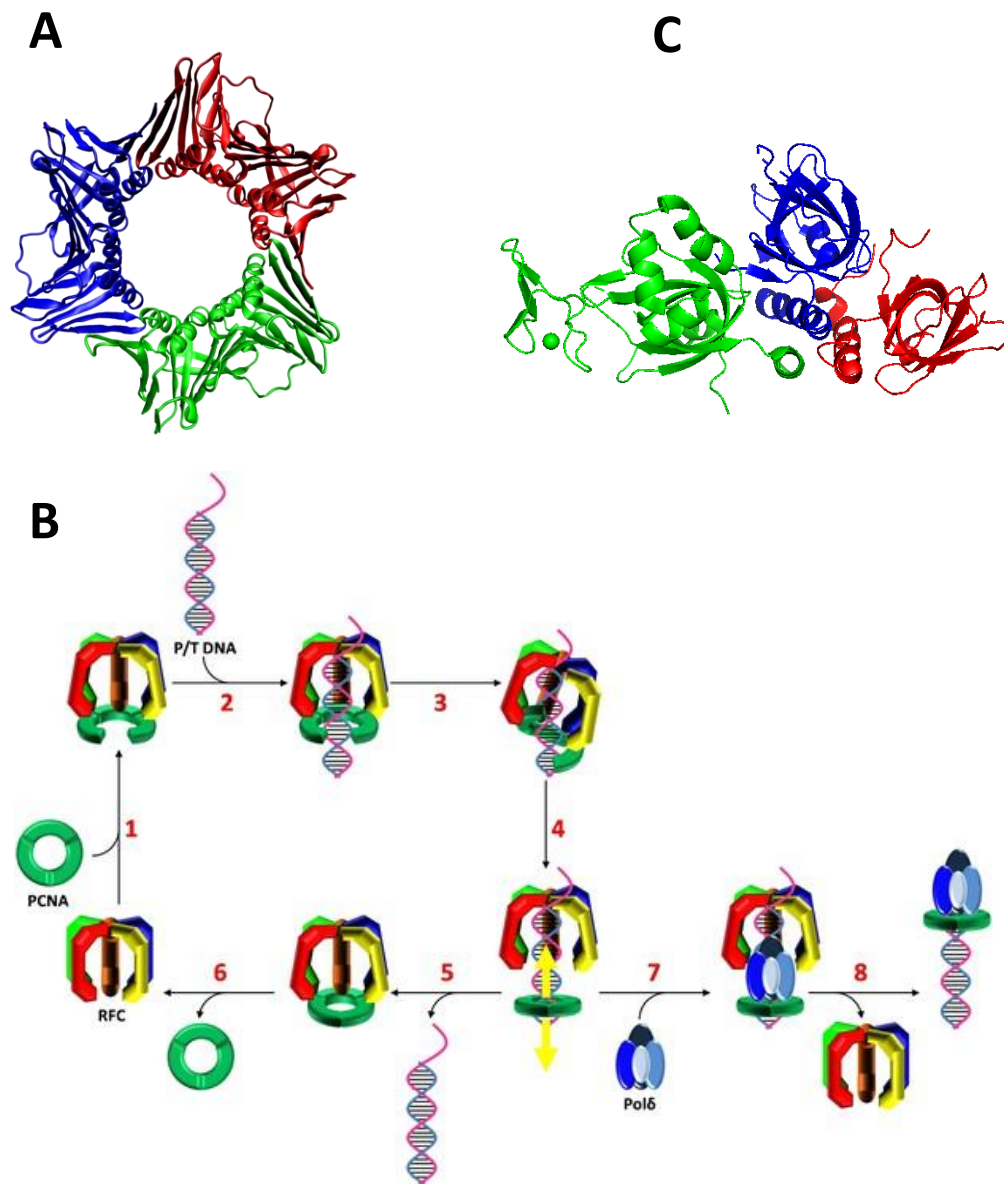


Figure 5.11: PCNA, RFC, and RPA. (A) The structure of homotrimer PCNA. Each monomer is shown in a different color. Figure from the reference (Gulbis et al., 1996, PDB:1AXC) (B) The scheme of PCNA loading to Primer/template (P/T) junction by RFC. Figure from the reference (Hedglin et al., 2013) (C) The structure of heterotrimer RPA. Green = RPA70, blue = RPA32, red = RPA14 subunits. Figure from the reference (Bochkareva et al., 2002, PDB: 1L1O).

5.4.3 Cryo-EM structure of yeast DNA polymerase δ (Pol δ)

Aggarwal group at Icahn School of Medicine at Mount Sinai recently determined the structure of *Saccharomyces cerevisiae* Pol δ (Jain et al., 2019) (**Figure 5.12**). The structure shows that N-terminally located Pol3 catalytic subunit engaging template-primer DNA duplex is connected by two cysteine-rich metal-binding modules (CysAD and CysBD, containing a 4Fe-4S cluster) at its C-terminal to regulatory subunit Pol31. The C-terminal of Pol31 is connected to the N-terminal of Pol32. C-terminal of Pol32 has a PCNA interaction motif (PIP) that aid in anchoring the Pol δ holoenzyme to PCNA for increased processivity. The density map of the Pol δ holoenzyme is weak for parts of Pol31 and Pol32, indicating the structural flexibility of the regulatory subunits located away from the catalytic subunit. Through multi-body refinement with the catalytic module and regulatory module as discrete bodies, the resolution improved for regions Pol31 and Pol32, which they were able to model these subunits. A previous small-angle X-ray scattering analysis also showed the flexibility between these two modules (Jain et al., 2009). Based on the dynamics of these subunits, they proposed a model of yeast Pol δ -PCNA complex which PIP motif of C-terminal of Pol32 swings toward PCNA upon PCNA bound to a duplex DNA (**Figure 5.13**).

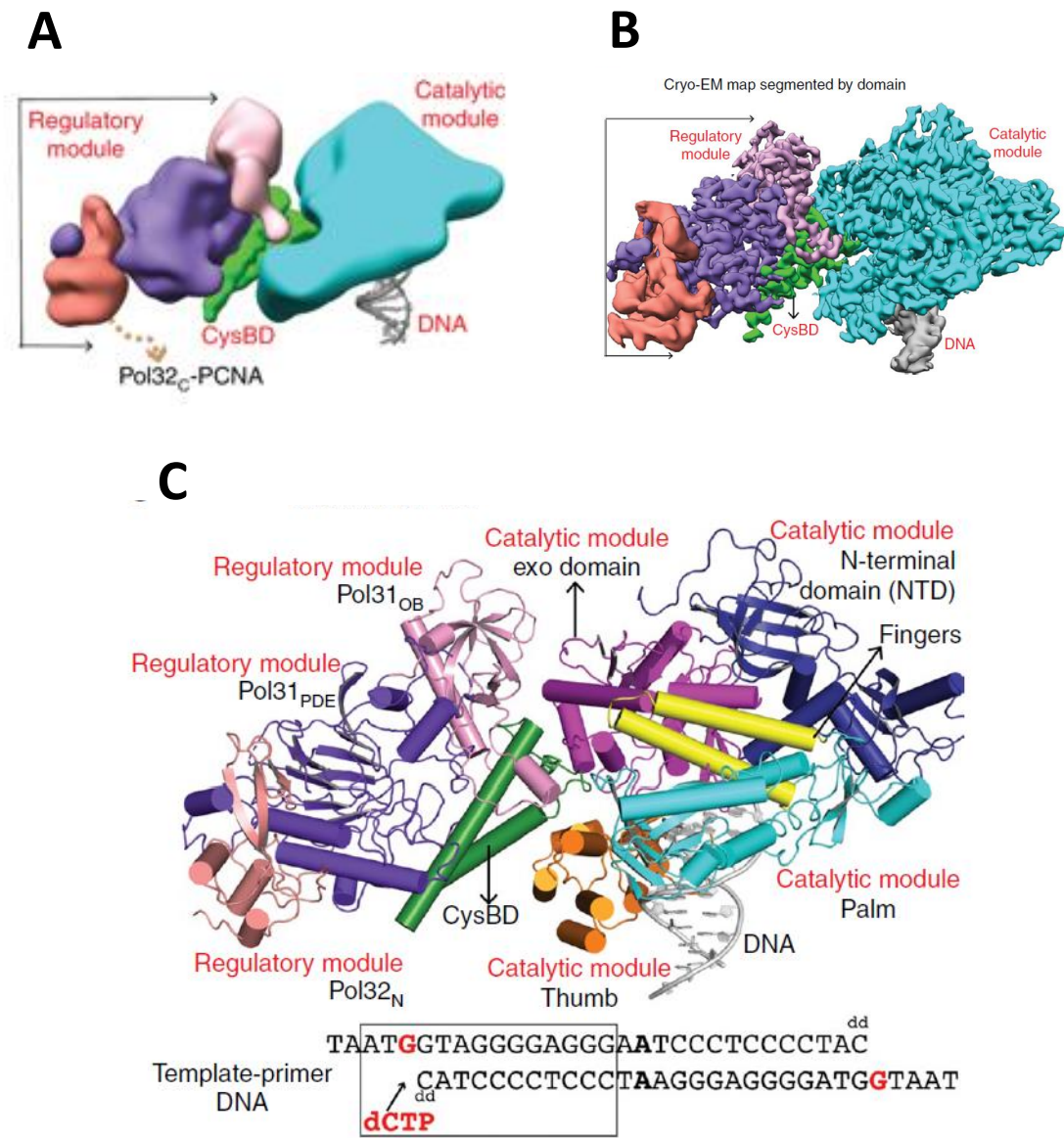


Figure 5.12: The cryo-EM structure of DNA bound yeast Pol δ holoenzyme. (A-B) Schematic of the structure of *S. cerevisiae* Pol δ . Blue indicates the largest catalytic subunit Pol3, purple and violet indicate regulatory subunits Pol31 and Pol32, green and pink indicate CysAD and CysBD modules which connect Pol3 to Pol31. (C) Structure of yeast Pol δ colored by domain. The template, primer and incoming dCTP used in this study is shown. Figure from the reference (Jain et al., 2019).

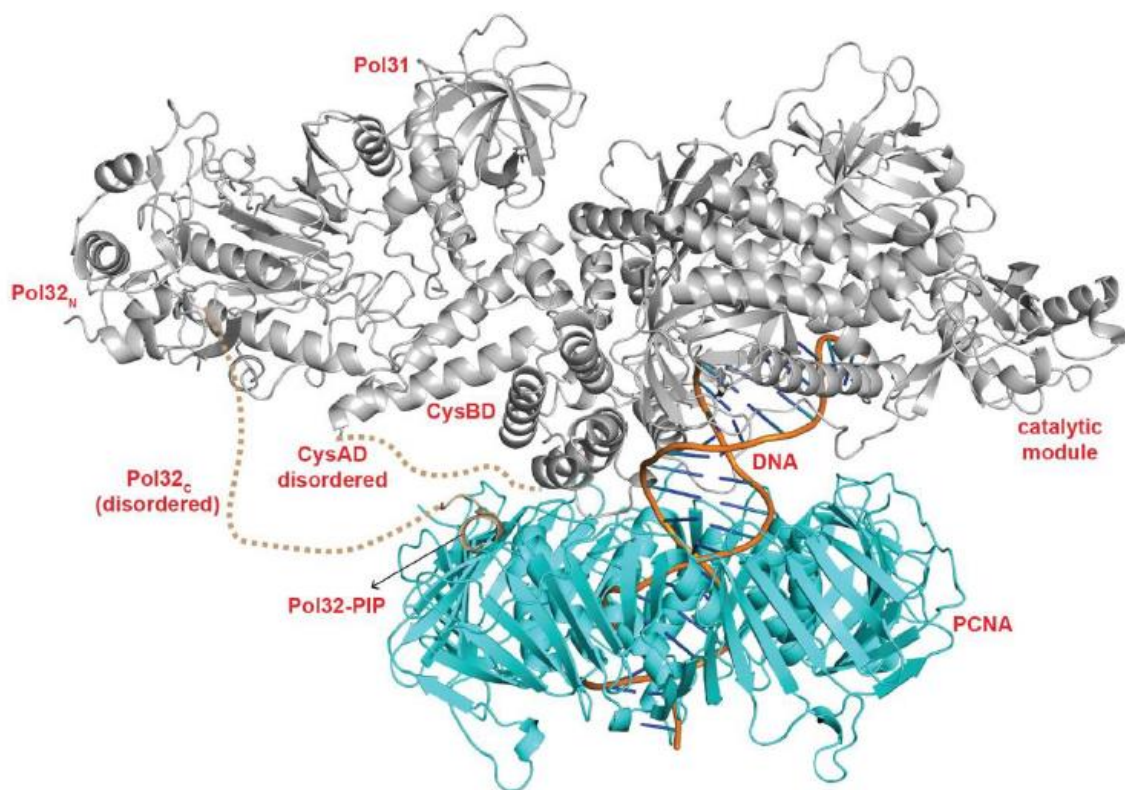


Figure 5.13: Proposed model of yeast Pol δ -PCNA complex. Yeast Pol δ from this study (Jain et al., 2019) was modeled with PCNA (PDB: 2OD8). The model was built based on the flexibility (weak density) of the PIP domain of Pol32 and CysAD domain. Figure from the reference (Jain et al., 2019).

5.4.4 Materials and Methods

5.4.4.1 Purification of human DNA polymerase δ (Pol δ)

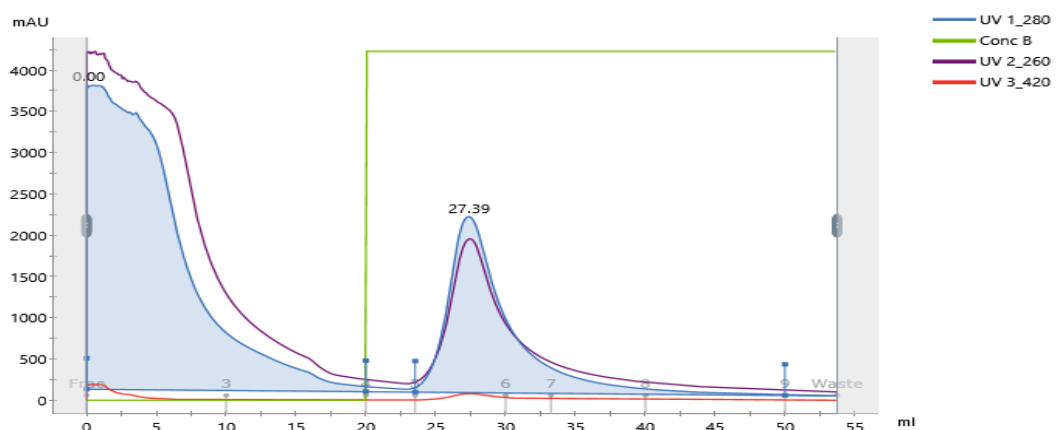
All human DNAP δ subunits were expressed in BL21-CodonPlus (DE3)-RP cells co-transformed with both pET-POLD4/1 and pGBM-POLD2/3. Cells were cultured by using a 100 L fermentor at the Penn State fermentation facility yielding about 600 g of cell paste. Briefly, cells

were grown in TB media supplemented with ampicillin (250 µg/ml), streptomycin (20 µg/ml) and chloramphenicol (30 µg/ml) at 37 °C until OD₆₀₀ reaches about 1.0. The temperature was reduced to 18 °C and maintained for at least 1 hr. Ferric ammonium citrate (0.5 mM), Cysteine (0.5 mM) and IPTG (0.2 mM) were added and cells were continued to grow for 15 hr before harvesting. About 50 g of cell paste was suspended in 117 ml lysis buffer (50 mM HEPES-NaOH, pH7.5 at R.T), 500 mM NaCl, 0.1 mM EDTA, 10 mM B-ME, 1 mM PMSF, protease inhibitor cocktail set) and lysed by sonication (5 cycles of 1 min continuous sonication, cool down between cycles). The lysate was centrifuged at 15,000 rpm/4 °C/15 min. To the supernatant, imidazole was added to the final concentration of 5 mM followed by ice-cold 10% (w/v) streptomycin sulfate drop-wise to 1% final. The solution was stirred in a cold room for 30 min and centrifuged at 20,000 rpm/4 °C/30 min. The supernatant was applied to 5 ml Ni Column equilibrated with buffer A /5 mM imidazole (50 mM HEPES-NaOH, pH7.5 at R.T), 0.5 M NaCl, 10% glycerol, 10 mM B-ME) and eluted with buffer A / 100 mM imidazole (**Figure 5.14A**)

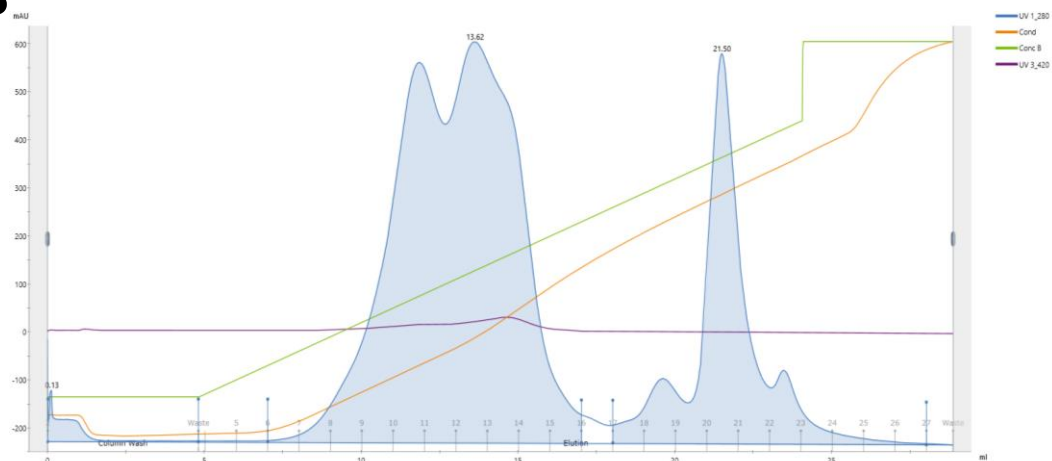
The eluted sample was diluted to 0.1 M NaCl with buffer B (50 mM HEPES-NaOH, pH7.5 at R.T, 10% glycerol, 10 mM B-ME) and applied to 1ml Q HP column equilibrated with buffer B/0.1 M NaCl. Protein was eluted by salt gradient (0.1M – 0.8M NaCl, 20 cv) (**Figure 5.14B**). The fractions containing Polδ (9-11) were combined (**Figure 5.14C**) and applied to the SD200 size-exclusion column. The fractions containing Polδ (13-17) were combined (**Figure 5.14D-E**). For an unknown reason, I could not completely get rid of one contaminant protein showing with the Polδ p50 subunit even after the gel filtration column (**Figure 5.14E**). Mass spectrometry analysis shows this protein is bifunctional polymyxin resistance protein ArnA OS in *E. coli*, which is irrelevant (or could possibly be overproduced from the presence of 3 antibiotics – stress response?). Therefore, the sample was diluted to 0.1 M NaCl with buffer B and applied to a 1ml heparin column equilibrated with buffer B/0.1M NaCl. Protein was eluted by stepwise elution at buffer B / 0.6 M NaCl. (**Figure 5.14F**). SDS-PAGE analysis indicates the contaminant

band was separated from Pol δ (**Figure 5.14G**). The Pol δ was concentrated to 5 mg/ml and saved at -80 °C.

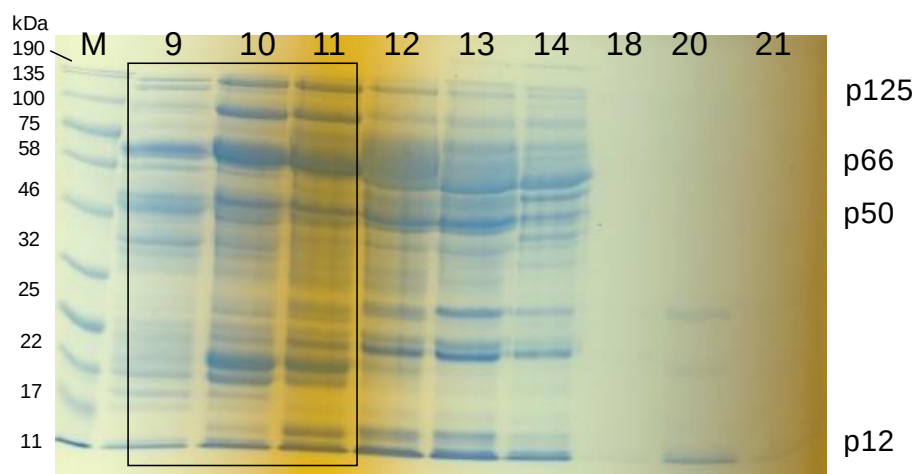
A



B



C



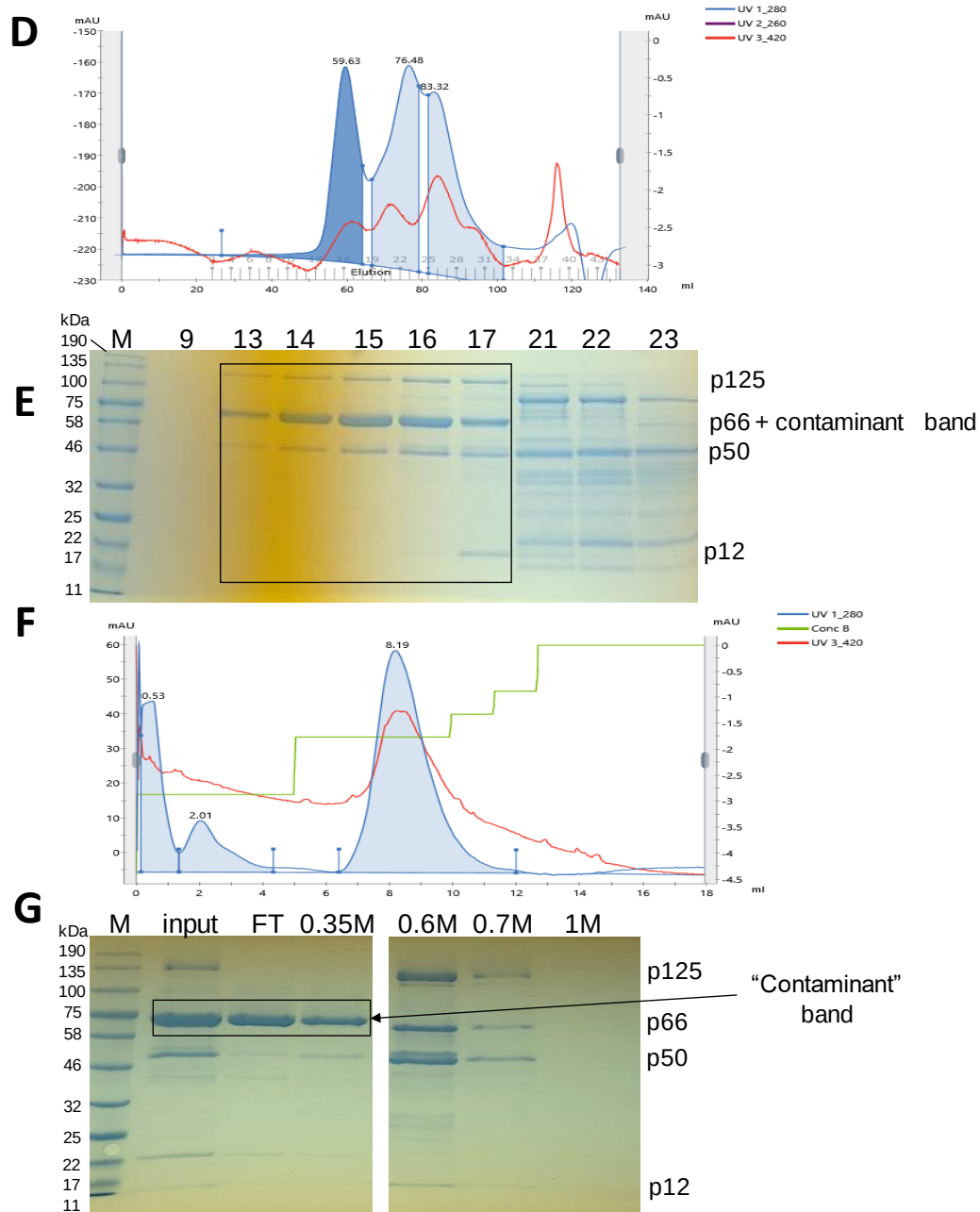


Figure 5.14: Purification of human Pol δ . (A) 5ml Ni column chromatography, (B) 5ml Q HP column chromatography (C) SDS-PAGE gel of the main fractions from Q HP column, (D) 120 ml Size-exclusion column chromatography, (E) SDS-PAGE gel of the main fractions size-exclusion column, (F) Heparin column chromatography, (G) SDS-PAGE gel of the main fractions from heparin column.

5.4.4.2 Vitrification of cryo-EM grids

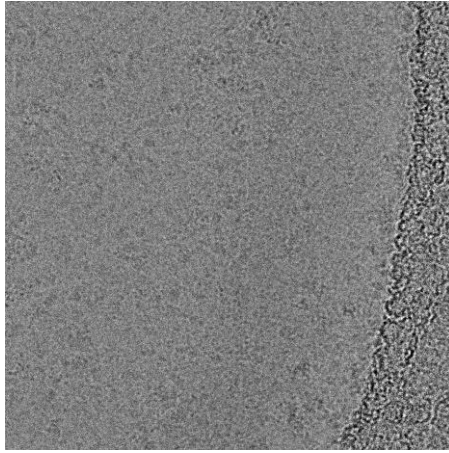
The first set of grids were prepared with 4.5 mg/ml and 3 mg/ml (**Figure 5.15A**) of Pol δ without decreasing the salt concentration (0.6 M NaCl) after the elution from the heparin column. Glycerol was removed and CHAPSO was not used. 3.5 μ L sample was applied to a glow-discharged C-Flat Holey Carbon grid (Cu 2/1, 400 mesh), blotted and plunge-frozen in liquid ethane using Vitrobot Mark IV (FEI, USA) at 95% humidity and 4°C. The screening of both grids indicates that both concentrations look a bit high. The particles do not look as clear as *E. coli* RNAP, possibly due to the high salt concentration and/or smaller size of particles. The first cryo-EM data was collected with a grid containing 3 mg/ml of Pol δ .

The second set of grids were prepared with 1.5 mg/ml of Pol δ with five-fold decreased salt concentration to 0.12M NaCl with the 2% final glycerol concentration. Grids were prepared and compared without (**Figure 5.15B**) or with CHAPSO (**Figure 5.15C**). Lowering the protein and salt concentrations and using CHAPSO helped in stabilizing and visualizing the particles.

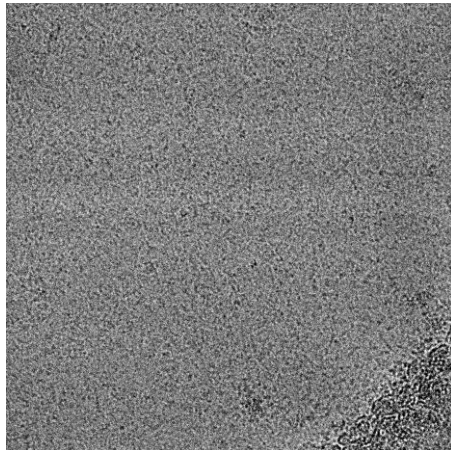
5.4.4.3 Cryo-EM data collection

Data was collected using the Titan Krios (Thermo Fisher) microscope equipped with a Falcon II direct electron detector (Gatan) at the Penn State Cryo-EM facility with the software EPU (ThermoFisher Scientific). Each sample grid was imaged at 300 kV, with a defocus range of 1.0 – 3.0 μ m. The magnification was 75,000X in electron counting mode (pixel size = 0.88 $\text{\AA}$ /pixel) with a total dose of 45 electrons / $\text{\AA}^2$ that is 1.02 electrons / frame / $\text{\AA}^2$.

A 3 mg/ml (600 mM NaCl, no glycerol)
- CHAPSO



B 1.5 mg/ml (120 mM NaCl, 2% glycerol)
- CHAPSO



C 1.5 mg/ml (120 mM NaCl, 2% glycerol)
+ CHAPSO (8 mM)

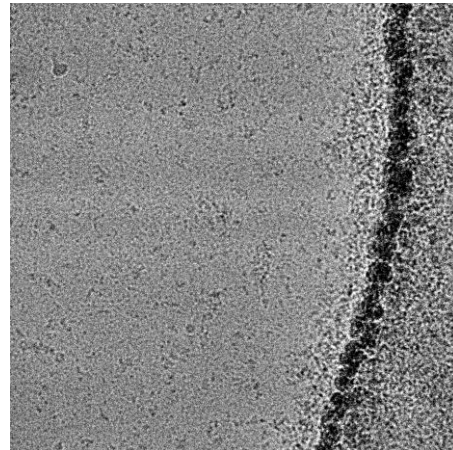


Figure 5.15: Representative images of human Pol δ holoenzyme on C-Flat Holey Carbon grids (Cu 2/1, 400 mesh). (A) 3 mg/ml (600 mM NaCl, no glycerol, no CHAPSO), (B) 1.5 mg/ml (120 mM NaCl, 2% glycerol, no CHAPSO), (C) 1.5 mg/ml (120 mM NaCl, 2% glycerol with 8 mM CHAPSO).

5.4.4.4 Processing cryo-EM data

Data were processed using cryoSPARC V2.9.0. From the collected movies, usable movies were aligned, dose weighted, and CTF fitted using Patch-Motion correction and Patch CTF. Initially, about 1000 particles were manually picked to generate a template followed by autopicking. The autopicked particles were 2D classified to approximately 10 classes and good 2D classes were selected for the subsequent 3D classification. After an initial 3D classification, heterogeneous and homogeneous refinements were performed to generate a cryo-EM density map.

5.4.5 Results

The low-resolution map ($\sim 10\text{\AA}$) was generated from the small dataset (**Figure 5.16**). The density map fits well with the previously published yeast Pol δ structure (PDB: 6P1H) (Jain et al., 2019). Some regions of the density map are not well-resolved, indicating structural flexibility. The multi-body refinement (catalytic subunit / regulatory subunit) will be performed to improve the quality of the map.

Yeast Pol δ

Human Pol δ (this study)

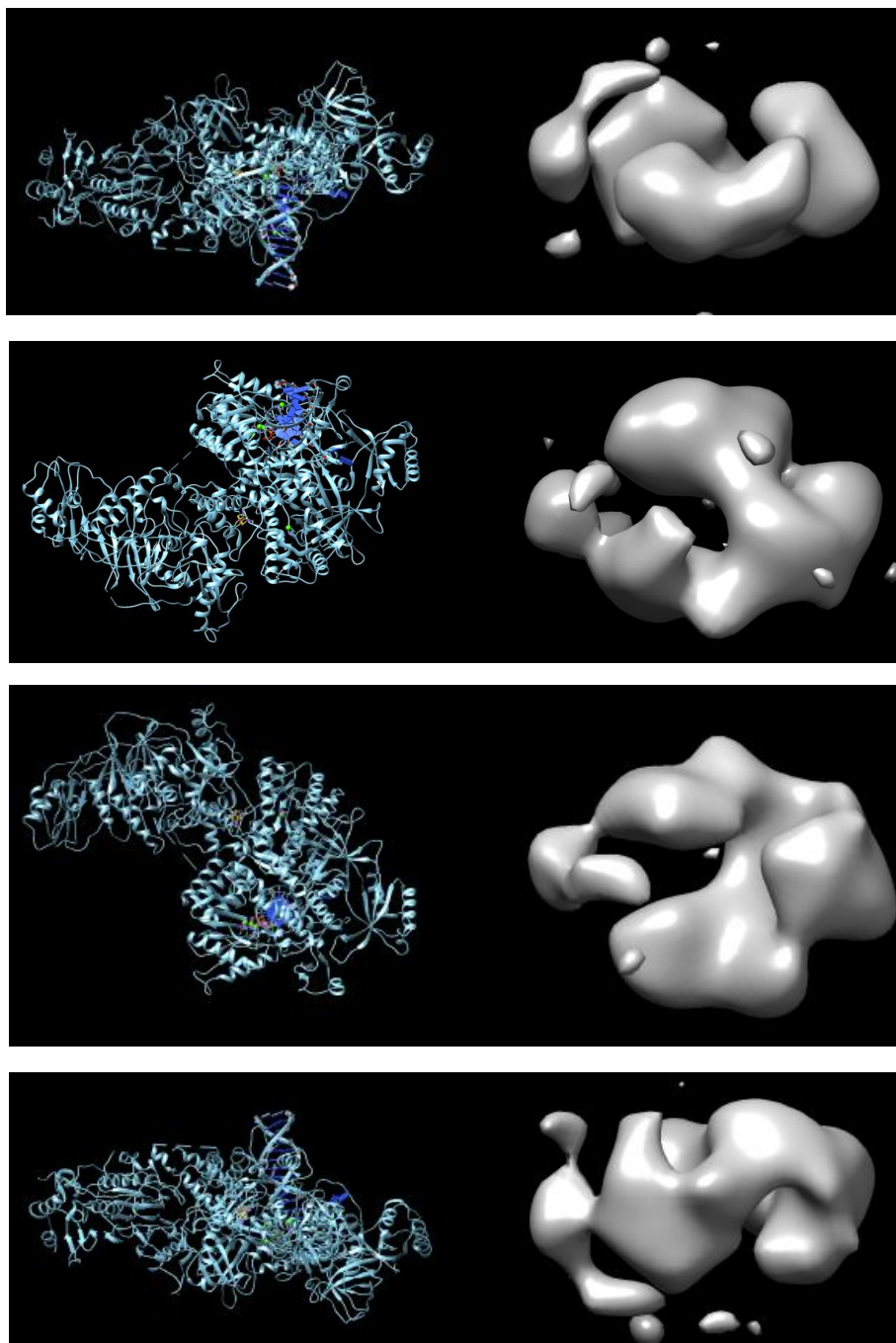


Figure 5.16: The yeast Pol δ -DNA complex structure (left, PDB: 6P1H) and human Pol δ density map at 10 Å resolution (right, this study). The overall density map fits well with the published yeast Pol δ structure.

5.4.6 Future directions

The cryo-EM structure of apo-form human Pol δ will be determined using the grid prepared with 1.5 mg/ml concentration sample (120 mM NaCl, 2% glycerol, 8 mM CHAPSO). To study the human Pol δ in the act of DNA synthesis, the Pol δ -primer/template complex (**Figure 5.17**) will be prepared as well using a known template DNA-primer sequences that stabilize the complex. After determining the Pol δ -P/T complex structure, the holoenzyme will be prepared by adding PCNA and RPA to the Pol δ -DNA complex.

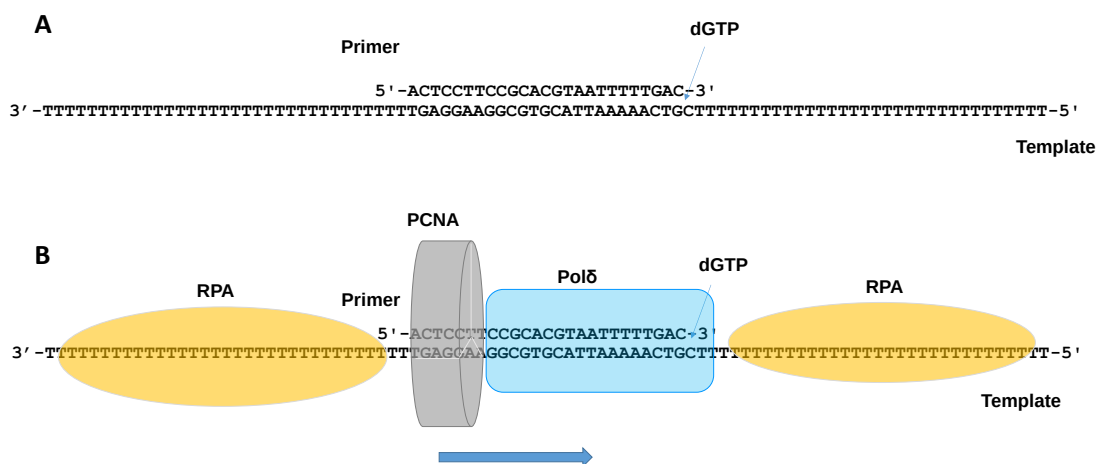


Figure 5.17: Schematic of Pol δ replication complex preparation (A) The sequence of the template DNA and the primer for preparing the Pol δ -primer/template complex. (B) The replication complex containing Pol δ -DNA-PCNA and RPAs.

5.5 Acknowledgements

I thank Carol Bator and Joseph Sunghyun Cho at the Penn State Huck Institute Cryo-EM facility for supporting the cryo-EM data collections. I thank Dr. Mark Hedglin at Penn State for providing me the plasmids pET-POLD4/1 and pGBM-POLD2/3. I thank Ashik Sathish at Penn State Fermentation Facility for supporting fermentation of cells and Tatiana Laremore at Penn State Proteomics and Mass Spectrometry Core Facility for mass spectrometry analysis.

5.6 References

- Bae, S.H., Bae, K.H., Kim, J.A., and Seo, Y.S. (2001). RPA governs endonuclease switching during processing of Okazaki fragments in eukaryotes. *Nature* 412, 456-61.
- Bremer, H., and Dennis, P.P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus* 3.
- Bochkareva, E., Korolev, S., Lees-Miller, S.P., and Bochkarev, A. (2002). Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J* 21, 1855-1863.
- Boulet, A., Simon, M., Faye, G., Bauer, G.A., and Burgers, P.M. (1989). Structure and function of the *Saccharomyces cerevisiae* CDC2 gene encoding the large subunit of DNA polymerase III. *EMBO J* 8, 1849-1854.
- Cashel, M., and Gallant, J. (1969). Two compounds implicated in the function of the RC gene of *Escherichia coli*. *Nature* 221, 838-841.

Gulbis, J.M., Kelman, Z., Hurwitz, J., O'Donnell, M., and Kuriyan, J. (1996). Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* 87, 297-306.

Hartwell, L.H. (1976). Sequential function of gene products relative to DNA synthesis in the yeast cell cycle. *J Mol Biol* 104, 803-817

Haugen, S.P., Berkmen, M.B., Ross, W., Gaal, T., Ward, C., and Gourse, R.L. (2006). rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* 125, 1069-1082.

Haugen, S.P., Ross, W., and Gourse, R.L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nat Rev Microbiol* 6, 507-519.

Hauryliuk, V., Atkinson, G.C., Murakami, K.S., Tenson, T., and Gerdes, K. (2015). Recent functional insights into the role of (p)ppGpp in bacterial physiology. *Nat Rev Microbiol* 13, 298-309.

Hedglin, M., and Benkovic, S.J. (2017). Replication protein A prohibits Diffusion of the PCNA sliding clamp along single-stranded DNA. *Biochemistry* 56, 1824-1835.

Hedglin, M., Perumal, S.K., Hu, Z., and Benkovic, S.J. (2013). Stepwise assembly of the human replicative polymerase holoenzyme. *eLife* 2, e00278.

Jain, R., Rice, W.J., Malik, R., Johnson, R.E., Prakash, L., Prakash, S., Ubarretxena-Belandia, I., and Aggarwal, A.K. (2019). Cryo-EM structure and dynamics of eukaryotic DNA polymerase delta holoenzyme. *Nat Struct Mol Biol* 26, 955-962.

Leibman, M., and Hochschild, A. (2007). A sigma-core interaction of the RNA polymerase holoenzyme that enhances promoter escape. *EMBO J* 26, 1579-1590.

Liu, C., Heath, L.S., and Turnbough, C.L., Jr. (1994). Regulation of *pyrBI* operon expression in *Escherichia coli* by UTP-sensitive reiterative RNA synthesis during transcriptional initiation. *Genes Dev* 8, 2904-2912.

Loeb, L.A., and Monnat R,J., Jr. (2008). DNA polymerases and human disease. *Nat Rev Genet* 9, 594-604.

Meng, Q., Turnbough, C.L., Jr., and Switzer, R.L. (2004). Attenuation control of *pyrG* expression in *Bacillus subtilis* is mediated by CTP-sensitive reiterative transcription. *Proc Natl Acad Sci U S A* 101, 10943-10948.

Molodtsov, V., Sineva, E., Zhang, L., Huang, X., Cashel, M., Ades, S.E., and Murakami, K.S. (2018). Allosteric Effector ppGpp Potentiates the Inhibition of Transcript Initiation by DksA. *Mol Cell* 69, 828-839 e825.

Mondol, T., Stodola, J.L., Galletto, R., and Burgers, P.M. (2019). PCNA accelerates the nucleotide incorporation rate by DNA polymerase delta. *Nucleic Acids Res* 47, 1977-1986.

Murakami, K.S., Shin, Y., Turnbough, C.L., Jr., and Molodtsov, V. (2017). X-ray crystal structure of a reiterative transcription complex reveals an atypical RNA extension pathway. *Proc Natl Acad Sci U S A* 114, 8211-8216.

Nomura, M. (1999). Regulation of ribosome biosynthesis in *Escherichia coli* and *Saccharomyces cerevisiae*: diversity and common principles. *J Bacteriol* 181, 6857-6864.

Paul, B.J., Barker, M.M., Ross, W., Schneider, D.A., Webb, C., Foster, J.W., and Gourse, R.L. (2004a). DksA: a critical component of the transcription initiation machinery that potentiates the regulation of rRNA promoters by ppGpp and the initiating NTP. *Cell* 118, 311-322.

Paul, B.J., Ross, W., Gaal, T., and Gourse, R.L. (2004b). rRNA transcription in *Escherichia coli*. *Annu Rev Genet* 38, 749-770.

Potrykus, K., and Cashel, M. (2008). (p)ppGpp: still magical? *Annu Rev Microbiol* 62, 35-51.

Simon, M., Giot, L., and Faye, G. (1991). The 3' to 5' exonuclease activity located in the DNA polymerase delta subunit of *Saccharomyces cerevisiae* is required for accurate replication. *EMBO J* 10, 2165-2170.

Shin, Y., Hedglin, M., and Murakami, K.S. (2020). Structural basis of reiterative transcription from the *pyrG* and *pyrBI* promoters by bacterial RNA polymerase.

Turnbough, C.L. (2011). Regulation of gene expression by reiterative transcription. *Current Opinion in Microbiology* 14, 142-147.

Turnbough, C.L., Jr., and Switzer, R.L. (2008). Regulation of pyrimidine biosynthetic gene expression in bacteria: repression without repressors. *Microbiol Mol Biol Rev* 72, 266-30.

Yuzhakov, A., Kelman, Z., Hurwitz, J., and O'Donnell, M. (1999). Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. *EMBO J* 18, 6189-99.

Appendix A

Fermentation of *T. thermophilus* HB27 cells

30 L Micros microbial fermentor (M30) at the fermentation facility was sterilized according to the protocol (SOP: FER30-02). Previously saved bacteria glycerol stock *T. thermophilus* strain HB27 cells (batch 2011) were grown as a seed culture in 500 ml media containing 8g/L tryptone, 4g/L yeast extract, 2g/L NaCl, 1ml/L of 0.35 M CaCl₂, 1ml/L of 0.4 M MgSO₄ and 2ml/L of 2M NaOH at 60 °C / 100 rpm / 26 hours. The next day, the seed culture was transferred to the M30 containing 20 L of media and grown for an additional 16 hours at 65 °C (OD600 ~ 1.5). Cells were harvested using the Sharples T-1-P laboratory presuritized centrifuge according to the procedure (SOP: EQP-012). It is important to note that cell pellets should be yellowish - mustard color to purify enough amount of RNAP. Pinkish cell pellet, which indicates overgrowth did not produce enough amount of RNAP.

Appendix B

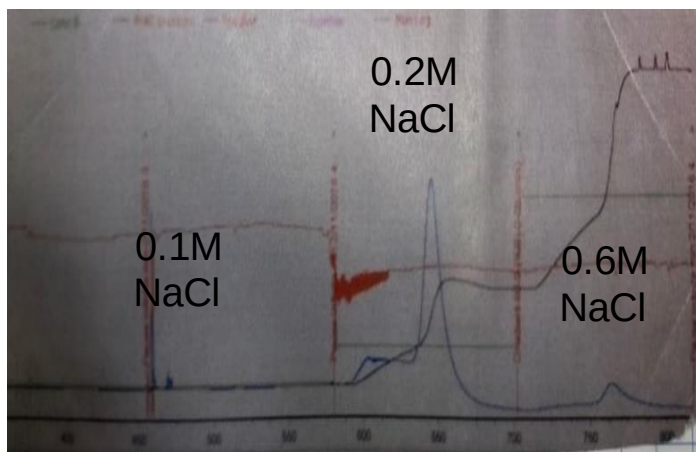
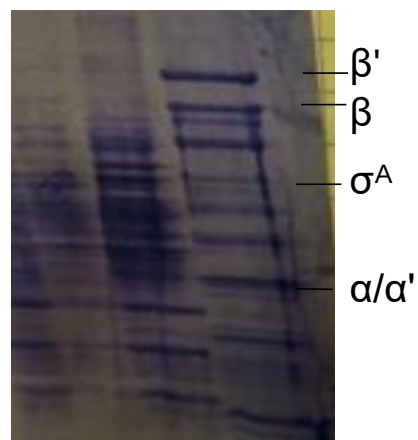
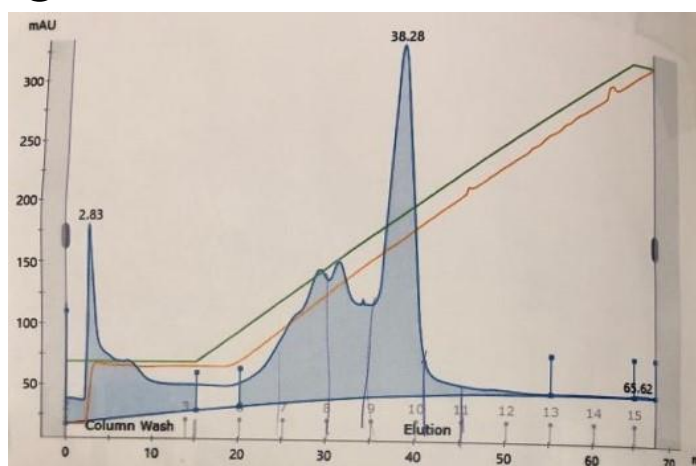
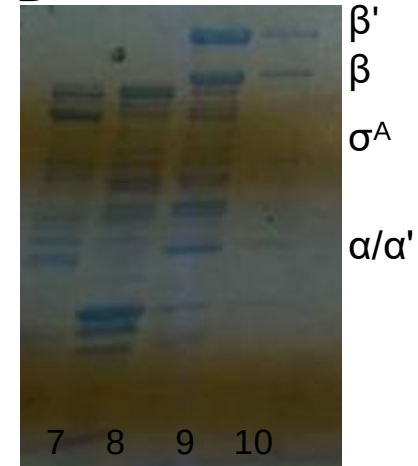
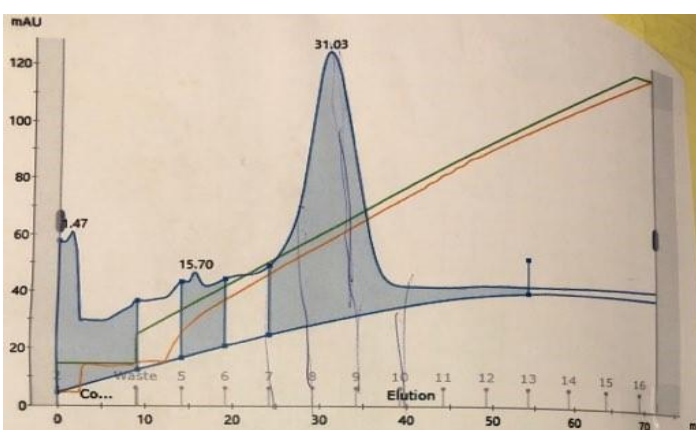
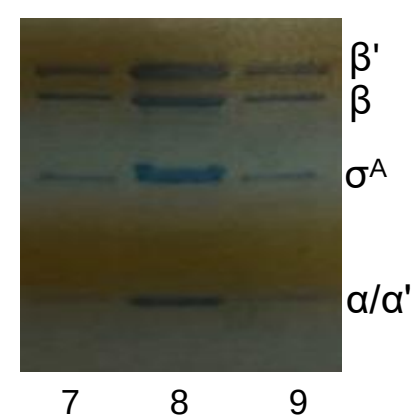
Purification of *T. thermophilus* RNAP σ^A holoenzyme

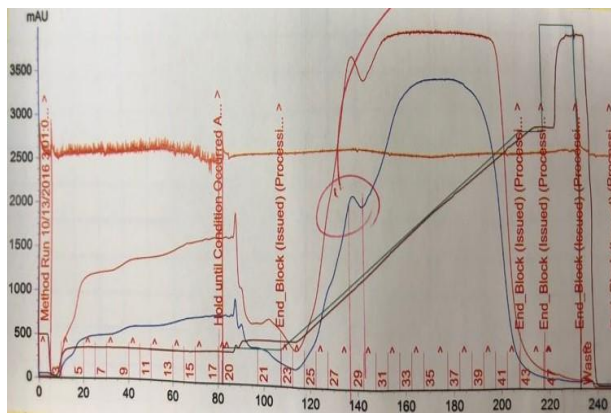
T. thermophilus HB27 cells were cultured by using a 30 L fermentor at the Penn State fermentation facility yielding about 100 g of cell paste. 100 g of cell paste was suspended in 300 ml of lysis buffer [40 mM Tris-HCl, pH8 at 4 °C, 100 mM NaCl, 10 mM EDTA, 10 mM B-ME (mercaptoethanol), 1 mM PMSF (phenylmethylsulfonyl fluoride), 1 protease inhibitor cocktail set) and the cells were lysed using microfluidizer with pressure $\geq 20,000$ psi. The lysate was centrifuged at 13,000 rpm/4 °C/1hr. To the supernatant, glycerol was added to a final concentration of 5% followed by polymine P to a final concentration of 0.5%. The lysate was kept stirring at 4 °C/15min and centrifuged 13,500 rpm/4 °C/20 min. The supernatant was discarded and 100 ml wash buffer (40 mM Tris-HCl, pH8 at 4 °C, 200 mM NaCl, 1 mM EDTA, 5% glycerol, 10 mM B-ME, 1mM PMSF) was added to each bottle to wash pellet (resuspend pellet and keep on ice for 10 min followed by centrifugation for 20 min, discard supernatant). 50 ml of extraction buffer (40 mM Tris-HCl, pH8 at 4 °C, 800 mM NaCl, 1 mM EDTA, 5% glycerol, 10 mM B-ME, 1mM PMSF) was added to each bottle and kept on ice for 10 min followed by centrifugation 15 min. The supernatant was collected and protein was precipitated with 45% saturated ammonium sulfate at 4 °C for 16 hours. The next day, the slurry was centrifuged at 14,000 rpm /4 °C/45min. The supernatant was discarded and the pellet was resuspended in 200 ml TGED buffer (10 mM Tris-HCl, pH8 at 4 °C, 5% glycerol, 0.1 mM EDTA, 1 mM DTT) with 60% ammonium sulfate. The precipitate was spun down at 14,000 rpm/4 °C/15min. The supernatant was discarded and the pellet was dissolved in roughly 600 ml TGED buffer to the conductivity of 0.1 M NaCl. The sample was applied to the heparin column (50 ml) equilibrated with TGED/0.1 M NaCl buffer. After washing the column with TGED/0.2 M

NaCl buffer until the background becomes plateau, the protein was eluted with TGED/0.6 M NaCl (**Appendix B Figure A and B**). The sample was dialyzed for 16 hours at 4 °C in TGED buffer, and the sample was applied to the Source 15Q column equilibrated with TGED/0.1M NaCl. The protein was eluted with a salt gradient (**Appendix B Figure C and D**). The fractions containing core subunits (9 and 10) were dialyzed in TGED buffer for 4 hours and diluted to 0.05 M NaCl. The sample was applied to the Source 15S column equilibrated with TGED/0.05 M NaCl. The fractions containing core subunits (7 – 9) were combined and concentrated (**Appendix B Figure E and F**).

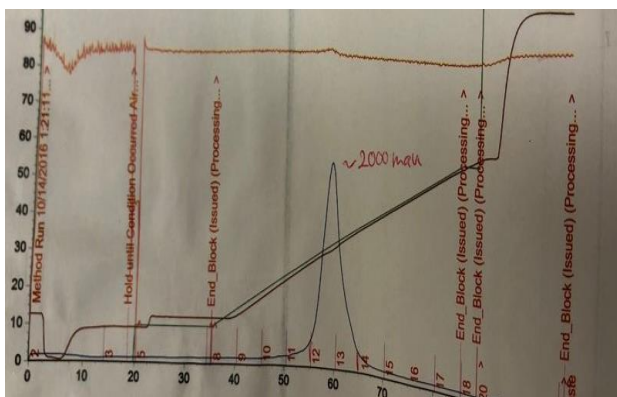
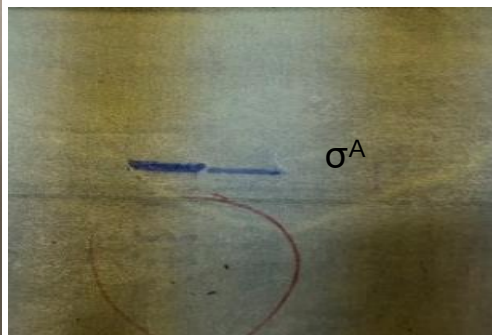
T. thermophilus SigA was expressed in BL21 (DE3) cells transformed with pET21a-TthsigA. Cells were grown in 1.5 L LB media containing 100 µg/ml ampicillin at 37 °C until OD₆₀₀ = 0.5, induced with 0.5 mM IPTG and grown for 3 hours at 37 °C. About 20 g of cell paste was suspended in 30 ml of lysis buffer (50 mM Tris-HCl, pH8 at 4 °C, 0.2 M NaCl, 1 mM EDTA, 5% glycerol, 2 mM DTT, 1 mM PMSF) and lysed by sonication. The lysate was centrifuged at 12,000 rpm for 10 mins followed by heating the supernatant in 65 °C water bath for 30 mins. The sample was centrifuged at 12,000 rpm for 10 mins and the supernatant (diluted twofold to make 0.1 M NaCl) was applied to the Q-sepharose column equilibrated with TGED/0.1 M NaCl (**Appendix B Figure G and H**). Fraction 29 was diluted to 0.1 M NaCl with TGED buffer and apply to the Source 15Q column. The fractions containing σ^A (12 -13) were combined and used (**Appendix B Figure I and J**).

RNAP σ^A holoenzyme was reconstituted by adding 3-fold molar excess of σ^A to core RNAP and incubated at 4 °C for O/N, followed by purification with Superdex200 size- exclusion column chromatography equilibrated Tth crystallization buffer (20 mM Tris-HCl, pH8 at 4 °C, 100 mM NaCl, 0.1 mM EDTA, 1% glycerol, 1 mM DTT). The fractions containing RNAP σ^A holoenzyme (6-8) were combined and concentrated to 10 mg/ml (23 µM) (**Appendix B Figure K and L**).

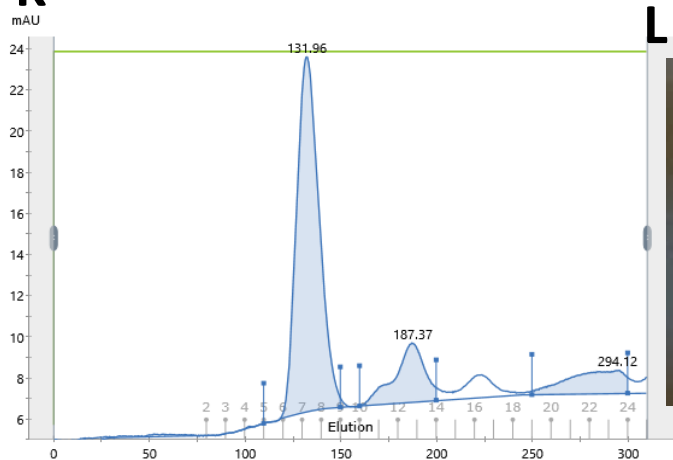
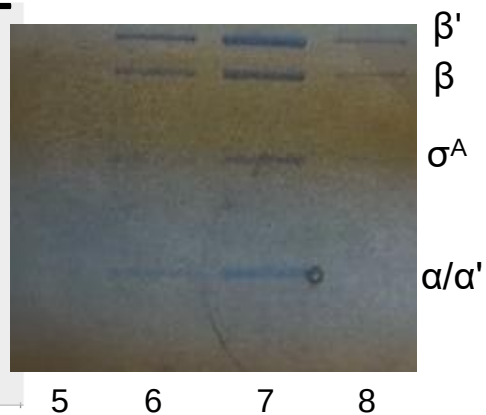
A**B****C****D****E****F**

G**H**

28 29 33 34 35 36 37 38 39 40

I**J**

10 11 12 13 14 15

K**L**

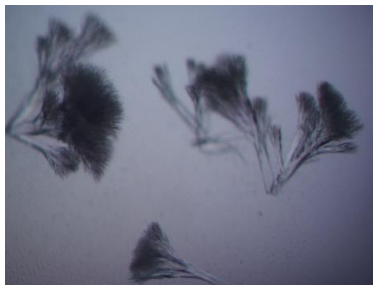
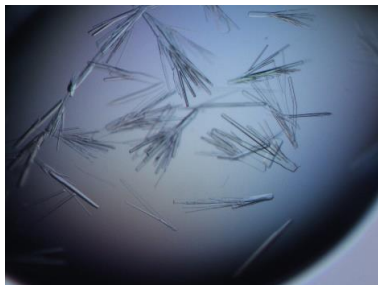
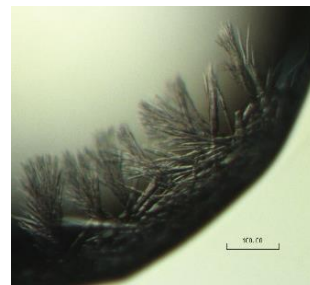
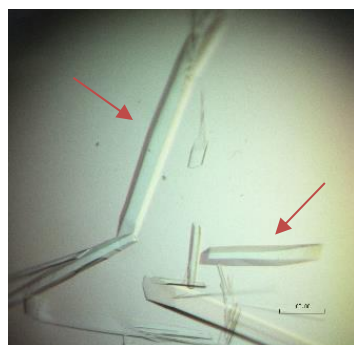
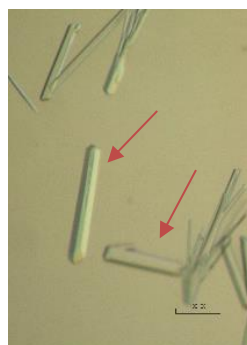
Appendix B Figure: Purification of *T. Thermophilus* RNA polymerase σ^A holoenzyme. (A)

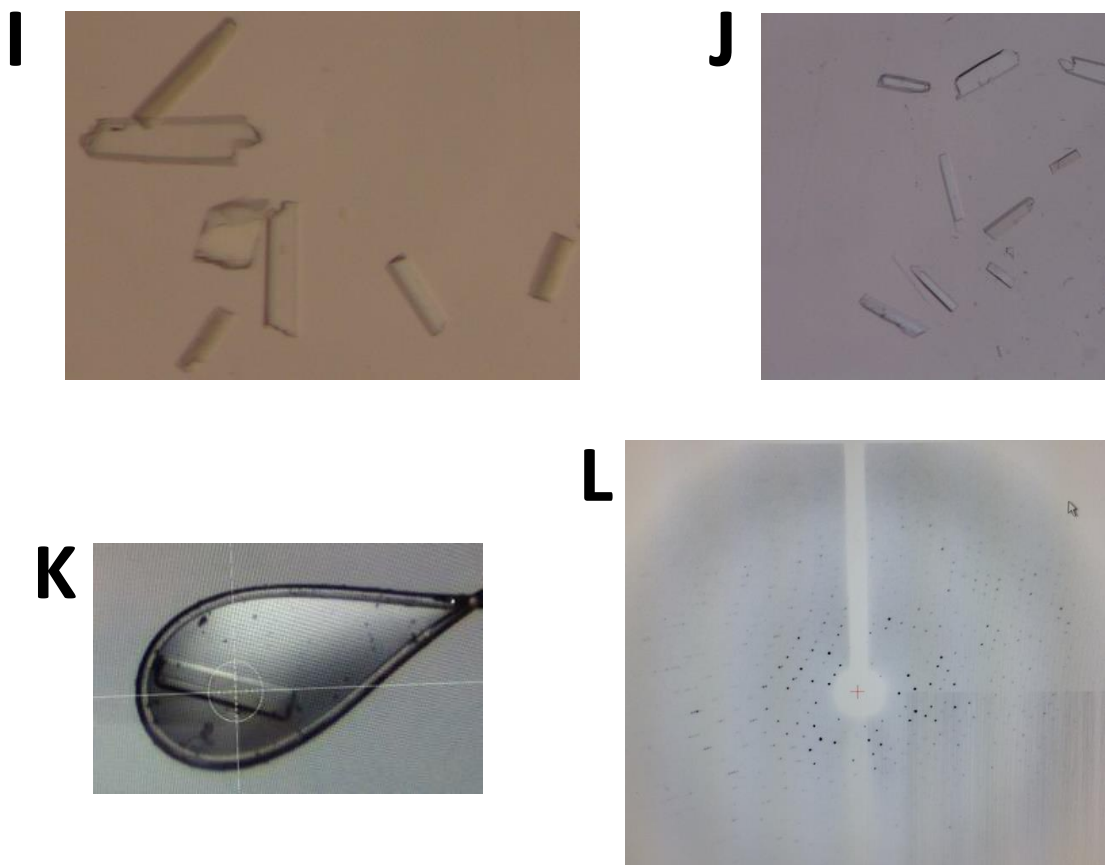
Heparin column chromatography, (B) SDS-PAGE gel of fractions from cell lysis to heparin column, (C) Source15Q column chromatography, (D) SDS-PAGE gel of fractions from Source15Q, (E) Source 15S column chromatography, (F) SDS-PAGE gel of fractions from Source 15S, (G) Q-sepharose column chromatography, (H) SDS-PAGE gel of fractions from Q-sepharose column, (I) Source15Q column chromatography, (J) SDS-PAGE gel of fractions from Source15Q, (K) Superdex200 size-exclusion column chromatography and (L) SDS-PAGE gel of main peak fractions from Superdex200 size- exclusion column.

Appendix C

Crystallization of *T. thermophilus* RNAP σ^A holoenzyme – Promoter DNA complex

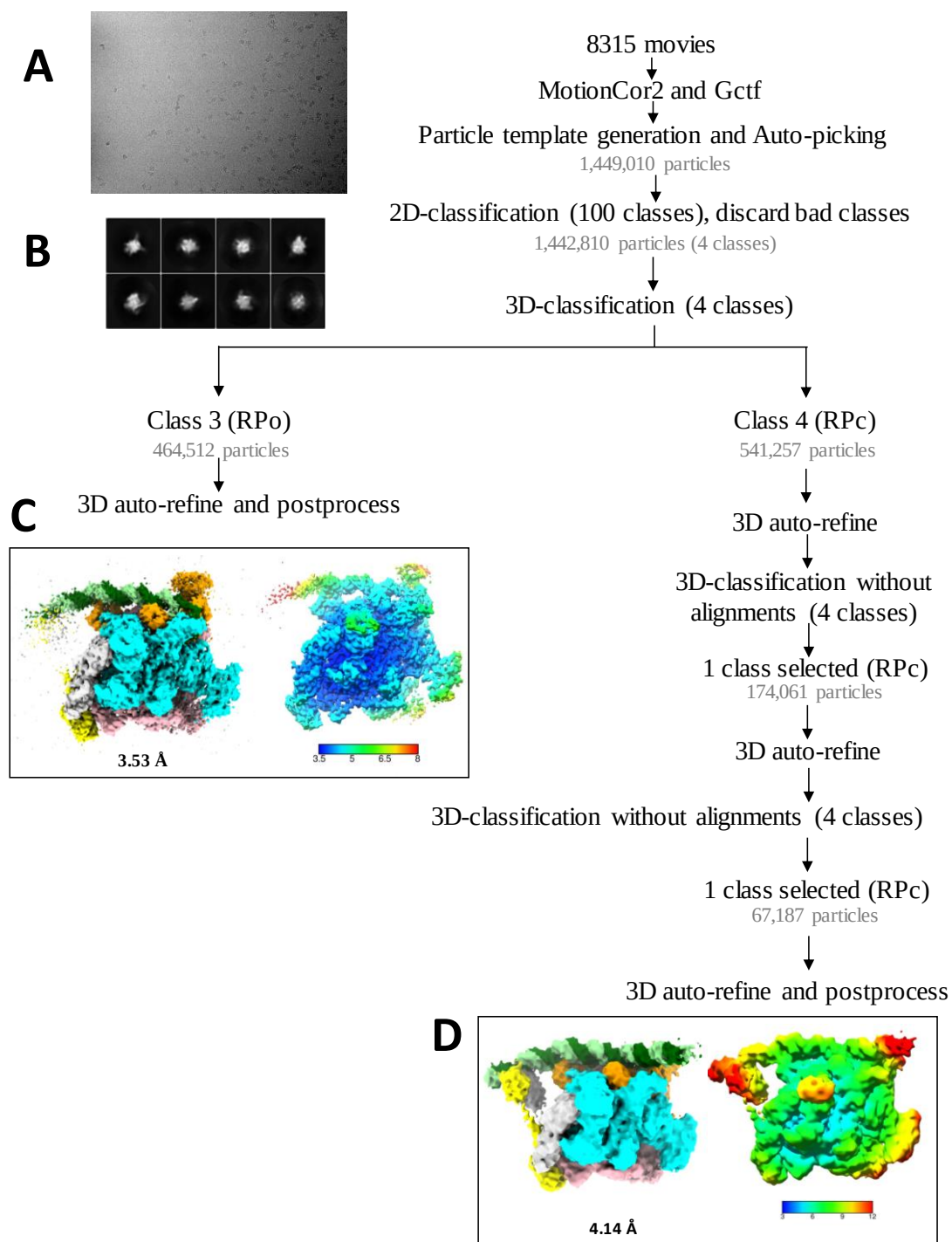
After preparing homogeneous RNAP, it is necessary to find optimal conditions (salt, buffer, pH, temperature, volume) that yield full-sized diffracting crystals. Initial trials were set by varying buffer pH and PEG4000 concentration. The RNAP-promoter DNA complex was prepared by mixing *T.thermophilus* RNAP σ^A holoenzyme (18 μ M) and DNA scaffold (27 μ M). Hanging drop vapor diffusion crystal tray was set up by mixing 1 μ l complex and 1 μ l crystallization buffer (0.1 M Tris-HCl, pH 7 – pH 9 at 22°C, 0.2 M KCl, 50 mM MgCl₂, 8% - 12% PEG4000). I have tried different total volumes (1 μ l to 3 μ l), but around 2 μ l produces the best crystals. Not optimized buffer conditions did not result in full-sized crystal, but clustered, brushed, needle-shaped crystals or microcrystals (**Appendix C Figure A-E**). These crystals are unsuitable for X-ray data collection but can be used as a seed crystal for new drops, which aids in the initial nucleation of proteins. A small portion of the microcrystal (**Appendix C Figure D, red circles**) was cut in the buffer, and 0.2 μ l was mixed with the RNAP-promoter DNA complex. Eventually, the full-sized crystal was obtained at pH 8 and 10% PEG4000 (**Appendix C Figure F-H**), harvested and cryoprotected by transferring in the solutions (0.1 M Tris-HCl, pH8 at 22°C, 0.2 M KCl, 50 mM MgCl₂, 15% PEG4000 and 15% butanediol) stepwise (**Appendix C Figure I and J**). The crystal was flash-frozen in liquid nitrogen and diffraction data was collected (**Appendix C Figure k and L**).

A**B****C****D****E****F****G****H**

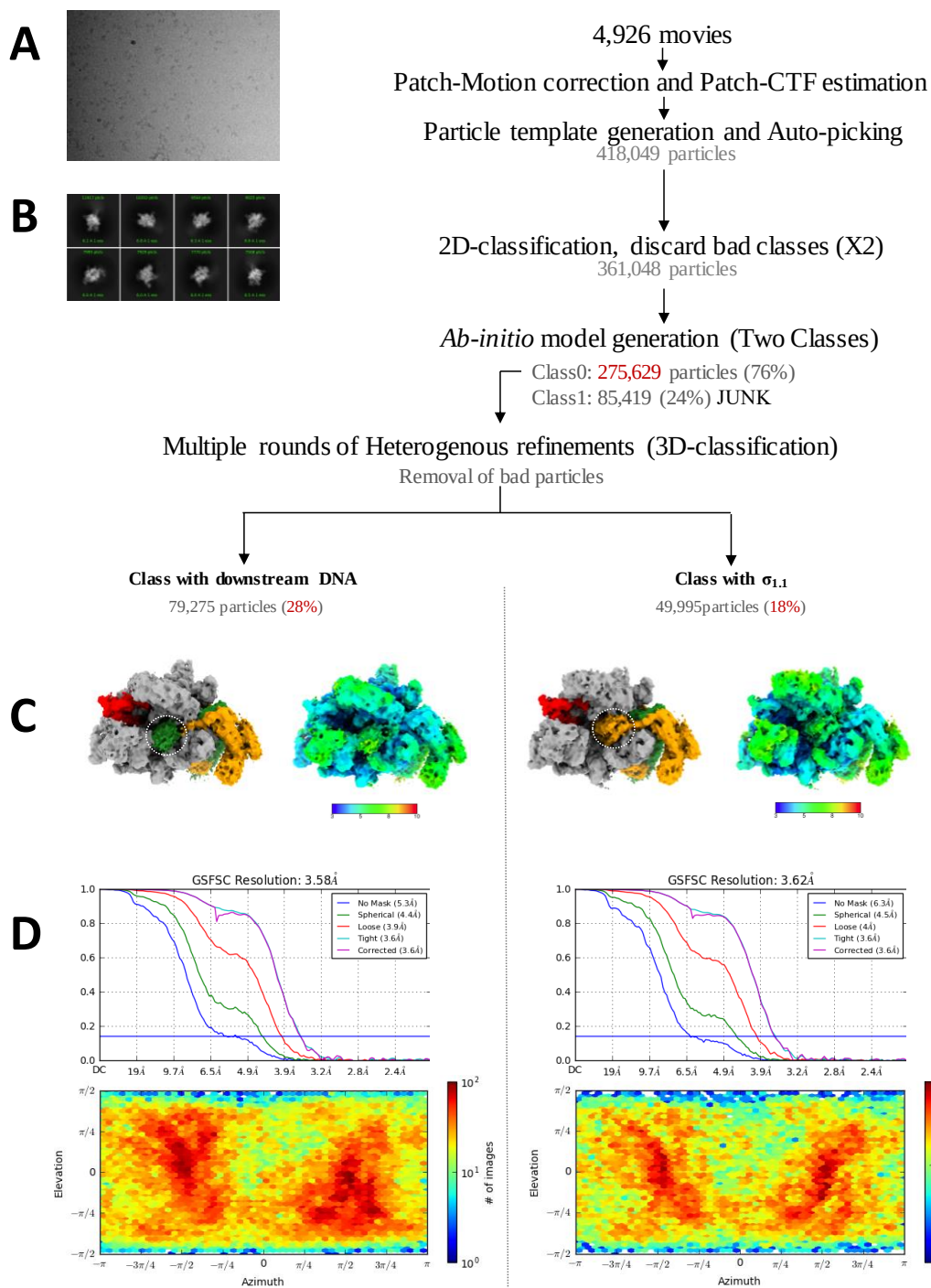


Appendix C Figure: Crystallization of *T. thermophilus* RNAP σ^A holoenzyme – promoter DNA complex. (A-E) clustered, brushed, needle-shaped crystals or microcrystals produced in not optimized conditions. A small portion of these crystals (red circles) can be used as a seed for new drops. (F-H) full-sized crystals (red arrows) in optimized conditions. (I-J) Crystals in a cryo-protectant solution. (K) A crystal in a cryo-loop for data collection. (L) A representative X-ray diffraction pattern from a *T.th* RNAP – DNA binary complex.

Appendix D



Appendix E



VITA

Yeonoh Shin, Ph.D.

Education

2015-2020 **Ph.D.** in Biochemistry, Microbiology and Molecular Biology
The Pennsylvania State University, University Park, PA

2013-2015 **M.S.** in Biomedical Sciences
Thomas Jefferson University, Philadelphia, PA

2013 **B.S.** in Bioscience Technologies
Thomas Jefferson University, Philadelphia, PA

Peer-Reviewed Research Publications

Shin, Y*, Qayyum, M.Z*, Pupov, D., Esyunina, D., Kulbachinskiy, A., Murakami, K.S. (2020). Structural basis of the ribosomal RNA transcription regulation. *Manuscript in preparation*.

Shin, Y., Hedglin, M., Murakami, K.S. Structural basis of reiterative transcription from the *pyrG* and *pyrBI* promoters by bacterial RNA polymerase (2020). *Nucleic Acids Research*, 48, 2144-2155.

Black, S.J., Ozdemir, A.Y., Kashkina, E., Kent, T., Rusanov, T., Ristic, D., **Shin, Y.**, Suma, A., Hoang, T., Chandramouly, G., Siddique, L.A., Borisonnik, N., Sullivan-Reed, K., Mallon, J.S., Skorski, T., Carnevale, V., Murakami, K.S., Wyman, C., Pomerantz, R.T. Molecular basis of microhomology-mediated end-joining by purified full-length pol θ . (2019). *Nature Communications*, 10, 4423.

Mosaei, H., Molodtsov, V., Kepplinger, B., Harbottle, J., Moon, C.W., Jeeves, R.E., Ceccaroni, L., **Shin, Y.**, Morton-Laing, S., Marrs, E.C.L., Wills, C., Clegg, W., Yuzenkova, Y., Perry, J.D., Bacon, J., Errington, J., Allenby, N.E.E., Hall, M.J., Murakami, K.S., Zenkin, N. (2018). Mode of action of kanglemycin A, an ansamycin natural product that is active against rifampicin-resistant *Mycobacterium tuberculosis*. *Molecular cell*, 72, 263-274.

Murakami, K.S*, **Shin, Y***, Turnbough, C.L., Jr., Molodtsov, V. (2017). X-ray crystal-structure of a reiterative transcription complex reveals an atypical RNA extension pathway. *Proc Natl Acad Sci U S A*, 114, 8211-8216. ***co-first authors**

Book Chapters and Reviews

Qayyum, M.Z., **Shin, Y.**, and Murakami, K.S. (2020). RNA polymerase reaction in bacteria. The Encyclopedia of Systems Biology. 3rd edition, Springer, 2020.