

Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

DocuSigned by:
Signature: 
056D06C3BA724FB...

Jin Qian
Name

4/23/2024 | 4:10 PM EDT
Date

Title Unraveling the Mechanisms of Bacterial Transcription through A Multi-tiered Lens

Author Jin Qian

Degree Doctor of Philosophy

Program Physics

Approved by the Committee

DocuSigned by:

C6D87D08901D464...

Laura Finzi
Advisor

DocuSigned by:

FD37C859211344D...

David Dunlap
Advisor

DocuSigned by:

EB6F716EF56E4A4...

Eric Weeks
Committee Member

DocuSigned by:

0A5D2ACAFB19444...

Daniel Weissman
Committee Member

DocuSigned by:

C771C7322C1D438...

Marco Cosentino Lagomarsino
Committee Member

DocuSigned by:

8E2307D5F831439...

Shashank Shekhar
Committee Member

Accepted by the Laney Graduate School:

Kimberly Jacob Arriola, Ph.D, MPH
Dean, James T. Laney Graduate School

Date

Unraveling the Mechanisms of Bacterial Transcription through A Multi-tiered Lens

By

Jin Qian
Ph.D. Candidate

Advisor: Laura Finzi, Ph.D.
Co-advisor: David Dunlap, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Department of Physics
2024

Abstract

Unraveling the Mechanisms of Bacterial Transcription through A Multi-tiered Lens

By Jin Qian

Bacterial transcription, a fundamental process essential for cellular function and regulation, presents a complex interplay of mechanisms operating at various hierarchical levels. The significance of this field lies in its broad implications ranging from understanding basic cellular processes to developing targeted therapeutic interventions. This dissertation provides a comprehensive analysis of bacterial transcription dynamics from the base-pair level to the interactions within crowded cellular environments. Employing single-molecule biophysics and molecular biology techniques, I developed a model elucidating transcription kinetics at the base-pair level, informed by the thermal dynamics of transcription bubbles and nascent RNA structures. This model not only predicts transcriptional dynamics with high precision but also sheds light on the effects of transcriptional tension and regulatory proteins on transcription pauses. Further, the study explores the interaction of RNA polymerase with DNA-bound roadblocks, revealing mechanisms of navigational flexibility under genomic constraints. The mechanisms highlight two distinct mechanisms - passive waiting for obstacle dissociation, and active backtracking, recovery, and forceful passage. Additionally, the study reveals the mechanism of a non-canonical, post-termination fate of RNA polymerase, termed as transcriptional recycling. Lastly, the research extends to understanding DNA behavior in crowded cellular milieus, thereby providing insights into chromatin dynamics *in vivo*. The findings from this research enhance our understanding of transcription processes at a molecular level and provide valuable insights in the realms of genetic regulation.

Unraveling the Mechanisms of Bacterial Transcription through A Multi-tiered Lens

By

Jin Qian
Ph.D. Candidate

Advisor: Laura Finzi, Ph.D.
Co-advisor: David Dunlap, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Department of Physics
2024

Acknowledgments

I am profoundly grateful for the support and guidance I received during the course of my PhD journey. First and foremost, I extend my deepest appreciation to my advisors, Dr. Laura Finzi and Dr. David Dunlap, whose expertise, understanding, and patience added considerably to my graduate experience. Their mentoring was both a personal and professional blessing.

I would also like to thank the members of my dissertation committee for their invaluable feedback and insightful critiques which greatly enhanced the quality of my research.

My sincere gratitude goes to my colleagues and peers at Emory University, who provided an environment of camaraderie and intellectual rigor. Their support and friendship made this journey more enjoyable and rewarding.

A special note of thanks to my family for their unwavering support and encouragement throughout my academic pursuits. Their belief in my capabilities has been a constant source of strength and motivation.

Lastly, I am thankful to everyone who directly or indirectly contributed to my academic and personal growth during my PhD. This dissertation stands as a testament to your collective impact on my life.

Contents

1	Introduction	1
2	Basis of Bacterial Transcription	4
2.1	Ternary Elongation Complex (TEC) Configuration	5
2.2	Transcriptional Pausing	7
2.2.1	The Elementally Paused Elongation Complex (ePEC)	8
2.2.2	Backtracked, Paused Elongation Complex (bPEC)	9
2.2.3	Hairpin-Stabilized Paused Elongation Complex (hsPEC)	11
2.3	Previous Attempts at Modeling Pause-Interspersed Transcription	12
3	A Model of the Kinetics of Pause-Interspersed Transcription	13
3.1	Model Description	13
3.1.1	TEC Configuration And State Transition	13
3.1.2	Forward Translocation	15
3.1.3	Backtracking	16
3.1.4	Hairpin-stabilized Pausing	19
3.1.5	The Effect of Tension and Transcriptional Factors	21
3.1.6	Model Training	22
3.2	Comparison of the Model with Experimental Data	24
3.3	Strengths and limitations of the model	29

4	Single-molecule Insights into Transcriptional Trafficking and Regulation of Gene Expression	34
4.1	DNA-binding proteins organize DNA and modulate transcription . . .	36
4.2	Single-molecule Techniques to Study in vitro Transcription	40
5	Magnetic Tweezers Experiments Revealed A Hybrid Mechanism for RNAP Transiting through Roadblocks	44
5.1	Magnetic Tweezers Assays	45
5.1.1	GreA and forces opposing or assisting RNAP translocation change pausing at roadblocks	46
5.1.2	GreA and tension reveal two paths through roadblocks	51
5.2	A hybrid transit model recapitulates the effects of force and GreA in the MT assays	56
5.3	Significance and Limitations of the Hybrid Transit Model	63
6	Transcription Recycling – An Overlooked Regulatory Mechanism	66
6.1	Force directs RNA polymerase diffusion and repetitive transcription .	67
6.2	Secondary transcription from a roadblock site	70
6.3	Sigma factor promoted sliding and secondary promoter recognition . .	72
6.4	Force biased re-initiation between converging or diverging promoters .	74
6.5	Summary of Transcriptional Recycling Mechanism	77
7	Transcription in Crowded Cellular Milieus	83
7.1	Compaction and Extension of DNA	84
7.2	Liquid-liquid Phase Separation (LLPS)	85
7.3	Theoretical Models To Explain LLPS	87
8	A Phenomenological Model of crowding effect on DNA under tension	90

8.1	Model DNA Polymers As the Shish-Kebab	91
8.2	Thermodynamics of a globular state DNA	93
8.3	Thermodynamics of a stretched chain state DNA	97
8.4	The model predicts the critical forces for DNA phase transition under different crowder conditions	99
8.5	Thermodynamics of a tadpole state DNA	100
8.6	DNA relaxation experiment proves the coexistence state	101
9	Conclusion	104
	Bibliography	107

List of Figures

2.1	Configuration of TEC	6
2.2	Pausing signals in transcriptional elongation	9
3.1	State transitions and statistical approach in the transcription model .	14
3.2	Model construction	16
3.3	Modeling hairpin structures	20
3.4	Model fitting and prediction	25
3.5	Comparison between dwell times on major pause sites from experi- ments and model	28
3.6	Comparison of RNA hairpin formation at different energy profiles and positions	32
4.1	Mechanisms to overcome chromatin roadblocks	39
4.2	Single-molecule techniques to study transcription	41
5.1	Illustration of Magnetic Tweezers	46
5.2	Features of the real-time experiments of RNAP transiting through roadblocks.	47
5.3	Experiments with heparin suggest roadblock rebinding negligibly af- fects pause times.	49
5.4	Pauses at LacI-O1 and LacI-O2 roadblocks under different conditions.	50

5.5	Pause times and fractions of passage at LacI-Os roadblocks under different conditions.	52
5.6	Pause times and the fraction of transit through EcoRI Q111 roadblocks under different conditions.	54
5.7	The hybrid transit model and simulation results	57
5.8	Simulations show the effects of forces and GreA in three different regimes	59
5.9	Experiments with different GreA concentrations show effects on pause times similar to those of simulations.	63
6.1	Force-directed sliding leads to repetitive transcription.	69
6.2	DNA-binding protein roadblocks, LacI and NusG affected post-termination sliding and repetitive transcription	71
6.3	Repetitive transcription on templates with two adjacent promoters . .	75
6.4	The post-termination fate of RNAP includes secondary transcription	79
7.1	Molecular components of an <i>E. coli</i> cell	85
7.2	Schematic representation of the physical basis for the Asakura–Oosawa model and the Kirkwood–Buff model	88
8.1	Illustration of the Beads-On-Sting model and the Shish-Kebab model for the DNA	92
8.2	A schematic of the sphero-cylinder model of a DNA globule	94
8.3	Geometry of two interacting rods in the Shish Kebab model	96
8.4	DNA pulling experiment results and model simulation	100
8.5	A comparison of free energies of DNA globule, tadpole, stretched chain	102
8.6	DNA relaxation experiment results and model simulation	103

List of Tables

3.1	Summary of experimental pause positions, durations and mechanisms.	24
3.2	Values of the optimized parameters in the model.	26
5.1	Kinetic parameters generated by fitting our model to the experimental pause time distributions.	62
6.1	Summary of records using single promoter, convergent promoter, and divergent promoter templates	78

List of Algorithms

1	Simulate RNAP pause time at a roadblock	58
---	---	----

Chapter 1

Introduction

In my doctoral research at Emory University under the mentorship of Drs. Laura Finzi and David Dunlap, I delve into the complexities of bacterial transcription. This field is not only fundamental to our understanding of molecular biology but also crucial for developing therapeutic applications. My dissertation presents a hierarchical exploration of transcriptional mechanisms, examining the interactions of RNA polymerase (RNAP) at levels ranging from individual base pairs to the broader cellular context.

First aspect in this work is a model that characterizes transcription kinetics at the base-pair level, crucial for deciphering the intricacies of gene expression. At this juncture, the sequence of the DNA template, the structure of the transcribing RNAP, the nascent RNA transcript, and the transcription factors (TFs) influencing RNAP-DNA interaction assume paramount importance. As documented in refs [1, 2], RNAP, an efficient motor enzyme, translocates along the DNA template, producing a transcript that mirrors the DNA sequence. This process occurs at a rate of 10-20 bp/second and is punctuated by pauses of varying durations [3, 4]. These pauses can be short, typically under a second and known as elemental pauses, believed to be precursors to longer pauses. The longer pauses, which can extend to tens of seconds, fall into

either class I (hairpin-stabilized) and class II (backtracked) categories. These pauses are not only structurally characterized but also mechanistically explored, with their regulation attributed to the DNA sequence, the nascent transcript's structure, and the availability of TFs [5]. Despite early efforts for a physical model that explains and predicts transcriptional pauses induced by DNA sequences, previous models have had limited success in predicting pausing types, durations, and the influence of external tension or TFs [6, 7, 8, 9, 10]. To address these gaps, I have developed a model based on contemporary biochemical understanding of transcription pausing mechanisms, optimizing it with high-resolution transcription data [11]. This thermodynamic model not only elucidates the roles of external tension and TFs but also accurately simulates observed pause sites and durations. Additionally, it predicts transcription dynamics on unfamiliar DNA sequences and is adaptable to include initiation and termination stages.

Progressing to the nucleo-protein interaction level, I investigate the regulatory roles of DNA-binding proteins in transcription [12, 13, 14]. The affinity of these proteins for specific or non-specific DNA sequences can vary significantly, influenced by physiological conditions [15]. Utilizing multiplexed magnetic tweezers [16, 17], my research examines RNAP's passage through genomic roadblocks of differing strengths. The findings indicate that RNAP can switch between passive and active, reciprocating pathways to surmount these roadblocks. Passively, RNAP waits for spontaneous dissociation of roadblocks, while actively, it engages in backtracking-recovery cycles to forcibly clear the roadblock from its path. The efficiency of these pathways can be influenced by tension and TFs, highlighting crucial aspects for enhancing gene expression efficiency and offering strategies to inhibit bacterial transcription in pathogenic organisms.

A significant discovery relates to the recycling behavior of post-termination RNAP. This behavior suggests a previously unrecognized regulatory mechanism with poten-

tial implications for gene expression patterns [18, 19, 20]. The recycling of RNAP post-termination, especially under external tension and influenced by TFs and RNAP mutants, reveals tight regulation of transcription patterns. This has implications for modifying bacterial gene expression, with potential impacts on antibiotic resistance and pathogen control.

Lastly, my research extends to the behavior of DNA in crowded cellular environments. The focus here shifts to understanding the physical properties of the DNA polymer, particularly how phase separation and environmental constraints affect DNA behavior, thereby providing insights into chromatin dynamics *in vivo* [21, 22, 23]. Experimental observations of DNA transitioning from a chain to a globular state under macromolecular crowding have led to a phenomenological model explaining DNA polymer dynamics in such environments. This model accounts for the critical crowder fraction and the effects of tension, making it essential for understanding how macromolecular crowding influences chromatin formation and gene expression within the nucleoid. This aspect of the research provides a deeper understanding of the complex interplay between DNA and its cellular environment, enhancing our comprehension of cellular processes.

Chapter 2

Basis of Bacterial Transcription

Bacterial transcription, a process where DNA's genetic information is transcribed into RNA molecules by RNAP [24], initiates with RNAP binding to specific DNA regions called promoters. These promoters have sequences signaling the start of transcription. The RNAP then unwinds the DNA double helix, exposing the template strand in what is known as the initiation stage. It synthesizes a complementary RNA molecule, adhering to base-pairing rules and utilizing nucleoside triphosphates ATP, GTP, CTP, and UTP (all four kinds together are termed as NTPs). As the process continues, RNAP elongates the RNA molecule by adding nucleotides while moving along the DNA template. This elongation stage sees the RNA strand gradually detaching from the DNA template, which allows the DNA double helix to reform. The process culminates when RNAP encounters a termination sequence, detaching from the DNA and releasing the complete RNA molecule [25, 26].

This section delves into the finer points of bacterial transcription, with a particular focus on the nuances present during the elongation stages. Here, we explore the contemporary understanding of the transcription machinery's structural configuration and the paused states that regulate the dynamics of RNAP elongation.

2.1 Ternary Elongation Complex (TEC) Configuration

RNAPs, as molecular motors, consume chemical energy to synthesize RNA by incorporating nucleotide triphosphates. This activity follows the initial stages of promoter recognition, double-strand opening, and the polymerization of an approximately 8–12 nucleotide long transcript. The elongation phase involves a ternary elongation complex (TEC), distinguished by a DNA bubble separating the upstream and downstream DNA duplexes, an 8–9 nucleotide DNA/RNA hybrid, and an emerging nascent RNA chain. Recent advancements in x-ray crystallography and cryo-electron microscopy have further elucidated the TEC's structural features. These include the trigger loop (TL) and the bridge helix (BH), implicated in translocation and the proofreading of the RNA's 3' end, as well as the lid and flap domains, which interact with the RNA and upstream DNA (Figure 2.1A & B) [27, 28, 29, 30, 31]. Single-molecule assays have further contributed to our understanding of transcript elongation [3, 32, 33, 34, 35]. Among the various models, the Brownian ratchet mechanism, which postulates that TEC's forward motion is stochastically generated by thermal fluctuation and then stabilized by molecular pawls, has gained prominence (Figure 2.1C). This model is favored due to its 1-nucleotide translocation step and its compatibility with observed force-velocity relationships in transcription under different NTP conditions [4, 7, 36].

However, this model may oversimplify the real transcription mechanism, including potentially overlooking an allosteric nucleotide binding site within the elongation complex, as suggested by Foster et al. [37]. Yin and Steitz's work, based on RNAP's crystal structures with NTP substrates, suggests that RNAP translocation may not be solely driven by NTP-binding affinity. They postulated that RNAPs, coordinated by specific residues with pyrophosphate, remain in the pre-translocation position until pyrophosphate's release promotes the O helix rotation, leading to RNAP's transloca-

tion to the post-translocation position [38]. Contradicting most models' mechanistic assumptions about transcriptional pauses, their findings suggest that pyrophosphate release, not NTP presence, determines the translocation register. Aligning with this, a dual-ratchet model has been proposed, where a helix acts as a reciprocating pawl, pushing RNAP forward relative to the nucleic acid scaffold, and the incoming NTP substrate prevents backward slippage of RNAP [39].

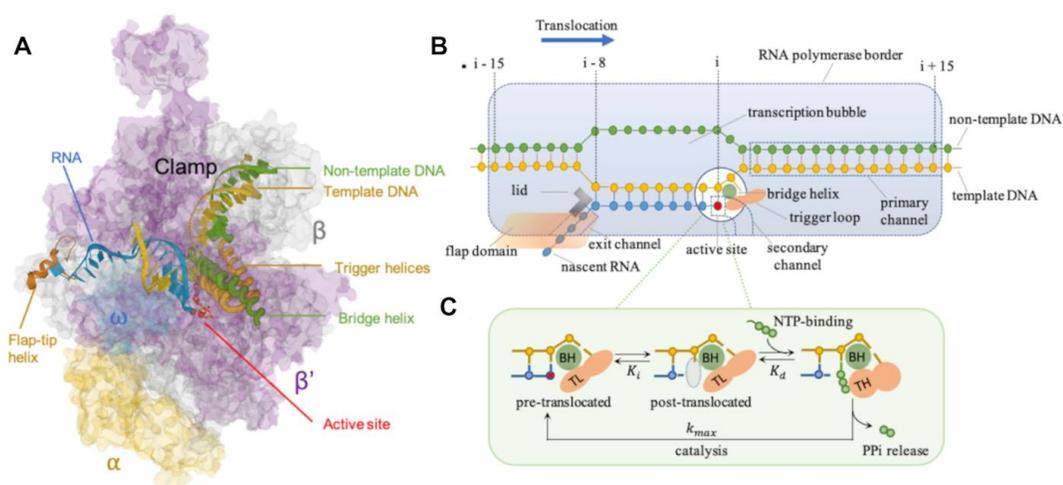


Figure 2.1: **(A)** Crystal structure of the *Thermus thermophilus* RNAP elongation complex with the bound NTP substrate (PDB: 205J). **(B)** Cartoon of an elongation complex. An elongation complex covers 30–35 base pairs (bp) of DNA, including ~ 14 bp of the DNA downstream of the active site, 10–11-bp of the transcription bubble, and nascent RNA which binds the template strand to form an 8–9-bp RNA–DNA hybrid. The RNAP primary channel accommodates the downstream dsDNA; the secondary channel serves as the site for NTP entry and RNA extrusion when RNAP backtracking occurs; the nascent RNA emerges through the exit channel near the flap after the RNA/DNA hybrid strand separates at the lid. The structure is stabilized by the interaction of specific RNAP domains, such as the bridge helix (BH) and trigger loop (TL) that comprise the active center, the lid and flap that interact with nascent RNA. **(C)** A nucleotide addition cycle involves a Brownian ratchet at the active center. RNAP shifts between the pre- and post-translocated registers until an incoming NTP (green) occupies the active site (gray). There, NTP reacts to form a phosphodiester linkage to 3' -OH group of the growing RNA chain and release inorganic pyrophosphate (PPi). During the process, the TL folds into trigger helices (TH) and positions an NTP for catalysis.

2.2 Transcriptional Pausing

TECs are highly processive, capable of synthesizing extensive nascent transcripts. However, the regulation of transcriptional elongation is critical, particularly in responding to abnormal events such as nucleotide misincorporation or encountering transcriptional roadblocks, coordinating with co-transcriptional events like translation and splicing, and ensuring biologically meaningful transcription termination at the correct positions. Regulation is achieved largely through pauses that interrupt forward translocation.

During elongation, RNAP at each nucleotide coding position may follow different kinetic pathways. Misincorporation can compromise forward translocation, leading to correction pathways involving backsliding and removal of the incorrect nucleotide. Alternatively, transcription may continue past the misincorporated base, albeit slowly, resulting in a mutated RNA. Additionally, transcript termination and release from the transcription complex are other potential outcomes. Various pausing mechanisms guide these divergent paths.

The roles of transcriptional pauses, while still under investigation, have been linked to various human pathologies in the Pol II system. In prokaryotic systems, transcriptional pauses, which vary greatly in duration from brief elemental pauses lasting mere milliseconds to extended pauses of several minutes, have been thoroughly studied [5, 40, 41, 42, 43, 44, 45]. Artsimovitch and Landick suggested that long pause states commence with the formation of an elemental paused state, which can be stabilized and converted into either a backtracked or hairpin-stabilized state [5]. These mechanistically diverse pauses are thought to serve different roles in transcription: backtracked pauses control gene expression at promoter-proximal sites and correct misincorporated bases, while hairpin-stabilized pauses guide the folding of leader RNA structures. Recent advances in single-molecule experimentation and cryo-EM structures have refined our understanding of factors influencing paused state entry

and escape [28, 35, 46, 47].

Transcriptional pausing is largely probabilistic; most pausing sites are not entirely efficient, making it challenging to fully assess the roles of different paused states in transcriptional regulation and the mechanisms integrating various transcriptional pauses to produce pause-interspersed transcription. Nonetheless, several studies have illuminated the dynamics of transcript elongation, leading to a key question in this dissertation: unraveling the dynamics of transitions between these probabilistic paused states.

2.2.1 The Elementally Paused Elongation Complex (ePEC)

In the Brownian-ratchet model, the ePEC originates from the RNA–DNA scaffold’s thermodynamics, allowing toggling between pre- and post-translocated states until NTP binding favors the post-translocated state (Figure 2.1C). This explanation, intuitively acceptable and first proposed by Yager & von Hippel and Guajardo & Sousa, underpins Bai et al.’s quantitative sequence-dependent transcription model that predicted pre-translocated pauses with lifetimes akin to those observed in ePECs [6, 10, 48, 49, 50]. Consensus elements for brief pauses include a GC-rich segment upstream of the RNA-DNA hybrid and a pyrimidine at the pause site followed by a G (G-10Y-1G+1). In the Brownian-ratchet model, G/C at positions –10 and +1 are less favorable for forward translocation than the less stable A/T [3, 44, 51]. Although elemental pausing also depends on the DNA–RNA hybrid and downstream DNA duplex sequences, contributions from sequences in the transcriptional bubble’s fork-junctions broadly align with the Brownian-ratchet model’s predictions regarding the ePEC [44, 52, 53].

However, experimental evidence points to various origins for the ePEC beyond those predicted by the Brownian ratchet model. Cryo-EM structures of the ePEC have shown a post-translocated RNA with pre-translocated DNA forming a tilted

RNA/DNA hybrid, accompanied by rearrangements in the RNAP trigger loop (Figure 2.2(2)). This half-translocated state, potentially an off-pathway state linked to elementally paused states, might exhibit extended dwell times [53, 54]. Other proposed conformational arrangements that might inhibit RNAP forward translocation include a frayed 3' end of the nascent RNA at the active site (Figure 2.2(1)) and sequence-dependent nucleic acid and RNAP interactions (Figure 2.2(4)), challenging the Brownian ratchet model's explanation for ePEC formation [55, 56, 57].

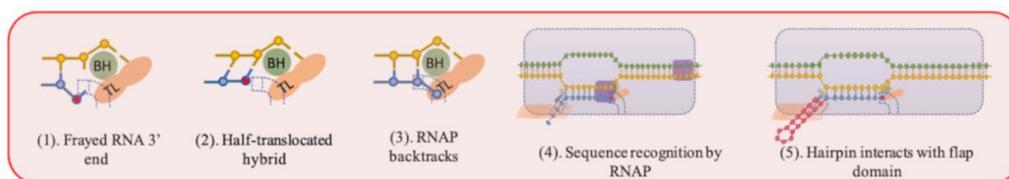


Figure 2.2: States proposed to explain elongational pauses include: (1) fraying of the DNA-RNA hybrid at the active site detected by crosslinking, although structural data are not consistent with this; (2) incomplete template DNA strand translocation, with a pre-translocated DNA strand and a post-translocated RNA strand, that precludes NTP addition; (3) RNAP backtracking upstream and extruding the 3' end of nascent RNA into/through the catalytic site; (4) RNAP recognizing pause signals encoded in DNA/RNA hybrid and/or downstream DNA sequences (purple); (5) a hairpin structure forming in nascent RNA that interacts with the RNAP exit channel and flap domain to cause a global conformational change that disrupts elongation. States 1, 2 and 4 are proposed as elemental pauses. State 3 represents the backtracked paused complex, and state 5 represents the hairpin-stabilized paused complex.

2.2.2 Backtracked, Paused Elongation Complex (bPEC)

Backtracking, RNAP's reverse motion, can induce pausing by extending the 3' end of a nascent transcript into or through the catalytic site, as seen in Figure 2.2(3). This can lead to the formation of a backtracked, paused elongation complex (bPEC). The potential resolution of this state involves an endonucleolytic cleavage event, either intrinsically or via accessory factors such as GreA and GreB for *Escherichia coli* (*E. coli*) or SII/TFIIS for Pol II. This process can rescue a bPEC from its stalled state [58, 59, 60]. *E. coli*'s backtracking is notably present at operon polarity suppressor (ops) sites, leading to promoter-proximal pausing, and a similar mechanism is thought

to be at play in eukaryotic Pol II, where high G/C content followed by A/T-rich sequences near DNA promoters induce unstable RNA–DNA hybridization, prompting polymerase backtracking [5, 61, 62]. Shaevitz et al., employing optical tweezers assays, observed that backtracking events associated with pauses in RNAP, lasting from 20 seconds to over 30 minutes, occurred not only at consensus sequences but also throughout various DNA template locations [40].

The formation of bPECs is sensitive to external forces, and the path leading to these complexes is relatively well understood. Assisting or opposing loads either inhibit or facilitate bPEC formation, indicating that bPECs are energetically stable states created by RNAP’s reverse translocation [63]. Saba et al. suggested that an ePEC could rapidly equilibrate among the pre-translocated, half-translocated, and one-base-pair backtracked states, given the minimal energy barrier associated with a single base pair [53]. High-resolution optical trapping assays also show the ease of forming a one-base-pair bPEC [50, 64]. Extended backtracking is less common but characterized by longer dwell times, suggesting larger activation barriers for deeper bPEC formation. Evidence also points to a possible conformational change associated with backtracking, pushing the bPEC into a state resistant to rescue by external loads and RNA cleavage events [65, 66, 67].

Recovery from backtracking is considered to occur either via 1D diffusion or cleavage of the RNA blocking the catalytic site. Optical trapping assays on RNAP II indicate that the distribution of backtracked pauses of under 10 seconds follows a $t^{3/2}$ power law. This suggests that backtracked RNAP II diffuses in a 1D unbiased random walk with one-nucleotide steps in the absence of RNA cleavage events [63]. Lisica et al. reported that the recovery mechanism choice is determined by a kinetic competition between the random walk and the RNA cleavage event. Shallow energetic traps, like one-base-pair bPECs, tend to recover through 1D diffusion, while deeper traps are more likely to recover via RNA cleavage [68]. Supporting this, re-

cent work using high-throughput magnetic tweezers showed that PECs backtracked by more than 4-bp predominantly recover through intrinsic cleavage in the absence of cleavage factors [67].

2.2.3 Hairpin-Stabilized Paused Elongation Complex (hsPEC)

A significant class of long-lived pausing signals is encoded within RNA secondary structures, as illustrated in Figures 2.2(5). Nascent RNA structures, such as pseudoknots or hairpins, can interact with RNAP’s flap domain near the exit channel and impede nucleotide addition in the active site, located ~ 65 Å away. Supported by cryo-EM reconstructions, an allosteric model posits that upon hairpin-flap interaction, RNAP undergoes a global conformational rearrangement. This stabilizes a swivel module and prevents trigger loop folding, disrupting the active nucleotide addition cycle [30, 54]. Transcriptional factor NusA is believed to enhance hairpin-mediated pausing by creating a positively charged cavity in the RNAP exit channel, facilitating RNA secondary structure formation and stabilizing RNA–RNAP interactions [30]. It’s worth noting that an RNA hairpin can also disrupt the RNA–DNA hybrid during intrinsic transcriptional termination, possibly inducing similar conformational rearrangements in polymerases during both hairpin-stabilized pausing and intrinsic termination [5, 69, 70, 71].

While the mechanisms underlying the hsPEC are not fully elucidated, experimental data have revealed many of its features. Recent studies have separately shown that the RNAP flap domain, trigger loop, and RNA hairpin are indispensable for hsPEC formation, aligning with the mechanism suggested by the cryoEM hsPEC structure [64, 72, 73, 74]. A four-nucleotide gap between the hairpin and the RNA–DNA hybrid is optimal for hsPEC formation, and reducing this spacer to 2 nt significantly diminishes pausing. The stability and size of the hairpin affect hsPEC formation, with a stable stem favoring formation while the size of the end-loop being less crucial [31, 74].

Overall dimensions matter, with longer stems reducing hsPEC formation. Intriguingly, an artificial hairpin formed by hybridizing an oligonucleotide to the nascent transcript cannot replicate the nascent RNA hairpin’s effect [75].

2.3 Previous Attempts at Modeling Pause-Interspersed Transcription

Substantial progress has been made in recent years in elucidating the mechanism of transcription elongation. Pioneered by Yager & von Hippel, a static sequence-dependent thermodynamic analysis of TEC stability in *E. coli* RNAP suggested that a more complete description of transcription would require kinetic analysis [10, 76]. Subsequent models have described each NTP incorporation cycle as a multi-step or even multi-branch reaction [48, 77, 78]. Bai et al. further characterized the NTP-dependent kinetic rates and proposed a model with a branched paused state to include the long-lived bPEC [6, 7].

Despite these early models shedding much light on the general transcription mechanism, a significant gap remains between theoretical modeling and the experimentally observed, probabilistic occurrence of transcriptional pauses. Firstly, a theoretical basis for sequence-dependent kinetics of transcription remains obscure. Secondly, while these models predict certain types of experimentally detected pauses (ePEC and bPEC), they fail to account for others (hsPEC). Thirdly, these early models demonstrate some predictive power on sequence-dependent pause locations but fall short in predicting observed pause durations. Lastly, these models do not consider external tension or transcription factors, which are known to significantly influence transcriptional pauses.

Chapter 3

A Model of the Kinetics of Pause-Interspersed Transcription

In this section, we propose a model grounded in our current biochemical understanding of transcription pausing mechanisms. This model has been optimized with high-resolution transcription data and provides a thermodynamic explanation for the influence of external tension and transcription factors. It accurately simulates observed pause sites and durations and is adept at predicting transcription dynamics on unfamiliar DNA sequences.

3.1 Model Description

3.1.1 TEC Configuration And State Transition

The model describes a TEC by its transcription position (m) and state (n). The position along the DNA template (m) represents the length of the RNA transcript. The TEC can exist in active ($n=0$) or backtracked ($n<0$) translocation states, or in a distinct hairpin-stabilized state (hsp). Figure 3.1a illustrates the interconnections among these states. An active TEC at position m ($m,0$) can either move forward to

the next active state $(m+1,0)$, transition into a backtracked state $(m, -1)$, or enter a hairpin-stabilized state (m, hsp) .

The TEC's energy is estimated as the sum of the free energies from several contributions: the transcription bubble, DNA-nascent RNA hybrid, nascent RNA, and the RNAP-DNA interaction [6, 9, 10], as given by Equation (3.1).

$$G_{TEC} = G_{bubble} + G_{hybrid} + G_{RNA} + G_{RNAP_binding}. \quad (3.1)$$

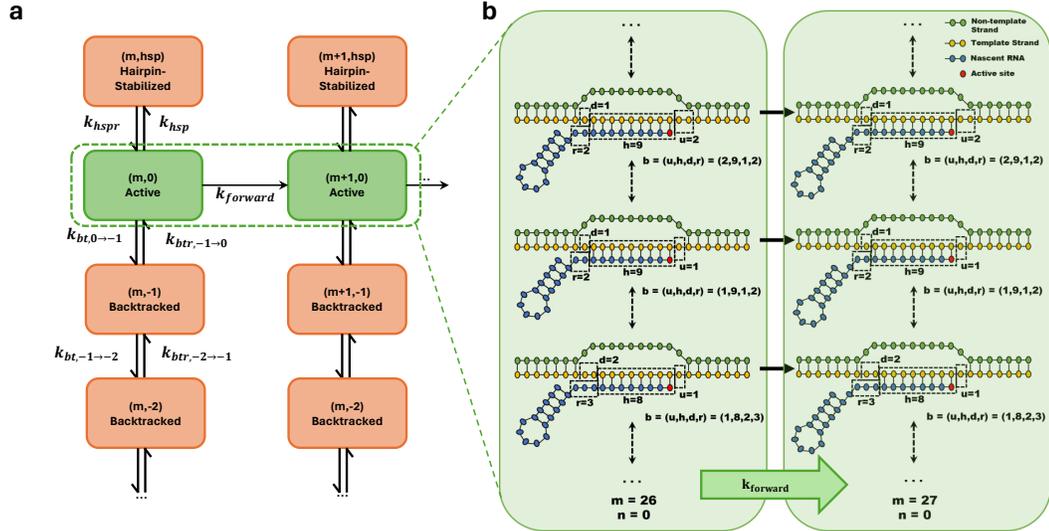


Figure 3.1: State transitions in the model and the statistical approach to the transcription bubble configuration (a) A diagram of transcriptional states considered in the model shows their interconnections. (b) An illustration of the statistical approach to characterize transcription bubble configurations including the forward translocation step. Dashed arrows indicate fast equilibrium and solid arrows indicate the allowed state transitions.

To detail the configuration of a transcription bubble and the energy profile of a TEC, we employ a statistical mechanics-based approach, founded by Tadigotla [9]. A transcription complex (m, n) exists in rapid equilibrium among various microstates, each defined by parameter (b) , depending on the number of unpaired DNA bases upstream (u) and downstream (d) of the DNA-RNA hybrid inside the RNAP enzyme, the length of the hybrid (h) , and the number of single-stranded RNA bases protected

by RNAP (r) (Figure 3.1b).

The model assumes that equilibrium among microstates is quickly reached, much faster than the time required for state transitions. Thus, for each transcription complex (m, n), the probability of a particular microstate b follows the Boltzmann distribution as described in Equations (3.2) to (3.4). Figure 3.1b shows how the model statistically treats the forward translocation step. All state transitions in the model are determined based on Equation (3.4).

$$P_m^b = Z_m^{-1} \exp\left(\frac{-G_{\text{TEC}}^{m,b}}{k_B T}\right) \quad (3.2)$$

$$Z_m = \sum_b \exp\left(\frac{-G_{\text{TEC}}^{m,b}}{k_B T}\right) \quad (3.3)$$

$$k_{m \rightarrow m+1} = \sum_b P_m^b k_{m \rightarrow m+1}^b \quad (3.4)$$

3.1.2 Forward Translocation

We model the forward (active) translocation of RNAP using the Michaelis-Menten (M-M) equation (Equation (3.5)). The forward translocation involves three steps: a fast equilibrium between position m and position $m+1$, recruitment of NTP at the active site, and catalysis followed by the release of pyrophosphate (Figure 3.2a). Parameters k_{max} and K_d of Equation (3.5) are fitted to experimental data to identify slow translocation sites that precede long-lived pauses. These slow translocation events are interpreted as pre-translocated, elemental pauses on the translocation pathway.

$$k_{\text{forward}} = \frac{k_{\text{max}} [\text{NTP}]}{K_d (1 + K_i) + [\text{NTP}]}, \quad (3.5)$$

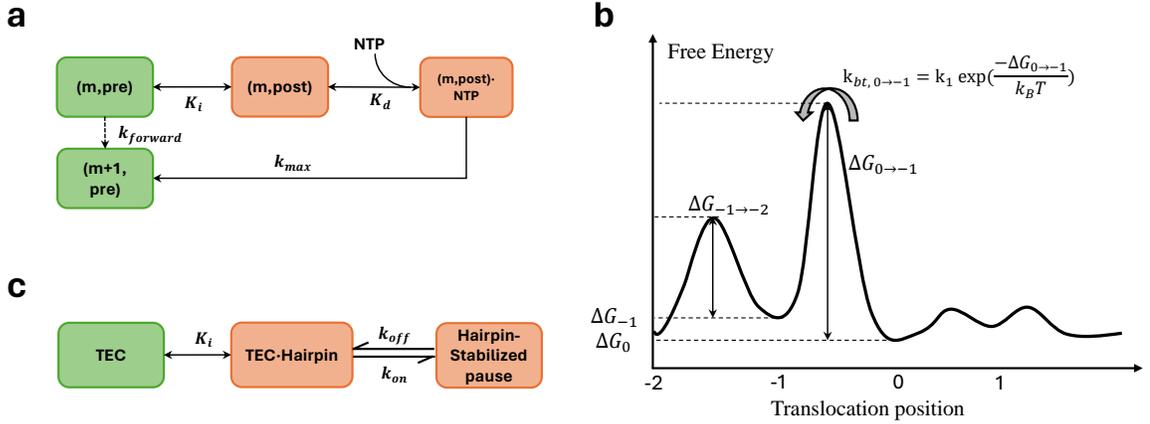


Figure 3.2: Model construction. (a) An illustration of RNAP forward translocation using the Michaelis-Menten equation. (b) The free energy landscape for the backtracking pathways. Note that the first backtracking step has a different energy barrier than the deeper backtracking steps. (c) The proposed kinetic mechanism for the hairpin-stabilized pause.

3.1.3 Backtracking

In our model, backtracking is represented following the framework of the Arrhenius Equation (3.6),

$$k_{bt} = k_1 \exp(-\Delta G/k_B T) \quad (3.6)$$

previously set with an activation barrier ranging from $40-50k_B T$ per step of backward translocation, as identified in earlier studies [6]. However, this estimate appears excessively high when considering that the free energy involved in base pairing within a transcription bubble generally does not exceed $-20k_B T$ [9].

To more realistically model this process, we adopt the Arrhenius approach but with a nuanced treatment of the initial backtracking step as distinct from the subsequent steps. This distinction stems from the understanding that the initial phase of backtracking involves the 3' end of the nascent transcript impeding the active site, subsequently encroaching into RNAP's secondary channel [79]. This initial interaction sets the stage for further backtracking steps, each of which further consolidates the

RNA's presence within the secondary channel. Figure 3.2b illustrates our modified approach, highlighting these distinct phases of the backtracking process.

We assume the energy barrier for an active TEC to enter the backtracked state to be:

$$\Delta G_{0 \rightarrow -1} = \Delta G_{bt} - G_0 \quad (3.7)$$

where ΔG_{bt} is a fixed activation energy specific for entering a backtracked state. We can assume that ΔG_{bt} will be limited to the energy available from complete collapse of the bubble, which is estimated to be in the range $-20 \sim -10k_B T$. ΔG_0 is the energy of a TEC at an active site. The rate constant to enter the backtracked state from the active state (0) would be

$$k_{0,bt} = k_1 \exp(-\Delta G_{0 \rightarrow -1}/k_B T) \quad (3.8)$$

where k_1 is the prefactor of backtracking.

For any further backward translocation of RNAP, the energy barrier should relate to the energy difference between two adjacent translocation states and the backtracked distance. Thus, for $n > 0$,

$$\Delta G_{-n \rightarrow -n-1} = \Delta G_{bt,increment} + 0.5(G_{-n} - G_{-n-1}) \quad (3.9)$$

and

$$k_{-n,bt} = k_1 \exp(-\Delta G_{-n \rightarrow -n-1}/k_B T) \quad (3.10)$$

where $\Delta G_{bt,increment}$ represents the backtracking energy barrier due to increase in the length of the transcript inserted into the secondary channel.

Our model contemplates two distinct pathways for recovering from a backtracked state, denoted as k_{btr} . These are the diffusive pathway and the cleavage pathway. The diffusive pathway is facilitated by RNAP diffusion, adhering to the principles of the Arrhenius equation, which takes into account the energy barriers as previously delineated.

$$k_{-n-1,btr} = k_1 \exp(-\Delta G_{-n-1 \rightarrow -n}/k_B T) \quad (3.11)$$

and

$$k_{-1,btr} = k_1 \exp(-\Delta G_{-1 \rightarrow 0}/k_B T) \quad (3.12)$$

Conversely, the cleavage pathway involves a mechanistic process wherein the nascent RNA, which has intruded into the secondary channel of RNAP during backtracking, is cleaved. This cleavage serves to reposition the 3' end of the nascent RNA into the RNAP's active site. Notably, this particular process is presumed to be independent of the RNA sequence. Consequently, in our model, it is treated as a process occurring at a uniform, constant rate.

In our model, we have chosen not to include hypertranslocation, a state defined as the forward translocation of RNAP without simultaneous RNA elongation at the active site. This decision is underpinned by two key considerations. Firstly, hypertranslocation is not universally recognized as a regular occurrence during transcription. As pointed out by Larson et al. [80], its presence in transcriptional processes is not consistently observed, leading to uncertainties about its general relevance. Additionally, in force spectroscopy assays, hypertranslocation events cannot be reliably distinguished from backtracking, which further complicates its inclusion in the model. Secondly, from an energetic standpoint, hypertranslocation does not present an advantage. In this process, the extent of base-pairing is decreased in comparison to the

active state, making it an energetically less favorable event.

3.1.4 Hairpin-stabilized Pausing

In our approach to modeling hairpin-stabilized pausing, we adopt an allosteric perspective. This view posits that the pause is initiated when an RNA hairpin interacts with a short α helix located at the tip of the RNAP flap domain, which oversees the RNA exit channel. This interaction is crucial in triggering the pause, as elucidated in studies by Touloukhonov et al. and Chauvier et al. [31, 74]. Our model conceptualizes this process as a swiftly reaching equilibrium between two structural states: one where the hairpin is absent and another where the hairpin is strategically positioned near the RNAP flap domain, influencing its functionality. This equilibrium, which shifts rapidly in comparison to the rate at which chemical bonds stabilize the inactive state, is followed by a rate-limiting catalytic step, as depicted in Figure 3.2c. The speed of reaching this equilibrium is considered rapid compared to the formation of chemical bonds that stabilize the inactive state.

We use Equation 3.13 and 3.14 to model the entry rate to the hairpin-stabilized pause,

$$k_{hsp} = k_{on}/(1 + K_{i,h}) \quad (3.13)$$

$$K_{i,h} = \exp\left(\frac{G_{lowest} - G_{hairpin.included}}{k_B T}\right), \quad (3.14)$$

where k_{on} is the catalytic rate of interaction between the RNA hairpin loop and the RNAP flap interaction, and $K_{i,h}$ is the fraction of hairpin formation. Equation (3.14) gives the expression for $K_{i,h}$, which represents the equilibrium among all possible RNA secondary structures. The secondary structure of RNA transcript rapidly transitions among many microstates, and the simulation of transitions among these microstates

is computationally expensive. We bypass this difficulty by simplifying the equilibrium to a two-state system of the lowest energy state and the hairpin-included state.

In scenarios where RNase A is absent, the model assumes that the lowest energy state of the RNA allows for a free folding of up to 100 nucleotides outside the exit channel. To identify a state that includes a hairpin, the model first searches for potential hairpin structures near the exit channel, starting from the 3' end of the RNA and covering up to 30 nucleotides. Following this, the model allows for the free folding of up to 100 remaining ribonucleotides of the transcript. This approach helps in estimating the equilibrium between the lowest energy state and the hairpin-included state, which in turn aids in calculating the fraction of hairpin formation, as illustrated in Figure 3.3.

On the other hand, the presence of RNase A alters this dynamic considerably. With RNase A in the mix, the length of the RNA that can freely fold is reduced to just 15 nucleotides. This constraint can either eliminate existing pause-stabilizing hairpins or lead to the formation of new ones.

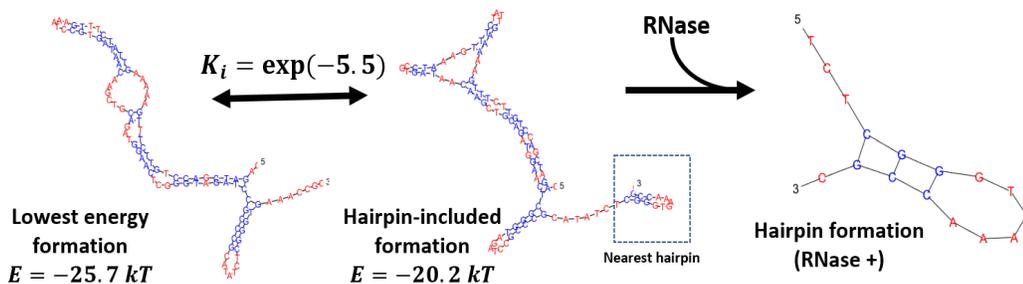


Figure 3.3: Comparison of the lowest energy conformation with one including a proximal (3') hairpin. Hairpin formation is unfavorable at this position without RNase because the hairpin-included structure is a less stable than the lowest-energy formation, denoted by K_i (Left and Middle panels). In the presence of RNase, the length of freely folded RNA is limited to 15 nt. The formation of hairpin is favored because the shortened RNA length ruled out other competing conformations (Right panels).

A chemical bond between the hairpin loop and the RNAP flap is required to stabilize the hairpin-flap interaction. The catalytic rate relates to the length of stem

and loop, and the fraction of G and C in the loop as shown below,

$$k_{on} = k_2 \exp\left(-\frac{D_{stem} * \Delta G_{stem} + D_{loop} * \Delta G_{loop} + F_{GC} * \Delta G_{GC}}{kT}\right) \quad (3.15)$$

where k_2 is the prefactor, D_{stem} and D_{loop} are the deviation from optimal lengths of stem (3 – 8 bases) and loop (4-20 bases), respectively, F_{GC} is the fraction of G and C nucleotides within the loop, and ΔG_{stem} , ΔG_{loop} , and ΔG_{GC} are the energy changes due to D_{stem} , D_{loop} , and F_{GC} .

The exit rate from a hairpin-stabilized paused state (k_{hspr}) must be much slower than the entry rate, and is determined by the rate of RNAP hairpin denaturation. For simplicity, the rate is taken to be a constant.

3.1.5 The Effect of Tension and Transcriptional Factors

In our model, the roles of external tension and transcription factors (TFs) are given significant attention, recognizing their critical influence on transcriptional outcomes. Both tension and TFs are known to modify the energy profile of the transcription complex and interact directly with the transcription machinery, as highlighted in research by Herbert et al. and Touloukhonov et al. [3, 31]. Our model incorporates these factors into the thermodynamics of the TEC, specifically in the pathways of forward translocation and backtracking.

The approach we adopted posits that both the equilibrium constant in the forward translocation step, represented as K_i , and the energy barrier for the backtracking step, denoted as $\Delta G_{n \rightarrow n-1}$, are influenced by the work generated by tension. This concept aligns with the effects of mechanical forces on molecular systems as described by Tinoco et al. [81]. Furthermore, the model takes into account the impact of GreB factors, which are known to affect the dynamics of transcription, as shown in Equations 3.16 and 3.17,

$$K_i^* = \exp\left(\frac{G_{post} - G_{pre} - F * L_{forward}}{k_B T}\right), \quad (3.16)$$

and

$$\Delta G_{n \rightarrow n-1}^* = \Delta G_{n \rightarrow n-1} + \Delta G_{GreB} + F * L_{bt}, \quad (3.17)$$

where G_{pre} and G_{post} are the energy of TEC in pre- and post-translocation state, respectively, $L_{forward}$ and L_{bt} are the effective lengths over which external tension acts in the forward translocation step and in the backtracking step, respectively, and ΔG_{GreB} is the energy barrier change due to GreB factor.

The hairpin-stabilized pause was assumed to be unaffected by any applied tension, since such pauses do not involve the translocation of RNAP. However, as discussed earlier, the presence of RNase A can restrict the length of freely folded RNA transcripts, impacting the formation and stability of hairpin structures within the RNA.

3.1.6 Model Training

To validate and refine our model, we utilized experimental data that encompassed a range of applied tensions, from -7pN to 25pN, and also considered the presence or absence of GreB factors and RNase A. This diverse dataset allowed us to quantitatively assess and integrate the impacts of tension and transcription factors into the model. By fitting the model against these varied experimental conditions, we were able to gauge and incorporate the nuanced effects these external elements have on the transcription process.

It is crucial to recognize that transcription is a process characterized by the involvement of only a very limited number of reactants. This particular aspect renders traditional approaches, such as the chemical law of mass action which assumes large

numbers of molecules, less effective for accurately determining reaction rates in transcription. To address this, our model incorporates two stochastic methods: (i) the continuous-time Markov chain and (ii) stochastic simulation.

The continuous-time Markov chain method is particularly advantageous as it allows for an analytical resolution of the expected time that the transcription complex spends in each state at a given position along the DNA. This approach is adept at handling the probabilistic nature of state transitions in transcription, which is critical given the small number of molecules involved.

On the other hand, stochastic simulation offers insights into the development of individual pausing events. This method is essential for understanding the randomness and variability inherent in transcription at the molecular level. By simulating individual events, we can observe and analyze the nuances of how pauses in transcription are initiated and resolved.

The model is encapsulated in a MATLAB class object, which can generate a predicted residence time histogram with the input of a template sequence and a guess of unknown parameters. This feature allows the model to be ‘trained’ using data derived from real-time, single-molecule experimental setups. We employed time series data from high-resolution optical tweezers transcription experiments conducted by Gabizon et al. [64]. These experiments were notable for their use of the 8XHis DNA template, which incorporates the T7A1 promoter followed by eight consecutive repeats of a 239 bp sequence. This sequence notably includes the *his*-leader pause site along with four other sequence-dependent pause sites, as detailed in the study by Herbert et al. [3]. A significant advantage of this approach is its high temporal resolution, which is adequate to identify pausing events that exceed 100 ms. Considering the transcription rates, which range from 10-20 bp/s, this resolution is effective enough to discern pauses with an accuracy of one base pair, a remarkable feat achieved using optical tweezers. Alignment of the traces under different forces and with dif-

ferent transcription factors generates the residence time histograms (Figure 3.4) as described previously.

3.2 Comparison of the Model with Experimental Data

The analysis of experimental data, under varying conditions with different accessory factors, has clarified the mechanisms behind transcriptional pauses. For example, pauses at position ‘a’ are likely pre-translocation, showing little change with GreB or RNase A addition. In contrast, pauses at ‘b’ respond to these factors, suggesting a mix of backtracking and hairpin-stabilization, as also indicated by backward RNAP translocation [64]. The ‘P1’, ‘d’, and ‘his’ pauses, identified as hairpin-stabilized, nearly vanish with RNase A, while the ‘P2’ pause emerges only in its presence, highlighting a unique mechanism. This data, summarized in Table 3.1, demonstrates how pauses differ in their formation and response to molecular factors.

Table 3.1: Summary of experimental pause positions, durations and mechanisms for 10pN under different transcriptional factor conditions.

Pause	Position of Peak (bp)	Averaged Duration (s)			Associated state(s)
		WT	+GreB	+RNase	
‘a’	9	0.66	0.58	0.64	Pre-translocated
‘b’	34	0.94	1.27	0.59	Backtracked + Hairpin-stabilized
‘c’	66	0.42	0.41	0.38	Unknown
‘d’	94	0.74	0.96	0.33	Hairpin-stabilized
‘his’	161	0.68	0.95	0.25	Hairpin-stabilized
‘P1’	16	0.41	0.40	0.25	Hairpin-stabilized
‘P2’	44	0.16	0.17	0.34	Hairpin-stabilized (with RNase)

We fine-tuned the model parameters, as shown in Table 3.2, to align our model’s output with actual experimental data. This optimization process enabled us to generate a dwell time histogram closely mirroring the experimental observations (see Figure 3.4a and b). The model successfully replicated the positions and durations of the transcriptional pauses identified in the experiments, with the notable exception

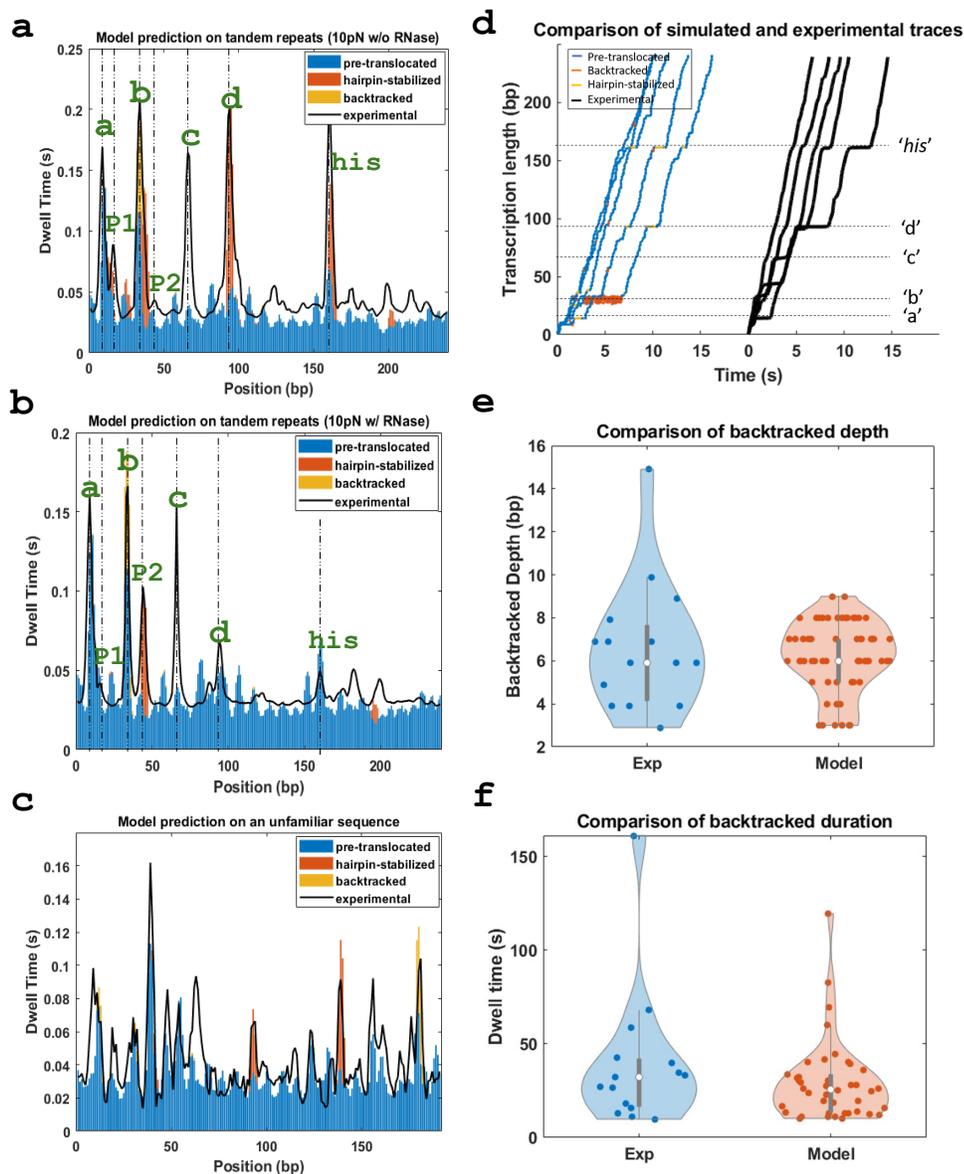


Figure 3.4: Model fitting and prediction. (a) Stacked histogram produced by the model for the condition of 10 pN in presence of RNase. The residence time due to different pausing mechanisms is represented by different colors. The experimental result is shown by the black line. Goodness of fitting is 0.948 for the major pause sites except for 'c' and 0.884 for the overall histogram; (b) Stacked histogram produced by the model for the condition of 10 pN in absence of RNase. Goodness of fitting is 0.959 for the major pause sites except for 'c' and 0.904 for the overall histogram; (c) Predicted histogram by the model on an unfamiliar sequence. Goodness of fitting is 0.871 for the overall histogram. (d) Examples of traces generated by Monte Carlo simulation. The simulated traces show similar pauses at sites 'a', 'b', 'd' and 'his' and generate comparable transcription rates to experimental data. (e) Distributions of backtrack depth observed experimentally and predicted by the model. (f) Distributions of backtrack duration observed in the experiments and predicted by the model.

Table 3.2: Values (95% confidence interval from 100 bootstrapped values) of the optimized parameters under 10 pN assisting force and WT conditions.

Parameters and descriptions		Symbol and Value	Note
Forward	Rate of NTP catalysis	$k_{max} = [85(9), 77(5), 82(9), 41(3)]s^{-1}$	Fitted (k_{max} and K_d values are in order of AUCG)
	Equilibrium constant	$K_d = [34(3), 96(9), 15(2), 26(4)]\mu M$	
	Effective length for forward translocation	$L_{forward} = 0.56(0.07)bp$	
Prefactor of backtracking		$k_1 = 1000s^{-1}$	Fixed
Energy barrier height of first base-pair backtracking		$G_{bt} = 9.8(0.8)k_B T$	
Backtracked pause	Energy barrier height of deeper backtracking	$G_{bt,incre} = 1.8(0.1)k_B T$	Fitted with fixed k_{max} and K_d
	Effective length for backtracking	$L_{bt} = 0.06(0.01)bp$	
	Energy change due to unlikely stem length	$\Delta G_{stem} = Inf$	
Energy change due to unlikely loop size		$\Delta G_{loop} = Inf$	
Hairpin-stabilized pause	Energy change due to GC fraction	$\Delta G_{GC} = 8.8(1.1)k_B T$	Fitted with fixed k_{max} and K_d and backtrack related parameters
	Hairpin-flap interaction rate	$k_{on} = 807(71)s^{-1}$	
	Hairpin denaturation rate	$k_{hspr} = 3.4(0.3)s^{-1}$	
	Allowed RNA-DNA hybrid length	$h = 7 \sim 9bp$	
TEC structure	Allowed upstream spacer length	$u = 1 \sim 3bp$	Fixed range
	Allowed downstream spacer length	$d = 1 \sim 3bp$	
	Allowed number of single-stranded RNA protected by RNAP	$r = 1 \sim 3bp$	

of pause ‘c’. In the following section, we delve into potential explanations for the model’s inability to accurately capture the characteristics of pause ‘c’.

The model not only replicates pause positions and durations but also accurately predicts their mechanisms, as indicated by experimental findings (Figure 3.5a). According to these findings, GreB increases the dwell time at pause ‘b’, a phenomenon the model mimics by altering the backtracking energy barrier [64]. The introduction of RNase A, on the other hand, shortens dwell times at ‘P1’, ‘d’, and ‘his’ sites, extends them at ‘P2’, and does not significantly affect other pauses. Our model accounts for these variations by adapting to shorter RNA lengths, which impacts hairpin stability.

This adjustment results in slower or destabilized hairpin formation at ‘P1’, ‘d’, and ‘his’, while enhancing it at ‘P2’ (Figure 3.3).

The effect of tension is modeled by defining two distinct effective lengths, $L_{forward}$ for forward translocation and L_{bt} for backtracking. As shown in Figure 3.5b and Table 3.2, the effective length for forward translocation is less than one base, while for backtracking, it’s under 0.1 base. These values are consistent with prior research findings [8, 55]. This differentiation suggests that external tension, particularly when opposing transcription, prolongs pauses by reducing transcription rates and encouraging entry into backtracked pausing. Moreover, it reinforces the notion that long-lived pauses, such as backtracked pauses, often succeed shorter-lived pauses.

The model’s ability to accurately forecast significant pauses in the transcription of an unfamiliar 200 base sequence highlights its strong predictive capacity. This particular sequence, located ahead of the repetitive region of the 8XHis template, was not part of the data set used for parameter optimization. Despite this, as depicted in Figure 3.4c, the model effectively identifies the primary pauses occurring near bases 15, 45, 140, and 180. These predictions align closely with experimental findings which are achieved through the alignment of transcription records and the analysis of dwell time histograms. This capability of the model to reliably predict transcriptional behavior in sequences not previously encountered underscores its robustness and utility in understanding transcription dynamics.

To further test the validity of the model, we employed Monte Carlo simulations to produce an extensive set of transcription traces. These simulated traces allowed us to delve into the dynamics of backtracking, offering a detailed comparative analysis with experimental traces. Specifically, we focused on the backtracking characteristics at pause site ‘b’, examining both the depth of backtracking and the duration of these pauses, as illustrated in Figure 3.4d-f. The agreement between simulated and experimental results lends further support to the model.

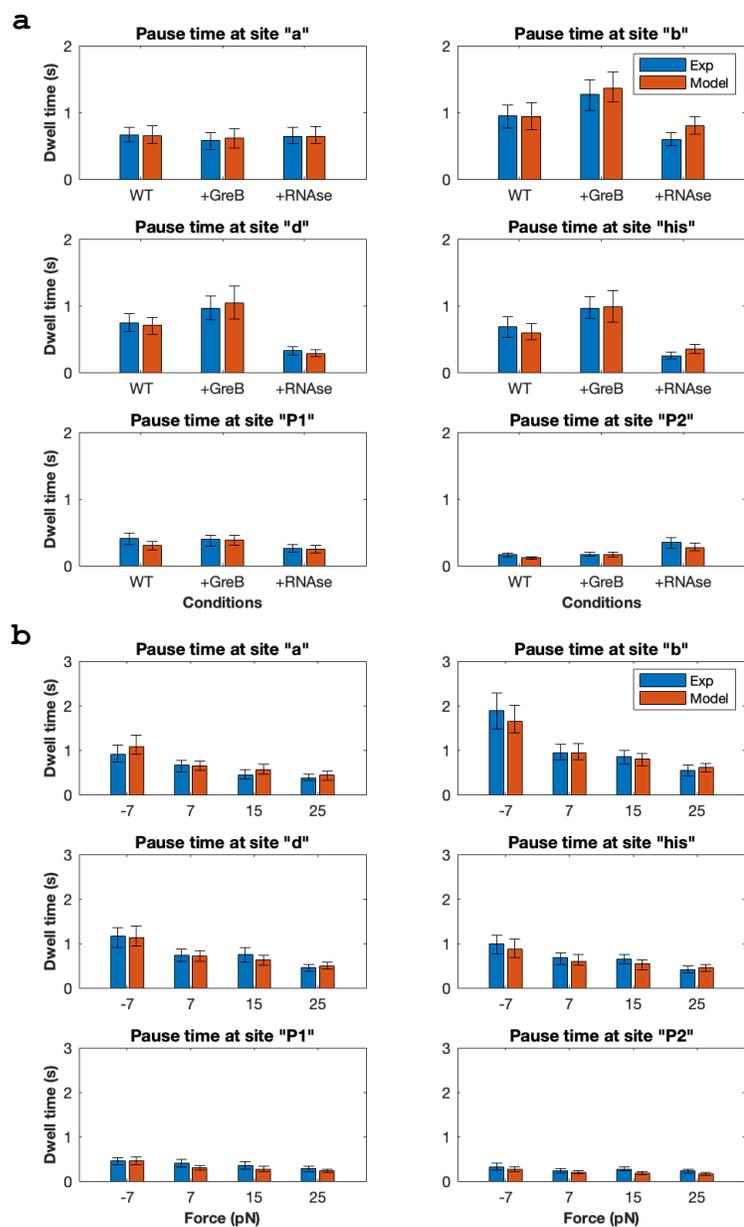


Figure 3.5: Averaged dwell times from experiments (blue) and model (red) at pause sites. (a) With various transcriptional factor conditions under 10pN assisting tension and (b) WT condition under different tensions. Error bars are the 25th and 75th percentile of 100 bootstrapped values.

3.3 Strengths and limitations of the model

In our model, short pre-translocated pauses, also known as ubiquitous or elemental pauses, are characterized as moments of slowed forward translocation, aligning with the Michaelis-Menten Brownian-ratchet model (Equation 3.5). These are considered on-pathway pauses. Contrarily, some studies categorize short pauses as off-pathway events, diverging from the active translocation path. However, our model finds a strong correlation at elemental pause ‘a’ and accurately predicts slower translocation rates at other prolonged pause sites when the parameters K_d and k_{max} are optimized in Equation 3.5. This observation leads us to treat an on-pathway elemental pause state as a viable representation for these short-lived pauses. Despite this, our model, like others, cannot conclusively determine whether elemental pauses are inherently on- or off-pathway. This is primarily due to the challenge in distinguishing between genuinely slow translocation and actual pausing events. It’s worth noting that our model demonstrates a higher degree of concordance between experimental data and longer-lived pause sites, compared to short-lived pause sites, as shown in Figure 3.4. This suggests that while the model is effective in capturing the dynamics of longer pauses, the nature and categorization of shorter pauses remain areas for further exploration and refinement.

There’s a noticeable discrepancy between the values of K_d and k_{max} reported in previous studies [7] and those derived from the experimental data we examined. This variation could stem from the methodologies used in earlier models, which perhaps did not precisely localize pauses during transcription. Alternatively, it might suggest that the on-pathway state, as we have modeled it, does not entirely capture the nuances of elemental pauses. However, the strong fit and predictive accuracy of our model imply that an on-pathway system, based on the Michaelis-Menten equation, has sufficient complexity to effectively represent elemental pauses. Given the possibility that off-pathway events could involve intricate rearrangements within the

RNAP active site, our approach to model fitting addresses the challenge of depicting off-pathway elemental pauses [28]. Notably, using our adjusted parameter values – but not those previously reported – the M-M equation predicts a translocation rate of 3.4 bp/sec at a consensus elemental pause site, as identified through NET-seq [44]. This agreement lends additional support to our method of fitting K_d and k_{max} for accurately identifying pause sites.

Our model effectively identifies and elucidates the mechanisms underlying transcriptional pausing. Consistent with previous theoretical analyses, such as the one by Artsimovitch et al. [5], our results reinforce the concept that long-lived pauses often evolve from shorter, more common pauses. For instance, at pause site ‘b’, the model indicates a preference for backtracking rather than forward translocation due to a lower rate of the latter. It predicts similar backtracking rates at both the 35 bp site (pause ‘b’) and the 190 bp site, but the quick forward translocation at the 190 bp site reduces the likelihood of backtracking. Utilizing the well-established Michaelis-Menten framework, we determined that the rate of forward translocation along the DNA template can vary significantly, ranging from under 3 nt/s to as high as 70 nt/s. This variability implies that the transcription complex, when transcribing slowly, is more likely to enter a prolonged pause at a certain site, even if the energy barrier for backtracking at this site is greater than at another site with a higher transcription rate. This aspect of our model highlights the dynamic interplay between translocation speed and pause development.

Our model estimates that the effective length impacted by tension during forward translocation is approximately half of a base pair, while for the backtracking pathway, it’s much smaller, less than 0.1 base. This finding suggests that the forces typically exerted during transcription have a minimal effect on the backtracking rate. When RNAP backtracks, it moves backwards along the DNA, disturbing the RNA-DNA hybrid near the active site. We theorize that this backtracking rate is largely governed

by the process of unwinding the most recently formed base pair. Therefore, external forces seem to have a limited role in altering the backtracking process, as opposed to their more significant impact on shifting the equilibrium constant in the forward translocation pathway.

The occurrence of a hairpin-stabilized pause is contingent upon an interaction between a transcript hairpin and the RNAP flap domain. While previous models have simulated nascent transcript folding using a lowest-energy approach [8, 9, 82], this method may not accurately pinpoint hairpin locations. This is because RNA folding happens concurrently with transcription and might not always achieve the lowest-energy state. Moreover, simulating co-transcriptional RNA folding is computationally intensive. To address these challenges, we have developed a novel method that evaluates the stability difference between a structure with a hairpin and the lowest-energy structure. This approach helps estimate the likelihood of hairpin formation more effectively. For instance, while hairpins at positions 101 and 178 are stable, they are less likely to interact with RNAP compared to the less stable hairpins at positions 94 and 161, corresponding to pauses at sites ‘d’ and ‘his’, respectively (Figure 3.6). Additionally, our method efficiently captures the dynamics of pause ‘P2’, which becomes significantly elongated in the presence of RNase, as shown in Figure 3.3. This new approach thus offers a more accurate representation of hairpin formation during transcription, enhancing the model’s predictive power.

It’s important to acknowledge that the current model might not account for all possible paused states. Notably, it doesn’t adequately explain pauses at site ‘c’ and other less prominent sites. The pause at ‘c’, for example, shows little response to the addition of GreB or RNase, indicating a pausing mechanism different from backtracking or hairpin-stabilization, which our model currently emphasizes. Recent research by Janissen et al. [67] identified three interlinked paused states in transcription: an elemental pause, a backtracked pause, and a backtrack-stabilized pause. Our model

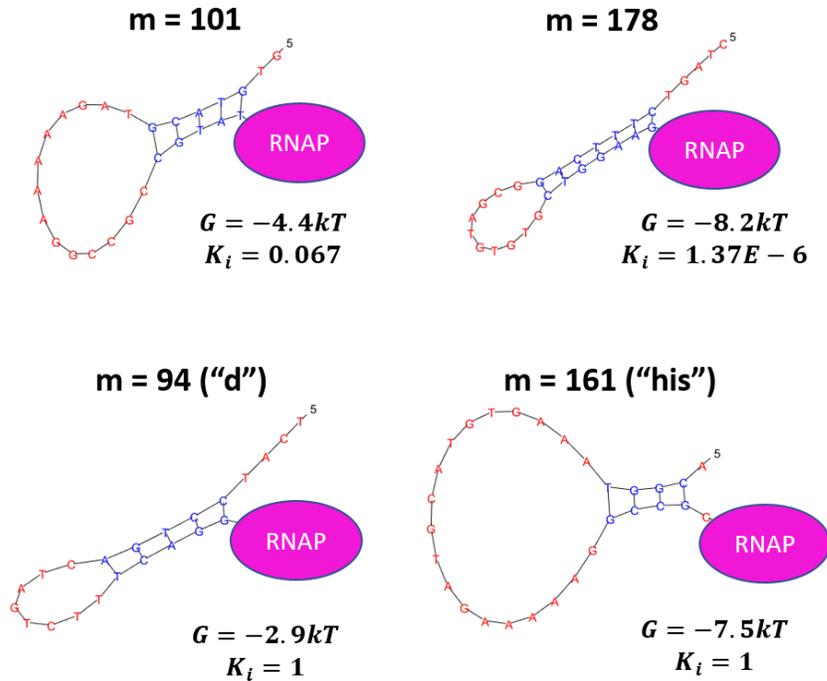


Figure 3.6: Comparison of energy and $K_{i,h}$ at different positions. Hairpin formation is unfavorable at position 101 and 178, although the hairpin structures at these positions are fairly stable. While at position 94 and 161, the hairpin structures can readily form and induce the hairpin-stabilized pauses.

currently does not include the backtrack-stabilized state. This exclusion is partly because the data used, derived from a 239 bp DNA sequence with tandem repeats, does not exhibit the extremely long pauses (around 100 seconds or more) typically associated with backtrack-stabilized states, as characterized by Janissen et al.

Our thermodynamically focused approach to modeling the transcription complex has proven adept at replicating transcription kinetics. The model's incorporation of both Class I and Class II pauses enhances our understanding of both the active and branched pathways in transcription, allowing for the prediction of these pause classes which play key roles in transcription regulation. By adapting specific kinetic parameters to match experimental conditions, the model not only aligns statistically with experimental outcomes but also provides detailed insights into the impacts of DNA sequences, external tension, and transcription factors.

There is, however, room for further enhancement. Advancements in our biochemical knowledge of transcriptional pauses, alongside improvements in experimental data quality and refinements in the model itself, could significantly boost its predictive capabilities. For instance, the model's ability to predict pauses at site 'c' could be greatly improved if the specific mechanism of this pause is identified and incorporated into the model. Additionally, the availability of extended high-resolution transcription data, offering a broader range of sequence variations, would facilitate more effective optimization of the model, leading to more accurate predictions. Such developments would make the model an even more powerful tool for understanding the complex processes governing transcription.

Chapter 4

Single-molecule Insights into Transcriptional Trafficking and Regulation of Gene Expression

Building upon our exploration of transcription kinetics at the base-pair level in previous chapters, which primarily involved interactions between RNAP and DNA, it's essential to consider the impact of higher-level factors on gene expression. These elements encompass the DNA's topology, interactions between RNAP and other proteins, environmental stimuli, and the cell's overall physiological state. Such high-level factors exert their influence through a variety of pathways. For instance, research conducted in our lab has shown how DNA topology can play a critical role in transcription regulation [83, 84]. In particular, we found that transcription can be suppressed within DNA regions stabilized by supercoiling. This suppression is due to the physical constraints imposed by the DNA's structural form, which can affect the accessibility and movement of transcriptional machinery. Additionally, the action of repressor proteins provides another layer of gene expression control [85]. These proteins can specifically block RNAP from binding to certain promoters, thus inhibiting the tran-

scription of targeted gene sequences. This mechanism is crucial in the cell's ability to selectively express genes in response to various needs and conditions. Moreover, extracellular signals like insulin or adrenaline are known to trigger changes in gene expression, particularly concerning genes involved in metabolic processes and stress responses [86, 87]. Such signals often initiate cascading effects that lead to alterations in transcriptional activity, highlighting the interconnected nature of cellular signaling and gene expression. These multi-tiered regulatory mechanisms collectively ensure efficient and appropriate gene expression and protein synthesis. They exemplify the complexity of transcriptional control, extending far beyond the interactions at the DNA-RNAP level to include a broader range of biological and environmental factors. Understanding this intricate interplay is crucial for a comprehensive grasp of gene expression regulation.

In this chapter, we delve into the complexities of transcriptional regulation facilitated by protein trafficking within cells. As the transcription machinery navigates the genome, it encounters various molecular “roadblocks” that can significantly influence both the efficiency and fidelity of transcription. These challenges encompass a diverse range of molecular interactions and structural impediments:

- **DNA Damage:** Factors like UV radiation, chemicals, or replication errors can damage DNA. Encounters with damaged DNA by RNA polymerase during transcription may result in transcriptional errors or stalling.
- **DNA-Binding Proteins:** A variety of proteins that bind to DNA can obstruct RNAP's progression along the DNA strand. These proteins serve different functions but share the potential to physically impede transcription.
- **Chromatin Structure:** In eukaryotes, DNA is wrapped into chromatin, a complex of DNA and proteins. The degree of chromatin condensation influences gene accessibility; for instance, tightly packed heterochromatin is typically less

transcriptionally active.

- **Regulatory Proteins:** These include transcription factors that can either facilitate or inhibit transcription. They play a crucial role in determining the spatiotemporal dynamics of gene expression.
- **R-loops Formation:** These structures occur when a newly synthesized RNA strand hybridizes with its DNA template, forming a three-stranded configuration. R-loops can interfere with transcription and contribute to genomic instability.
- **Supercoiling of DNA:** Transcription induces changes in DNA topology, leading to either overwinding or underwinding in regions adjacent to the RNA polymerase. This supercoiling can have significant effects on the transcription process.

Cells have mechanisms to deal with these challenges, such as DNA repair pathways, chromatin remodeling, helicases and topoisomerases that resolve R-loops and DNA supercoiling. Among these mechanisms, some are well studied while others remains insufficiently explored. Here, we dedicate to the regulation caused by DNA-binding proteins, which includes a wide range of roadblock proteins such as DNA-binding transcription factors, replication/transcription proteins, nuclease, repair proteins, chromatin remodelers, etc.

4.1 DNA-binding proteins organize DNA and modulate transcription

The nucleoid in bacterial cells, a densely packed structure of DNA, RNA, and proteins, exhibits remarkable efficiency in condensing the bacterial genome. Genomic sizes in

bacteria range widely, from as small as 0.1 Mbp in certain symbionts to as much as 14 Mbp in some myxobacteria [88, 89]. Despite this variability, the genomic material is effectively condensed to occupy just about 15-25% of the cell's volume [90, 91], a stark contrast to its potential uncondensed volume, which could exceed the cell's volume by a thousandfold. The compact nucleoid is nestled within a cytoplasm rich in proteins and RNA [92].

A diverse array of DNA-binding proteins plays a crucial role in shaping the structure of the bacterial nucleoid. These proteins, which include HU, IHF, Fis, H-NS, StpA, Dps, Lrp, CRP, MukBEF, and MatP, bind to the genome with varying specificity and are involved in functions like DNA bending, wrapping, and bridging [93]. Their interactions with DNA can vary from simple binding to more intricate actions like looping DNA or stabilizing complex structures such as plectonemes [94, 95, 96]. The full spectrum of these proteins' roles in DNA organization and gene expression regulation is still being explored.

These proteins exert regulatory influence in several ways. A fundamental method is by occluding promoters, thereby preventing RNAP from initiating transcription [97]. For instance, in *E. coli* and similar bacteria, chromatin proteins like H-NS may effectively compete with RNAP for binding at the σ 10 element of promoters, especially in the absence of activators, due to the similarity between the consensus σ 10 sequence and these proteins' binding motifs [97]. Chromatin proteins can also inhibit transcription factor binding to maintain gene silencing. Fis and H-NS, particularly well-studied in this regard, are known to occlude RNAP binding and regulate genes crucial for virulence and stress response [98, 99, 100, 101, 102] .

DNA-binding proteins can significantly influence transcription by modifying DNA topology, with their effects manifesting in the regulation of transcription initiation, elongation, and termination. This influence is due to the supercoiling generated by transcription, both positive (+) and negative (-) ahead of and behind TECs,

respectively. Chromatin proteins can either alleviate or intensify these supercoils, consequently assisting or hindering transcription. For instance, Fis [103, 104] and HU [105, 106, 107] proteins are known to facilitate transcription by impacting supercoiling, whereas H-NS [108, 109] can induce transcriptional pauses by bridging topological domains.

During transcription through chromatin, RNAP inevitably encounters DNA-binding proteins. Although such encounters are frequent, only a select few proteins linger on DNA long enough to substantially impede transcript elongation. Notable examples include catalytically inactive CRISPR Cas9 [110], *E. coli lac* repressor (LacI) [111, 112], *E. coli* GalR [113], *B. subtilis* CodY [114], and *B. subtilis* CcpA [115, 116]. Biochemical studies using these roadblocking proteins have delineated the prerequisites for effective transcriptional obstruction and several bypass mechanisms. RNAP can navigate past a roadblock by several means, including transcribing around the protein (as seen with eukaryotic nucleosomes [117]), actively displacing the roadblocking protein, or advancing forward when the protein temporarily detaches from DNA. The efficiency of these mechanisms can be influenced by a range of factors (illustrated in Figure 4.1), such as:

1. The concentration and binding dynamics of the roadblocking protein: Higher concentrations and slower dissociation rates of the roadblocking protein can increase the likelihood of transcriptional interference.
2. The number of RNAPs simultaneously transcribing a gene: Multiple RNAPs can cooperatively translocate along the DNA, potentially overcoming roadblocks more effectively.
3. The stability of backtracked or arrested RNAP at the roadblock: The more stable the halted RNAP, the harder it is for transcription to proceed.
4. The presence of ribosomes on the nascent RNA: Ribosomes can prevent RNAP

backtracking and facilitate elongation.

5. The action of proteins that alleviate topological stress: These proteins can reduce supercoiling-related transcription challenges.
6. The presence of transcription rescue proteins like Mfd or Gre factors: Such proteins can help RNAP bypass or overcome roadblocks.

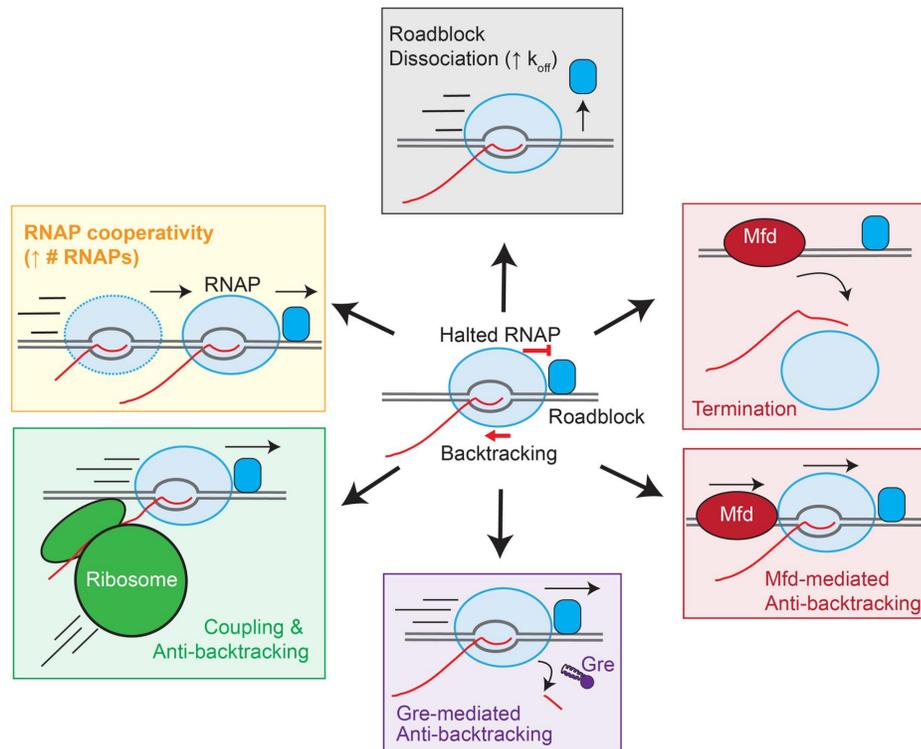


Figure 4.1: Mechanisms to overcome chromatin roadblocks (Adapted from [93]). Strong protein roadblocks (blue rounded rectangle) can induce RNAP pausing, leading to backtracking (center; red “T” barrier). At least five mechanisms exist to help RNAP escape backtracking and transcribe past the roadblock (black arrows): roadblock protein dissociation, high EC density at the roadblock, transcription–translation coupling, Gre factors, and Mfd. A higher DNA-binding protein off-rate will decrease the roadblock strength (black box). Trailing ECs can help the leading EC transcribe through a DNA-binding protein (yellow box). During coupling, ribosomes inhibit EC backtracking (green box). Gre factors stimulate cleavage of the backtracked RNA to restore an active EC (purple box). Mfd translocates on DNA and binds RNAP. Mfd can either help ECs through a roadblock by preventing backtracking or, if the roadblock is strong, can dissociate RNAP from DNA (i.e., terminate transcription; red boxes).

4.2 Single-molecule Techniques to Study *in vitro* Transcription

Transcription processes can be studied using a variety of techniques, each offering distinct insights into different aspects of the mechanism. Tethered particle motion (TPM) is one such technique, where the movement of a bead attached to DNA is related to the DNA's contour length. TPM is particularly useful for observing loop formation, breakdown, and transcription under conditions of minimal tension, as seen in Figure 4.2A. This method has been a valuable tool in various transcription studies [118, 119, 120, 121, 122, 123, 124]. Atomic Force Microscopy (AFM), shown in Figure 4.2B, offers nanometer resolution and is adept at providing detailed static snapshots of transcription-related structures [125, 126]. However, its capability for dynamic observation is limited. Magnetic Tweezers (MTs), depicted in Figures 4.2C-E, can apply a wide range of forces from sub-picoNewtons to about a hundred picoNewtons. MTs are unique in their ability to conduct simultaneous measurements of multiple molecules, although they offer lower temporal and spatial resolution compared to optical tweezers [127, 128, 129, 130, 131]. Optical Tweezers (OTs), illustrated in Figures 4.2F-H, are highly suitable for dynamic studies due to their high temporal (over 1000Hz) and spatial (less than 1nm) resolution. The limitation of OTs lies in their restriction to single-molecule manipulation [132, 133, 134]. Other methods include nanofluidic confinement, which is beneficial for investigating DNA and transcription complexes in confined spaces [135], and *in vivo* single-molecule imaging, allowing the observation of transcription within living cells [136, 137]. The choice of technique depends on several factors. TPM, for instance, is tension-free, while AFM and OTs can apply significant force. The decision is also influenced by whether high-resolution dynamic data is sought (as with OTs) or if the study requires observing many molecules at once, albeit with lower resolution (as with MTs). Each method presents its own

advantages and constraints, which should be considered in relation to the study's specific goals.

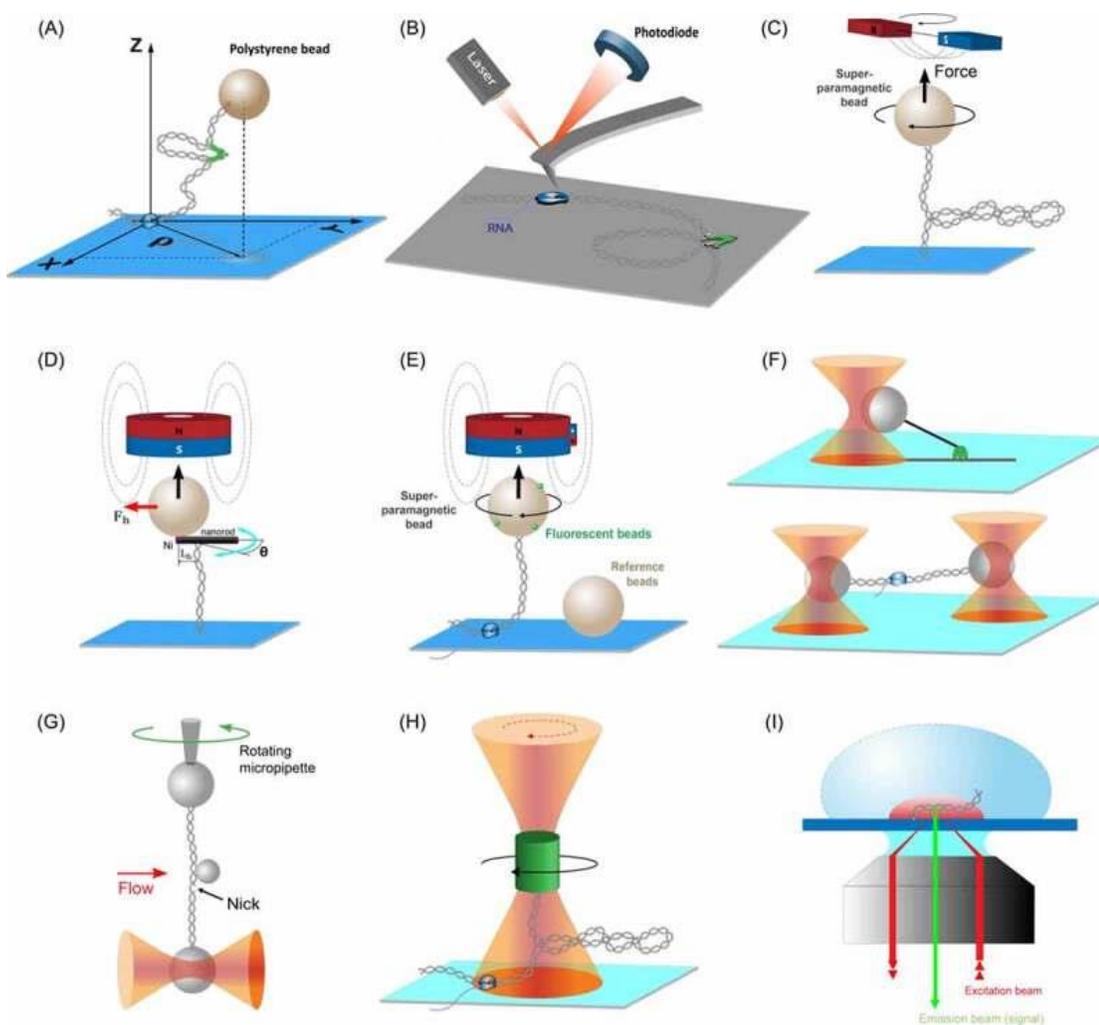


Figure 4.2: Single molecule techniques used to study transcription, transcription through roadblocks, or against torsion. The tethered particle motion technique (a), atomic force microscopy (b), magnetic tweezers (c), magnetic tweezers with nanorod (d), magnetic torque tweezers (e), single and dual trap optical tweezers (f), optical tweezers and rotating pipette (g), angular optical trap (h), total internal fluorescence and FISH (i)

Previous single-molecule studies in our group, utilizing AFM and biochemical techniques, have revealed that the formation of a loop mediated by the *lac* repressor can block transcription elongation more effectively than the repressor binding at a single site [112]. These experiments were conducted without torsional constraints on the DNA. More recently, it has been shown that elongation across a few hundred base

pairs takes longer when the DNA is looped. This delay may be alleviated by positive supercoiling generated during transcription, which likely aids in displacing the *lac* repressor to exit the loop [83]. The influence of supercoiling on the binding affinity of nucleoid-associated proteins (NAPs) and other transcription factors has also been documented [138, 139].

Pertinent questions about RNAP traversing roadblocks include whether RNAP actively dislodges these obstacles or simply capitalizes on their spontaneous dissociation [140, 141, 142]. There's also curiosity about whether the activity of RNAP and the binding affinity of roadblocks are influenced by tension in the DNA template. Definitive answers are lacking, but several studies have started to delineate the 'rules of the road' for traffic on DNA. For instance, in eukaryotes, diffusion of DNA loops around nucleosomes has been observed to facilitate repositioning of the histone octamer roadblock [143]. However, such a mechanism is unlikely for site-specific repressors as discussed by Vörös et al. [112]. The widely accepted Brownian-ratchet model of transcript elongation posits that single base pair sliding movements interconvert the elongation complex between pre- and post-translocated states. Therefore, conditions favoring NTP binding might increase the pressure of RNAP on DNA-binding roadblocks, potentially dislodging them [144]. Research has shown that increasing the binding affinity of a roadblock (decreasing the off-rate) creates a more formidable obstacle than simply increasing the concentration of roadblocks (increasing the on-rate), hinting that roadblock transit likely involves their dissociation [111]. Moreover, a burst of transcription from a strong promoter can enhance transit through a roadblock, suggesting cooperation among transcription elongation complexes and supporting the active dislodgement mechanism [142, 144, 145]. However, these findings don't entirely rule out the possibility of spontaneous dissociation of roadblocks and subsequent reactivation of stalled RNAP complexes by trailing RNAPs.

In summary, while initial investigations have shed light on how RNAPs navigate

various roadblocks in some contexts, a universal understanding of the kinetics and dynamics of these interactions remains out of reach. A key question is how these diverse mechanisms seamlessly collaborate to ensure precise transcription through various roadblock proteins. The following chapter aims to delve deeper into this question, using single-molecule magnetic tweezers assays to uncover the complex interplay of these interactions and their overall impact on transcription fidelity.

Chapter 5

Magnetic Tweezers Experiments Revealed A Hybrid Mechanism for RNAP Transiting through Roadblocks

In our quest to unveil the fundamental principles that guide RNAP movement through transcriptional roadblocks, we employed magnetic tweezers to observe the progression of *E. coli* elongation complexes (ECs) on DNA templates in the presence of specific roadblock proteins. These proteins included the *lac* repressor (LacI), bound at its operator sites with varying affinities, and a mutant variant of the EcoRI endonuclease, EcoRI Q111, which binds to DNA but lacks cutting activity. Our experiments were designed to apply forces up to 5 piconewtons, either opposing or aiding RNAP translocation, and were conducted with and without the presence of GreA, the primary factor involved in resolving backtracking in *E. coli*.

Magnetic tweezers, particularly in their multiplexed form, proved to be an excellent tool for capturing a substantial number of instances where RNAP navigates

past these roadblocks. This large dataset allowed us to analyze and understand the dynamics of such a probabilistic process in detail. In this chapter, we will begin by detailing the experimental setup used for these observations. Following this, we will summarize the key findings from these experiments. To conclude, we will introduce a model that effectively explains the observed results, shedding light on how RNAP interacts with and overcomes transcriptional roadblocks. This model not only elucidates the specifics of our experiments but also contributes to the broader understanding of transcriptional regulation in the presence of DNA-binding proteins.

5.1 Magnetic Tweezers Assays

Figure 5.1 illustrates the standard setup for magnetic tweezers, an essential tool in the study of molecular interactions at the single-molecule level. Central to this setup is a magnet positioned strategically above a flow cell, which is mounted on an inverted microscope. The magnetic field, critical to the operation of magnetic tweezers, is typically generated using a pair of permanent magnets. However, alternative configurations employing electromagnets [146, 147, 148] or utilizing the near-field effect from a single permanent magnet [149] are also prevalent in various experimental designs.

In this system, a paramagnetic bead placed within the magnetic field becomes magnetized, experiencing a force that is directly proportional to the gradient of the applied magnetic field. By adjusting the position and strength of the magnets, forces ranging from sub-picoNewton to over 100 picoNewtons can be applied with precision to paramagnetic beads, typically ranging in diameter from 1 to 3 micrometers [150, 151]. The design allows for the magnets to be placed conveniently outside the flow chamber, facilitating ease of control and adjustment.

A distinctive feature of the magnetic tweezers setup is the large characteristic length scale over which the magnetic field gradient changes, usually about 1 millime-

ter. This means that the force exerted on the bead remains approximately constant over the distance it moves, making magnetic tweezers an effective infinite bandwidth force clamp. This design eliminates the need for the complex active force feedback mechanisms required in optical tweezers [152], simplifying the experimental setup. While this feature of magnetic tweezers may not be pivotal in studies focused exclusively on the dynamics of RNAP navigating through transcriptional roadblocks, this unique advantage of MTs has enabled the observation and study of transcription recycling, a process that had previously gone unnoticed, in Chapter 6.

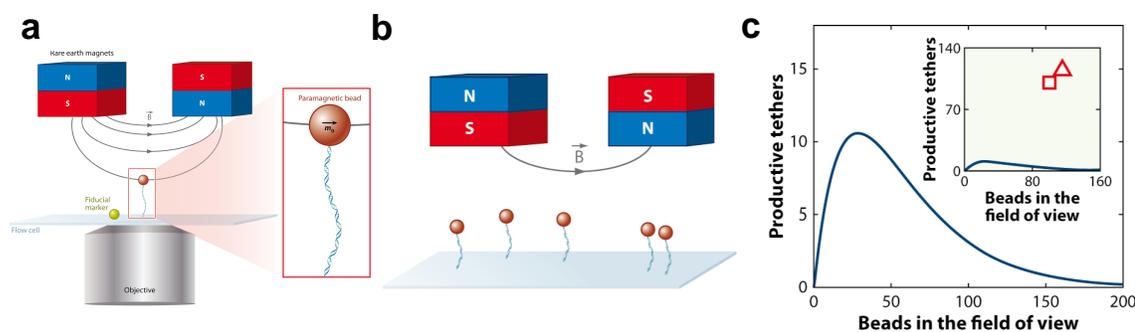


Figure 5.1: Illustration of Magnetic Tweezers (Adapted from [153]). (a) Schematic of basic implementation of magnetic tweezers. A molecule is tethered between the surface of a flow cell and a paramagnetic bead. The magnetic field generated by a pair of magnets induces a magnetic moment m_0 in the paramagnetic bead. The bead experiences a force proportional to the gradient of the field. The molecule can be coiled by rotating the external magnet. (b) Both the magnetic-force-based manipulation and the camera-based detection used in magnetic tweezers are compatible with multiplexed measurements. (c) A random distribution of DNA-bead tethers leads to poor occupation of the field of view space. More than 10 productive DNA-bead tethers can be found in a $100 \times 100 \mu\text{m}$ field of view (criterion chosen here: nearest neighbor spacing $>10 \mu\text{m}$). Abbreviations: N, magnetic north pole; S, magnetic south pole.

5.1.1 GreA and forces opposing or assisting RNAP translocation change pausing at roadblocks

In our study, we quantified pauses encountered by RNAP during transcription elongation at specific protein roadblocks using carefully designed DNA templates. These templates were labeled with digoxigenin and featured a T7A1 promoter, a binding site for either the LacI protein at different operators - Os ($K_d = 10 \text{ pM}$), O1 ($K_d =$

0.05 nM), or O2 ($K_d = 0.1$ nM), or the mutant EcoRI endonuclease, EcoR1 Q111 ($K_d = 5$ pM) [154, 155, 156]. Each template also included a λ T1 terminator, as illustrated in Figure 5.2A.

To conduct these measurements, we used biotin-labeled RNAP holoenzyme attached to streptavidin-coated magnetic beads. These complexes were introduced into flow chambers containing the tethered DNA templates, and transcription elongation was monitored using a magnetic tweezer microscope setup, depicted in Figure 5.2B.

In this setup, whether the force opposed or assisted transcription depended on which end of the DNA template was labeled with digoxigenin and anchored to the glass surface of the flow chamber. The magnitude of the external force applied to the RNAP was determined by the distance between the permanent magnets positioned above the flow cell. By adjusting this separation, we could effectively control the force exerted on the RNAP, allowing us to analyze its behavior under different mechanical conditions.

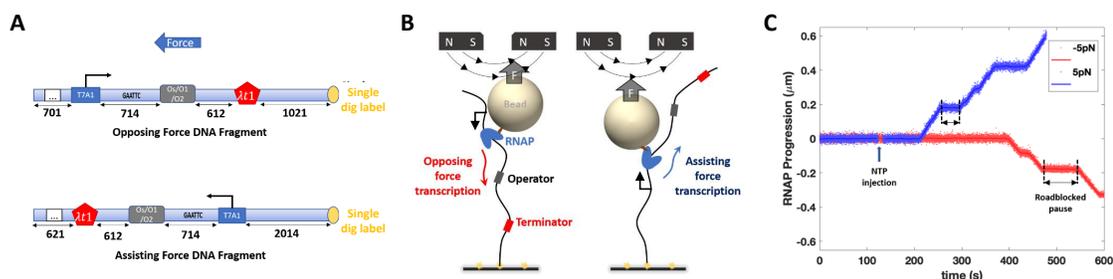


Figure 5.2: Features of the real-time experiments. **(A)** DNA templates for opposing and assisting force experiments have identical transcribed sequences. The numbers indicate distances in base pairs. **(B)** A schematic illustration shows force opposing (left) or assisting (right) transcription. **(C)** Representative records of transcription template length as a function of time under opposing (red) and assisting (blue) force conditions show pauses at LacI roadblock sites (bracketed by black dashed lines).

When we examined transcription elongation in the presence of LacI, we observed a distinct pattern: ECs tended to pause near the LacI operator sites, indicative of the LacI protein binding there. These ECs eventually managed to transit past this

roadblock, as shown in Figure 5.2C. Notably, the pauses associated with the LacI roadblock were significantly longer than more random, ubiquitous pauses encountered during transcription. We pinpointed these roadblock-specific pauses as occurring within ± 20 nanometers (about 60 base pairs) of the LacI binding site.

A small subset of the transcription traces, less than 10%, did not display any pauses associated with the roadblock. This observation likely reflects instances where the roadblock proteins did not fully bind to the DNA tethers. For our analysis, we distinguished these roadblock-associated pauses from shorter, more ubiquitous pauses by setting a threshold: pauses shorter than 20 seconds were categorized as ubiquitous and were thus not included in our roadblock-focused analysis.

To ensure the validity of our observations, we conducted control experiments in the presence of heparin. Heparin is known to bind free proteins in solution; hence, its presence would sequester any unbound or dissociated roadblock proteins. This step was crucial to confirm that re-binding of the roadblock proteins to the DNA did not influence the observed pause durations (Figure 5.3).

For templates featuring O1 or O2 sites, we observed that nearly all ECs were able to successfully navigate past these roadblocks within an hour. The distribution of pause times in these scenarios typically followed an exponential pattern under all tested conditions, as shown in Figure 5.4A and C. An exception was noted for LacI-O2 under assisting force conditions, which will be discussed subsequently. Intriguingly, we found that the duration of pauses was influenced by the direction of the applied force (either assisting or opposing transcription) but was not significantly affected by the magnitude of that force, as depicted in Figure 5.4B and D.

As anticipated, the LacI-O1 roadblocks, possessing intermediate affinity, resulted in longer pause durations compared to the lower-affinity LacI-O2 roadblocks. On templates that included the high-affinity, artificially created symmetric binding site Os, we noted that a fraction of ECs were indefinitely paused at the roadblock (Figure

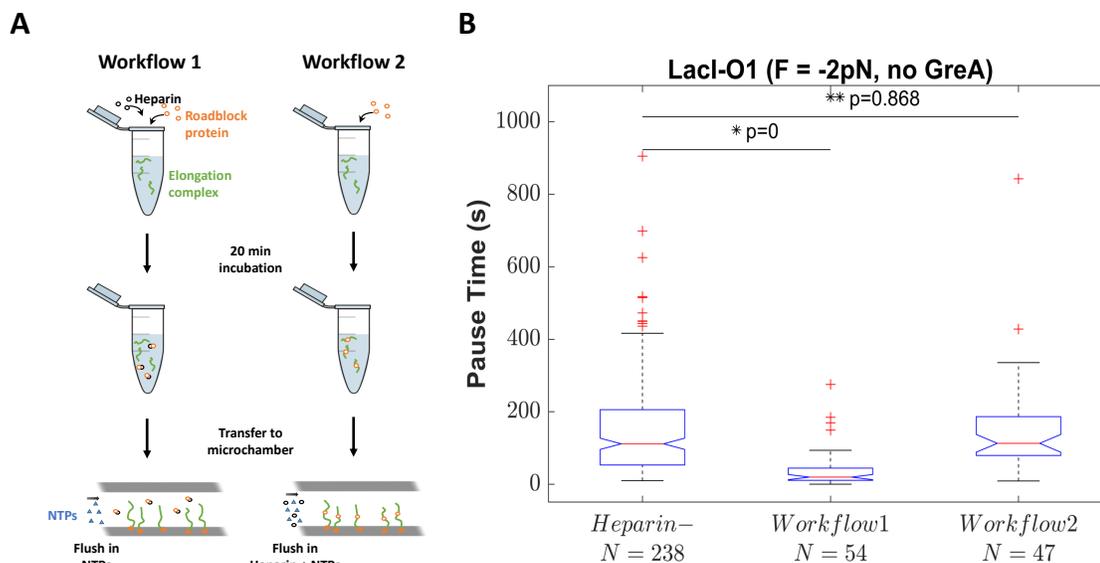


Figure 5.3: Experiments with heparin suggest roadblock rebinding negligibly affects pause times. (A) Experiments with heparin were performed to determine whether roadblock proteins might dissociate but rapidly re-associate at DNA binding sites to extend pauses. (Left) In Workflow 1 (positive control), heparin at a concentration of 10 $\mu\text{g}/\text{mL}$ was introduced with LacI to diminish roadblock formation before adding NTPs to initiate transcription. (Right) In Workflow 2, after incubation to produce LacI roadblocks on stalled EC-DNA templates, the same concentration of heparin was introduced with NTPs to sequester LacI dissociating during transcription. (B) Shorter pauses measured following workflow 1 (middle column) versus no heparin (left column) indicate that heparin diminished the formation of LacI-DNA roadblocks. Using the same heparin concentration to sequester LacI dissociating during transcription did not significantly shorten pauses at roadblocks formed via workflow 2. This indicates that re-association of LacI at roadblock sites did not influence pause measurements.

5.5A). This observation led us to analyze both the lifetimes of these pauses (Figure 5.5B) and the percentage of RNAPs that successfully transited past the LacI-Os roadblock (Figure 5.5C). Similar to the findings with LacI-O1/O2 roadblocks, the pause durations and the rates of successful transit at LacI-Os roadblocks varied with the direction of the applied force. However, again, the actual force magnitude had minimal impact.

The phenomenon of backtracking during transcription, particularly when ECs encounter roadblocks, is a well-documented occurrence [142, 157]. Backtracking involves the reverse movement of ECs, potentially leading to pauses in transcription. The application of force that assists EC translocation may prevent backtracking or facilitate

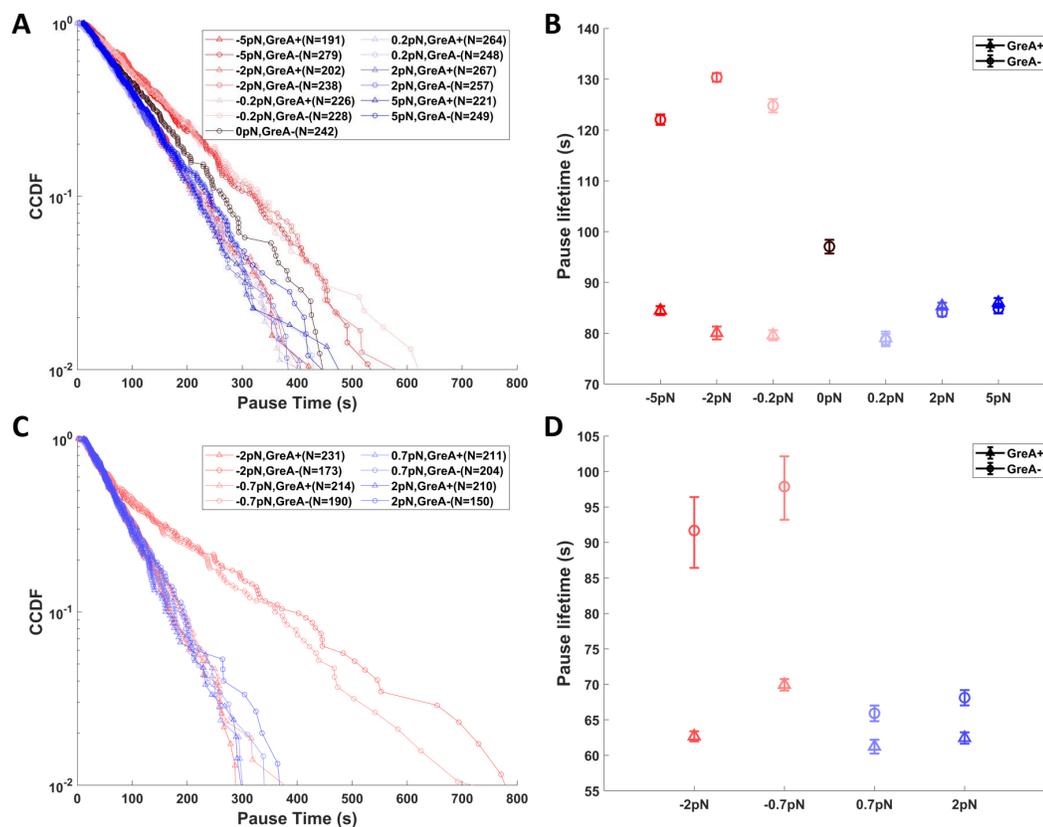


Figure 5.4: Pauses at LacI-O1 and LacI-O2 roadblocks under conditions of opposing (red) or assisting (blue) force and with (triangles) or without (circles) GreA. (A) The complementary cumulative distribution function (fraction of pauses longer than a given time, CCDF) and (B) characteristic times of pauses at LacI-O1 roadblocks shows that the longest pauses were associated with opposing force without GreA, followed by shorter pauses with no force, and even shorter pauses with assisting force or opposing force with GreA. (C) The CCDF and (D) characteristic times of pauses at LacI-O2 roadblocks shows that the longest pauses were associated with opposing force without GreA followed by shorter pauses with assisting force or opposing force with GreA. In (A) and (C), N represents the number of tethered templates that exhibited a transcription event examined under different force and GreA conditions. Data in (B) and (D) represent the exponentially fitted characteristic times \pm the 90% confidence intervals

recovery from backtracked states, while an opposing force could have the opposite effect, promoting backtracking or impeding recovery.

To investigate the role of backtracking in transcriptional pausing at roadblocks, we introduced GreA, a factor known for its role in rescuing backtracked ECs. GreA achieves this by catalyzing the cleavage of the nascent RNA's 3' end, which can obstruct the active site of the RNA polymerase [5]. Our results indicated that GreA did not affect pausing when the force was assisting transcription (Figure 5.4B & D,

5.5B). This outcome suggests that under these conditions, backtracking, and consequently the need for GreA's intervention, was minimal. Conversely, under conditions of opposing force, GreA appeared to expedite the passage through roadblocks by enhancing recovery from backtracked states (Figure 5.4B & D, 5.5B). This finding aligns with previous research indicating that GreA promotes successful navigation through LacI roadblocks [158].

Notably, we observed that even a modest assisting force of 0.2pN significantly reduced backtracking in roadblocked ECs. This effect contrasts with sequence-induced backtracking, where such gentle forces have less impact [55]. Our observations are in harmony with the established understanding that the energy barriers to backtracking are relatively low, as has been previously suggested for other polymerases like Rpo41 and PolII [159].

5.1.2 GreA and tension reveal two paths through roadblocks

The impact of force direction and the addition of GreA on transcription pauses varied notably depending on the *lac* repressor binding sites (Os, O1, or O2) involved. With DNA templates containing either O1 or O2, an opposing force extended the duration of pauses compared to when the force assisted transcription. However, when GreA was introduced, it brought the pause durations back down to levels comparable to those observed under assisting-force conditions (Figure 5.4B & D). These findings corroborate the widely held view that backtracking, in the absence of Gre factor-mediated recovery, impedes effective RNA synthesis.

In contrast, on templates containing the Os site, opposing force, surprisingly, reduced pause times below those seen under assisting force. This reduction was even more pronounced with the addition of GreA (Figure 5.5B). This observation suggests an intriguing phenomenon where rapid cycles of backtracking and recovery—a sort of reciprocating motion—might enable RNAP to force its way through roadblocks that

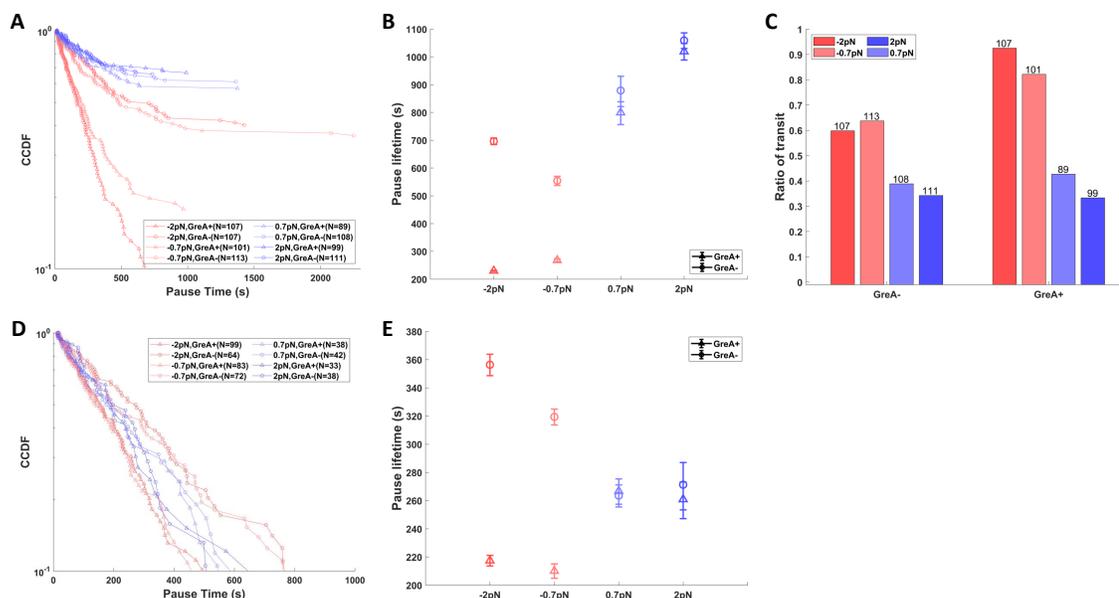


Figure 5.5: Pause times and fractions of passage at LacI-Os roadblocks under conditions of opposing (red) or assisting (blue) force and with (triangles) or without (circles) GreA. **(A)** The CCDF and **(B)** characteristic times of all recorded pauses at LacI-Os roadblocks show that the longest pauses were associated with assisting force with and without GreA, followed by shorter pauses with opposing force without GreA, and even shorter pauses with opposing force plus GreA. Note that all distributions except that for opposing force plus GreA include a significant fraction of indefinitely paused ECs. **(C)** Passage through LacI-Os roadblocks was more frequent under opposing than assisting force and was enhanced by the addition of GreA. The number of transcription events in each condition are listed above each bar. **(D)** The CCDF and **(E)** characteristic times of all except indefinite pauses at LacI-Os roadblocks show that the longest pauses were associated with opposing force without GreA, followed by shorter pauses with assisting force without GreA, and even shorter pauses with opposing force plus GreA. In **(A)** and **(D)**, N represents the number of tethered templates that exhibited a transcription event examined under different force and GreA conditions. Data in **(B)** and **(E)** represent the exponentially fitted characteristic times \pm the 90% confidence intervals.

dissociate slowly. The presence of GreA appears to hasten these repetitive encounters between RNAP and the roadblock, which could be particularly crucial if interactions with RNAP delayed the dissociation of the roadblock protein.

To further investigate our hypothesis regarding the interaction dynamics between RNAP and transcriptional roadblocks, we conducted transcription assays using another strong roadblock: the mutant EcoRI Q111 endonuclease. This variant binds DNA with high affinity but does not cleave its recognition site, 5'-GAATTC. Under 50 mM potassium ion ($[K^+]$) buffer conditions, EcoRI Q111 effectively obstructed nearly all transcription ECs, irrespective of the force direction. Intriguingly, when

GreA was introduced, the passage rate through this roadblock increased to 20% but only under conditions of opposing force (Figure 5.6A, 50 mM [K+]).

The timing of events was also noteworthy. The addition of GreA to a roadblocked EC was frequently followed shortly by successful passage through the roadblock (Figure 5.6B). This observation suggests that the combination of opposing force and GreA action significantly enhances the likelihood of successful transit through these long-lived roadblocks. This result provides additional evidence supporting the idea that backtracking and subsequent recovery facilitated by GreA, especially under opposing force conditions, are crucial mechanisms by which RNAP can navigate through persistent roadblocks during transcription.

When we elevated the salt concentration to 150 mM potassium ions ([K+]), the affinity of EcoRI Q111 protein for its DNA recognition site diminished. This change allowed a substantial proportion of ECs to successfully navigate past these roadblocks, as evidenced in Figure 5.6A for the 150 mM [K+] condition [160]. Consequently, we evaluated both the distribution of pause durations induced by the roadblock and the percentage of ECs that managed to pass through.

In all tested conditions, with the exception of those involving opposing force and GreA, we observed that approximately 30-60% of ECs were able to disengage from the EcoRI Q111 roadblocks in the high salt environment (Figure 5.6C, the long tails in the CCDF not reaching 0). This substantial fraction of indefinite blockages produced elongated tails in the distribution of pause durations, yet the average pause duration under assisting force was approximately 80 seconds. This duration is shorter than the pauses observed at LacI-O1 sites and comparable to those at LacI-O2 sites (Figure 5.6D). Given that the processivity of bacterial ECs and the stability of DNA tethers remain largely unaffected by high monovalent salt concentrations [161], it is plausible that the high salt concentration prompted dissociation of ECs stalled at roadblocks. This phenomenon might have obscured the observation of more transit

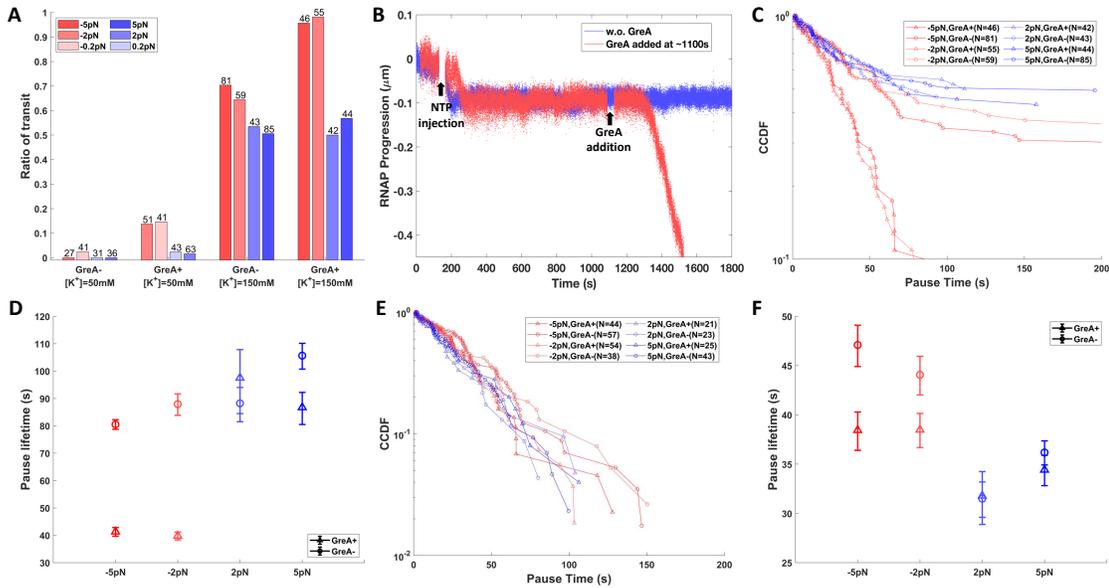


Figure 5.6: Pause times and the fraction of transit through EcoRI Q111 roadblocks under conditions of opposing (red) or assisting (blue) force and with (triangles) or without (circles) GreA. **(A)** Transit through EcoRI Q111 roadblocks was rare in 50 mM [K⁺] but increased dramatically in 150 mM [K⁺] especially upon the addition of GreA. **(B)** Without GreA most ECs in 50 mM [K⁺] buffer paused indefinitely at EcoRI roadblocks (blue), but adding GreA (red, ~1100 s) rescued paused ECs that resumed transcription (red, ~1300 s). **(C)** The CCDFs and **(D)** characteristic times of all recorded pauses at EcoRI Q111 roadblocks show that the longest pauses were associated with assisting force with and without GreA or opposing force without GreA, and shorter pauses with opposing force plus GreA. Note that all distributions except that for opposing force plus GreA include a significant fraction of indefinitely paused ECs. **(E)** The CCDFs and **(F)** characteristic times of pauses including only ECs that eventually pass through EcoRI Q111 roadblocks in 150 mM [K⁺] are shown. In **(C)** and **(E)**, N represents the number of tethered templates that exhibited a transcription event examined under different force and GreA conditions. Data in **(D)** and **(F)** represent the exponentially fitted characteristic times \pm the 90% confidence intervals.

events. Intriguingly, the introduction of GreA reduced the duration of pauses under opposing force to levels even lower than those observed under assisting force (Figure 5.6D). Furthermore, applying opposing force increased the fraction of ECs that successfully transited past EcoRI Q111 roadblocks, an effect further amplified by adding GreA (Figure 5.6A, 150 mM [K⁺]). The measured pause durations align with the estimated dissociation constant of EcoRI Q111 under high salt conditions, $K_d = 0.12$ nM, assuming a linear relationship between $\ln(K_a)$ and $\ln([M^+])$ [162].

The study revealed distinct responses of ECs to changes in force direction and the addition of GreA when navigating through LacI-O2/O1 roadblocks as opposed

to LacI-Os/EcoRI Q111 roadblocks. With LacI-O2/O1, an assisting force, whether or not combined with GreA, established a baseline pause duration before transit through the roadblock. In contrast, the addition of GreA was necessary to achieve this baseline under opposing force. Differently, for LacI-Os and EcoRI Q111 roadblocks, an opposing force expedited passage more effectively than an assisting force, especially when GreA was present.

To investigate whether these differing responses were associated with the strength of the roadblocks, subsets of data for LacI-Os and EcoRI Q111, excluding the indefinitely stalled ECs, were examined. This approach effectively focused on a subset of ECs encountering shorter-lived, lower affinity roadblocks. For these selected LacI-Os roadblocks, an opposing force induced longer pauses compared to the baseline set by an assisting force. Notably, the introduction of GreA significantly shortened the duration of pauses under opposing force, bringing them below the baseline level established by assisting force (Figure 5.5D & E). Similarly, for short-lived EcoRI Q111 roadblocks, GreA reduced the duration of pauses under opposing force to levels comparable with those observed under assisting force (Figure 5.6E & F). This pattern mirrors the behavior seen with LacI-O1 and LacI-O2 roadblocks. These observations underscored the role of the roadblock's lifetime in determining the efficiency of transcriptional transit under varying force conditions. The results indicate that the strength or longevity of a roadblock can influence the transit efficiency under opposing forces, as compared to the baseline established under assisting forces.

The insensitivity of pauses at roadblocks under assisting force to the addition of GreA suggests that assisting force prevents backtracking, leading ECs to adopt a passive, transcriptionally active state. This state allows ECs to progress as soon as the roadblock dissociates. This passive pathway is effective for transit through LacI-O1 and LacI-O2 roadblocks, but results in a substantial number of ECs becoming indefinitely stalled at stronger roadblocks such as LacI-Os and EcoRI Q111 (Figure

5.5A & C, 5.6A & C).

Conversely, the presence of GreA significantly enhances transit through roadblocks when force opposes transcription. This enhancement aligns with an active, reciprocating pathway involving EC backtracking and subsequent recovery facilitated by GreA. Intriguingly, this active pathway seems to expedite transit through stronger LacI-Os and EcoRI Q111 roadblocks (Figure 5.5B & C, 5.6C & D), while it decelerates transit through the weaker LacI-O1 and O2 roadblocks (Figure 5.4B & D). The variation in the reduction of pause times by GreA across different roadblocks suggests that the active pathway may encompass multiple cycles of backtracking and recovery. This cycle potentially allows ECs to eventually overcome the roadblock through repeated interactions.

5.2 A hybrid transit model recapitulates the effects of force and GreA in the MT assays

A hybrid model including the reciprocating/active and passive pathways is consistent with the data. Figure 5.7A depicts the progression through roadblocks via different states along these pathways. The passive pathway progresses through states ① → ② → ④ → ⑥, and the active pathway through ① → (② → ③ → ②)_n → ⑤ → ⑥ with n cycles of backtracking and recovery. The model includes three kinetic parameters k_1 , k_2 and k_3 , which represent the backtrack rate, backtrack recovery rate, and roadblock dissociation rate, respectively. Parameter $P1$ represents the probability of dislodging the roadblock at each encounter. Therefore, the transit rate of the passive pathway is simply $k_{\text{passive}} = k_3$, and rate of active pathway is $k_{\text{active}} = k_1/(1 + k_1/k_2)$.

Elongation complexes navigating transcriptional roadblocks appear to undergo multiple backtracking and recovery cycles before effectively dislodging these obstructions. Interestingly, roadblocks might also spontaneously dissociate during these cy-

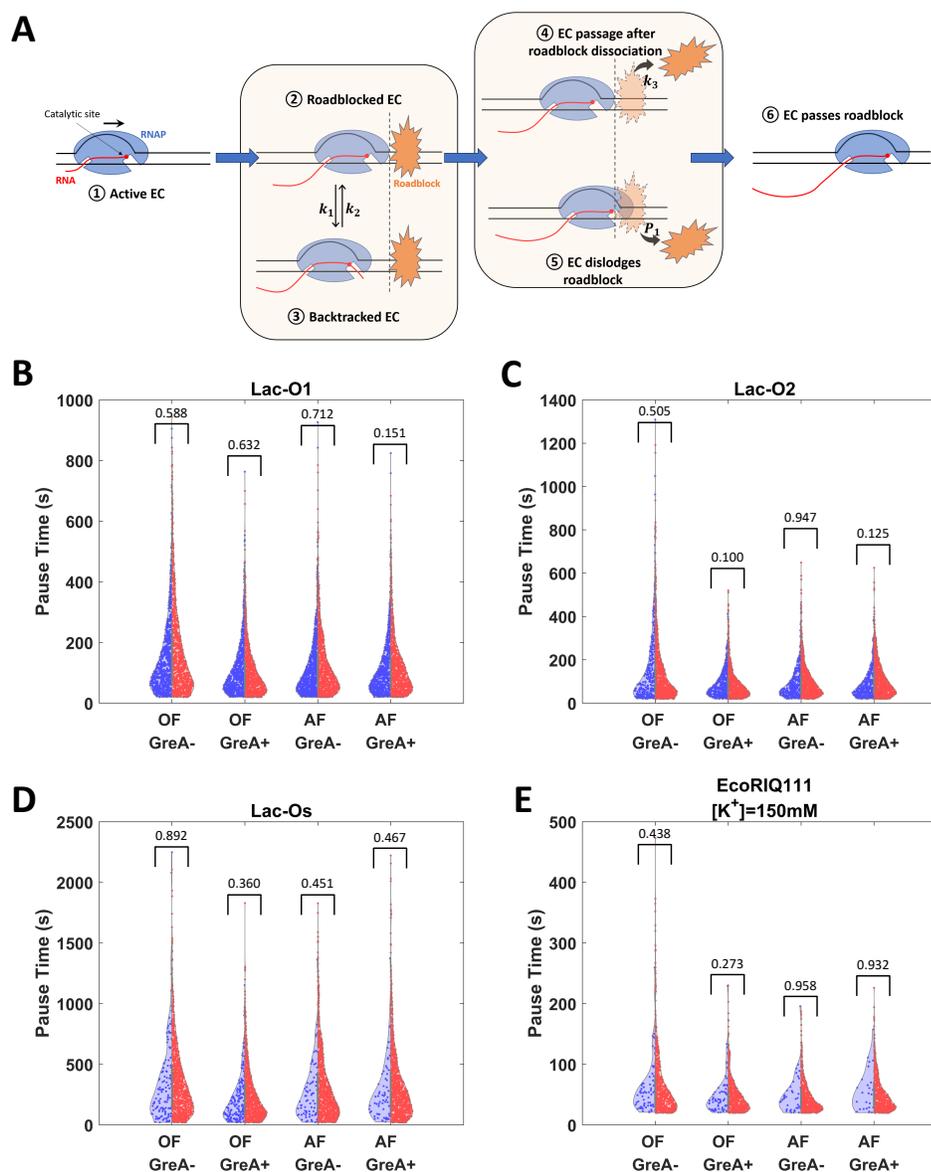


Figure 5.7: Model and simulation results. (A) A proposed model of EC transit through a roadblock includes six states: State 1: Transcription prior to the encounter with roadblock; 2: An EC encounters the roadblock; 3: At a roadblock an EC backtracks with a backtracking rate k_1 and recovery rate k_2 ; 4: A roadblock dissociates from DNA spontaneously with a dissociation rate k_3 ; 5: An actively transcribing EC, including a recently backtracked EC, has a probability P_1 of dislodging the roadblock; 6: An EC transits through the roadblock either by actively dislodging the roadblock or after spontaneous dissociation of the roadblock. (B) - (E) Simulations (red) produced pause time distributions very similar to those observed (blue) in LacI-O1, LacI-O2, LacI-Os and EcoRI Q111 experiments under different conditions. OF and AF represent opposing force and assisting force, respectively. P values (> 0.05 in all conditions) from two-sided two sample t-test are shown in figures.

cles. Given the stochastic movement of ECs across various states, Monte Carlo simulation is an apt choice for modeling the distributions of pause times under diverse conditions (Algorithm 1). Since Algorithm 1 can simulate a pause time distribution from a set of arbitrary values of parameters, we can search for an optimized set of values that minimize the difference between the simulated and experimental pause time distributions. This simulation approach successfully replicates the observed influences of both opposing and assisting forces, as well as the impact of adding GreA (Figure 5.7B-E).

Algorithm 1 Simulate RNAP pause time t_c at a roadblock. States ① – ⑥ are explained in Figure 5.7A

Require: RNAP state: ①; Roadblock state: on; $t_{max} = 5000s$; $dt = 1s$; $t_c = 0s$

```

1: while RNAP state not ⑥ and  $t_c < t_{max}$  do
2:   if Roadblock state is on then
3:     set Roadblock state to off with probability  $1 - \exp(-k_3 * dt)$ 
4:   end if
5:   if RNAP state is ① then
6:     set RNAP state to ⑤
7:   else if RNAP state is ② then
8:     if Roadblock state is on then
9:       set RNAP state to ③ with probability  $1 - \exp(-k_1 * dt)$ 
10:    else
11:      set RNAP state to ⑥
12:    end if
13:   else if RNAP state is ③ then
14:     set RNAP state to ⑤ with probability  $1 - \exp(-k_2 * dt)$ 
15:   else if RNAP state is ⑤ then
16:     if Roadblock state is on then
17:       set RNAP state to ⑥ with probability  $P1$ , otherwise set state to ②
18:     else
19:       set RNAP state to ⑥
20:     end if
21:   end if
22:   update current time -  $t_c = t_c + dt$ 
23: end while

```

Our model adeptly predicts changes in transcription pause times and EC transit frequencies in response to GreA and varying forces. This relies on balancing the

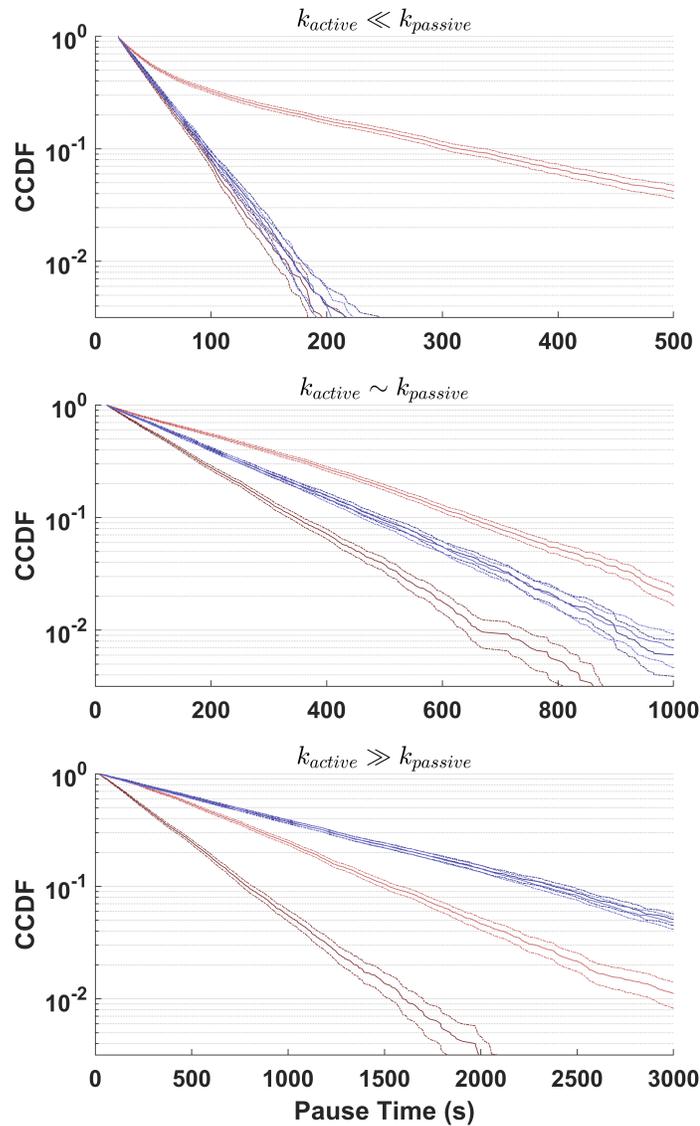


Figure 5.8: Simulations show the effects of forces and GreA in different regimes. Rate constants used were as follows: $k_1 = 0.01s^{-1}$, $k_2 = 0.005s^{-1}$, $k_3 = 0.03s^{-1}$, $P_1 = 0.2$ (passive transit); $k_1 = 0.01s^{-1}$, $k_2 = 0.005s^{-1}$, $k_3 = 0.005s^{-1}$, $P_1 = 0.2$ (hybrid of passive and reciprocating/active transit); $k_1 = 0.01s^{-1}$, $k_2 = 0.005s^{-1}$, $k_3 = 0.001s^{-1}$, $P_1 = 0.2$ (reciprocating/active transit). The model produced CCDF showing: **(A)** $\tau_{(OF, GreA-)} > \tau_{(OF, GreA+)} \sim \tau_{(AF, GreA-)} \sim \tau_{(OF, GreA+)}$ ($k_{active} \ll k_{passive}$, passive transit); **(B)** $\tau_{(OF, GreA-)} > \tau_{(AF, GreA-)} \sim \tau_{(AF, GreA+)} > \tau_{(OF, GreA+)}$ ($k_{active} \sim k_{passive}$, hybrid of passive and reciprocating/active transit) and **(C)** $\tau_{(AF, GreA-)} \sim \tau_{(AF, GreA+)} > \tau_{(OF, GreA-)} > \tau_{(OF, GreA+)}$ ($k_{active} \gg k_{passive}$, reciprocating/active transit).

dynamics between RNAP backtracking-recovery cycles (k_{active}) and the natural dissociation rate of roadblocks ($k_{passive}$). The simulation uncovers three distinct behavioral regimes:

- In the ‘passive route regime’ ($k_{active} \ll k_{passive}$), ECs embroiled in a backtracking-

recovery cycle due to opposing forces may not complete a cycle before the roadblock spontaneously dissociates. The model here predicts relatively short pauses for both assisting/Gre- and assisting/Gre+, and longer pauses for opposing/Gre- conditions. This mirrors findings from experiments using LacI-O2, LacI-O1, and EcoRI Q111 roadblocks at high salt concentrations (Figures 5.4B & D, 5.6F, 5.8A). Notably, in this regime, a double exponential fit better represents the distribution of pauses for opposing/Gre- conditions, suggesting two distinct stochastic processes at play, as evidenced by the longer tail in the pause distribution for LacI-O2 roadblocks under opposing/Gre- conditions (Figure 5.4D).

- In the ‘hybrid route regime’ ($k_{\text{active}} \sim k_{\text{passive}}$), opposing force could prolong or reduce dwell times depending on whether RNAP is in a backtracked state at the moment of roadblock dissociation. This regime’s simulated pause distribution (Figure 5.8B) aligns with observations for high-salt EcoRI Q111 roadblocks and a sub-population of low-affinity LacI-Os roadblocks, where opposing/Gre- conditions resulted in longer pauses compared to assisting force conditions, and opposing/Gre+ led to shorter pauses (Figures 5.5E, 5.6D). Here, active and passive pathways proceed at comparable rates, with all distributions fitting well to a single exponential curve.
- In the ‘active route regime’ ($k_{\text{active}} \gg k_{\text{passive}}$), the passive pathway is less efficient, and the active/backtracking-recovery pathway proves more effective. Here, opposing forces favor backtracking-recovery cycles, leading to shorter pauses and higher EC transit rates, particularly when enhanced by GreA. The simulated outcomes (Figure 5.8C) correspond with experimental data for LacI-Os and low-salt EcoRI roadblocks, where opposing force yielded shorter pauses and higher transit frequencies than assisting force, and GreA further decreased pause times and increased transit rates (Figure 5.5B & C, 5.6A).

Table 5.1 concisely captures the kinetic parameters derived from fitting our model to the experimental data. These parameters reflect the relationship between roadblock affinity and the efficiency of transcriptional transit. Specifically, we observe that the passive rates of transit align with the affinities of the respective roadblocks, with $k_{\text{passive}}(\text{high salt EcoRI})$ and $k_{\text{passive}}(\text{O2})$ being higher than $k_{\text{passive}}(\text{O1})$, which in turn is higher than $k_{\text{passive}}(\text{Os})$.

Moreover, the model suggests that transit through high-salt EcoRI Q111 and LacI-Os roadblocks (excluding instances of indefinite stalls) follows what we term a ‘hybrid route’. This implies that, for these particular roadblocks, the active backtracking-recovery cycling rate (k_{active}) is roughly equivalent to the passive roadblock dissociation rate (k_{passive}). In contrast, transit through LacI-O1 and LacI-O2 roadblocks appears to predominantly follow a passive route, with a significantly lower k_{active} compared to k_{passive} .

This divergence in transit dynamics between EcoRI and LacI roadblocks raises intriguing questions about the underlying mechanisms. We hypothesize that variations in the DNA sequences upstream of the EcoRI and LacI binding sites could be a contributing factor. For instance, using the thermodynamic model described in Chapter 3, the energy profiles calculated for the EC are less stable upstream of the EcoRI roadblock, potentially facilitating faster backtracking and recovery cycles. Additionally, the interaction between RNAP and the LacI repressor (as suggested by the co-partitioning behavior between LacI and RNAP) might slow the rate of backtracking and recovery at LacI roadblocks.

The experimental findings, coupled with our model, indicate that the typical duration of a backtracking-recovery cycle for LacI roadblocks is approximately 260 seconds, and around 160 seconds for EcoRI Q111 roadblocks in high salt conditions, as detailed in Table 5.1. These durations exceed those generally associated with sequence-induced backtracked pauses, as described by Toulme et al. and Neuman et al. [55, 158]. Inter-

Table 5.1: Kinetic parameters generated by fitting our model to the experimental pause time distributions. (* k_3 values postulated from pause time CCDFs, rather than model fitting, due to the large fraction of indefinite stalls in the experimental data set.)

	k_{passive} (s^{-1})	k_1 (s^{-1})	k_2 (s^{-1})	P_1	k_{active} (s^{-1})
LacI-O2	0.0119	0.0087	0.0070	0.2549	0.0038
LacI-O1	0.0094	0.0087	0.0070	0.2549	0.0038
LacI-Os (Indefinite stalls excluded)	0.0034	0.0087	0.0070	0.2549	0.0038
EcoRI(150mM[KGlu]) (Indefinite stalls excluded)	0.0331	0.0103	0.0158	0.1267	0.0062
LacI-Os	0.001*	0.0087	0.0070	0.2549	0.0038
EcoRI(150mM[KGlu])	0.01*	0.0103	0.0158	0.1267	0.0062

estingly, similar prolonged pauses are a common observation in diverse experimental contexts and are categorized as stabilized-backtracked pauses [64, 163]. These enduring, force-independent pauses lend credibility to the idea that roadblock-induced backtracking might encompass a force-independent, rate-determining intermediate state.

Given that the backtracking recovery rate, denoted as k_2 , is expected to be dependent on the concentration of GreA, our model anticipates that the efficiency of RNAP transit is influenced by GreA levels, particularly in scenarios involving reciprocating motion. To further substantiate our model, we sought to determine the values of k_2 as a function of GreA concentration. We adopted a Michaelis-Menten approach, positing a relationship between the rate of GreA-facilitated cleavage and the concentration of GreA. Subsequently, simulations were executed to produce distributions of pause times across a range of GreA concentrations. The results from these simulations, reflecting the pause time's dependency on GreA concentration, were in harmony with the characteristic pause lifetimes observed experimentally under varying GreA con-

centrations. This congruence between simulated and experimental data reinforces the validity of our proposed model, as illustrated in Figure 5.9.

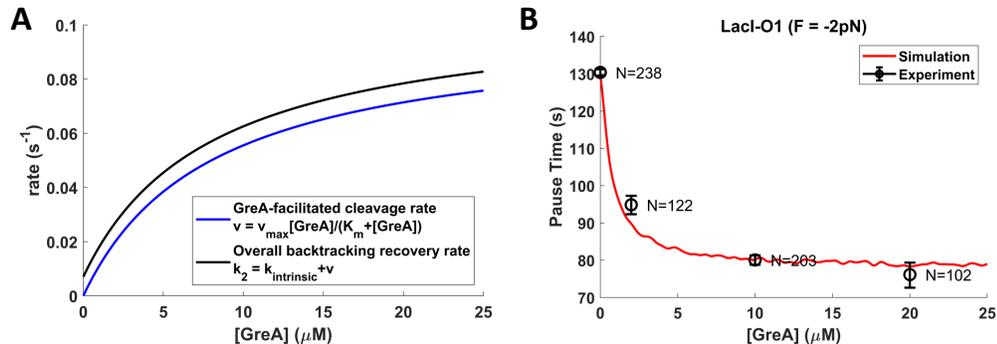


Figure 5.9: Experiments with different GreA concentrations show effects on pause times similar to those of simulations. **(A)** Assuming a Michaelis-Menten relationship, the rate of GreA-induced cleavage of nascent RNA in backtracked RNAP complexes is shown in blue. The rate is estimated using parameters $K_m = 8 \mu\text{M}$ (a reasonable guess according to the empirical results [164]) and $v_{\max} = 0.1 \text{s}^{-1}$. The overall backtrack recovery rate k_2 (black curve) is evaluated as the sum of the intrinsic cleavage rate (from fitting backtrack recovery rate from experimental data in $[\text{GreA}] = 0$ condition, see Table 5.1) and the GreA-facilitated cleavage rate. **(B)** Simulation was executed by iterating Algorithm 1 in the main text to generate pause time distributions at various GreA concentrations, using the estimated k_2 values in panel (A) and fitted k_1 , k_3 and P_1 values in Table 5.1. The range of GreA concentration was taken as 0 – 25 μM with 0.1 μM increment, and Algorithm 1 was repeated 10,000 times at each GreA concentration. The red curve shows the characteristic lifetime of simulated pause time distributions as a function of GreA concentration. Exponential fits $\pm 90\%$ confidence interval of pause times collected in experiments at 0, 2, 10, and 20 μM GreA concentrations (black) agree with the simulation results.

5.3 Significance and Limitations of the Hybrid Transit Model

Our experimental data and model advocate for a dual-pathway model in RNAP navigation through transcriptional roadblocks. This model posits that RNAP can employ two distinct mechanisms: a passive mode, where ECs pause and proceed once roadblocks dissociate, and a reciprocating mode involving cycles of backtracking and recovery to overcome the roadblocks. The choice of pathway hinges on the roadblock’s duration, influencing the EC’s transit strategy. In instances of encountering LacI-

O1 and LacI-O2 roadblocks, ECs predominantly utilized the passive route, whereas the reciprocating pathway actively facilitated passage through the more persistent LacI-Os and EcoRI Q111 roadblocks by abbreviating the pause durations.

Surprisingly, ECs exhibited an acute sensitivity to mechanical forces, with even marginal forces like 0.2 pN significantly impacting their transit, a phenomenon echoed in physiological conditions due to genome architecture [165, 166]. Conversely, EC transit through roadblocks showed an unexpected insensitivity to force magnitudes ranging from 0.2 pN to 5 pN. This finding contrasts with earlier studies where greater opposing forces extended backtracked pauses [46, 56]. Our hypothesis is that roadblock-associated backtracking includes a force-sensitive intermediate state, succeeded by a force-resistant rate-limiting phase. This speculation aligns with previous findings in chapter 3 suggesting that external forces primarily influence the backtracking by altering the forward transcription rate from backtracked positions, while the bidirectional fluctuations of backtracked ECs remain largely unresponsive to external forces [11, 55].

Backtracking, traditionally viewed as an impediment to transcription, emerges as a facilitative mechanism in our study, expediting the displacement of DNA-bound proteins that dissociate slowly. The active reciprocating pathway, characterized by cycles of backtracking and recovery, is shown to enhance efficient passage through enduring roadblocks, crucial in preventing RNAP from stalling and becoming vulnerable to exonucleolytic activity [167]. This understanding also clarifies the distinct responses of RNAP and helicase RecBCD when confronting the EcoRI roadblock [168]. While helicases utilize the full energy of ATP hydrolysis for movement, RNAP, limited to generating chemo-mechanical force only during NTP incorporation [11], depends on Brownian-ratchet translocation and consequently exerts less force on nucleoprotein obstacles, necessitating repeated backtracking-recovery cycles to destabilize and eventually remove high-affinity roadblocks.

An intriguing aspect of our findings is the implications for internal RNAP dynamics. Stable protein-DNA contacts, not only from extrinsic proteins in RNAP's path but also from intrinsic contacts like those between promoter DNA elements and the initiation σ factor, pose similar challenges to RNAP. These contacts lead to the synthesis and release of short abortive RNAs or the formation of arrested complexes [169]. Gre factors, known to facilitate promoter escape [170], suggest that cycles of backtracking and RNA cleavage are necessary to disrupt σ -DNA interactions. Likewise, transcription elongation factors like RfaH, which promote backtracking and rely on Gre factors for escape from recruitment sites [171], indicate that backtracking and re-extension cycles are essential for continuous RNA synthesis whenever strong DNA-protein interactions obstruct RNA chain extension.

This study, while illuminating, acknowledges certain limitations and avenues for future research. The focus here excluded torsional aspects of transcription, even though reciprocal movement along a few base pairs in torsionally anchored templates might generate sufficient torsion to stall RNA polymerase [83, 172]. Additionally, our analysis revealed that backtracking and recovery cycles upstream of EcoRI roadblocks are faster than those upstream of lac binding sites, suggesting potential exploration of sequences with broader energy profile ranges.

In conclusion, our study uncovers a hybrid mechanism by which elongation complexes traverse protein roadblocks. ECs may either passively await roadblock dissociation or actively combat them, and backtracking plays a pivotal role in modulating pause durations at roadblocks based on their persistence and interactions with ECs. The influence of tension and the transcript cleavage factor GreA highlights that structural (roadblock affinity) and dynamic (applied tension) elements can shape the efficiency of these pathways and the route taken by ECs in overcoming roadblocks. The nuanced interplay of various *in vivo* elongation factors might finely tune these pathways, rendering distinct and deterministic biological responses.

Chapter 6

Transcription Recycling – An Overlooked Regulatory Mechanism

A transcription cycle consists of distinct phases: RNAP recognizes a promoter, initiates transcription, elongates the transcript, and ultimately terminates the process. Traditionally, it's believed that RNAP disengages from the DNA template upon reaching a terminator, diffusing away predominantly in three dimensions [173, 174]. However, recent observations indicate alternative RNAP behaviors post-termination. Notably, RNAP may linger on the DNA, engaging in one-dimensional diffusion back to a promoter, potentially reinitiating transcription [175, 176]. This suggests that RNAP could repetitively transcribe the same DNA sequence, efficiently amassing transcripts without recruiting additional RNAP molecules, thus minimizing transcriptional conflicts.

Though this non-canonical termination pathway has been reported [18, 19, 20, 176], its biological significance and precise mechanism remain uncertain. Factors like external force could influence RNAP's sliding behavior, with potential impediments from DNA-bound proteins. The role of the sigma factor (σ) in this repetitive transcription is also not fully understood. While σ is crucial for initiating transcription

[177], there is evidence it may stay associated with RNAP during secondary transcription events [178]. Furthermore, the specific RNAP domains necessary for initiating transcription in the opposite direction after completing a primary elongation round are yet to be identified [18].

Our previous MTs study on roadblocked transcription revealed that a significant proportion of RNAPs displaying this atypical termination. Through extensive MT experiments, we examined the influence of various factors — external force, roadblock proteins, and different terminators — on repetitive transcription. These experiments suggested that external force could direct RNAP’s diffusion, biasing its search for a secondary promoter. The results indicated a likely necessity for the σ factor in the relatively rapid initiation of repetitive transcription from promoters. Additionally, our findings showed that while RNAP could re-initiate transcription from promoters in either orientation relative to the previous elongation direction, deleting the α subunits’ C-terminal domains, known for DNA interactions, markedly reduced the ability to change transcription direction.

6.1 Force directs RNA polymerase diffusion and repetitive transcription

To investigate the mechanisms of repetitive transcription by RNAP, we utilized MTs to exert mechanical forces on single transcription complexes along DNA templates. These templates, containing either one or two promoters, a high-affinity *lac* repressor binding site (Lac O1), and a terminator, allowed for simultaneous observation of multiple DNA molecules. The experimental setup (Figure 6.1a) facilitated control over both the direction and magnitude of the force exerted on RNAP, with forces ranging from -5 pN (opposing transcription) to +5 pN (aiding transcription).

Our observations revealed that upon encountering a terminator, RNAPs under

opposing forces frequently slid back towards the promoter. In several instances, these RNAPs would then re-initiate transcription (Figure 6.1b, cycles 1 & 2). Notably, single RNAP molecules were observed to repeatedly transcribe the same sequence, with some doing so up to seven times. Typically, these repetitive cycles concluded with RNAP sliding back to the promoter, pausing briefly, and then either detaching directly from the promoter or sliding off the DNA template's end (Figure 6.1b, cycle 3). Our assay could not distinguish between these two pathways of RNAP release, as in either case, the bead detached from the DNA swiftly exited the focal plane.

Repetitive transcription presents an intriguing deviation from the classic RNAP dissociation post-termination model, showing distinctive behaviors under mechanical force. Observations indicate that over 20% of RNAP molecules slid backward from single λ T1 or T500 terminators when subjected to opposing mechanical force. Remarkably, placing two consecutive terminators (as shown in Figure 6.1c) nearly doubled this probability to about 50%, suggesting that the efficiency of termination affects the RNAP's transition from a transcribing to a sliding state. While sliding generally started at terminators, it occasionally began beyond the terminator (Figure 6.1b, Cycle 3), hinting at gradual conformational changes during elongation that facilitate sliding, in line with the notion of a two-stage post-elongation complex [179].

The sensitivity of repetitive transcription to the force direction, but not its magnitude, is striking. Up to 25% of post-termination complexes quickly slid backward to a previously used promoter under opposing forces (Figure 6.1d), with about 10% re-initiating transcription (Table 6.1, Single promoter records). Even under minimal forces like 0.2 pN, post-termination complexes predominantly followed the force direction. This behavior was confirmed with a reversed DNA construct (Figure 6.1d, right): post-termination complexes seldom slid backward against the force after transcription with assisting force. This force responsiveness implies that post-termination complexes traverse the DNA template rather than diffusing in space or hopping

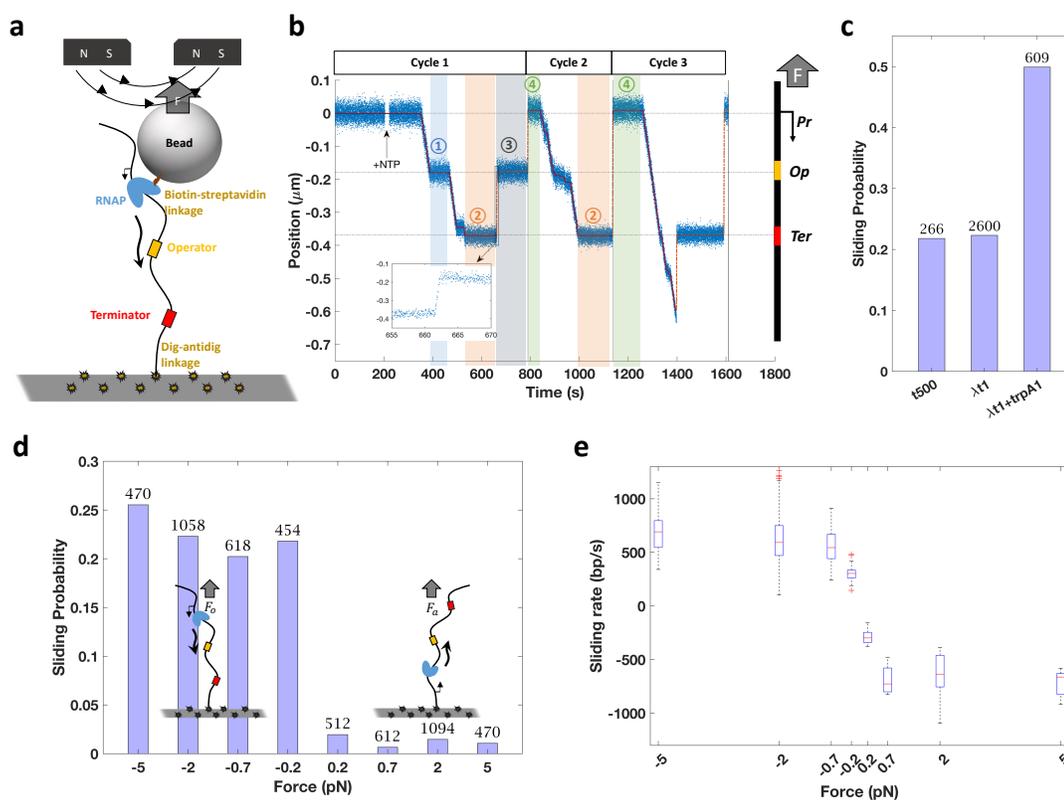


Figure 6.1: Force-directed sliding leads to repetitive transcription. (a) A diagram of the experimental setup for transcription against opposing force. (b) A representative recording of multiple rounds of transcription under opposing force includes a temporary roadblock-associated pause during transcription (shaded region 1), pauses at the terminator (shaded regions 2), RNAP temporarily roadblocked during backward sliding (shaded region 3), and pauses at the promoter prior to re-initiation (shaded regions 4). The inset shows data points corresponding to RNAP sliding back from the terminator in cycle 1. (c) On templates with a dual terminator sequence, the percentage of RNAP that slid backward was twice that on templates with single terminators. The total number of events are listed above each bar. (d) Opposing force (negative values) significantly raised the probability that the post-terminator complex slid toward the promoter from which the previous cycle of transcription initiated. The total number of events are listed above each bar. (e) RNAP sliding rates increased rapidly as opposing (-) or assisting (+) force increased from 0 to 0.7 pN but plateaued thereafter. Red crosses indicate outliers.

between DNA segments. Transcription rates were consistent with existing studies [66, 180, 181], and the rate distributions for successive cycles on a single template closely matched but varied from those on another template, suggesting the same RNAP enzyme repeatedly transcribed in each case.

Unidirectional RNAP sliding, likely influenced by force-directed one-dimensional diffusion, occurred swiftly, needing less than 3 seconds even under low force like 0.2

pN. Although the applied force did not alter the sliding probability, higher forces increased the sliding speed, estimated at 200 bp/sec at the weakest forces and plateauing at higher forces (Figure 6.1e). This speed limit might be associated with the entropic force required for the DNA template to uncoil into the sliding RNAP, estimated around 0.7-1.9 pN for DNA entering a small pore [182], matching the observed inflection point in our sliding velocity data.

6.2 Secondary transcription from a roadblock site

The introduction of LacI protein in our experiments provided clear evidence of its function as a transcriptional roadblock. During transcription, RNAP frequently paused at the O1 binding site in the presence of LacI, underscoring its roadblocking effect (Figure 6.2a, Cycle 1). Interestingly, LacI also impeded post-termination sliding of RNAP, resulting in a pause followed by continued sliding towards the promoter (Figure 6.2a, gray-shaded region of Cycle 2, Figure 6.2g) or re-initiation of transcription directly from the roadblock site (Figure 6.2a, gray-shaded region of Cycle 1, Figure 6.2g). Notably, the dwell times of these roadblock-induced pauses were unaffected by the force magnitude (Figure 6.2b), lending further support to the notion that force-driven post-termination complexes travel along the DNA, temporarily halting at roadblocks until they dissociate from the DNA. This observation reinforces the idea that LacI, as a DNA-bound protein, can significantly alter RNAP dynamics both during and after transcription termination.

Upon encountering a roadblock, most post-termination complexes in our study proceeded past it to reach the promoter. Approximately 4% of these complexes then initiated another cycle of transcription, regardless of the presence of LacI protein (Figure 6.2c). Interestingly, the presence of LacI influenced the behavior of these post-termination complexes: around 2.4% recommenced transcription right at the

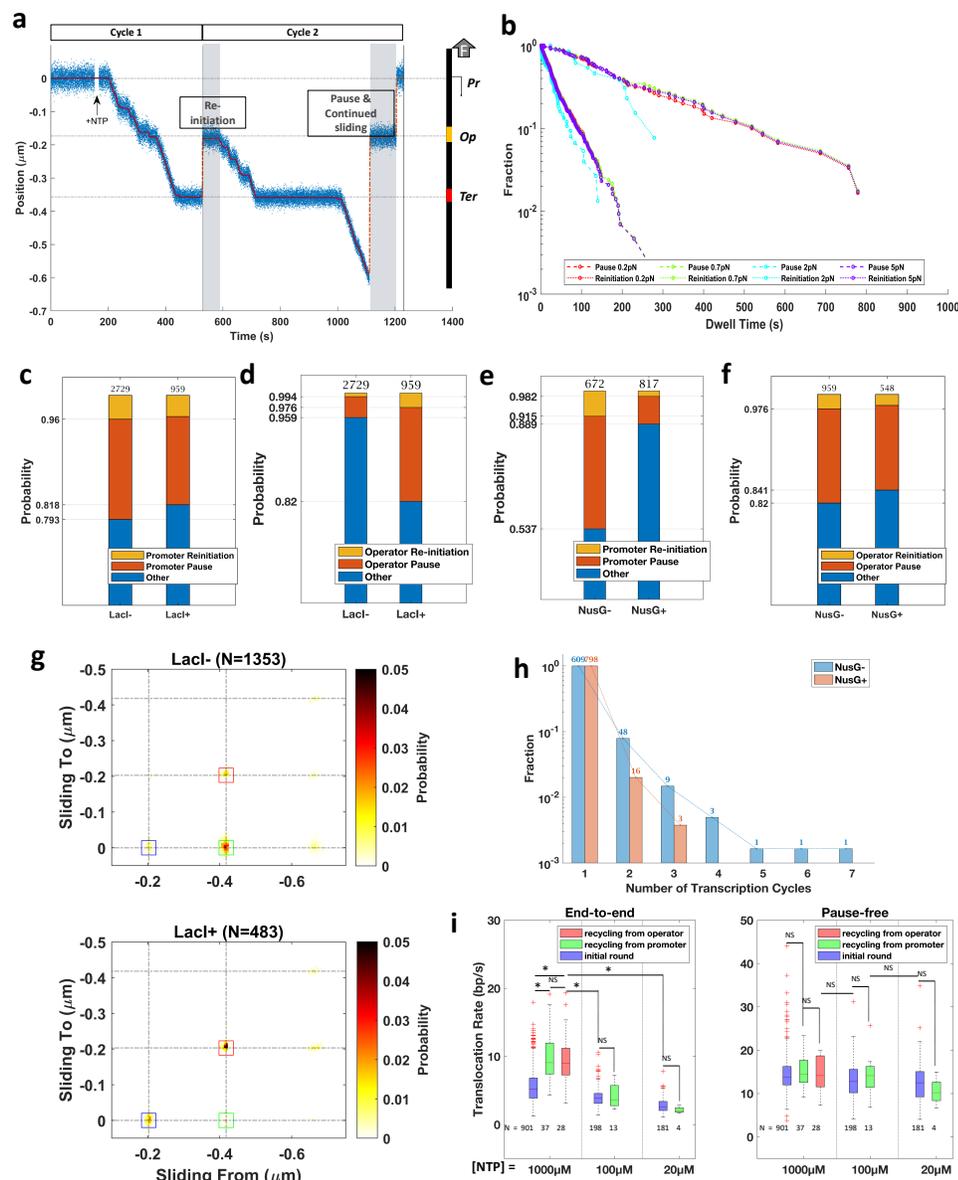


Figure 6.2: DNA-binding protein roadblocks, LacI and NusG affected post-termination sliding and repetitive transcription. (a) In this representative recording, the shaded regions indicate pauses at the roadblock during post-termination sliding followed by re-initiation ① or continued sliding past the roadblock to the promoter (recapture) ②. (b) Dwell times at the lac repressor binding site prior to re-initiation (dotted lines) were longer and produced a much broader distribution than dwell times preceding continued sliding events (dashed lines). Remarkably, the magnitude of force had a negligible effect on these distributions. (c) The presence of LacI has a negligible effect on the probability of promoter pause and subsequent re-initiation. (d) Adding LacI to the buffer increased the probability of roadblocking and re-initiation at the roadblock. (e) Although NusG diminished the probability of sliding to promoters, it did not change the ratio of re-initiation, ~ 0.1 (yellow/red). (f) The presence of NusG has a negligible effect on the probability of roadblocking at the lac repressor site or subsequent re-initiation. (g) Heat maps indicate probabilities associated with locations at which post-termination complexes started and stopped sliding under opposing force in buffer without (upper) or with (lower) LacI. Sliding primarily began at the terminator ($\sim -0.4 \mu\text{m}$) and ended at promoter ($\sim 0 \mu\text{m}$) (green box), unless LacI was present to block sliding and induce terminator-to-LacI binding site (red box) and LacI binding site-to-promoter (blue box) sliding events. (h) A minor population of RNAPs, less than 10%, exhibited multiple cycles of transcription. (i) sub-saturating level of NTPs (100 μM and 20 μM) reduced average end-to-end transcription rate, but not pause-free rate, of both initial and re-initiating transcriptions, compared to saturating NTP level (1000 μM). N: number of events; *: p<0.1; NS: p>0.1.

LacI roadblock, a notable contrast to the significantly lower 0.6% observed without LacI (Figure 6.2d). These findings were further substantiated by the observation that post-termination complexes that restarted transcription at the roadblock site exhibited considerably longer dwell times compared to those that merely paused before continuing to slide (Figure 6.2b).

These results indicate that, given enough time at a roadblock, post-termination complexes are capable of forming open complexes and resuming transcription. This demonstrates the dual impact of protein roadblocks like LacI on DNA templates. Not only can they impede elongation complexes, causing delays in transcription and potentially leading to 3'-truncations of transcripts, but they also influence the behavior of sliding post-termination complexes. These complexes can result in the production of 5'-truncated transcripts, underscoring the complex interplay between transcriptional machinery and DNA-bound proteins and highlighting the diverse outcomes that can arise from such interactions [11].

6.3 Sigma factor promoted sliding and secondary promoter recognition

Promoter recognition by RNAP and its transition to an open complex fundamentally involve σ factors. In a scenario devoid of free σ^{70} in the buffer, secondary initiation from the promoter would necessitate that the σ factor remains bound to RNAP through the entirety of primary transcription and subsequent sliding of the post-termination complex to the promoter. To investigate this hypothesis, we introduced NusG, a transcription elongation factor known to compete with σ^{70} for RNAP binding [183]. Our observations revealed that NusG reduced the frequency of repetitive transcription from the promoter (Figure 6.2e), aligning with expectations from a competitive interaction leading to the release of σ^{70} from elongation

and post-termination complexes. Notably, NusG seemed to diminish the likelihood of post-termination complexes sliding along the DNA template, though a similar proportion of sliding complexes proceeded to re-initiate transcription upon reaching the promoter. This aligns with other studies suggesting that σ^{70} often remains associated with post-termination complexes on DNA [19, 178], and contrasts with findings from optical tweezers experiments which reported only brief, force-sensitive dwell times at certain terminators [80] without any evidence of sliding or secondary transcription. Our experiments also showed that the probability of re-initiation dramatically dropped after one round of transcription (Figure 6.2h), with a dissociation constant of $3.8 \pm 0.8 \times 10^{-3}$ [178] suggesting that few RNAPs retained σ^{70} through multiple lengthy transcription cycles.

The presence of NusG had little effect on re-initiation at the roadblock (Figure 6.2f), although post-termination complexes tended to pause longer at roadblocks than at promoters before re-initiating transcription (Figure 6.2b). This indicates that regardless of the presence of an associated sigma factor, sliding post-termination complexes can slowly form transcription bubbles suitable for elongation at non-promoter sequences.

To confirm that the observed secondary cycles of RNAP translocation indeed involved transcription, we analyzed the translocation rates of RNAP in relation to NTP concentration (Figure 6.2i). Under saturating NTP conditions, initial transcription was slower than subsequent cycles due to pauses at LacI roadblocks. Yet, the pause-free translocation rates remained consistent across both cycles, whether translocation originated from promoters or roadblocks. When LacI was absent and NTP levels were sub-saturating, the pause-free rates remained unaffected, but the overall translocation rate for both initial and secondary cycles dropped, characterized by more frequent and extended pauses. This pattern is indicative of transcription, with the pause-free translocation rates and sensitivity to NTP concentration validating the transcriptional

nature of both primary and secondary cycles. The similarity in translocation rates from promoters and roadblocks under saturating NTP conditions, and the decrease in re-initiation probability at lower NTP concentrations, suggest that re-initiation involves a canonical initiation stage followed by transcript synthesis, mirroring the mechanics of the initial transcription cycle.

6.4 Force biased re-initiation between converging or diverging promoters

The application of mechanical force played a pivotal role in directing the movement of post-termination RNAP complexes and consequently influenced the location of transcription re-initiation. In our experiments, we observed that post-termination complexes slid back and re-initiated transcription multiple times at promoters oriented against the direction of the applied force. However, when the promoter was aligned with the force, post-termination complexes were directed away from it, as depicted in Figure 6.3a-d.

Specifically, in DNA templates designed with convergent promoters, the promoter aligned with the applied force was engaged by RNAP only if it was located in the initial round of transcription (Figure 6.3a&c). Contrastingly, in templates featuring divergent promoters, the promoter in alignment with the force was utilized either during the initial transcription round or in the final cycle of recycling (Figure 6.3b&d).

These findings indicate that even minimal forces impacting the sliding of post-termination complexes could significantly influence the selection of transcription initiation sites from nearby, yet oppositely oriented, promoters in a biological context. This suggests a nuanced interplay between mechanical forces and transcriptional regulation within the cellular environment, where slight directional biases could have considerable implications for gene expression.

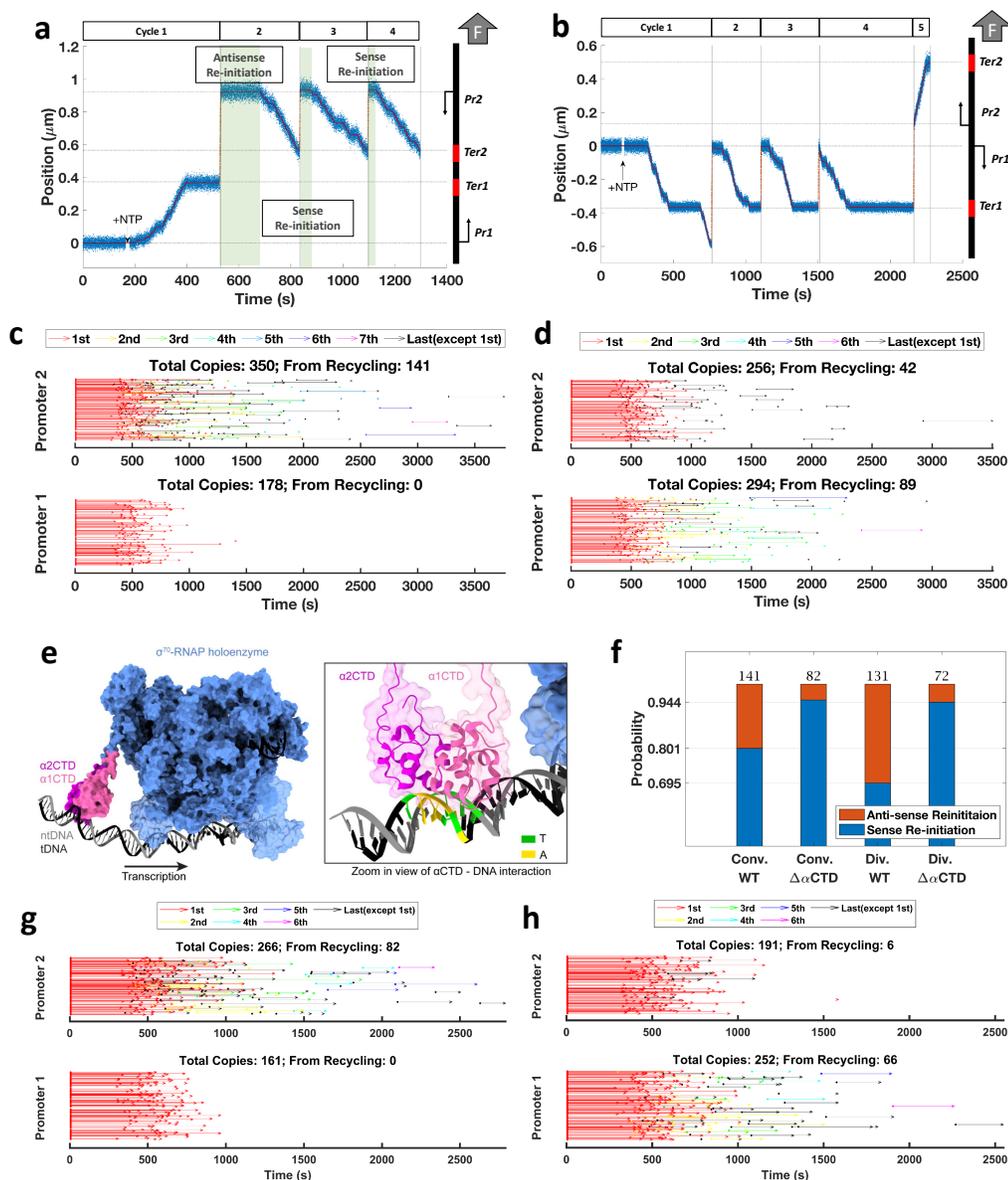


Figure 6.3: Repetitive transcription on templates with two adjacent promoters. (a) In a recording for a template with converging promoters, after one transcription event from P1 under assisting force, RNAP slid to P2 and completed three cycles of transcription opposing force before finally dissociating from the promoter. (b) In a recording for a template with diverging promoters, transcription from P1, opposing the direction of force, repeated four times before RNAP slid past P1 to re-initiate once from P2 assisted by force. (c) A time-based catalog of the transcription events involving promoter P2 observed along the templates with converging promoters shows the beginning and end of each transcription event with different rounds depicted in different colors. Transcription from P1, the promoter oriented in the direction of force, never repeated. (d) A time-based catalog of the transcription events involving promoter P2 for templates with diverging promoters shows repetitive events from promoter P1, oriented against the force but only single primary or final repetitive events from P2 oriented with the force. (e) In figures prepared with Chimera version 1.2, α -CTDs are shown to interact with promoter DNA in a closed *rrnB* promoter complex [184] (left). An enlarged view of the α -CTD shows contacts with the UP element (right). (f) Deletion of α -CTD diminished the probability that RNAP turned around to re-initiate in the direction opposite to the primary transcription event (anti-sense re-initiation) on templates with converging or diverging promoters. (g) A summary of $\Delta\alpha$ -CTD shows RNAP transcription events on templates with convergent or (h) divergent promoters. Compared to wild-type RNAP in panel (c) and (d), the deletion of α -CTD reduced the ability of RNAP to turn around and re-initiate from a secondary promoter oriented in the the direction opposite to the preceding transcription event.

The behavior of sliding post-termination RNAP complexes, as reported in earlier studies, showed an intriguing ability to recognize and interact with promoters regardless of their orientation [20]. Our recent experiments, involving templates with either convergent or divergent pairs of promoters, demonstrated that RNAP complexes displayed similar dwell times before re-initiating transcription in either direction.

In scenarios where secondary transcription was initiated from a promoter oriented in the opposite direction to the preceding transcription, the post-termination RNAP complex, likely having lost part or all of its transcription bubble [185], needed to rapidly slide in the direction of the applied force. This motion continued until the RNAP recognized a promoter, whereupon it switched polarity to use what was previously the non-template DNA strand to form a new open bubble. This switch necessitates the temporary loss of nonspecific DNA contacts made by the core RNAP and the sigma factor. A critical question arises: how does the enzyme maintain its association with DNA while executing this U-turn?

The C-terminal domains of the α subunits (α -CTD) could mediate the polarity switch. The α -CTDs are connected to the α N-terminal domains (NTDs) via long flexible linkers; the NTDs interact with the β ($\alpha 1$) and β' ($\alpha 2$) subunits, serving as a scaffold for core enzyme assembly [186]. The α -CTDs make direct contacts to AT-rich sequences (UP elements) upstream of the core promoter (Figure 6.3e) [184], and are required for the exceptional strength of rRNA promoters, but are dispensable for initiation at many promoters [187]. A “consensus” UP element is composed of T- and A-tracks centered at the -50 and -40 promoter positions [187]. T-tracks are also key signature motifs of intrinsic terminators, which may also contain a matching A-track upstream of a hairpin [69]. Thus, the α -CTDs may maintain contacts with their preferred DNA elements in the course of polarity switch, anchoring RNAP on the DNA despite the loss of other protein-DNA contacts.

To test this hypothesis, we employed an RNAP variant lacking the α -CTD and

observed its behavior on transcription from divergent or convergent secondary promoters. Remarkably, this mutant RNAP seldom switched direction to re-initiate transcription at promoters oriented opposite to the direction of the preceding transcription event (Figure 6.3f-h). However, the absence of the α -CTD did not impede the RNAP's ability to re-initiate transcription from a promoter aligned with the direction of the preceding transcription event (Table 6.1, convergent and divergent promoter records summary). These findings highlight the pivotal role of α -CTDs in facilitating the complex mechanics of RNAP polarity switching during transcription.

6.5 Summary of Transcriptional Recycling Mechanism

In this study, we demonstrate an alternative to canonical RNAP termination. The ability of RNAP to slide along DNA and re-initiate transcription without fully dissociating can efficiently accumulate transcripts and reduce the need to recruit RNAPs, especially in conditions where rapid and repeated transcription from a single promoter is advantageous, or where protein synthesis needs to be tightly regulated. The influence of external forces on post-termination behavior of RNAP, and the interplay between RNAP and transcription factors like sigma, could have significant implications in cellular environments where mechanical forces are present. For example, molecular crowding, and phase separation or intracellular tethering [188], could produce forces that influence RNAP behavior, therefore affecting gene expression patterns. Additionally, this study offers insights into how cells might exploit mechanical forces to regulate transcription dynamically, potentially influencing cellular responses to environmental changes or stress. Notably, we demonstrated that roadblock proteins aligned with uni-directional tension can be a source of incomplete, cryptic transcripts from non-promoter sites, which play important role in cell aging [189].

Single Promoter Records

Conditions				Summary			
Terminator	Force (pN)	LacI	NusG	Post-termination sliding	Promoter re-initiation	Operator re-initiation	Total number
λ t1	5	–	–	5(1.1%)	0	0	470
λ t1	2	–	–	16(1.5%)	3(0.3%)	1(0.1%)	1099
λ t1	0.7	–	–	4(0.7%)	1(0.2%)	0	613
λ t1	0.2	–	–	10(1.9%)	1(0.2%)	0	515
λ t1	-0.2	–	–	104(21.8%)	11(2.3%)	2(0.4%)	477
λ t1	-0.7	–	–	131(20.2%)	18(2.8%)	6(0.9%)	648
λ t1	-2	–	–	242(21.9%)	39(3.5%)	4(0.4%)	1105
λ t1	-5	–	–	128(25.7%)	22(4.4%)	4(0.8%)	499
λ t1	-2	+	–	224(23.4%)	29(3%)	23(2.4%)	959
λ t1	-2	+	+	93(17.0%)	0	10(1.8%)	548
t500	-2	–	–	60(21.1%)	10(3.5%)	2(0.7%)	284
λ t1+trpA1	-2	–	–	330(49.1%)	40(6.0%)	6(0.9%)	672
λ t1+trpA1	-2	–	+	137(16.8%)	12(1.5%)	4(0.5%)	817

Convergent Promoter Records

Conditions				Summary			
Terminator	Force (pN)	LacI& NusG	RNAP	Post-termination sliding	Sense re-initiation	Anti-sense re-initiation	Total number
λ t1+trpA1	-2	–	WT	303(57.4%)	113(21.4%)	28(5.3%)	528
λ t1+trpA1	-2	–	$\Delta\alpha$ -CTD	245(57.4%)	78(18.3%)	4(0.9%)	427

Divergent Promoter Records

Conditions				Summary			
Terminator	Force (pN)	LacI& NusG	RNAP	Post-termination sliding	Sense re-initiation	Anti-sense re-initiation	Total number
λ t1+trpA1	-2	–	WT	246(44.7%)	91(16.5%)	40(7.3%)	550
λ t1+trpA1	-2	–	$\Delta\alpha$ -CTD	212(47.9%)	68(15.3%)	4(0.9%)	443

Table 6.1: Summary of records using single promoter, convergent promoter, and divergent promoter templates. Each record was analyzed based on the post-termination fate of RNAP. Records that exhibited post-termination sliding and subsequent re-initiation of transcription were included in both the post-termination sliding and re-initiation categories.

These experiments shed light on a mechanism of repetitive transcription that may contribute significantly to gene regulation. A four-step mechanism from the completion of elongation of an RNA transcript to the start of a second round of transcription consists of (1) RNAP remaining associated with the DNA template after elongation, (2) RNAP sliding or diffusing along the DNA, (3) RNAP stopping

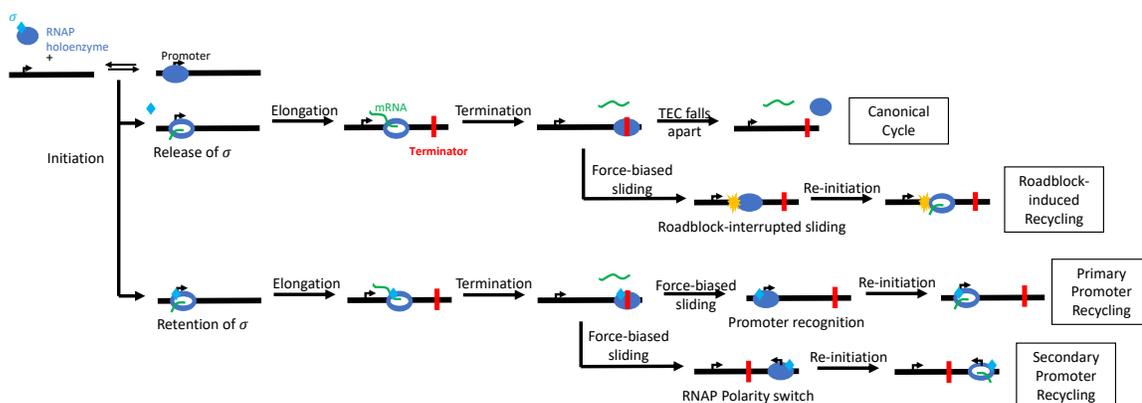


Figure 6.4: The post-termination fate of RNAP includes secondary transcription. After termination, RNAP (blue) may release mRNA (green) and dissociate from the DNA, or slide along the template. If blocked by a protein roadblock (yellow), RNAP may re-open a transcription bubble at the roadblock site and exhibit roadblock-induced recycling, a process independent of the presence of σ factor (cyan). Upon recognition of a distant promoter with the help of σ factor, wild-type RNAP may re-initiate transcription at promoters oriented in either direction.

at obstacles or promoters, and (4) RNAP re-initiating transcription at these sites. In step (1), the RNAP-DNA complex likely undergoes some conformational change that allows the RNAP to slide freely along DNA. This may resemble a final stage of transcription termination in which the DNA strands of the transcription bubble re-hybridize [185]. In step (2), RNAP either diffuse along the DNA as previously reported [18, 20], or slide rapidly in one direction driven by as little as sub-piconewton levels of force. In step (3), RNAP (i) recognized promoters in either orientation or only aligned in the direction of the previous elongation if the α -CTDs are missing or (ii) become blocked by DNA-bound proteins. In step (4), RNAP can form open complexes and initiate transcription at promoters with a characteristic delay of about a minute, or even at a non-promoter site if roadblocked for 5-fold longer. The presence of DNA-bound proteins, σ factors and tension would significantly bias the products of the secondary transcription, as illustrated in Figure 6.4.

Subtle force opposing or assisting RNAP translocation and DNA-bound proteins (roadblocks) along the template could greatly affect the efficiency and output of repetitive transcription. Even a tiny amount of force greatly reduces the time required to

slide to a re-initiation site. Moreover, force restricts sliding to a single direction along a DNA sequence. Opposing force increases the frequency of repetitive transcription of a single promoter by an individual RNAP and necessarily diminishes gene expression from other promoters further downstream. Assisting force prohibits repetitive transcription and drives a sliding post-termination complex to a downstream promoter. Thus the direction of force acting on RNAP might modulate the expression of one or more genes. Indeed, it biases transcription between convergent promoters without transcriptional interference between RNAPs initiating from opposing promoters [190, 191]. The collision model would not predict similar levels of interference between divergent genes. In contrast, force-directed RNAP diffusion produces preferential expression of one transcript independently of the arrangement of promoters or additional regulatory factors [192]. Even factors like RapA, which was recently shown to accelerate dissociation of post-termination complexes [179] and would therefore diminish post-elongation sliding, would not weaken preferential, force-driven expression.

In absence of free σ factor in solution, less than 10% of post-termination complexes exhibited repetitive transcription at a promoter site, which is likely the portion of post-termination complexes retaining σ factor [178]. Adding NusG reduced that ratio to $\sim 2\%$ suggesting that retention of σ factor enhanced sliding and repetitive transcription. In the absence of force, high affinity roadblocks would limit the sequences accessible to diffusing RNAPs and increase dwell times. However, a sliding RNAP driven by force against a high affinity roadblock might pause long enough to start transcription with or without σ factor. This mechanism may differ from canonical transcription initiation in which the σ factor assists promoter recognition and DNA strand separation. How a core RNAP enzyme might form an open bubble remains to be investigated.

Indeed, these experiments indicate that post-termination sliding occurs without a transcription bubble. First, if a sliding complex were to include a transcription

bubble, free NTPs in the buffer would have access to the catalytic site in which they could base pair with the template and limit threading of the template through the enzyme to the rate of nucleotide addition. However, sliding rates far exceeded measured *in vitro* transcription rates even under the lowest external mechanical forces, and a transcription bubble travelling at these speeds would consume a prodigious amount of chemical energy to unzip double-stranded DNA. Second, sliding of a transcription bubble cannot explain our findings that (a) sigma factor, which helps form the transcription bubble, is crucial for promoter re-initiation, and (b) re-initiation from a non-promoter site requires 5-fold longer dwell time (very likely for opening a transcription bubble). Third, RNAPs re-initiated with the same delay at forward and reverse promoters, indicating RNAPs formed transcription bubbles to re-initiate in either direction.

It is curious that post-termination sliding has not been previously reported in force spectroscopy assays of transcription [40, 80, 66, 181, 193] in which post-termination complexes mostly dissociated from the template. One might hypothesize that in these experiments, RNAP slid too fast to seize the promoter and ran off the end of the template. However, the current experiments show that post-termination complexes often pause for several seconds at the terminator before sliding at as much as 500 bp/sec driven by forces as high as 5 pN. At this sliding rate they also readily seize promoters and re-initiate transcription. This finding contradicts the previous assumption that 3 pN of tension could make sliding too fast and transient to be detected [18]. Alternatively, higher forces employed in much of the previous work might disrupt weak interactions between a sliding RNAP and the DNA backbone, although optically resolved sliding was not observed in transcription interference assays without force either [194]. Most previous measurements utilized optical trapping with a feedback mechanism to maintain constant force. This may be suitable to monitor slower processive steps of molecular motors (transcribing RNAPs), but high band-

width feedback may be critical for faster, continuous, non-processive events (sliding RNAPs) [195].

It remains unclear if other RNA polymerases exhibit post-termination sliding. There is no experimental evidence of such sliding by eukaryotic polymerases. However, eukaryotic cells evolved a specialized polymerase (Pol I) with a higher loading rate to amplify rRNA production [196], which might be an efficient substitute.

Force significantly directed sliding of post-termination complexes to accelerate repetitive transcription and modulate the relative utilization of adjacent promoters. Stringent control of the constituents afforded by the single molecule assembly revealed that σ^{70} often remained associated with the RNAP core enzyme to enhance re-initiation after sliding. Furthermore, DNA-binding proteins acting as roadblocks to sliding established non-promoter locations at which RNAP re-initiated transcription with five-fold greater delays than those observed for promoters. Sliding post-termination complexes indiscriminately utilized promoters oriented in either direction but could not easily switch template strands once the α -CTDs were deleted. These experiments highlight how very slight forces affecting post-termination diffusion of RNAP significantly impact transcription regulation.

Chapter 7

Transcription in Crowded Cellular Milieus

The cellular environment, characterized by its high level of molecular crowding, plays a crucial role in influencing a myriad of biological processes. Up to 40% of the cellular volume is occupied by various macromolecules, creating a dense, crowded milieu within the limited space of a few cubic micrometers in bacteria and around a hundred cubic micrometers in eukaryotic cells [197, 198, 199, 200, 201]. This crowding, coupled with high compartmentalization, leads to elevated local concentrations of specific molecules, intensifying the crowded conditions even in areas without membrane-enclosed organelles [202].

The impact of such a crowded environment is profound, influencing DNA compaction and extension, the kinetics of DNA-related activities, and interactions between DNA and proteins. However, replicating these intricate cellular conditions *in vitro* remains a significant challenge. Most laboratory studies are conducted in dilute buffer solutions, creating a discrepancy between *in vitro* experiments and the reality of cellular activities. Recent efforts have focused on narrowing this gap by introducing synthetic or organic macromolecules into experimental setups, aiming to simulate the

cellular environment more accurately.

Nevertheless, the complex nature of biological cells, filled with structures like the cytoskeleton, various salts, proteins, and nucleic acids of differing sizes and configurations, makes perfect emulation difficult [203]. This crowded cellular environment crucially influences a wide array of biological functions, from viral infection and gene expression to chromosomal compaction, replication, and transcription [197, 198, 199, 204, 205].

In this chapter, we synthesize the current understanding of how macromolecular crowding affects DNA behavior and protein-DNA interactions, review theoretical models related to DNA in crowded environments, and identify gaps in the literature. Notably, the need for biologically relevant crowding agents, the simultaneous use of multi-sized crowders, and empirical links between macromolecular crowding and the liquid-liquid phase separation of nucleic materials are areas ripe for further exploration and investigation.

7.1 Compaction and Extension of DNA

In the context of an *E. coli* bacterium, a commonly examined model organism, the physical constraints within the cell are remarkable. An *E. coli* cell, typically around one micron in length, houses DNA that, if stretched out, would extend over a millimeter, indicating a substantial degree of compaction [206]. Additionally, the cell's internal milieu is densely packed, with solutes accounting for 30–40% of the intracellular volume (Figure 7.1) [207]. This degree of molecular crowding is critical for the formation and maintenance of various dense DNA structures such as rods, fibers, flexible rings, toroids, and hierarchical coils [22], and it significantly influences DNA packaging and overall architecture.

DNA compaction and decompaction are vital processes for the effective storage,

maintenance, and processing of genetic information within a cell. In prokaryotes, the regulation of DNA organization involves a delicate balance between DNA supercoiling [208], macromolecular crowding, and interactions with nucleoid-associated proteins (NAPs) [209, 210, 211].

Various agents, such as multivalent cations [212], cationic lipids [213], detergents [214], peptides [215], and non-interacting polymers like polyethylene glycol (PEG), dextran, and bovine serum albumin (BSA), are known to induce the collapse of DNA chains. This phenomenon is particularly relevant in the context of gene therapy, where DNA packaging efficiency is crucial. Moreover, the ability to modulate the condensation and decondensation equilibrium of DNA is fundamental to gene regulation, as it controls the accessibility of the DNA double helix.

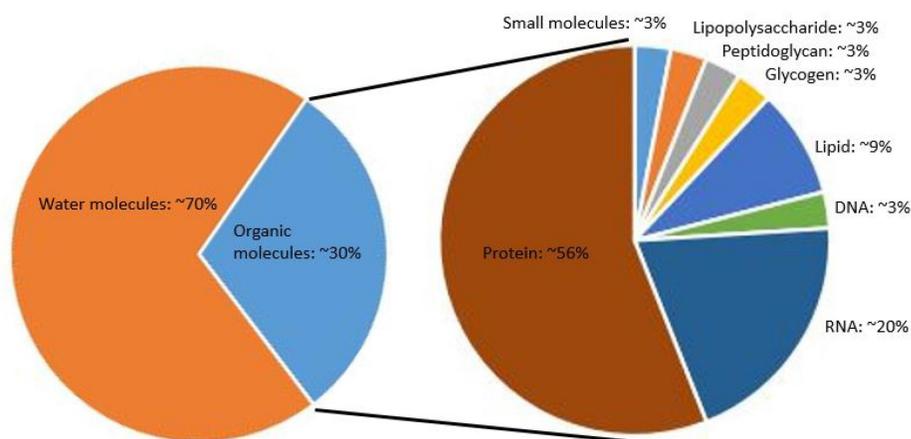


Figure 7.1: Molecular components of an *E. coli* cell. Adapted from [203, 216]

7.2 Liquid-liquid Phase Separation (LLPS)

The concept of phase separation in the cytoplasm due to macromolecular crowding has been a topic of interest for several decades, with early hypotheses suggesting its role in microcompartmentalization as far back as 1995 by Walter and Brooks [217].

Recent advancements have solidified the significance of liquid–liquid phase separation (LLPS) in various cellular functions. For instance, Levone et al. demonstrated the crucial role of the DNA/RNA-binding protein FUS in initiating DNA repair through LLPS [218]. Similarly, Shakya et al.’s study on the LLPS of histone proteins shed light on their role in chromatin organization, revealing that histones contribute to heterochromatin formation through reversible LLPS with DNA [219].

These developments have been complemented by investigations into the role of polyethylene glycol (PEG) in promoting LLPS. Park et al. combined experimental approaches and field-theory simulations to demonstrate that PEG drives LLPS by dehydrating polymers [220]. This has significant implications for our understanding of how crowding agents like PEG shift binding equilibria and extend the range of conditions under which molecular interactions occur. Such findings are reinforced by studies indicating that macromolecular crowding can significantly impact DNA packaging, the association and aggregation of polymers, and the folding of extended polypeptides [221, 222, 223].

Moreover, the crowding effects of PEG and other polymers have been shown to influence the thermal stability of both long and short double-stranded DNA (dsDNA) structures, as well as triple-stranded DNA [224, 225]. For instance, PEG increases the melting point of dsDNA and has a more pronounced effect on the stability of triple-stranded DNA, even in the presence of mismatched bases [224, 226, 227].

The emerging understanding of LLPS and droplet formation in the highly crowded cellular environment highlights the need for further systematic studies in this area. As Kohata et al. point out, the selective interactions and specific functions of biomolecules in these crowded environments exhibit complex temporal and spatial patterns [203]. This complexity is evident in the formation of various membraneless organelles or droplets within cells, which serve essential biological functions related to the storage and regulation of RNA and transcription factors [228, 229, 230, 231, 232].

The ability of these droplets to dynamically exchange contents with their surroundings underscores their importance in cellular regulation, especially in response to external signals and stress [233].

7.3 Theoretical Models To Explain LLPS

Theoretical models that delve into the effects of macromolecular crowding on DNA and other polymers primarily focus on the phenomenon of volume exclusion and how it's influenced by the relative sizes of the crowders. Two fundamental models commonly employed in these studies are the Asakura–Oosawa (AO) model and the Kirkwood–Buff (KB) model.

The Asakura–Oosawa model is instrumental in analyzing the phase behavior of polymers and colloidal particles within a solvent. In this model, polymers are conceptualized as hard spheres, exerting exclusion forces not only on each other but also on solvent molecules (Figure 7.2 left). This exclusion creates osmotic pressure (Π) and results in depletion forces, contributing to a change in the free energy (ΔG), which is represented as $\Delta G = \Pi$. A key aspect of the AO model is its coarse-grained approach, which effectively captures the excluded volume effect among polymers and colloidal particles. However, it falls short in considering specific interactions, such as those between polymer branches or interactions between polymers and solvent molecules.

The Kirkwood–Buff model, in contrast, delves deeper into the spatial distribution of polymer molecules and their distinct interactions with other polymers and the solvent (Figure 7.2 right). It approaches the calculation of the chemical potential between two components in a solution by integrating the spatial distribution functions that define the interactions between these components. For basic two-component systems involving unbranched polymers, the KB model arrives at a conclusion of free energy that aligns with the predictions made by the AO model.

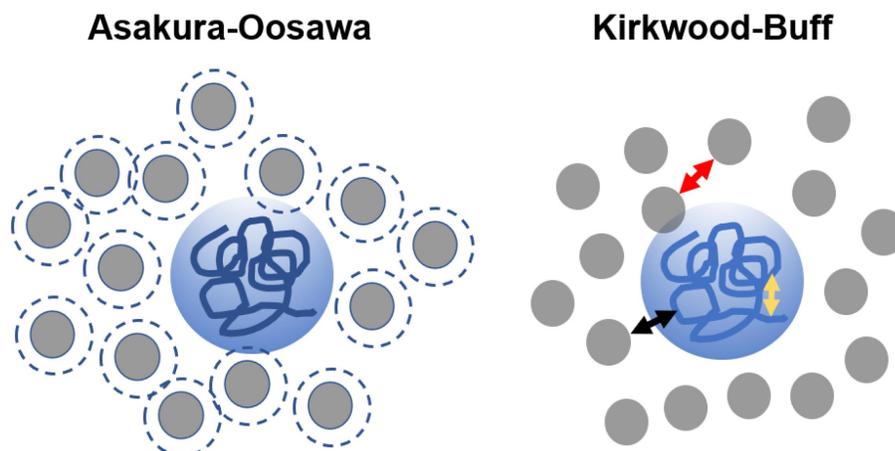


Figure 7.2: Schematic representation of the physical basis for the Asakura–Oosawa (AO; left) and the Kirkwood–Buff (KB; right) models. In the AO model, the molecules involved (here, DNA and crowder) are considered as hard spheres. The dashed line around each crowder “sphere” identifies the particle’s excluded volume. In the KB model, the arrows indicate the different types of molecular interactions (red: crowder–crowder; yellow: DNA–DNA; Black: crowder–DNA). Adapted from [216].

The study by Cao et al. using Langevin simulations explored the intriguing dynamics of a semi-flexible polymer chain in a concentrated solution filled with active (self-propelling) spherical crowders [234]. A standout observation from their research was the shrinkage-to-swelling transition observed in polymers of low rigidity. This transition was subject to a non-monotonic dependency on the dynamic persistence path of the active crowders, as revealed by a phase diagram constructed in the parameter space of active force and crowder size. The research revealed a nuanced interplay between crowder size and motile activity (force) on the polymer’s conformation: While smaller crowders amplified the crowding-induced shrinkage due to increased motile activity, larger crowders, conversely, limited this effect and led to the polymer’s swelling. In the case of large crowders, the swelling effect driven by motile activity outweighed the traditional crowding effects.

The intriguing phenomenon of effective attractive interactions between and within macromolecules in cosolute-containing solutions has been well-reviewed in the past [235]. Sukenik et al. have explored how cosolutes that are preferentially excluded

from macromolecular interfaces can lead to such interactions. Although the impact of cosolutes is often overlooked in crowding studies, incorporating them into models like the Asakura–Oosawa model and Kirkwood–Buff solution theory can enrich our understanding of crowding effects. By modifying the steric repulsion core with a ‘soft’ repulsive shell in these models, an enthalpic component is added to the depletion force, allowing for a more comprehensive understanding of cosolute effects. This approach also allows for a temperature-dependent depletion force that balances enthalpy and entropy, capturing the essence of osmolyte impacts on macromolecules [236, 237, 238].

Further, a theoretical model based on the Flory–Huggins approximation, which considers cosolutes in terms of size and two temperature-dependent interaction parameters, effectively describes protein stabilization in crowded cosolute solutions. This model not only aligns well with experimental findings but also offers insights into the entropic and enthalpic aspects of the depletion forces induced by specific cosolutes. The resulting depletion attraction is conceptualized as an effect of an ‘effective’ volume, shaped by the interplay of solvent, cosolute, and macromolecular interactions.

Finally, Monte Carlo simulations conducted by Shin et al. provided valuable insights into the impact of crowder size on polymer dynamics [239]. They observed that while small crowders tend to increase the solution’s effective viscosity and slow down chain dynamics, larger crowders facilitate coiling through confinement effects. These studies collectively enhance our understanding of the complex interactions within crowded environments.

Chapter 8

A Phenomenological Model of crowding effect on DNA under tension

In the preceding chapters, we investigated the relationship between DNA polymer properties and crowding forces, emphasizing the biological significance of understanding the physics behind these interactions. While prior research largely focused on simulations, an analytical theory could offer more insightful and direct connections between variables like crowder fractions and tension, and outcomes such as DNA polymer compaction or extension, and phase separation. Previous modeling approaches often faced a dilemma: either they were too coarse-grained to accurately predict specific scenarios, or they were overly complex for practical application.

In this chapter, our goal is to develop a phenomenological model grounded in mean-field theories. This model aims to predict the behavior of DNA polymers, such as collapse, pulling, and condensation, under various crowding and tension conditions. The mean-field approach simplifies the complex interactions in crowded environments by averaging the effects of the crowders over the space, thus offering a more tractable

yet insightful analysis.

This model starts with considering the DNA polymer as a flexible chain in a solution filled with crowding agents. The crowders are treated as spherical objects that exert entropic forces on the DNA polymer due to volume exclusion. The DNA polymer's response to these forces depends on its intrinsic properties, such as stiffness and length, as well as external factors like the concentration and size of the crowding agents. The model incorporates tension as a key variable. In the cellular environment, DNA polymers often experience mechanical tension due to processes such as transcription and replication. This tension can influence how the DNA polymer interacts with the crowding agents and its consequent behavior, like extension or collapse.

A key aspect of our phenomenological model is its ability to predict phase transitions in the DNA polymer, such as the transition from a coiled to an extended state or vice versa. By adjusting parameters like crowder concentration, DNA polymer stiffness, and applied tension, the model can simulate various scenarios and offer predictions that are experimentally testable.

Overall, this model seeks to bridge the gap between simplified theoretical approaches and the complexity of real biological systems. It aims to provide a tool for understanding the fundamental physics of DNA polymers in crowded environments, with implications for our understanding of various cellular processes.

8.1 Model DNA Polymers As the Shish-Kebab

In the quest to accurately model DNA polymer behavior in crowded environments, our first task is to find a representation of DNA that adequately reflects its physical properties. DNA is notably stiff, with a persistence length (the length over which it maintains its directionality) substantially larger than its diameter. Specifically, DNA

has a diameter (d) of approximately 2 nm and a persistence length (l_p) of about 35 nm. This significant difference poses a challenge in conceptualizing DNA for modeling purposes.

A common approach is the bead-on-string model, where the polymer is envisioned as a series of beads (monomer units) connected by flexible strings (Figure 8.1A). This model is often employed to represent chromosomal structures, with DNA-wrapped histones (nucleosomes) as the beads and the intervening DNA as the string. This analogy works well for modeling the basic structure of chromatin; however, when it comes to representing a linear DNA polymer, this model faces a critical limitation.

Given DNA's slender nature and the disproportionate ratio between its diameter and persistence length, a bead-on-string conceptualization would have to envision DNA as a string peppered with sparse and small beads. Such a representation would significantly underestimate the physical volume of DNA, leading to a miscalculation of the pairwise potential between adjacent beads. This underestimation becomes particularly critical when considering DNA's interactions in crowded environments, as the physical volume of DNA directly influences how it interacts with surrounding macromolecules.

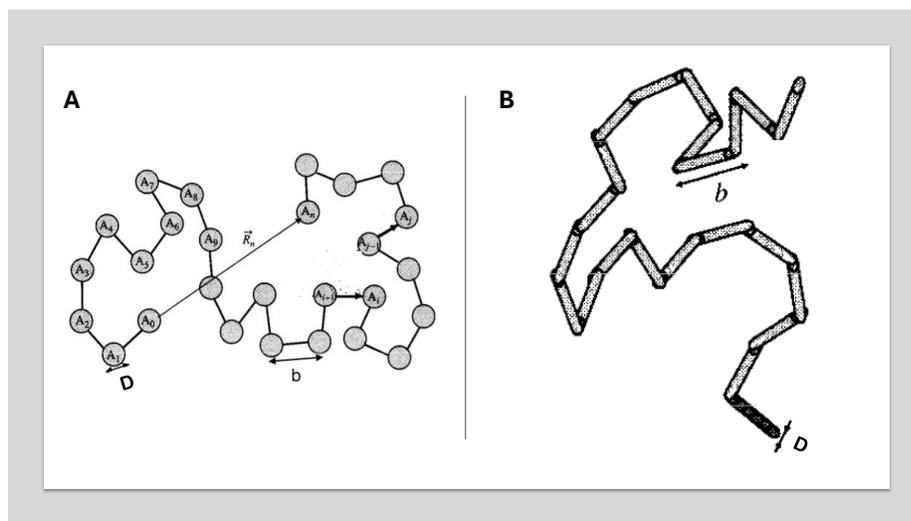


Figure 8.1: Illustration of (A) the Beads-On-String model and (B) the Shish-Kebab model for the DNA. Adapted from [240]

The Shish-Kebab model offers a more sophisticated approach to represent DNA polymers, balancing the need to account for both the slenderness and the physical volume of DNA. Unlike the bead-on-string model, which uses spherical beads, the Shish-Kebab model employs elongated cylindrical elements to mimic the monomers of DNA (Figure 8.1B). This model is visually akin to a series of kebabs on a skewer, with the skewer representing the backbone of the DNA and the kebabs symbolizing the stacked base pairs.

8.2 Thermodynamics of a globular state DNA

To develop a comprehensive understanding of the free energy of a polymer globule in a crowded environment, we must consider various factors. Based on the description provided, we can formulate an expression for the total free energy of the globule. The key elements to consider include the internal potential energy of monomers within the globule, the surface energy due to monomer-solvent interactions, and the volume of the globule as defined in the Shish-Kebab model (Figure 8.2).

1. Volume of the Globule: In the Shish-Kebab framework, the volume of the globule is given by $V = N_g b D^2$, where N_g is the number of monomers, b is a characteristic length scale, and D is the diameter of a cylinder (monomer).
2. Potential Energy Inside the Globule: Within the bulk of the globule, each monomer has a constant potential energy, which we take to be negative due to unfavorable interactions in a poor solvent. Let's denote this energy per monomer as $-u$.
3. Energy of Monomers on the Surface: Monomers on the surface of the globule have different energetics because they are exposed to the solvent. For a monomer at the surface, roughly half of its area is in contact with the solvent. We can approximate its potential energy as $u/2$.

4. Total Free Energy of the Globule: Considering these aspects, the total free energy of the globule in solution can be expressed by accounting for both the bulk and surface contributions.

$$F_g = -N_g u + \frac{A}{bD} \frac{u}{2} \quad (8.1)$$

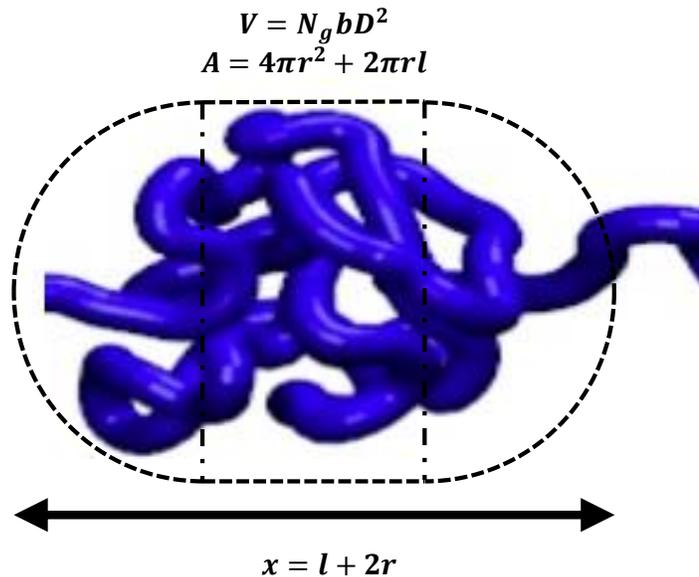


Figure 8.2: A schematic of the sphero-cylinder model of a globule with N_g monomers.

We now consider a situation where it is possible to apply a force f to the ends of the polymer. For a uni-axial force applied in the x -direction, we model the deformed polymer as a sphero-cylinder of a fixed volume determined by N_g (an alternative is to model as a uni-axial ellipsoid, but we find the algebra much easier for the spherocylinder with the results not significantly different). Evaluating the area of this sphero-cylinder, we obtain the free energy, after some simple maths, in the form

$$F_g = -N_g u + \left(\frac{2N_g D}{3x} + \sqrt{\frac{\pi N_g x}{b}} \right) u \quad (8.2)$$

When working with constant force, we need to use the Gibbs free energy for such

a stretched globule, which is given by:

$$G_{glo}(f) = F_g(x_{eq}(f)) - fx_{eq}(f), \quad (8.3)$$

where $x_{eq}(f)$ is defined by the condition of mechanical equilibrium:

$$\left. \frac{\partial F_g(x)}{\partial x} \right|_{x=x_{eq}(f)} = f, \quad (8.4)$$

the analysis can be simplified in the limit of small deformations from a sphere,

$$x_{eq}(f) = \frac{4}{3} \left(\frac{3N_g b D^2}{4\pi} \right)^{1/3} + \frac{16bD}{9\pi u} f \quad (8.5)$$

A natural attempt to estimate the parameter u as the the thermal energy multiplied to the normalized 2nd virial coefficient, to which only the attractive contribution is kept:

$$u \sim k_B T \frac{v_0 - v}{v_0} \quad (8.6)$$

Mayer formula can calculate the 2nd virial coefficient:

$$v = \int (1 - e^{-\beta E(\vec{r})}) d\vec{r} \quad (8.7)$$

However, estimate the 2nd virial coefficient under the Shish-Kebab framework is quite challenging. The first difficulty is how to estimate the inter-rods interaction energy under crowded environment. Another difficulty is how to estimate the hard-core excluded volume and excluded volume for the crowders, since adjacent rods might arrange in different angles, which contribute differently to the average excluded volume.

In the simplest case where two rods are aligned parallel, we can approximate the

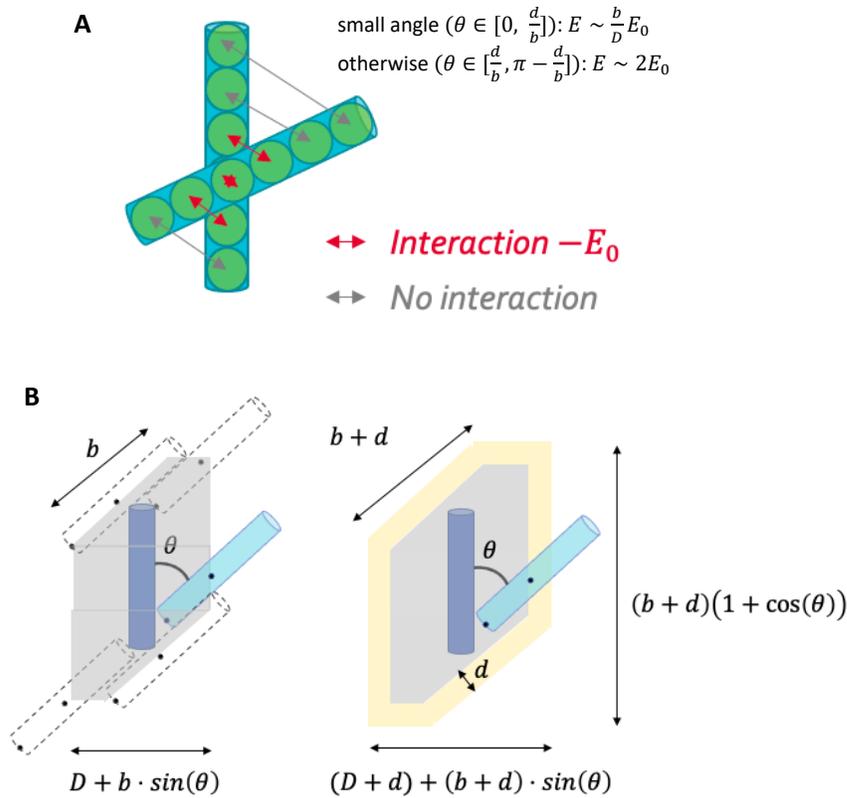


Figure 8.3: Geometry of two interacting rods in the Shish Kebab model. (A) Interaction between two rods can be evaluated by two parts. (B) Excluded volume for two rods hard-core (gray) and the shell area (yellow), as a function of a relative angle θ .

inter-rods interaction energy as the inter-beads interaction energy times the number of beads in a rod, in presence of crowder of size (diameter) d . The interaction energy between rods were previously quantified in a phenomenological formula [241],

$$\begin{aligned}
 \beta E_0 &= \frac{b}{D} \beta E_0 \\
 &= \frac{\phi}{1.25} \frac{b}{d} \ln \left[1 + \frac{1.35}{(1+b/D)(1+b/D)^2 - 1} \right] \\
 &\approx 0.22 \frac{\phi}{1.25} \frac{bD}{d^2}
 \end{aligned} \tag{8.8}$$

Regarding the second difficulty, as an approximation, consider that if θ is not too small, there are always two pairs of beads interacting, hence the interaction energy is $\epsilon_0 = 2E_0$. On the contrary, for small angles, almost all the beads are interacting, and $\epsilon_0 = \frac{b}{D} E_0$ (Figure 8.3A). The limit between these two situations is given by the

angle θ_{lim} , such that it satisfies the condition:

$$\begin{aligned}\sin(\theta_{lim}) &\approx \frac{b}{d} \\ \theta_{lim} &\approx \frac{b}{d}, \pi - \frac{b}{d}\end{aligned}\tag{8.9}$$

Therefore, we can dissect the overall angle distribution into two regimes and integrate energy on a shell of width D (the diameter of monomer crowders) around the hard-core excluded volume v_0 (Figure 8.3B). It yields the following integral:

$$\begin{aligned}\bar{v} &= \bar{v}_0 + \frac{1-e^{2E_0}}{2} \left[S_1 \int_{d/b}^{\pi-d/b} \sin(\theta) d\theta + S_2 \int_{d/b}^{\pi-d/b} \sin^2(\theta) d\theta \right] \\ &\quad + (1 - e^{bE_0/D}) \left[S_1 \int_0^{d/b} \sin(\theta) d\theta + S_2 \int_0^{d/b} \sin^2(\theta) d\theta \right] \\ &= \bar{v}_0 + (1 - e^{2E_0}) \left[S_1 \cos(d/b) + \frac{S_2}{2} \left(\frac{\pi}{2} - \sin(d/b) \cos(d/b) - d/b \right) \right] \\ &\quad + (1 - e^{bE_0/D}) \left[S_1 (1 - \cos(d/b)) + \frac{S_2}{2} (d/b - \sin(d/b) \cos(d/b)) \right],\end{aligned}\tag{8.10}$$

with the shell expression $S_1 = (b+d)(D+b)^2 - bD^2$ and $S_2 = (b+d)^2(D+b) - b^2D$.

With equation 8.10, we can now easily numerically evaluate Equations 8.3 - 8.6.

8.3 Thermodynamics of a stretched chain state DNA

To describe the extended chain phase, we first acknowledge that the stretched-out segment is where all monomers are exposed to the solvent, and the extension is a large proportion of the contour length of the exposed chain. Many classical models exist, from the Langevin function arising in the freely jointed chain model [242], to the widely used extrapolation formula for semiflexible chains by Marko and Siggia [243]. All of these early models have limits of applicability, but recently there has been a new formula introduced that is valid in all regimes of extension and chain flexibility [244]. For an inextensible semiflexible chain of contour length L , and persistence length l_p , the free energy of extension z combines the contribution from the bending stiffness

of the molecule (which is related to the persistence length l_p) and the entropic effects, which give the divergence as $z \rightarrow L$. The full expression for the chain free energy is

$$F_{ch}(z) = \frac{\pi^2 k_B T l_p}{2L} \left(1 - \left(\frac{z}{L} \right)^2 \right) + \frac{2k_B T L}{\pi l_p (1 - (z/L)^2)} \quad (8.11)$$

In the case of fully stretched long DNA polymer, we can take the approximation $l_p \approx b$ and $L \gg l_p$, and take the simplified expression

$$F_{ch}(z) \approx \frac{2k_B T L}{\pi l_p (1 - (z/L)^2)} - \frac{2k_B T L}{\pi l_p} \quad (8.12)$$

where the second term is added to fix the value of free energy at zero extension to zero.

Naturally, the next step is to calculate the Gibbs free energy of a stretched chain when working with constant force, which is given by:

$$G_{ch}(f) = F_{ch}(z_{eq}(f)) - f z_{eq}(f), \quad (8.13)$$

where $z_{eq}(f)$ is defined by the condition of mechanical equilibrium:

$$\left. \frac{\partial F_{ch}(z)}{\partial z} \right|_{z=z_{eq}(f)} = f. \quad (8.14)$$

Notice that the crowder fraction ϕ insignificantly affect the free energy of stretched chain state polymer compared to the globular state. Up to this point, we can use this model to predict some behaviors of DNA polymers under different tension and crowder fractions.

8.4 The model predicts the critical forces for DNA phase transition under different crowder conditions

The experimental measurements of the DNA phase transition under different force and crowder conditions have been carried on in Mantegazza group, by Cristofalo et al. [209]. It used the Magnetic Tweezers to study the phase transition of a 6kbp DNA fragment under forces ranging from 6 fN to 20pN, and PEG2000 crowders from 0% to 22%. MTs allow to measure in a straightforward manner the end-to-end extension as a function of the applied force (Figure 8.4A). The results are shown on Figure 8.4B. Note that each point is the average of a time series for extension under constant force. For crowder volume fractions $\phi > 10\%$, a critical force can be clearly identified, while below this fraction a smooth increase in end-to-end extension *versus* force is observed. This critical force f_c depends on crowding. From this result, together with evidence in literatures, one can confirm a tricritical behavior: at low crowding, the transition between a globule phase is continuous (2nd order transition) and there is no critical force. There would exist a critical volume fraction ϕ_c (tricritical point), above which the transition becomes of the 1st order, and thus a critical force can be identified, as well as a phase coexistence region.

To test the validity of our model, we use it to predict the experimentally observed critical force f_c . To find the force at the equilibrium thermodynamic phase transition, f_c , we equate the Gibbs energies of the globule and the chain, $G_{glo}(f) = G_{ch}(f)$ from Equations 8.3 and 8.13. Although it is very difficult to analytically solve for the critical force f_c , numerical solution is easy to compute for modern computers. A comparison between the model predicted and experimentally detected f_c is shown in Figure 8.4C&D.

We note that, in our model, only the bare DNA (0% crowder fraction) could exhibit

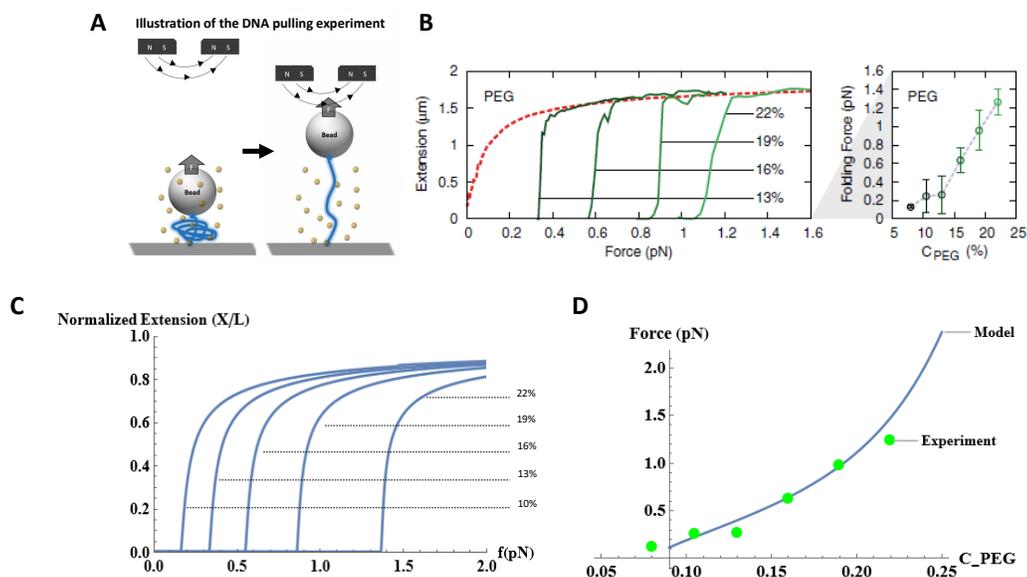


Figure 8.4: DNA pulling experiment results and model simulation. (a) Schematic illustration of the DNA pulling experiment. In presence of crowder, DNA was in a collapsed state in absence of tension. The end-to-end tether length increases as a function of tension. (b) Experimental results show that DNA polymer undergoes phase transition at critical force above a threshold crowder fraction. (c) The phenomenological model reproduces the tether length *versus* tension curve under different crowder fractions, and (D) the predicted critical tensions (blue curve) agree well with the experiment results (green dots).

continuous transition, which is in contradiction with the experimentally observed tricritical behavior starting at $> 10\%$. A way to justify this discrepancy is that there might exist a phase coexistence regime, which is neglected when we equilibrate the globular and chain state free energy.

8.5 Thermodynamics of a tadpole state DNA

A phase coexistence of globular and chain state DNA is characterized as a tadpole state, where a portion of DNA condensates to a globule while the rest of the segment remain stretched. We can examine the free energy of this tadpole state by mixing up the free energy of globular and chain parts. For a tadpole state DNA polymer with s fraction in the globular state and $1 - s$ in the chain state. We need to modify the

expression of Gibbs free energy of both states. For the globular state part, we have

$$\begin{aligned}
 F_g(x, s) &= -sN_g u + \left(\frac{2sN_g D}{3x} + \sqrt{\frac{\pi s N_g x}{b}} \right) u \\
 x_{eq}(f, s) &= \frac{4}{3} \left(\frac{3sN_g b D^2}{4\pi} \right)^{1/3} + \frac{16bD}{9\pi u} f \\
 G_{glo}(f, s) &= F_g(x_{eq}(f, s), s) - f x_{eq}(f, s),
 \end{aligned} \tag{8.15}$$

and for the chain state part, we have

$$\begin{aligned}
 F_{ch}(s) &= \frac{2k_B T L}{\pi l_p (1 - (1 - s^2))} - \frac{2k_B T L}{\pi l_p} \\
 G_{ch}(f, s) &= F_{ch}(s) - f(1 - s)L.
 \end{aligned} \tag{8.16}$$

The total Gibbs free energy of the tadpole state DNA is therefore

$$G_{tad}(f, s) = G_{ch}(f, s) + G_{glo}(f, s). \tag{8.17}$$

Note that, similar to the globular state Gibbs free energy in Equation 8.3, the effect of crowder fraction is achieved through the potential energy u in the expression of G_{glo} .

Figure 8.5 summarizes the free energies of DNA in the globular, the stretched chain and the tadpole state as described in Equations 8.3, 8.13 and 8.17. It clearly show that under a particular crowder fraction and force, the tadpole state is more energetically favored than both the globular and the chain states, therefore supporting the coexistence phase.

8.6 DNA relaxation experiment proves the coexistence state

We seek more supports of the coexistence state from experiment. Besides the smooth first order phase transition observed in the DNA pulling experiment (Figure 8.4A),

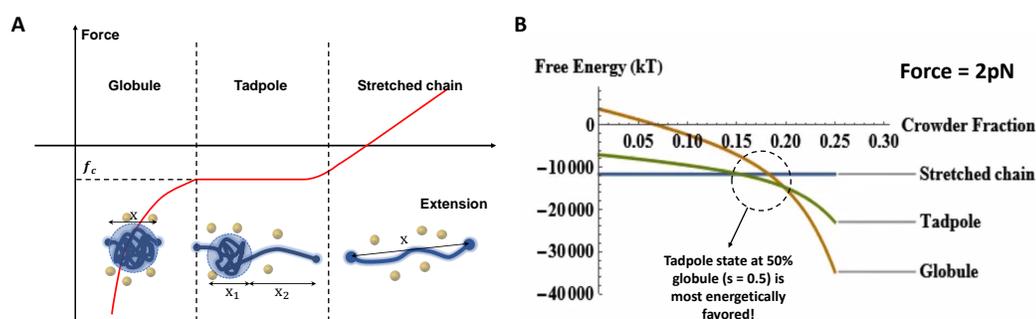


Figure 8.5: A comparison of free energies of DNA globule, tadpole, stretched chain. (A) DNA polymers undergo phase transition from a globule to a stretched chain, amid is a tadpole state. (b) The tadpole DNA can be more energetically favored than the globule and stretched chain in presence of crowder.

we proceed another set of DNA relaxation experiment using the Lumicks C-trap dual optical tweezers. Figure 8.6A demonstrates the setup of the experiment. At the beginning, a lambda DNA (48502 bp, $\sim 16\mu\text{m}$) is stretched to full length by the dual optical traps. Then the tension suddenly released to and maintained at a small level, allowing relaxation (condensation) of DNA. Under different crowder fractions, we measured the final end-to-end distance DNA reached, and observed a smooth decrease in the end-to-end distance of relaxed DNA *versus* the increase in crowder fractions. This smooth, first order transition strongly proves that DNA underwent the coexistence, tadpole state. Moreover, the model with the tadpole state accurately predicts the the relationship between end-to-end distance and tension, as shown in Figure 8.6B. Also, in Figure 8.4C, we show that the model can simulate the force *versus* DNA tether length curves obtained by the Mantegazza group. This good agreement across different experiment data set lends further support to the validity of our model.

Overall, this model shows that the globule could undergo a discontinuous jump in extension at a threshold force above a threshold crowder fraction, while undergo a continuous, smooth transition to the tadpole and eventually chain state with the

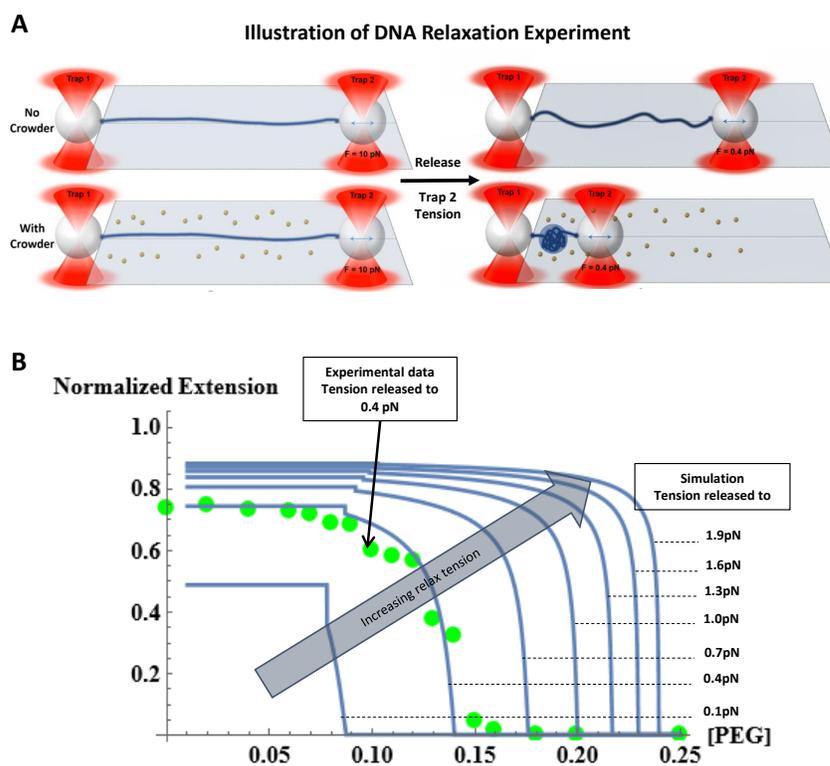


Figure 8.6: DNA relaxation experiment results and model simulation. (a) Schematic illustration of the DNA relaxation experiment. DNA polymer, both end attached to a polystyrene bead, was stretched to the full-length state in absence/presence of crowders, using the dual optical trap setup. The tension was suddenly released to a minimal value and maintained constant, allowing DNA polymer to relax under constant tension. We measured the normalized extension, the ratio of the final end-to-end distance to the fully stretched length. (b) The model predicts the normalized extension under different crowder fraction and tension (blue curves). The experiment 0.4pN data (green dots) agree well with the model.

increase of force in low crowder fraction. The model provides testable predictions of polymers states under a given condition of tension and crowder fraction.

Chapter 9

Conclusion

This dissertation has explored the intricate mechanisms of bacterial transcription through a multi-tiered analytical approach, employing both single-molecule biophysics and molecular biology techniques. The comprehensive analysis undertaken has not only expanded our understanding of transcription dynamics at the base-pair level but also elucidated the broader implications of transcription in crowded cellular environments.

The core findings of this research highlight the complexity and adaptability of the transcription process under various genomic constraints. Firstly, the development of a detailed model elucidating transcription kinetics at the base-pair level has provided significant insights. This model, informed by the thermal dynamics of transcription bubbles and nascent RNA structures, has proven effective in predicting transcriptional dynamics with high precision, thereby enhancing our understanding of genetic regulation.

Secondly, the interaction of RNA polymerase with DNA-bound roadblocks revealed mechanisms of navigational flexibility that are crucial for understanding transcription under genomic constraints. The study identified two distinct mechanisms: passive waiting for obstacle dissociation and active backtracking, recovery, and force-

ful passage, which are vital for RNA polymerase to overcome physical barriers along the DNA template.

Additionally, the concept of transcriptional recycling emerged as a novel finding from this study. The transcription recycling is identified as a force-biased regulatory mechanism that could significantly influence gene expression patterns. This discovery, involving the cycling of transcription by the same RNA polymerase enzyme, highlights an efficient method of transcript accumulation, reduces the probability of collisions among different RNA polymerases.

Lastly, the examination of DNA behavior in crowded cellular milieus provided valuable insights into chromatin dynamics and the physical properties of DNA under tension. The findings from these studies are crucial for understanding the molecular basis of gene expression and regulation within the complex cellular environment.

The implications of these findings are broad and significant. By providing a clearer picture of the transcriptional dynamics at a molecular level, this research contributes to the foundational understanding necessary for developing targeted therapeutic interventions. The knowledge gained could lead to novel strategies for manipulating transcription in bacterial cells.

Moreover, the insights into the mechanics of transcriptional recycling and the behavior of DNA in crowded environments open new avenues for research in genetic regulation and cellular biology. These findings could influence future studies in other organisms, including eukaryotes, thereby broadening the scope of transcriptional research.

While this dissertation has addressed several key aspects of bacterial transcription, further research is needed to deepen our understanding and verify the new models proposed. Future studies might aim to:

1. Explore the quantitative impact of transcriptional factors and external tensions more deeply to refine the current models of transcription kinetics.

2. Investigate the role of transcriptional recycling in other bacterial species and possibly in eukaryotic systems to assess its universality and functional implications.
3. Develop more sophisticated *in vivo* experiments to observe and measure transcription dynamics within living cells, which could validate the findings obtained from *in vitro* studies.

In conclusion, this dissertation not only furthers our understanding of the biochemical mechanisms underlying bacterial transcription but also sets the stage for innovative research in molecular biology. The insights from this study underscore the intricate interplays of molecular components in gene expression and regulation. The methodologies developed and findings reported herein provide a robust framework for future investigations aiming to unravel the complexities of gene expression and regulation.

Finally, I would like to express my deepest gratitude to my advisor, Dr. Laura Finzi and Dr. David Dunlap, committee members, colleagues at Emory University, and my family for their unwavering support and guidance throughout this academic endeavor. This dissertation stands as a testament to the collaborative spirit and intellectual rigor that have shaped my academic journey.

Bibliography

- [1] J. P. Richardson and J. Greenblatt, “Control of rna chain elongation and termination,” *Escherichia coli and Salmonella: cellular and molecular biology*, vol. 1, pp. 822–848, 1996.
- [2] S. Uptain, C. Kane, and M. Chamberlin, “Basic mechanisms of transcript elongation and its regulation,” *Annual review of biochemistry*, vol. 66, no. 1, pp. 117–172, 1997.
- [3] K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. M. Block, “Sequence-resolved detection of pausing by single rna polymerase molecules,” *Cell*, vol. 125, no. 6, pp. 1083–1094, 2006.
- [4] E. A. Abbondanzieri, W. J. Greenleaf, J. W. Shaevitz, R. Landick, and S. M. Block, “Direct observation of base-pair stepping by rna polymerase,” *Nature*, vol. 438, no. 7067, pp. 460–465, 2005.
- [5] I. Artsimovitch and R. Landick, “Pausing by bacterial rna polymerase is mediated by mechanistically distinct classes of signals,” *Proceedings of the National Academy of Sciences*, vol. 97, no. 13, pp. 7090–7095, 2000.
- [6] L. Bai, A. Shundrovsky, and M. D. Wang, “Sequence-dependent kinetic model for transcription elongation by rna polymerase,” *Journal of molecular biology*, vol. 344, no. 2, pp. 335–349, 2004.

- [7] L. Bai, R. M. Fulbright, and M. D. Wang, “Mechanochemical kinetics of transcription elongation,” *Physical review letters*, vol. 98, no. 6, p. 068103, 2007.
- [8] D. Ó. Maoiléidigh, V. R. Tadigotla, E. Nudler, and A. E. Ruckenstein, “A unified model of transcription elongation: what have we learned from single-molecule experiments?,” *Biophysical journal*, vol. 100, no. 5, pp. 1157–1166, 2011.
- [9] V. R. Tadigotla, D. Ó. Maoiléidigh, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein, “Thermodynamic and kinetic modeling of transcriptional pausing,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 12, pp. 4439–4444, 2006.
- [10] T. D. Yager and P. H. Von Hippel, “A thermodynamic analysis of rna transcript elongation and termination in escherichia coli,” *Biochemistry*, vol. 30, no. 4, pp. 1097–1118, 1991.
- [11] J. Qian, D. Dunlap, and L. Finzi, “Thermodynamic model of bacterial transcription,” *Phys. Rev. E*, vol. 106, p. 044406, Oct 2022.
- [12] J. T. Kadonaga, “Regulation of rna polymerase ii transcription by sequence-specific dna binding factors,” *Cell*, vol. 116, no. 2, pp. 247–257, 2004.
- [13] D. P. Ramji and P. Foka, “Ccaat/enhancer-binding proteins: structure, function and regulation,” *Biochemical Journal*, vol. 365, no. 3, pp. 561–575, 2002.
- [14] B. J. Wilkins, N. A. Rall, Y. Ostwal, T. Kruitwagen, K. Hiragami-Hamada, M. Winkler, Y. Barral, W. Fischle, and H. Neumann, “A cascade of histone modifications induces chromatin condensation in mitosis,” *Science*, vol. 343, no. 6166, pp. 77–80, 2014.

- [15] S. E. Halford and J. F. Marko, “How do site-specific dna-binding proteins find their targets?,” *Nucleic acids research*, vol. 32, no. 10, pp. 3040–3052, 2004.
- [16] N. Ribeck and O. A. Saleh, “Multiplexed single-molecule measurements with magnetic tweezers,” *Review of Scientific Instruments*, vol. 79, no. 9, 2008.
- [17] J. G. Piccolo, J. Méndez Harper, D. McCalla, W. Xu, S. Miller, J. Doan, D. Kovari, D. Dunlap, and L. Finzi, “Force spectroscopy with electromagnetic tweezers,” *Journal of Applied Physics*, vol. 130, no. 13, 2021.
- [18] T. T. Harden, K. S. Herlambang, M. Chamberlain, J.-B. Lalanne, C. D. Wells, G.-W. Li, R. Landick, A. Hochschild, J. Kondev, and J. Gelles, “Alternative transcription cycle for bacterial rna polymerase,” *Nature Communications*, vol. 11, no. 1, p. 448, 2020.
- [19] W. Kang, K. S. Ha, H. Uhm, K. Park, J. Y. Lee, S. Hohng, and C. Kang, “Transcription reinitiation by recycling rna polymerase that diffuses on dna after releasing terminated rna,” *Nature Communications*, vol. 11, no. 1, p. 450, 2020.
- [20] W. Kang, S. Hwang, J. Y. Kang, C. Kang, and S. Hohng, “Hopping and flipping of rna polymerase on dna during recycling for reinitiation after intrinsic termination in bacterial transcription,” *International Journal of Molecular Sciences*, vol. 22, no. 5, p. 2398, 2021.
- [21] V. F. Scolari and M. C. Lagomarsino, “Combined collapse by bridging and self-adhesion in a prototypical polymer model inspired by the bacterial nucleoid,” *Soft matter*, vol. 11, no. 9, pp. 1677–1687, 2015.
- [22] V. B. Teif and K. Bohinc, “Condensed dna: condensing the concepts,” *Progress in biophysics and molecular biology*, vol. 105, no. 3, pp. 208–222, 2011.

- [23] T. Odijk, “Osmotic compaction of supercoiled dna into a bacterial nucleoid,” *Biophysical chemistry*, vol. 73, no. 1-2, pp. 23–29, 1998.
- [24] S. Bustin, “Molecular biology of the cell, ; isbn: 9780815344643; and molecular biology of the cell, the problems book; isbn 9780815344537,” 2015.
- [25] J. Huang and R. Sousa, “T7 rna polymerase elongation complex structure and movement,” *Journal of molecular biology*, vol. 303, no. 3, pp. 347–358, 2000.
- [26] R. A. Ikeda and C. C. Richardson, “Interactions of the rna polymerase of bacteriophage t7 with its promoter during binding and initiation of transcription.,” *Proceedings of the National Academy of Sciences*, vol. 83, no. 11, pp. 3614–3618, 1986.
- [27] D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, and I. Artsimovitch, “Structural basis for transcription elongation by bacterial rna polymerase,” *Nature*, vol. 448, no. 7150, pp. 157–162, 2007.
- [28] I. Touloukhonov, J. Zhang, M. Palangat, and R. Landick, “A central role of the rna polymerase trigger loop in active-site rearrangement during transcriptional pausing,” *Molecular cell*, vol. 27, no. 3, pp. 406–419, 2007.
- [29] A. Mazumder, M. Lin, A. N. Kapanidis, and R. H. Ebright, “Closing and opening of the rna polymerase trigger loop,” *Proceedings of the National Academy of Sciences*, vol. 117, no. 27, pp. 15642–15649, 2020.
- [30] X. Guo, A. G. Myasnikov, J. Chen, C. Crucifix, G. Papai, M. Takacs, P. Schultz, and A. Weixlbaumer, “Structural basis for nusa stabilized transcriptional pausing,” *Molecular cell*, vol. 69, no. 5, pp. 816–827, 2018.
- [31] I. Touloukhonov, I. Artsimovitch, and R. Landick, “Allosteric control of rna

- polymerase by a site that contacts nascent rna hairpins,” *Science*, vol. 292, no. 5517, pp. 730–733, 2001.
- [32] M. H. Larson, R. Landick, and S. M. Block, “Single-molecule studies of rna polymerase: one singular sensation, every little step it takes,” *Molecular cell*, vol. 41, no. 3, pp. 249–262, 2011.
- [33] M. Dangkulwanich, T. Ishibashi, L. Bintu, and C. Bustamante, “Molecular mechanisms of transcription through single-molecule experiments,” *Chemical reviews*, vol. 114, no. 6, pp. 3203–3223, 2014.
- [34] M. L. Kireeva and M. Kashlev, “Mechanism of sequence-specific pausing of bacterial rna polymerase,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 22, pp. 8900–8905, 2009.
- [35] A. Weixlbaumer, K. Leon, R. Landick, and S. A. Darst, “Structural basis of transcriptional pausing in bacteria,” *Cell*, vol. 152, no. 3, pp. 431–441, 2013.
- [36] P. Thomen, P. J. Lopez, and F. Heslot, “Unravelling the mechanism of rna-polymerase forward motion by using mechanical force,” *Physical Review Letters*, vol. 94, no. 12, p. 128102, 2005.
- [37] J. E. Foster, S. F. Holmes, and D. A. Erie, “Allosteric binding of nucleoside triphosphates to rna polymerase regulates transcription elongation,” *Cell*, vol. 106, no. 2, pp. 243–252, 2001.
- [38] Y. W. Yin and T. A. Steitz, “The structural mechanism of translocation and helicase activity in t7 rna polymerase,” *Cell*, vol. 116, no. 3, pp. 393–404, 2004.
- [39] G. Bar-Nahum, V. Epshtein, A. E. Ruckenstein, R. Rafikov, A. Mustaev, and E. Nudler, “A ratchet mechanism of transcription elongation and its control,” *Cell*, vol. 120, no. 2, pp. 183–193, 2005.

- [40] J. W. Shaevitz, E. A. Abbondanzieri, R. Landick, and S. M. Block, “Backtracking by single rna polymerase molecules observed at near-base-pair resolution,” *Nature*, vol. 426, no. 6967, pp. 684–687, 2003.
- [41] N. Komissarova and M. Kashlev, “Transcriptional arrest: Escherichia coli rna polymerase translocates backward, leaving the 3 end of the rna intact and extruded,” *Proceedings of the National Academy of Sciences*, vol. 94, no. 5, pp. 1755–1760, 1997.
- [42] S. Naftelberg, I. E. Schor, G. Ast, and A. R. Kornblihtt, “Regulation of alternative splicing through coupling with transcription and chromatin structure,” *Annual review of biochemistry*, vol. 84, pp. 165–198, 2015.
- [43] T. Saldi, M. A. Cortazar, R. M. Sheridan, and D. L. Bentley, “Coupling of rna polymerase ii transcription elongation with pre-mrna splicing,” *Journal of molecular biology*, vol. 428, no. 12, pp. 2623–2635, 2016.
- [44] M. H. Larson, R. A. Mooney, J. M. Peters, T. Windgassen, D. Nayak, C. A. Gross, S. M. Block, W. J. Greenleaf, R. Landick, and J. S. Weissman, “A pause sequence enriched at translation start sites drives transcription dynamics in vivo,” *Science*, vol. 344, no. 6187, pp. 1042–1047, 2014.
- [45] F. Gong and C. Yanofsky, “A transcriptional pause synchronizes translation with transcription in the tryptophanase operon leader region,” *Journal of bacteriology*, vol. 185, no. 21, pp. 6472–6476, 2003.
- [46] J. Zhou, K. S. Ha, A. La Porta, R. Landick, and S. M. Block, “Applied force provides insight into transcriptional pausing and its modulation by transcription factor nusA,” *Molecular cell*, vol. 44, no. 4, pp. 635–646, 2011.
- [47] J. Y. Kang, T. V. Mishanina, R. Landick, and S. A. Darst, “Mechanisms of tran-

- scriptional pausing in bacteria,” *Journal of molecular biology*, vol. 431, no. 20, pp. 4007–4029, 2019.
- [48] R. Guajardo and R. Sousa, “A model for the mechanism of polymerase translocation,” *Journal of molecular biology*, vol. 265, no. 1, pp. 8–19, 1997.
- [49] M. Kireeva, C. Trang, G. Matevosyan, J. Turek-Herman, V. Chasov, L. Lubkowska, and M. Kashlev, “Rna–dna and dna–dna base-pairing at the upstream edge of the transcription bubble regulate translocation of rna polymerase and transcription rate,” *Nucleic acids research*, vol. 46, no. 11, pp. 5764–5775, 2018.
- [50] M. Dangkulwanich, T. Ishibashi, S. Liu, M. L. Kireeva, L. Lubkowska, M. Kashlev, and C. J. Bustamante, “Complete dissection of transcription elongation reveals slow translocation of rna polymerase ii in a linear ratchet mechanism,” *Elife*, vol. 2, p. e00971, 2013.
- [51] I. O. Vvedenskaya, H. Vahedian-Movahed, J. G. Bird, J. G. Knoblauch, S. R. Goldman, Y. Zhang, R. H. Ebright, and B. E. Nickels, “Interactions between rna polymerase and the “core recognition element” counteract pausing,” *Science*, vol. 344, no. 6189, pp. 1285–1289, 2014.
- [52] M. Imashimizu, H. Takahashi, T. Oshima, C. McIntosh, M. Bubunencko, *et al.*, “Visualizing translocation dynamics and nascent transcript errors in paused rna polymerases in vivo,” *Genome Biol*, vol. 16, p. 98, 2015.
- [53] J. Saba, X. Y. Chua, T. V. Mishanina, D. Nayak, T. A. Windgassen, R. A. Mooney, and R. Landick, “The elemental mechanism of transcriptional pausing,” *Elife*, vol. 8, p. e40981, 2019.
- [54] J. Y. Kang, T. V. Mishanina, M. J. Bellecourt, R. A. Mooney, S. A. Darst, and R. Landick, “Rna polymerase accommodates a pause rna hairpin by global

- conformational rearrangements that prolong pausing,” *Molecular cell*, vol. 69, no. 5, pp. 802–815, 2018.
- [55] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block, “Ubiquitous transcriptional pausing is independent of rna polymerase backtracking,” *Cell*, vol. 115, no. 4, pp. 437–447, 2003.
- [56] M. Depken, E. A. Galburt, and S. W. Grill, “The origin of short transcriptional pauses,” *Biophysical journal*, vol. 96, no. 6, pp. 2189–2193, 2009.
- [57] Y. X. Mejia, H. Mao, N. R. Forde, and C. Bustamante, “Thermal probing of e. coli rna polymerase off-pathway mechanisms,” *Journal of molecular biology*, vol. 382, no. 3, pp. 628–637, 2008.
- [58] N. Opalka, M. Chlenov, P. Chacon, W. J. Rice, W. Wriggers, and S. A. Darst, “Structure and function of the transcription elongation factor greb bound to bacterial rna polymerase,” *Cell*, vol. 114, no. 3, pp. 335–345, 2003.
- [59] A. C. Cheung and P. Cramer, “Structural basis of rna polymerase ii backtrack-ing, arrest and reactivation,” *Nature*, vol. 471, no. 7337, pp. 249–253, 2011.
- [60] M. T. Marr and J. W. Roberts, “Function of transcription cleavage factors grea and greb at a regulatory pause site,” *Molecular cell*, vol. 6, no. 6, pp. 1275–1285, 2000.
- [61] S. A. Perdue and J. W. Roberts, “A backtrack-inducing sequence is an essen-tial component of escherichia coli σ 70-dependent promoter-proximal pausing,” *Molecular microbiology*, vol. 78, no. 3, pp. 636–650, 2010.
- [62] S. Nechaev, D. C. Fargo, G. dos Santos, L. Liu, Y. Gao, and K. Adelman, “Global analysis of short rnas reveals widespread promoter-proximal stalling

- and arrest of pol ii in drosophila,” *Science*, vol. 327, no. 5963, pp. 335–338, 2010.
- [63] E. A. Galburt, S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante, “Backtracking determines the force sensitivity of rnap ii in a factor-dependent manner,” *Nature*, vol. 446, no. 7137, pp. 820–823, 2007.
- [64] R. Gabizon, A. Lee, H. Vahedian-Movahed, R. H. Ebright, and C. J. Bustamante, “Pause sequences facilitate entry into long-lived paused states by reducing rna polymerase transcription rates,” *Nature communications*, vol. 9, no. 1, pp. 1–10, 2018.
- [65] S. Mukherjee, L. G. Brieba, and R. Sousa, “Discontinuous movement and conformational change during pausing and termination by t7 rna polymerase,” *The EMBO Journal*, 2003.
- [66] N. R. Forde, D. Izhaky, G. R. Woodcock, G. J. Wuite, and C. Bustamante, “Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by escherichia coli rna polymerase,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 18, pp. 11682–11687, 2002.
- [67] R. Janissen, B. Eslami-Mossallam, I. Artsimovitch, M. Depken, and N. H. Dekker, “A unifying mechanistic model of bacterial transcription with three interconnected pause states and non-diffusive backtrack recovery,” *Biophysical Journal*, vol. 118, no. 3, p. 543a, 2020.
- [68] A. Lisica, C. Engel, M. Jahnel, É. Roldán, E. A. Galburt, P. Cramer, and S. W. Grill, “Mechanisms of backtrack recovery by rna polymerases i and ii,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 11, pp. 2946–2951, 2016.

- [69] J. W. Roberts, “Mechanisms of bacterial transcription termination,” *Journal of molecular biology*, vol. 431, no. 20, pp. 4030–4039, 2019.
- [70] A. V. Yakhnin and P. Babitzke, “Mechanism of nusg-stimulated pausing, hairpin-dependent pause site selection and intrinsic termination at overlapping pause and termination sites in the bacillus subtilis trp leader,” *Molecular microbiology*, vol. 76, no. 3, pp. 690–705, 2010.
- [71] A. Ray-Soni, M. J. Bellecourt, and R. Landick, “Mechanisms of bacterial transcription termination: all good things must end,” *Annual review of biochemistry*, vol. 85, pp. 319–347, 2016.
- [72] I. Touloukhonov and R. Landick, “The flap domain is required for pause rna hairpin inhibition of catalysis by rna polymerase and can modulate intrinsic termination,” *Molecular cell*, vol. 12, no. 5, pp. 1125–1136, 2003.
- [73] T. A. Windgassen, R. A. Mooney, D. Nayak, M. Palangat, J. Zhang, and R. Landick, “Trigger-helix folding pathway and si3 mediate catalysis and hairpin-stabilized pausing by escherichia coli rna polymerase,” *Nucleic acids research*, vol. 42, no. 20, pp. 12707–12721, 2014.
- [74] A. Chauvier, J.-F. Nadon, J. P. Grondin, A.-M. Lamontagne, and D. A. Lafontaine, “Role of a hairpin-stabilized pause in the escherichia coli thic riboswitch function,” *RNA biology*, vol. 16, no. 8, pp. 1066–1073, 2019.
- [75] I. Artsimovitch and R. Landick, “Interaction of a nascent rna structure with rna polymerase is required for hairpin-dependent transcriptional pausing but not for transcript release,” *Genes & development*, vol. 12, no. 19, pp. 3110–3122, 1998.
- [76] P. H. von Hippel, “An integrated model of the transcription complex in elongation, termination, and editing,” *Science*, vol. 281, no. 5377, pp. 660–665, 1998.

- [77] F. Jülicher and R. Bruinsma, “Motion of rna polymerase along dna: a stochastic model,” *Biophysical journal*, vol. 74, no. 3, pp. 1169–1185, 1998.
- [78] H.-Y. Wang, T. Elston, A. Mogilner, and G. Oster, “Force generation in rna polymerase,” *Biophysical journal*, vol. 74, no. 3, pp. 1186–1202, 1998.
- [79] B. E. Nickels and A. Hochschild, “Regulation of rna polymerase through the secondary channel,” *Cell*, vol. 118, no. 3, pp. 281–284, 2004.
- [80] M. H. Larson, W. J. Greenleaf, R. Landick, and S. M. Block, “Applied force reveals mechanistic and energetic details of transcription termination,” *Cell*, vol. 132, no. 6, pp. 971–982, 2008.
- [81] I. Tinoco Jr and C. Bustamante, “The effect of force on thermodynamics and kinetics of single molecule reactions,” *Biophysical chemistry*, vol. 101, pp. 513–533, 2002.
- [82] A. Bochkareva, Y. Yuzenkova, V. R. Tadigotla, and N. Zenkin, “Factor-independent transcription pausing caused by recognition of the rna–dna hybrid sequence,” *The EMBO journal*, vol. 31, no. 3, pp. 630–639, 2012.
- [83] W. Xu, Y. Yan, I. Artsimovitch, D. Dunlap, and L. Finzi, “Positive supercoiling favors transcription elongation through lac repressor-mediated dna loops,” *Nucleic Acids Research*, vol. 50, no. 5, pp. 2826–2835, 2022.
- [84] Y. Yan, W. Xu, S. Kumar, A. Zhang, F. Leng, D. Dunlap, and L. Finzi, “Negative dna supercoiling makes protein-mediated looping deterministic and ergodic within the bacterial doubling time,” *Nucleic Acids Research*, vol. 49, no. 20, pp. 11550–11559, 2021.
- [85] K. S. Matthews and J. C. Nichols, “Lactose repressor protein: functional prop-

- erties and structure,” *Progress in nucleic acid research and molecular biology*, vol. 58, pp. 127–164, 1997.
- [86] R. O’Brien, R. Streeper, J. Ayala, B. Stadelmaier, and L. Hornbuckle, “Insulin-regulated gene expression,” *Biochemical Society Transactions*, vol. 29, no. 4, pp. 552–558, 2001.
- [87] L. N. Sutherland, M. R. Bomhof, L. C. Capozzi, S. A. Basaraba, and D. C. Wright, “Exercise and adrenaline increase pgc-1 α mrna expression in rat adipose tissue,” *The Journal of physiology*, vol. 587, no. 7, pp. 1607–1617, 2009.
- [88] M. Land, L. Hauser, S.-R. Jun, I. Nookaew, M. R. Leuze, T.-H. Ahn, T. Karpinets, O. Lund, G. Kora, T. Wassenaar, *et al.*, “Insights from 20 years of bacterial genome sequencing,” *Functional & integrative genomics*, vol. 15, pp. 141–161, 2015.
- [89] G. C. Dicenzo and T. M. Finan, “The divided bacterial genome: structure, function, and evolution,” *Microbiology and Molecular Biology Reviews*, vol. 81, no. 3, pp. 10–1128, 2017.
- [90] M. Joyeux, “Compaction of bacterial genomic dna: clarifying the concepts,” *Journal of Physics: Condensed Matter*, vol. 27, no. 38, p. 383001, 2015.
- [91] P. A. Levin and E. R. Angert, “Small but mighty: cell size and bacteria,” *Cold Spring Harbor perspectives in biology*, vol. 7, no. 7, p. a019216, 2015.
- [92] S. Bakshi, H. Choi, and J. C. Weisshaar, “The spatial biology of transcription and translation in rapidly growing escherichia coli,” *Frontiers in microbiology*, vol. 6, p. 145705, 2015.
- [93] B. A. Shen and R. Landick, “Transcription of bacterial chromatin,” *Journal of molecular biology*, vol. 431, no. 20, pp. 4040–4066, 2019.

- [94] C. J. Dorman and M. J. Dorman, “Dna supercoiling is a fundamental regulatory principle in the control of bacterial gene expression,” *Biophysical reviews*, vol. 8, pp. 209–220, 2016.
- [95] S. Meyer, S. Reverchon, W. Nasser, and G. Muskhelishvili, “Chromosomal organization of transcription: in a nutshell,” *Current genetics*, vol. 64, pp. 555–565, 2018.
- [96] N. P. Higgins, “Rna polymerase: chromosome domain boundary maker and regulator of supercoil density,” *Current opinion in microbiology*, vol. 22, pp. 138–143, 2014.
- [97] R. Landick, J. T. Wade, and D. C. Grainger, “H-ns and rna polymerase: a love–hate relationship?,” *Current opinion in microbiology*, vol. 24, pp. 53–59, 2015.
- [98] D. F. Browning, D. C. Grainger, and S. J. Busby, “Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression,” *Current opinion in microbiology*, vol. 13, no. 6, pp. 773–780, 2010.
- [99] C. Kahramanoglou, A. S. Seshasayee, A. I. Prieto, D. Ibberson, S. Schmidt, J. Zimmermann, V. Benes, G. M. Fraser, and N. M. Luscombe, “Direct and indirect effects of h-ns and fis on global gene expression control in escherichia coli,” *Nucleic acids research*, vol. 39, no. 6, pp. 2073–2091, 2011.
- [100] W. W. Navarre, S. Porwollik, Y. Wang, M. McClelland, H. Rosen, S. J. Libby, and F. C. Fang, “Selective silencing of foreign dna with low gc content by the h-ns protein in salmonella,” *Science*, vol. 313, no. 5784, pp. 236–238, 2006.
- [101] S. S. Singh, N. Singh, R. P. Bonocora, D. M. Fitzgerald, J. T. Wade, and D. C. Grainger, “Widespread suppression of intragenic transcription initiation by h-ns,” *Genes & development*, vol. 28, no. 3, pp. 214–219, 2014.

- [102] S. Lucchini, G. Rowley, M. D. Goldberg, D. Hurd, M. Harrison, and J. C. D. Hinton, “H-ns mediates the silencing of laterally acquired genes in bacteria,” *PLoS pathogens*, vol. 2, no. 8, p. e81, 2006.
- [103] R. Schneider, A. Travers, T. Kutateladze, and G. Muskhelishvili, “A dna architectural protein couples cellular physiology and dna topology in escherichia coli,” *Molecular microbiology*, vol. 34, no. 5, pp. 953–964, 1999.
- [104] H. Auner, M. Buckle, A. Deufel, T. Kutateladze, L. Lazarus, R. Mavathur, G. Muskhelishvili, I. Pemberton, R. Schneider, and A. Travers, “Mechanism of transcriptional activation by fis: role of core promoter structure and dna topology,” *Journal of molecular biology*, vol. 331, no. 2, pp. 331–344, 2003.
- [105] A. Lal, A. Dhar, A. Trostel, F. Kouzine, A. S. Seshasayee, and S. Adhya, “Genome scale patterns of supercoiling in a bacterial chromosome,” *Nature communications*, vol. 7, no. 1, p. 11055, 2016.
- [106] S. Kar, R. Edgar, and S. Adhya, “Nucleoid remodeling by an altered hu protein: reorganization of the transcription program,” *Proceedings of the National Academy of Sciences*, vol. 102, no. 45, pp. 16397–16402, 2005.
- [107] P. Koli, S. Sudan, D. Fitzgerald, S. Adhya, and S. Kar, “Conversion of commensal escherichia coli k-12 to an invasive form via expression of a mutant histone-like protein,” *MBio*, vol. 2, no. 5, pp. 10–1128, 2011.
- [108] M. V. Kotlajich, D. R. Hron, B. A. Boudreau, Z. Sun, Y. L. Lyubchenko, and R. Landick, “Bridged filaments of histone-like nucleoid structuring protein pause rna polymerase and aid termination in bacteria,” *Elife*, vol. 4, p. e04970, 2015.
- [109] B. A. Boudreau, D. R. Hron, L. Qin, R. A. Van Der Valk, M. V. Kotlajich, R. T. Dame, and R. Landick, “Stpa and hha stimulate pausing by rna polymerase by

- promoting dna–dna bridging of h–ns filaments,” *Nucleic acids research*, vol. 46, no. 11, pp. 5525–5546, 2018.
- [110] L. S. Qi, M. H. Larson, L. A. Gilbert, J. A. Doudna, J. S. Weissman, A. P. Arkin, and W. A. Lim, “Repurposing crispr as an rna-guided platform for sequence-specific control of gene expression,” *Cell*, vol. 152, no. 5, pp. 1173–1183, 2013.
- [111] N. Hao, S. Krishna, A. Ahlgren-Berg, E. E. Cutts, K. E. Shearwin, and I. B. Dodd, “Road rules for traffic on dna—systematic analysis of transcriptional roadblocking in vivo,” *Nucleic acids research*, vol. 42, no. 14, pp. 8861–8872, 2014.
- [112] Z. Vörös, Y. Yan, D. T. Kovari, L. Finzi, and D. Dunlap, “Proteins mediating dna loops effectively block transcription,” *Protein Science*, vol. 26, no. 7, pp. 1427–1438, 2017.
- [113] D. E. Lewis, N. Komissarova, P. Le, M. Kashlev, and S. Adhya, “Dna sequences in gal operon override transcription elongation blocks,” *Journal of molecular biology*, vol. 382, no. 4, pp. 843–858, 2008.
- [114] B. R. Belitsky and A. L. Sonenshein, “Roadblock repression of transcription by bacillus subtilis cody,” *Journal of molecular biology*, vol. 411, no. 4, pp. 729–743, 2011.
- [115] S.-K. Choi and M. H. Saier Jr, “Regulation of sigl expression by the catabolite control protein ccpa involves a roadblock mechanism in bacillus subtilis: potential connection between carbon and nitrogen metabolism,” *Journal of bacteriology*, vol. 187, no. 19, pp. 6856–6861, 2005.
- [116] J. M. Zalieckas, L. V. Wray, Jr, A. E. Ferson, and S. H. Fisher, “Transcription–repair coupling factor is involved in carbon catabolite repression of the bacillus

- subtilis hut and gnt operons,” *Molecular microbiology*, vol. 27, no. 5, pp. 1031–1038, 1998.
- [117] S. S. Teves, C. M. Weber, and S. Henikoff, “Transcribing through the nucleosome,” *Trends in biochemical sciences*, vol. 39, no. 12, pp. 577–586, 2014.
- [118] S. Ucuncuoglu, K. L. Engel, P. K. Purohit, D. D. Dunlap, D. A. Schneider, and L. Finzi, “Direct characterization of transcription elongation by rna polymerase i,” *PLoS One*, vol. 11, no. 7, p. e0159527, 2016.
- [119] D. A. Schafer, J. Gelles, M. P. Sheetz, and R. Landick, “Transcription by single molecules of rna polymerase observed by light microscopy,” *Nature*, vol. 352, no. 6334, pp. 444–448, 1991.
- [120] C. Zurla, A. Franzini, G. Galli, D. Dunlap, D. Lewis, S. Adhya, and L. Finzi, “Novel tethered particle motion analysis of ci protein-mediated dna looping in the regulation of bacteriophage lambda,” *Journal of Physics: Condensed Matter*, vol. 18, no. 14, p. S225, 2006.
- [121] J. Milstein, Y. Chen, and J.-C. Meiners, “Bead size effects on protein-mediated dna looping in tethered-particle motion experiments,” *Biopolymers*, vol. 95, no. 2, pp. 144–150, 2011.
- [122] L. Finzi and J. Gelles, “Measurement of lactose repressor-mediated loop formation and breakdown in single dna molecules,” *Science*, vol. 267, no. 5196, pp. 378–380, 1995.
- [123] Y.-F. Chen, J. Milstein, and J.-C. Meiners, “Protein-mediated dna loop formation and breakdown in a fluctuating environment,” *Physical Review Letters*, vol. 104, no. 25, p. 258103, 2010.

- [124] S. Johnson, M. Lindén, and R. Phillips, “Sequence dependence of transcription factor-mediated dna looping,” *Nucleic acids research*, vol. 40, no. 16, pp. 7728–7738, 2012.
- [125] P. Kaur, D. Wu, J. Lin, P. Countryman, K. C. Bradford, D. A. Erie, R. Riehn, P. L. Opresko, and H. Wang, “Enhanced electrostatic force microscopy reveals higher-order dna looping mediated by the telomeric protein trf2,” *Scientific reports*, vol. 6, no. 1, p. 20513, 2016.
- [126] Y. L. Lyubchenko, L. S. Shlyakhtenko, T. Aki, and S. Adhya, “Atomic force microscopic demonstration of dna looping by galr and hu,” *Nucleic acids research*, vol. 25, no. 4, pp. 873–876, 1997.
- [127] D. Normanno, F. Vanzi, and F. S. Pavone, “Single-molecule manipulation reveals supercoiling-dependent modulation of lac repressor-mediated dna looping,” *Nucleic acids research*, vol. 36, no. 8, pp. 2505–2513, 2008.
- [128] Y. Yan, Y. Ding, F. Leng, D. Dunlap, and L. Finzi, “Protein-mediated loops in supercoiled dna create large topological domains,” *Nucleic Acids Research*, vol. 46, no. 9, pp. 4417–4424, 2018.
- [129] G. Lia, D. Bensimon, V. Croquette, J.-F. Allemand, D. Dunlap, D. E. Lewis, S. Adhya, and L. Finzi, “Supercoiling and denaturation in gal repressor/heat unstable nucleoid protein (hu)-mediated dna looping,” *Proceedings of the National Academy of Sciences*, vol. 100, no. 20, pp. 11373–11377, 2003.
- [130] G. Lia, E. Praly, H. Ferreira, C. Stockdale, Y. C. Tse-Dinh, D. Dunlap, V. Croquette, D. Bensimon, and T. Owen-Hughes, “Direct observation of dna distortion by the rsc complex,” *Molecular cell*, vol. 21, no. 3, pp. 417–425, 2006.
- [131] G. Lia, S. Semsey, D. E. Lewis, S. Adhya, D. Bensimon, D. Dunlap, and L. Finzi,

- “The antiparallel loops in gal dna,” *Nucleic acids research*, vol. 36, no. 12, pp. 4204–4210, 2008.
- [132] K. Sakata-Sogawa, M. Kurachi, K. Sogawa, Y. Fujii-Kuriyama, and H. Tashiro, “Direct measurement of dna molecular length in solution using optical tweezers: detection of looping due to binding protein interactions,” *European biophysics journal*, vol. 27, pp. 55–61, 1998.
- [133] J. Ma, L. Bai, and M. D. Wang, “Transcription under torsion,” *Science*, vol. 340, no. 6140, pp. 1580–1583, 2013.
- [134] J. Ma and M. D. Wang, “Dna supercoiling during transcription,” *Biophysical reviews*, vol. 8, no. Suppl 1, pp. 75–87, 2016.
- [135] M. Roushan, P. Kaur, A. Karpusenko, P. J. Countryman, C. P. Ortiz, S. Fang Lim, H. Wang, and R. Riehn, “Probing transient protein-mediated dna linkages using nanoconfinement,” *Biomicrofluidics*, vol. 8, no. 3, 2014.
- [136] Z. Hensel, X. Weng, A. C. Lagda, and J. Xiao, “Transcription-factor-mediated dna looping probed by high-resolution, single-molecule imaging in live e. coli cells,” *PLoS biology*, vol. 11, no. 6, p. e1001591, 2013.
- [137] S. Mohapatra, C.-T. Lin, X. A. Feng, A. Basu, and T. Ha, “Single-molecule analysis and engineering of dna motors,” *Chemical reviews*, vol. 120, no. 1, pp. 36–78, 2019.
- [138] C. J. Dorman, “Co-operative roles for dna supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription,” *Biochemical Society Transactions*, vol. 41, no. 2, pp. 542–547, 2013.
- [139] G. B. Koudelka, “Recognition of dna structure by 434 repressor,” *Nucleic acids research*, vol. 26, no. 2, pp. 669–675, 1998.

- [140] J. R. Widom, V. Rai, C. E. Rohlman, and N. G. Walter, “Versatile transcription control based on reversible dcas9 binding,” *Rna*, vol. 25, no. 11, pp. 1457–1469, 2019.
- [141] E. J. Strobel, K. E. Watters, Y. Nediaikov, I. Artsimovitch, and J. B. Lucks, “Distributed biotin–streptavidin transcription roadblocks for mapping cotranscriptional rna folding,” *Nucleic acids research*, vol. 45, no. 12, pp. e109–e109, 2017.
- [142] V. Epshtein, F. Toulmé, A. R. Rahmouni, S. Borukhov, and E. Nudler, “Transcription through the roadblocks: the role of rna polymerase cooperation,” *The EMBO journal*, 2003.
- [143] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, “Nucleosomal fluctuations govern the transcription dynamics of rna polymerase ii,” *Science*, vol. 325, no. 5940, pp. 626–628, 2009.
- [144] C. Mosrin-Huaman, C. L. Turnbough Jr, and A. R. Rahmouni, “Translocation of escherichia coli rna polymerase against a protein roadblock in vivo highlights a passive sliding mechanism for transcript elongation,” *Molecular microbiology*, vol. 51, no. 5, pp. 1471–1481, 2004.
- [145] J. Jin, L. Bai, D. S. Johnson, R. M. Fulbright, M. L. Kireeva, M. Kashlev, and M. D. Wang, “Synergistic action of rna polymerases in overcoming the nucleosomal barrier,” *Biophysical Journal*, vol. 98, no. 3, p. 67a, 2010.
- [146] J. K. Fisher, J. Cribb, K. V. Desai, L. Vicci, B. Wilde, K. Keller, R. M. Taylor, J. Haase, K. Bloom, E. T. O’Brien, *et al.*, “Thin-foil magnetic force system for high-numerical-aperture microscopy,” *Review of Scientific Instruments*, vol. 77, no. 2, 2006.

- [147] C. Gosse and V. Croquette, “Magnetic tweezers: micromanipulation and force measurement at the molecular level,” *Biophysical journal*, vol. 82, no. 6, pp. 3314–3329, 2002.
- [148] C. Haber and D. Wirtz, “Magnetic tweezers for dna micromanipulation,” *Review of Scientific instruments*, vol. 71, no. 12, pp. 4561–4570, 2000.
- [149] J. Yan, D. Skoko, and J. F. Marko, “Near-field-magnetic-tweezer manipulation of single dna molecules,” *Physical Review E*, vol. 70, no. 1, p. 011905, 2004.
- [150] P. Kollmannsberger and B. Fabry, “Bahigh-force magnetic tweezers with force feedback for biological applications,” *Review of Scientific Instruments*, vol. 78, no. 11, 2007.
- [151] J. Lipfert, X. Hao, and N. H. Dekker, “Quantitative modeling and optimization of magnetic tweezers,” *Biophysical journal*, vol. 96, no. 12, pp. 5040–5049, 2009.
- [152] K. C. Neuman and A. Nagy, “Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy,” *Nature methods*, vol. 5, no. 6, pp. 491–505, 2008.
- [153] I. De Vlaminck and C. Dekker, “Recent advances in magnetic tweezers,” *Annual review of biophysics*, vol. 41, pp. 453–472, 2012.
- [154] J. R. Sadler, H. Sasmor, and J. L. Betz, “A perfectly symmetric lac operator binds the lac repressor very tightly.,” *Proceedings of the National Academy of Sciences*, vol. 80, no. 22, pp. 6785–6789, 1983.
- [155] J. Romanuka, G. E. Folkers, N. Biris, E. Tishchenko, H. Wienk, A. M. Bonvin, R. Kaptein, and R. Boelens, “Specificity and affinity of lac repressor for the auxiliary operators o2 and o3 are explained by the structures of their protein–

- dna complexes,” *Journal of molecular biology*, vol. 390, no. 3, pp. 478–489, 2009.
- [156] B. Terry, W. Jack, R. Rubin, and P. Modrich, “Thermodynamic parameters governing interaction of *ecori* endonuclease with specific and nonspecific dna sequences,” *Journal of Biological Chemistry*, vol. 258, no. 16, pp. 9820–9825, 1983.
- [157] S. Proshkin, A. R. Rahmouni, A. Mironov, and E. Nudler, “Cooperation between translating ribosomes and rna polymerase in transcription elongation,” *Science*, vol. 328, no. 5977, pp. 504–508, 2010.
- [158] F. Toulmé, C. Mosrin-Huaman, J. Sparkowski, A. Das, M. Leng, and A. R. Rahmouni, “Grea and greb proteins revive backtracked rna polymerase in vivo by promoting transcript trimming,” *The EMBO journal*, 2000.
- [159] B. Zamft, L. Bintu, T. Ishibashi, and C. Bustamante, “Nascent rna structure modulates the transcriptional dynamics of rna polymerases,” *Proceedings of the National Academy of Sciences*, vol. 109, no. 23, pp. 8948–8953, 2012.
- [160] P. A. Pavco and D. Steege, “Elongation by *escherichia coli* rna polymerase is blocked in vitro by a site-specific dna binding protein,” *Journal of Biological Chemistry*, vol. 265, no. 17, pp. 9960–9969, 1990.
- [161] E. Cavac, L. E. Ramírez-Tapia, and C. T. Martin, “High-salt transcription of dna cotethered with t7 rna polymerase to beads generates increased yields of highly pure rna,” *Journal of Biological Chemistry*, vol. 297, no. 3, 2021.
- [162] V. K. Misra, J. L. Hecht, K. A. Sharp, R. A. Friedman, and B. Honig, “Salt effects on protein-dna interactions: the λ ci repressor and *ecori* endonuclease,” *Journal of molecular Biology*, vol. 238, no. 2, pp. 264–280, 1994.

- [163] R. Janissen, B. Eslami-Mossallam, I. Artsimovitch, M. Depken, and N. H. Dekker, “High-throughput single-molecule experiments reveal heterogeneity, state switching, and three interconnected pause states in transcription,” *Cell Reports*, vol. 39, no. 4, 2022.
- [164] L. L. d. Maddalena, H. Niederholtmeyer, M. Turtola, Z. N. Swank, G. A. Belogurov, and S. J. Maerkl, “Grea and greb enhance expression of escherichia coli rna polymerase promoters in a reconstituted transcription–translation system,” *ACS synthetic biology*, vol. 5, no. 9, pp. 929–935, 2016.
- [165] K. Amar, F. Wei, J. Chen, and N. Wang, “Effects of forces on chromatin,” *APL bioengineering*, vol. 5, no. 4, 2021.
- [166] P. M. Rullens and J. Kind, “Attach and stretch: Emerging roles for genome–lamina contacts in shaping the 3d genome,” *Current Opinion in Cell Biology*, vol. 70, pp. 51–57, 2021.
- [167] J. Wiedermannová and L. Krásný, “ β -casp proteins removing rna polymerase from dna: when a torpedo is needed to shoot a sitting duck,” *Nucleic Acids Research*, vol. 49, no. 18, pp. 10221–10234, 2021.
- [168] I. J. Finkelstein, M.-L. Visnapuu, and E. C. Greene, “Single-molecule imaging reveals mechanisms of protein disruption by a dna translocase,” *Nature*, vol. 468, no. 7326, pp. 983–987, 2010.
- [169] L. M. Hsu, “Promoter clearance and escape in prokaryotes,” *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, vol. 1577, no. 2, pp. 191–207, 2002.
- [170] L. M. Hsu, N. V. Vo, and M. J. Chamberlin, “Escherichia coli transcript cleavage factors grea and greb stimulate promoter escape and gene expression in vivo

- and in vitro.," *Proceedings of the National Academy of Sciences*, vol. 92, no. 25, pp. 11588–11592, 1995.
- [171] G. A. Belogurov, A. Sevostyanova, V. Svetlov, and I. Artsimovitch, "Functional regions of the n-terminal domain of the antiterminator rfah," *Molecular microbiology*, vol. 76, no. 2, pp. 286–301, 2010.
- [172] S. Forth, M. Y. Sheinin, J. Inman, and M. D. Wang, "Torque measurement at the single-molecule level," *Annual review of biophysics*, vol. 42, pp. 583–604, 2013.
- [173] O. G. Berg, R. B. Winter, and P. H. Von Hippel, "Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. models and theory," *Biochemistry*, vol. 20, no. 24, pp. 6929–6948, 1981.
- [174] F. Wang, S. Redding, I. J. Finkelstein, J. Gorman, D. R. Reichman, and E. C. Greene, "The promoter-search mechanism of escherichia coli rna polymerase is dominated by three-dimensional diffusion," *Nature structural & molecular biology*, vol. 20, no. 2, pp. 174–181, 2013.
- [175] K. M. Arndt and M. J. Chamberlin, "Transcription termination in escherichia coli: measurement of the rate of enzyme release from rho-independent terminators," *Journal of molecular biology*, vol. 202, no. 2, pp. 271–285, 1988.
- [176] M. J. Bellecourt, A. Ray-Soni, A. Harwig, R. A. Mooney, and R. Landick, "Rna polymerase clamp movement aids dissociation from dna but is not required for rna release at intrinsic terminators," *Journal of molecular biology*, vol. 431, no. 4, pp. 696–713, 2019.
- [177] M. S. Paget and J. D. Helmann, "The σ 70 family of sigma factors," *Genome biology*, vol. 4, pp. 1–6, 2003.

- [178] T. T. Harden, C. D. Wells, L. J. Friedman, R. Landick, A. Hochschild, J. Kondev, and J. Gelles, “Bacterial rna polymerase can retain $\sigma 70$ throughout transcription,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 602–607, 2016.
- [179] K. Inlow, D. Tenenbaum, L. J. Friedman, J. Kondev, and J. Gelles, “Recycling of bacterial rna polymerase by the swi2/snf2 atpase rapa,” *Proceedings of the National Academy of Sciences*, vol. 120, no. 28, p. e2303849120, 2023.
- [180] K. M. Herbert, J. Zhou, R. A. Mooney, A. La Porta, R. Landick, and S. M. Block, “E. coli nusg inhibits backtracking and accelerates pause-free transcription by promoting forward translocation of rna polymerase,” *Journal of molecular biology*, vol. 399, no. 1, pp. 17–30, 2010.
- [181] K. Adelman, A. La Porta, T. J. Santangelo, J. T. Lis, J. W. Roberts, and M. D. Wang, “Single molecule analysis of rna polymerase elongation reveals uniform kinetic behavior,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 21, pp. 13538–13543, 2002.
- [182] L. Chen and A. Conlisk, “Forces affecting double-stranded dna translocation through synthetic nanopores,” *Biomedical microdevices*, vol. 13, pp. 403–414, 2011.
- [183] M. NandyMazumdar, Y. Nedialkov, D. Svetlov, A. Sevostyanova, G. A. Belogurov, and I. Artsimovitch, “Rna polymerase gate loop guides the nontemplate dna strand in transcription complexes,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 52, pp. 14994–14999, 2016.
- [184] Y. Shin, M. Z. Qayyum, D. Pupov, D. Esyunina, A. Kulbachinskiy, and K. S. Murakami, “Structural basis of ribosomal rna transcription regulation,” *Nature communications*, vol. 12, no. 1, p. 528, 2021.

- [185] L. You, E. O. Omollo, C. Yu, R. A. Mooney, J. Shi, L. Shen, X. Wu, A. Wen, D. He, Y. Zeng, *et al.*, “Structural basis for intrinsic transcription termination,” *Nature*, vol. 613, no. 7945, pp. 783–789, 2023.
- [186] S. Murayama, S. Ishikawa, O. Chumsakul, N. Ogasawara, and T. Oshima, “The role of α -ctd in the genome-wide transcriptional regulation of the bacillus subtilis cells,” *PloS one*, vol. 10, no. 7, p. e0131588, 2015.
- [187] R. L. Gourse, W. Ross, and T. Gaal, “Ups and downs in bacterial transcription initiation: the role of the alpha subunit of rna polymerase in promoter recognition,” *Molecular microbiology*, vol. 37, no. 4, pp. 687–695, 2000.
- [188] H. A. Shaban, R. Barth, and K. Bystricky, “Navigating the crowd: visualizing coordination between genome dynamics, structure, and transcription,” *Genome biology*, vol. 21, no. 1, p. 278, 2020.
- [189] B. S. McCauley, L. Sun, R. Yu, M. Lee, H. Liu, D. S. Leeman, Y. Huang, A. E. Webb, and W. Dang, “Altered chromatin states drive cryptic transcription in aging mammalian stem cells,” *Nature aging*, vol. 1, no. 8, pp. 684–697, 2021.
- [190] N. Hao, A. C. Palmer, A. Ahlgren-Berg, K. E. Shearwin, and I. B. Dodd, “The role of repressor kinetics in relief of transcriptional interference between convergent promoters,” *Nucleic Acids Research*, vol. 44, no. 14, pp. 6625–6638, 2016.
- [191] J. A. Brophy and C. A. Voigt, “Antisense transcription as a tool to tune gene expression,” *Molecular systems biology*, vol. 12, no. 1, p. 854, 2016.
- [192] A. C. Wu, H. Patel, M. Chia, F. Moretto, D. Frith, A. P. Snijders, and F. J. van Werven, “Repression of divergent noncoding transcription by a sequence-specific transcription factor,” *Molecular cell*, vol. 72, no. 6, pp. 942–954, 2018.

- [193] R. John, n. Davenport, G. J. Wuite, R. Landick, and C. Bustamante, “Single-molecule study of transcriptional pausing and arrest by e. coli rna polymerase,” *Science*, vol. 287, no. 5462, pp. 2497–2500, 2000.
- [194] L. Wang, J. W. Watters, X. Ju, G. Lu, and S. Liu, “Head-on and co-directional rna polymerase collisions orchestrate bidirectional transcription termination,” *Molecular Cell*, vol. 83, no. 7, pp. 1153–1164, 2023.
- [195] S. E. Rice, T. J. Purcell, and J. A. Spudich, “[6] building and using optical traps to study properties of molecular motors,” in *Methods in enzymology*, vol. 361, pp. 112–133, Elsevier, 2003.
- [196] B. Albert, I. Léger-Silvestre, C. Normand, M. K. Ostermaier, J. Pérez-Fernández, K. I. Panov, J. C. Zomerdijk, P. Schultz, and O. Gadal, “Rna polymerase i-specific subunits promote polymerase clustering to enhance the rrna gene transcription cycle,” *Journal of cell biology*, vol. 192, no. 2, pp. 277–293, 2011.
- [197] S. Scott, C. Shaheen, B. McGuinness, K. Metera, F. Kouzine, D. Levens, C. J. Benham, and S. Leslie, “Single-molecule visualization of the effects of ionic strength and crowding on structure-mediated interactions in supercoiled dna molecules,” *Nucleic acids research*, vol. 47, no. 12, pp. 6360–6368, 2019.
- [198] R. J. Ellis and A. P. Minton, “Join the crowd,” *Nature*, vol. 425, no. 6953, pp. 27–28, 2003.
- [199] D. Miyoshi and N. Sugimoto, “Molecular crowding effects on structure and stability of dna,” *Biochimie*, vol. 90, no. 7, pp. 1040–1051, 2008.
- [200] S. B. Zimmerman and S. O. Trach, “Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of escherichia coli,” *Journal of molecular biology*, vol. 222, no. 3, pp. 599–620, 1991.

- [201] B. Bohrmann, M. Haider, and E. Kellenberger, “Concentration evaluation of chromatin in unstained resin-embedded sections by means of low-dose ratio-contrast imaging in stem,” *Ultramicroscopy*, vol. 49, no. 1-4, pp. 235–251, 1993.
- [202] Y. Shin and C. P. Brangwynne, “Liquid phase condensation in cell physiology and disease,” *Science*, vol. 357, no. 6357, p. eaaf4382, 2017.
- [203] K. Kohata and D. Miyoshi, “Rna phase separation–mediated direction of molecular trafficking under conditions of molecular crowding,” *Biophysical Reviews*, vol. 12, no. 3, pp. 669–676, 2020.
- [204] S.-i. Nakano, D. Miyoshi, and N. Sugimoto, “Effects of molecular crowding on the structures, interactions, and functions of nucleic acids,” *Chemical reviews*, vol. 114, no. 5, pp. 2733–2758, 2014.
- [205] W. M. Mardoum, S. M. Gorczyca, K. E. Regan, T.-C. Wu, and R. M. Robertson-Anderson, “Crowding induces entropically-driven changes to dna dynamics that depend on crowder structure and ionic conditions,” *Frontiers in physics*, vol. 6, p. 53, 2018.
- [206] N. R. Council, D. on Engineering, P. Sciences, S. S. Board, C. on Physical Sciences, Mathematics, Applications, and S. G. for the Workshop on Size Limits of Very Small Microorganisms, “Size limits of very small microorganisms: Proceedings of a workshop,” 1999.
- [207] L. E. Baltierra-Jasso, M. J. Morten, L. Laflor, S. D. Quinn, and S. W. Magennis, “Crowding-induced hybridization of single dna hairpins,” *Journal of the American Chemical Society*, vol. 137, no. 51, pp. 16020–16023, 2015.
- [208] L. Finzi and W. K. Olson, “The emerging role of dna supercoiling as a dynamic player in genomic structure and function,” *Biophysical reviews*, vol. 8, no. Suppl 1, pp. 1–3, 2016.

- [209] M. Cristofalo, C. Marrano, D. Salerno, R. Corti, V. Cassina, A. Mammola, M. Gherardi, B. Sclavi, M. C. Lagomarsino, and F. Mantegazza, “Cooperative effects on the compaction of dna fragments by the nucleoid protein h-ns and the crowding agent peg probed by magnetic tweezers,” *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1864, no. 12, p. 129725, 2020.
- [210] C. J. Dorman, “H-ns: a universal regulator for a dynamic genome,” *Nature Reviews Microbiology*, vol. 2, no. 5, pp. 391–400, 2004.
- [211] S. C. Dillon and C. J. Dorman, “Bacterial nucleoid-associated proteins, nucleoid structure and gene expression,” *Nature Reviews Microbiology*, vol. 8, no. 3, pp. 185–195, 2010.
- [212] J. Widom and R. L. Baldwin, “Monomolecular condensation of λ -dna induced by cobalt hexammine,” *Biopolymers: Original Research on Biomolecules*, vol. 22, no. 6, pp. 1595–1620, 1983.
- [213] A. D. Miller, “Cationic liposomes for gene therapy,” *Angewandte Chemie International Edition*, vol. 37, no. 13-14, pp. 1768–1785, 1998.
- [214] G. Krishnamoorthy, G. Duportail, and Y. Mély, “Structure and dynamics of condensed dna probed by 1, 1’-(4, 4, 8, 8-tetramethyl-4, 8-diazaundecamethylene) bis [4-[[3-methylbenz-1, 3-oxazol-2-yl] methyldine]-1, 4-dihydroquinolinium] tetraiodide fluorescence,” *Biochemistry*, vol. 41, no. 51, pp. 15277–15287, 2002.
- [215] G. Krishnamoorthy, B. Roques, J.-L. Darlix, and Y. Mély, “Dna condensation by the nucleocapsid protein of hiv-1: a mechanism ensuring dna protection,” *Nucleic Acids Research*, vol. 31, no. 18, pp. 5425–5432, 2003.
- [216] D. Collette, D. Dunlap, and L. Finzi, “Macromolecular crowding and dna:

- Bridging the gap between in vitro and in vivo,” *International Journal of Molecular Sciences*, vol. 24, no. 24, p. 17502, 2023.
- [217] H. Walter and D. E. Brooks, “Phase separation in cytoplasm, due to macromolecular crowding, is the basis for microcompartmentation,” *FEBS letters*, vol. 361, no. 2-3, pp. 135–139, 1995.
- [218] B. R. Levone, S. C. Lenzken, M. Antonaci, A. Maiser, A. Rapp, F. Conte, S. Reber, J. Mechttersheimer, A. E. Ronchi, O. Mühlemann, *et al.*, “Fus-dependent liquid–liquid phase separation is important for dna repair initiation,” *Journal of Cell Biology*, vol. 220, no. 5, p. e202008030, 2021.
- [219] A. Shakya, S. Park, N. Rana, and J. T. King, “Liquid-liquid phase separation of histone proteins in cells: role in chromatin organization,” *Biophysical journal*, vol. 118, no. 3, pp. 753–764, 2020.
- [220] S. Park, R. Barnes, Y. Lin, B.-j. Jeon, S. Najafi, K. T. Delaney, G. H. Fredrickson, J.-E. Shea, D. S. Hwang, and S. Han, “Dehydration entropy drives liquid-liquid phase separation by molecular crowding,” *Communications Chemistry*, vol. 3, no. 1, p. 83, 2020.
- [221] S. B. Zimmerman and A. P. Minton, “Macromolecular crowding: biochemical, biophysical, and physiological consequences,” *Annual review of biophysics and biomolecular structure*, vol. 22, no. 1, pp. 27–65, 1993.
- [222] M. Tabaka, T. Kalwarczyk, J. Szymanski, S. Hou, and R. Holyst, “The effect of macromolecular crowding on mobility of biomolecules, association kinetics, and gene expression in living cells,” *Frontiers in Physics*, vol. 2, p. 54, 2014.
- [223] A. P. Minton, “The effect of volume occupancy upon the thermodynamic activity of proteins: some biochemical consequences,” *Molecular and cellular biochemistry*, vol. 55, pp. 119–140, 1983.

- [224] R. Goobes, N. Kahana, O. Cohen, and A. Minsky, “Metabolic buffering exerted by macromolecular crowding on dna- dna interactions: Origin and physiological significance,” *Biochemistry*, vol. 42, no. 8, pp. 2431–2440, 2003.
- [225] P. Woolley and P. R. Wills, “Excluded-volume effect of inert macromolecules on the melting of nucleic acids,” *Biophysical chemistry*, vol. 22, no. 1-2, pp. 89–94, 1985.
- [226] C. Spink and J. Chaires, “Selective stabilization of triplex dna by poly (ethylene glycols),” *Journal of the American Chemical Society*, vol. 117, no. 51, pp. 12887–12888, 1995.
- [227] C. H. Spink and J. B. Chaires, “Effects of hydration, ion release, and excluded volume on the melting of triplex and duplex dna,” *Biochemistry*, vol. 38, no. 1, pp. 496–508, 1999.
- [228] S. Alberti, “The wisdom of crowds: regulating cell function through condensed states of living matter,” *Journal of cell science*, vol. 130, no. 17, pp. 2789–2796, 2017.
- [229] A. H. Fox, Y. W. Lam, A. K. Leung, C. E. Lyon, J. Andersen, M. Mann, and A. I. Lamond, “Paraspeckles: a novel nuclear domain,” *Current Biology*, vol. 12, no. 1, pp. 13–25, 2002.
- [230] A. I. Lamond and J. E. Sleeman, “Nuclear substructure and dynamics,” *Current biology*, vol. 13, no. 21, pp. R825–R828, 2003.
- [231] M. Cioce and A. I. Lamond, “Cajal bodies: a long history of discovery,” *Annu. Rev. Cell Dev. Biol.*, vol. 21, pp. 105–131, 2005.
- [232] T. Pederson, “The nucleus introduced,” *Cold Spring Harbor perspectives in biology*, vol. 3, no. 5, p. a000521, 2011.

- [233] C. P. Brangwynne, C. R. Eckmann, D. S. Courson, A. Rybarska, C. Hoegge, J. Gharakhani, F. Jülicher, and A. A. Hyman, “Germline p granules are liquid droplets that localize by controlled dissolution/condensation,” *Science*, vol. 324, no. 5935, pp. 1729–1732, 2009.
- [234] X. Cao, B. Zhang, and N. Zhao, “Crowding-activity coupling effect on conformational change of a semi-flexible polymer,” *Polymers*, vol. 11, no. 6, p. 1021, 2019.
- [235] S. Sukenik, L. Sapir, and D. Harries, “Balance of enthalpy and entropy in depletion forces,” *Current opinion in colloid & interface science*, vol. 18, no. 6, pp. 495–501, 2013.
- [236] L. Sapir and D. Harries, “Macromolecular stabilization by excluded cosolutes: mean field theory of crowded solutions,” *Journal of chemical theory and computation*, vol. 11, no. 7, pp. 3478–3490, 2015.
- [237] L. Sapir and D. Harries, “Macromolecular compaction by mixed solutions: Bridging versus depletion attraction,” *Current opinion in colloid & interface science*, vol. 22, pp. 80–87, 2016.
- [238] L. Sapir and D. Harries, “Is the depletion force entropic? molecular crowding beyond steric interactions,” *Current opinion in colloid & interface science*, vol. 20, no. 1, pp. 3–10, 2015.
- [239] J. Shin, A. G. Cherstvy, and R. Metzler, “Kinetics of polymer looping with macromolecular crowding: effects of volume fraction and crowder size,” *Soft matter*, vol. 11, no. 3, pp. 472–488, 2015.
- [240] M. Rubinstein and R. Colby, “Polymer physics oxford univ,” *Press, NY*, 2003.

- [241] Q. Chaboche, G. Campos-Villalobos, G. Giunta, M. Dijkstra, M. C. Lagomarsino, and V. F. Scolari, “A mean-field theory for predicting single polymer collapse induced by neutral crowders,” *Soft Matter*, 2024.
- [242] H. M. James and E. Guth, “Theory of the increase in rigidity of rubber during cure,” *The Journal of Chemical Physics*, vol. 15, no. 9, pp. 669–683, 1947.
- [243] J. F. Marko and E. D. Siggia, “Stretching dna,” *Macromolecules*, vol. 28, no. 26, pp. 8759–8770, 1995.
- [244] J. Blundell and E. Terentjev, “Stretching semiflexible filaments and their networks,” *Macromolecules*, vol. 42, no. 14, pp. 5388–5394, 2009.