Title
Molecular Mechanisms of Factors that Control RNA Polymerase II Transcription Elongation Dynamics

Author
Dangkulwanich, Manchuta

Publication Date
2015

Peer reviewed|Thesis/dissertation

Molecular Mechanisms of Factors that Control RNA Polymerase II Transcription Elongation Dynamics

by

Manchuta Dangkulwanich

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos J. Bustamante, Chair
Professor Judith P. Klinman
Professor Evangelina Nogales De La Morena

Summer 2015

Molecular Mechanisms of Factors that Control RNA Polymerase II Transcription Elongation Dynamics

Abstract

Molecular Mechanisms of Factors that Control RNA Polymerase II Transcription Elongation Dynamics

by

Manchuta Dangkulwanich

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Carlos J. Bustamante, Chair

The expression of a gene begins by transcribing a target region on the DNA to form a molecule of messenger RNA. As transcription is the first step of gene expression, it is therefore highly regulated. The regulation of transcription is essential in fundamental biological processes, such as cell growth, development and differentiation. The process is carried out by an enzyme, RNA polymerase, which catalyzes the addition of a nucleotide complementary to the template and moves along the DNA one base pair at a time. To complete its tasks, the enzyme functions as a complex molecular machine, possessing various evolutionarily designed parts.

In eukaryotes, RNA polymerase has to transcribe through DNA wrapped around histone proteins forming nucleosomes. These structures represent physical barriers to the transcribing enzyme. In chapter 2, we investigated how each nucleosomal component—the histone tails, the specific histone-DNA contacts, and the DNA sequence—contributes to the strength of the barrier. Removal of the tails favors progression of RNA polymerase II into the entry region of the nucleosome by locally increasing the wrapping-unwrapping rates of the DNA around histones. In contrast, point mutations that affect histone-DNA contacts at the dyad abolish the barrier to transcription in the central region by decreasing the local wrapping rate. Moreover, we showed that the nucleosome amplifies sequence-dependent transcriptional pausing, an effect mediated through the structure of the nascent RNA. Each of these nucleosomal elements controls transcription elongation by distinctly affecting the density and duration of polymerase pauses, thus providing multiple and alternative mechanisms for control of gene expression by additional factors.

During transcription elongation, RNA polymerase has been assumed to attain equilibrium between pre- and post-translocated states rapidly relative to the subsequent catalysis. Under this assumption, a branched Brownian ratchet mechanism that necessitates a putative secondary nucleotide binding site on the enzyme was proposed. In chapter 3, we challenged

individual yeast RNA polymerase II (Pol II) with a nucleosome as a "road block", and separately measured the forward and reverse translocation rates with our single-molecule transcription elongation assay. Surprisingly, we found that the forward translocation rate is comparable to the catalysis rate. This finding reveals a linear, non-branched ratchet mechanism for the nucleotide addition cycle in which translocation is one of the rate-limiting steps. We further determined all the major on- and off-pathway kinetic parameters in the elongation cycle. This kinetic model provides a framework to study the influence of various factors on transcription dynamics.

To further dissect the operation of Pol II, we focused on the trigger loop, a mobile element near the active site of the enzyme. Biochemical and structural studies have demonstrated that the trigger loop makes direct contacts with substrates and promotes nucleotide incorporation. It is also an important regulatory element for transcription fidelity. In chapter 4, we characterized the dynamics of a trigger loop mutant RNA polymerase to elucidate the roles of this element in transcription regulation, and applied the above kinetic framework to quantify the effects of the mutation. In comparison to the wild-type enzyme, we found that the mutant is more sensitive to force, faster at substrate sequestration, and more efficient to return from a pause to active transcription. This work highlighted important roles of regulatory elements in controlling transcription dynamics and fidelity.

Moreover, RNA polymerase interacts with various additional factors, which add layers of regulation on transcription. Transcription factors IIS (TFIIS) and IIF (TFIIF) are known to interact with elongating RNA polymerase directly and stimulate transcription. In chapter 5, we studied the effects of these factors on elongation dynamics using our single molecule assay. We found that both TFIIS and TFIIF enhance the overall transcription elongation by reducing the lifetime of transcriptional pauses and that TFIIF also decreases the probability of pause entry. Furthermore, we observed that both factors enhance the efficiency of nucleosomal transcription. Our findings helped elucidate the molecular mechanisms of gene expression modulation by transcription factors.

In summary, we have dissected the mechanisms by which the nucleosomal elements regulate transcription, and derived a quantitative kinetic model of transcription elongation in a linear Brownian ratchet scheme with the slow translocation of the enzyme. The corresponding translocation energy landscape shows that the off-pathway states are favored thermodynamically but not kinetically over the on-pathway states. This observation confers the enzyme its high propensity to pause, thus allowing additional regulatory mechanisms during pausing. TFIIS and TFIIF, for example, regulate transcription dynamics by shortening the lifetime of Pol II pauses. On the other hand, the trigger loop of Pol II regulates both the active elongation and pausing. These examples illustrate molecular mechanisms of cis- and trans-acting factors regulate the dynamics of transcription elongation.

To my awesome family, who always believes in me

Contents

List of Figures

List of Tables

Acknowledgments

Throughout my pursuit of this dissertation, I have had support and assistance from many talented individuals. First, I would like to thank Lacramioara Bintu, who introduced me to the world of single-molecule biophysics and trained me to work with DNA, RNA polymerase, nucleosomes, and, more importantly, optical tweezers. She was also a tremendous help when it comes to data analysis and scientific discussions. I am grateful to Toyotaka Ishibashi for teaching me nucleosome biochemistry and instilling in me the hardworking spirit. I would like to thank Yves Coello and Troy Lionberger for their help with instrumentations, calculations, and simulations. I am indebted to Shixin Liu, who taught me to see things from different angles and craft an interesting story. Shixin always knows everything both the science and the magic of having a good time. I cannot forget shared memories with my fellow comrades from the chemistry department: Maya Sen and Daniel Goldman. Their supports and discussions helped me overcome various barriers to this dissertation.

I am also indebted to our collaborators: Dr. Maria Kireeva, Dr. Lucyna Lubkowska and Dr. Mikhail Kashlev from the gene regulation and chromosome laboratory at the National Cancer Institute not only for providing purified yeast RNA polymerase II, but also for their expertise in kinetics of Pol II transcription. We had fruitful conversations dissecting single molecule data and comparing it with invaluable information obtained from bulk studies. I would like to thank another collaborator, Prof. Alfred S. Ponticelli from University at Buffalo, for providing purified yeast TFIIF protein, expression plasmid and detailed protocols. I am also thankful for Prof. Karolin Luger and Daniel Krzizike from Colorado State University for training me their techniques of nucleosome reconstitution and analytical ultracentrifugation analysis of nucleosome samples.

In addition to the people mentioned above, I also would like to thank the past and current members of the Bustamante lab: Craig Hetherington, Gheorghe Chistol, Ninning Liu, Lisa Alexander, Varsha Desai, Bibiana Onoa, Shannon Yan, Jae Yen Shin, Sara Tafoya, Sam Leachman, Antony Lee, Sanghyuk Lee, Brad Zamft, Alyssa Rosenbloom, Maurizio Righini, Liang Meng Wee, Cesar Diaz-Celis, Ronen Gabizon, Luis Ramirez, and Filipp Frank for creating a multi-cultural laboratory environment, and providing supports in various aspects of my projects, my life and my journey through graduate school. I would like to also thank our gifted undergraduate students: Gloria Wu, Tran Do, Lian Lash-Rosenberg and Vianca Vianzon for their experimental assistance.

My time at UC Berkeley would not be as pleasant without the support of my fellow Thai students. I would like to thank the Thai Students Association, especially my fellow graduate students: Chayut Thanapirom, Natth Bejraburnin, Chawita Netirojjanakul, Chao Kusollerschariya, Parichart Munsgool, Penporn Koanantakool, Waruntorn Kanitpanyacharoen, and Pongsin Poosankam to name a few. These talented individuals have been a second home away from home to me. I would like to thank them for organizing gastronomic gatherings,

adventurous trips both in the bay area and beyond, and most importantly, providing support, friendships and advice throughout the ups and downs of graduate school.

Finally, I would like to thank my advisor, Prof. Carlos J. Bustamante, for his constant support and guidance throughout my graduate study. He has taught me to express my ideas clearly both through writing and public speaking. His passion for biophysics and mechanistic views of biological processes has pushed me over many steep energetic barriers of research.

Chapter 1

Introduction

1.1 Introduction to gene expression and transcription

DNA is a molecule of life—a storage of information for biological systems. A good analogy of DNA is a book that contains four letters of A, T, C, and G in various combinations. Even though we can read the strings of these letters, we only understand the meaning of some words or sentences in this biological language. Some regions of the DNA encode for proteins, which carry out essential functions like maintaining cellular structure, delivering cargos to various parts of the cells, or catalyzing reactions to break down food. Scientists labeled these protein-coding regions as genes. To express a gene, information in a DNA molecule is transcribed into RNA molecules that are then used to translate the genetic information into proteins (Figure 1.1). Transcription represents the first step in gene expression. Therefore, it is not surprising that the process is highly regulated and its control is essential for the flow and processing of information to maintain cellular homeostasis. Conserved in all life forms, this logical pattern of information flow renders transcription an essential and fundamental cellular process.

Why has transcription evolved into such an essential cellular process? Why not directly express the information encoded in the DNA genome into proteins? There are several reasons to justify the evolution of transcription as an intermediate step for the synthesis of proteins. First, transcription expands the variety of gene products by allowing for splicing. Second, copying the information within DNA into many RNA molecules increases the rate of total protein synthesis in the cell and avoids the bottleneck that would result from expression of a gene directly from the DNA. Third, the number of RNA molecules available at any given time to synthesize proteins can be precisely regulated to give a burst of products. The

Portions of this chapter were published in M. Dangkulwanich, T. Ishibashi, L. Bintu, and C.J. Bustamante. "Molecular Mechanisms of Transcription through Single-Molecule Experiments." *Chemical Reviews*; 114: 3203-3223. 2014. DOI:10.1021/cr400730x. Used with permission.

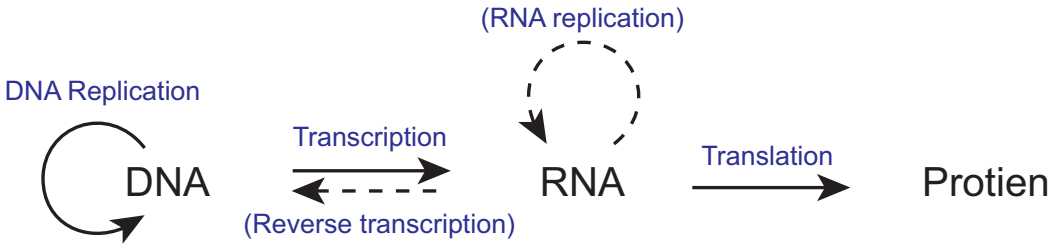

Figure 1.1: **The central dogma of molecular biology.** The process in parentheses and dashed line are not common in all organisms.

signal amplification implicit in the transcription process increases the dynamic range of the expression and it gives rise to stochasticity in gene expression, which causes variabilities among genetically identical cells.

Despite performing various functions in the body, proteins are encoded by only about 1.1% of the human genome as exons and 24% is in introns, which are transcribed but get removed prior to translation into proteins [133]. The other 75% may encode for some regulatory elements, or non-coding RNAs, such as transfer RNAs, which bring amino acids to translate mRNAs into proteins, and micro RNAs which repress gene expression by RNA silencing. The production of these RNAs requires transcription from the target region in the genomic DNA.

The key player in transcription is enzyme RNA polymerase (RNAP). Eukaryotes express three different RNA polymerases: RNA polymerase I, II and III, which transcribe different classes of genes. RNA polymerase I transcribes ribosomal RNAs, and RNA polymerase III transcribes 5s ribosomal RNAs for the small subunit of the ribosomes, transfer RNAs and other small RNAs. RNA polymerase II (Pol II), a 12-subunit enzyme of 550 kDa (Figure 1.2), is responsible for transcribing the pre-cursor of messenger RNAs, some snRNAs and microRNAs.

The process of transcription can be divided into three main stages: initiation, elongation and termination. Whereas single-subunit viral polymerases such as T7 and SP6 RNAP can start transcription at a promoter region without additional cofactors, multi-subunit bacterial and eukaryotic RNAPs require transcription factors that aid the enzyme to recognize and bind to the promoter. Together they form an initiation complex that unwinds the DNA at the promoter and produces a nascent RNA transcript that stabilizes the complex and primes the enzyme for the processive synthesis of a full-length RNA transcript. Although similar events occur in prokaryotes and eukaryotes, different factors participate in the initiation process. The elongation phase of transcription starts once the polymerase has produced a long enough RNA chain and has cleared the promoter region. In this phase, the polymerase incorporates nucleoside triphosphate (NTP) – building blocks – into the growing RNA chain and uses the energy of NTP incorporation to advance on DNA. When the RNAP finishes synthesizing the full-length transcript, it must stop at a specific location and release the transcript in a controlled manner during the termination stage of transcription. In cells, various regulatory

players are present at all three stages of transcription. In this dissertation, I focused on the elongation stage, where the conversion of the chemical energy from NTP substrates into mechanical movement of the enzyme on the DNA takes place. The process is highly dynamic and much of transcription fidelity controls takes place here.

1.2 Single-molecule approaches to transcription

How does a complex molecular machinery like RNA polymerase perform its task? How long does it take to synthesize an RNA transcript? The answers to these questions require tools that are capable of following the progression of transcription in real time. In bulk, one can hope to follow, at most, the progression of transcription as an average of unsynchronized contributions from individual molecules within a population. This averaging obscures crucial information contained in the time-dependent behavior of individual molecules. Single-molecule methods overcome the limitations inherent to the ensemble averaging of bulk methods by allowing one to follow the trajectories of individual molecules in real time. The picture that emerges from single-molecule studies of transcription is that of a rich and complex process that provides many checkpoints for regulation throughout transcription.

Over the past two decades, various methods of single-molecule manipulation and detection have been employed to characterize all three stages of transcription. In the first stage of transcription initiation, RNA polymerase (RNAP) must locate specific promoter sites on the genome in the densely packed cellular environment. Single-molecule methods, such as atomic

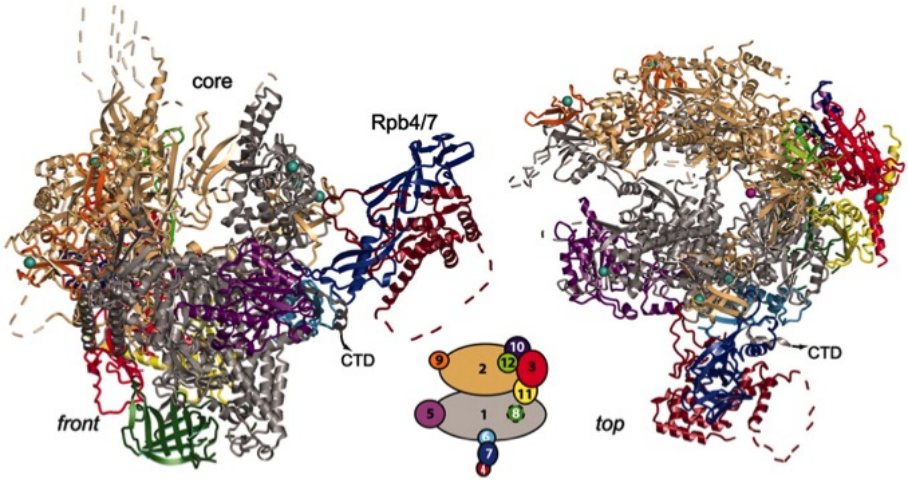

Figure 1.2: **Crystal structure of RNA polymerase II.** The front and top views of the RNA polymerase II from *S. cerevisiae* are shown [5]. The 12 subunits Rpb1-Rpb12 are colored according to the key below the views. Dashed lines represent disordered loops. The magnesium ion at the active site and eight zinc ions are depicted as a pink and cyan spheres, respectively. PDB ID: 1WCM.

force microscopy (AFM) and fluorescence-based approaches, have provided insights into how RNAP locates its promoter and unwinds the DNA duplex. Because of the DNA helical structure, unwinding of the duplex is accompanied by changes in its twist. Through the use of magnetic tweezers, it has been possible to both apply torque and follow the torsional states of individual initiating RNAP complexes. During the second stage of elongation, RNAP operates as a molecular motor, converting difference between high-energy phosphoanhydride bonds and lower energy phosphodiester bonds into mechanical work, through the generation of force (in piconewton range) and displacement (in subnanometer scale). Methods of single-molecule manipulation, such as optical tweezers, are ideally suited to precisely measure forces and displacements on this scale; thus, optical tweezers are capable of providing unique insight on the mechanochemical conversion in the transcription process as well as the mechanisms by which transcription factors regulate the dynamics and the progress of the enzyme. When the RNAP finishes synthesizing the full-length transcript, it must stop at a specific location and release the transcript in a controlled manner. Single-molecule techniques make it possible to selectively apply loads on either the DNA template or the RNA transcript, and to dissect regulatory elements in the final stage of transcription, termination.

Optical tweezers (or optical trapping) is ideally suited to characterize the dynamics of RNA polymerase during transcription elongation as it precisely measures piconewton ranges of forces and nanometer ranges of displacement.

1.2.1 A brief introduction to optical trapping

The technique was first reported in 1970 by Ashkin who showed a stable three-dimensional trapping of dielectric particles using radiation pressure from a single laser beam [7]. We briefly discuss the general principles of optical trapping here (see [120, 100, 95, 101, 20] for more details).

Light carries momentum; thus, it can exert forces on particles. In optical traps, there are two components of forces: scattering forces, which can be thought of as a stream of photons pushing the objects along the propagation direction of the light and gradient forces, which pull objects towards a more intense spot in the gradient [8]. When a ray travels through an interface with a higher index of reflection (i.e., dielectric particles, such as polystyrene beads) than the medium, it bendts towards the normal to the boundary, resulting in changes in the momentum of the light. The change of momentum over time is equivalent to force: $\Delta\overrightarrow{p}/\Delta\mathrm{t} = \mathrm{m}\Delta\overrightarrow{v}/\Delta\mathrm{t} = \mathrm{m}\overrightarrow{a} = \overrightarrow{F}$.

As shown in Figure 1.3A, an incident beam with a momentum of p_{in} is refracted by the particle to yield the new momentum of p_{out}, and the difference in momentum of Δp (blue). Since the momentum is always conserved, the particle experiences a force of equal magnitude, but in opposite direction (F). In an intensity gradient increasing from left to right, the force produces by the more intense ray (F_a) is larger than that produces by (F_b)(Figure 1.3B); thus, the particle is attracted to the point of the highest intensity. To stably trap a particle in three dimensions, axial forces that balance the scattering force is required [100]. Using a high numerical aperture (NA) objective, which sharply focuses the laser, generates gradient

forces that attracts the particle toward the focal point, and stably trap a particle just slightly down-beam from the focal point (Figure 1.3C)[100]. In the vicinity of the focus, the optical trap acts a "Hookean" spring, exerting forces that are proportional to the displacement of the object from the equilibrium position.

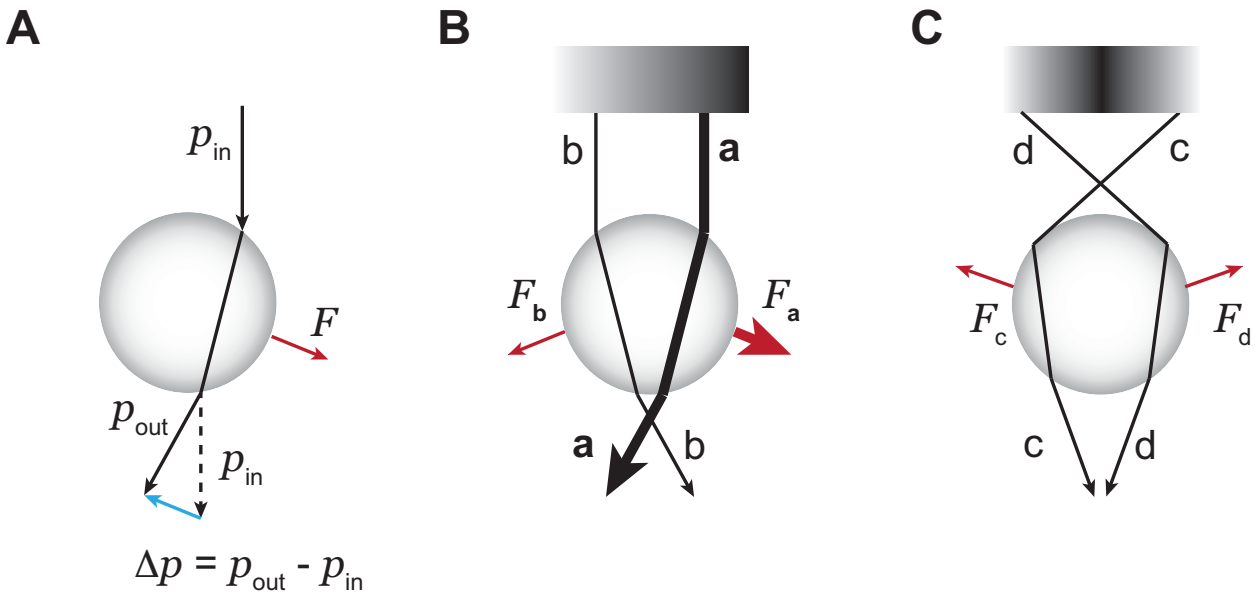

Figure 1.3: **Simplified illustration of optical trapping.** (**A**) An incident ray is refracted by a particle, generating an optical force (F). (**B**) In an intensity gradient of light, the force generated by the more intense light is larger; thus, the particle is attracted to the point of the highest intensity. (**C**) In a three-dimensional gradient, the particle is stably trapped just below the focus. Two representative rays are refracted by the particle and produce restoring forces, pushing the particle toward the focus. The lateral components of the forces balance each other and the axial force balances the scattering force [100].

For biological applications, trapping lasers are usually in the near infrared wavelengths (800–1,100 nm) to minimize sample damage. Our setup used a diode-pumped neodymium yttrium aluminum garnet (Nd:YAG) laser with a wavelength of 1,064 nm. In a typical optical tweezers experiment, molecules of interest, such as DNA, or molecular motors, are attached to micron-sized polystyrene beads, which can be held in an optical trap. To manipulate the molecule, the second attachment on the other end is required. It can be attached to a glass surface, a second bead held in a trap by suction, or a bead in a second optical trap. We can move the trapped bead from the second attachment point to apply force and stretch the molecule. The beads held in optical traps also serve as probes for the movements of the molecules. One of the techniques to measure the movement of the beads is back-focal plane interferometry, which uses the interference pattern formed between the trapping laser and the scattered light [95]. The recorded position of the beads are then analyzed to extract

dynamic information from the system of interest.

1.2.2 Optical tweezers based transcription assay

For transcription elongation studies, we attached a stalled biotinylated RNA polymerase elongation complex on streptavidin coated polystyrene beads. One end of the template DNA is labeled with a digoxygenin molecule, which can interact with an antidigoxygenin antibody coated bead and form a tether between the two beads. One can exert a desired amount of force on the complex. The force can be applied to either oppose or assist the enzyme by changing the location of the digoxygenin label, either upstream (assisting force, Figure 1.4A), or downstream (opposing force, Figure 1.4B). Then, the starting position of the beads are recorded. Upon introduction of NTP into the flow chamber, active molecules of polymerase will start transcribing within a few seconds, thus, changing the distance between the two beads. Real-time trajectories of the enzyme are recorded. The "Materials and methods" section of each chapter described specific protocols.

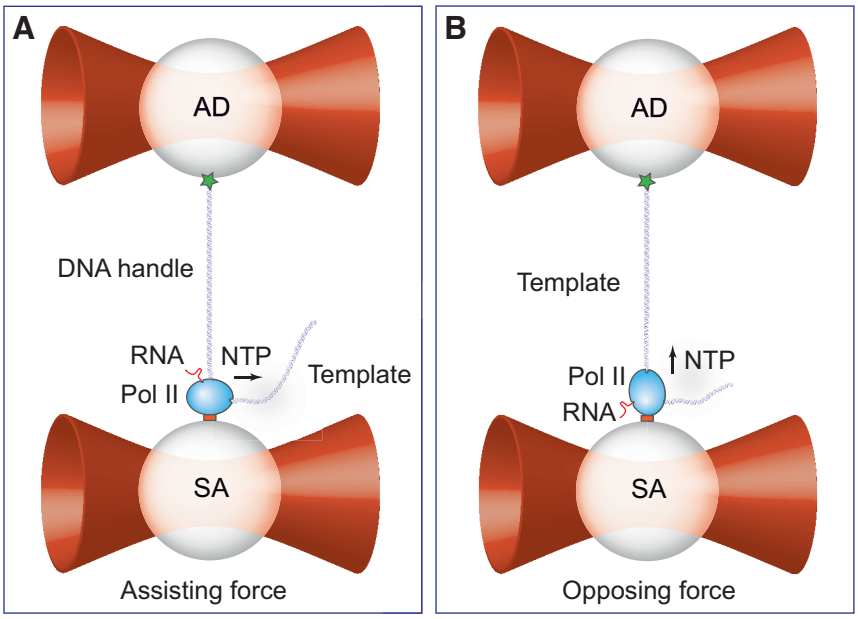

Figure 1.4: **Optical tweezers based transcription assay** in assisting (A) and opposing force geometry (B).

1.3 The mechanistic views of transcription elongation

During the elongation phase of transcription, Pol II adopts various conformations to orchestrate the incorporation of cognate nucleoside triphosphate (NTP) to the growing RNA chain and advances along the DNA one nucleotide at a time. In doing so, Pol II acts as a molecular motor that converts the energy difference of breaking the high-energy phosphoanhydride bonds and forming the phosphodiester bonds into mechanical translocation along the DNA template (Figure 1.5). Various kinetic models of transcription have been proposed to describe this conversion of chemical into mechanical energy (mechanochemistry). In general, the enzyme translocates along the DNA by thermal energy and the incoming NTP stabilizes the post-translocated state before hydrolysis and forming a bond with the growing RNA chain. The enzyme also pauses for various durations. Pausing and active elongation are competing pathways, but the mechanistic details of the process are still a matter of much debate.

A closer inspection of the structure shows that the active site (where the magnesium atom is located in Figure 1.2) is buried deep in the structure, and surrounded by several structural elements that can adopt different conformations throughout the nucleotide addition cycle [67]. Two prominent features of RNA polymerases are the bridge helix (BH) and the trigger loop (TL), which have been proposed to function together to drive translocation. The BH is a long helix that spans the active site adjacent to the RNA/DNA hybrid and the substrate loading site [42, 11], and adjacent to the BH is the TL (Figure 1.6).

During the nucleotide addition cycle, an NTP molecule first binds to an open active center in a preinsertion state [64, 132]. Then, the TL folds into alpha helices and closes on the active center, which moves the NTP into the insertion site [25, 132, 141]. The folded TL makes contacts with the correct NTP in the active site and leads to catalysis and bond formation with the growing RNA chain [141], followed by the release of pyrophosphate, which may unfold the TL and open the active center for translocation [37]. The translocation has been proposed to occur in two steps: the TL and the BH shift the recently added base out of the active site, and then the BH relaxes and allows for NTP entry to the active center [16]. Snapshots of various steps have been captured in crystal structures (Figure 1.6) [17, 26]. However, the time domain in the sequence of these snapshots are missing. Methods of single-molecule manipulation, such as optical tweezers, promise high temporal and spatial resolution information that will complete the mechanistic understanding of transcription elongation.

1.4 Nucleosomes: Barriers to gene expression

In vivo, eukaryotic DNA in the nucleus is organized by wrapping around histone proteins into nucleosomes. An array of nucleosomes further folds into chromatin fibers, chromatin domains, and chromosome (Figure 1.7A). The nucleosomes impose physical barriers to any processes that require access to the genetic code, such as DNA replication, DNA repair and

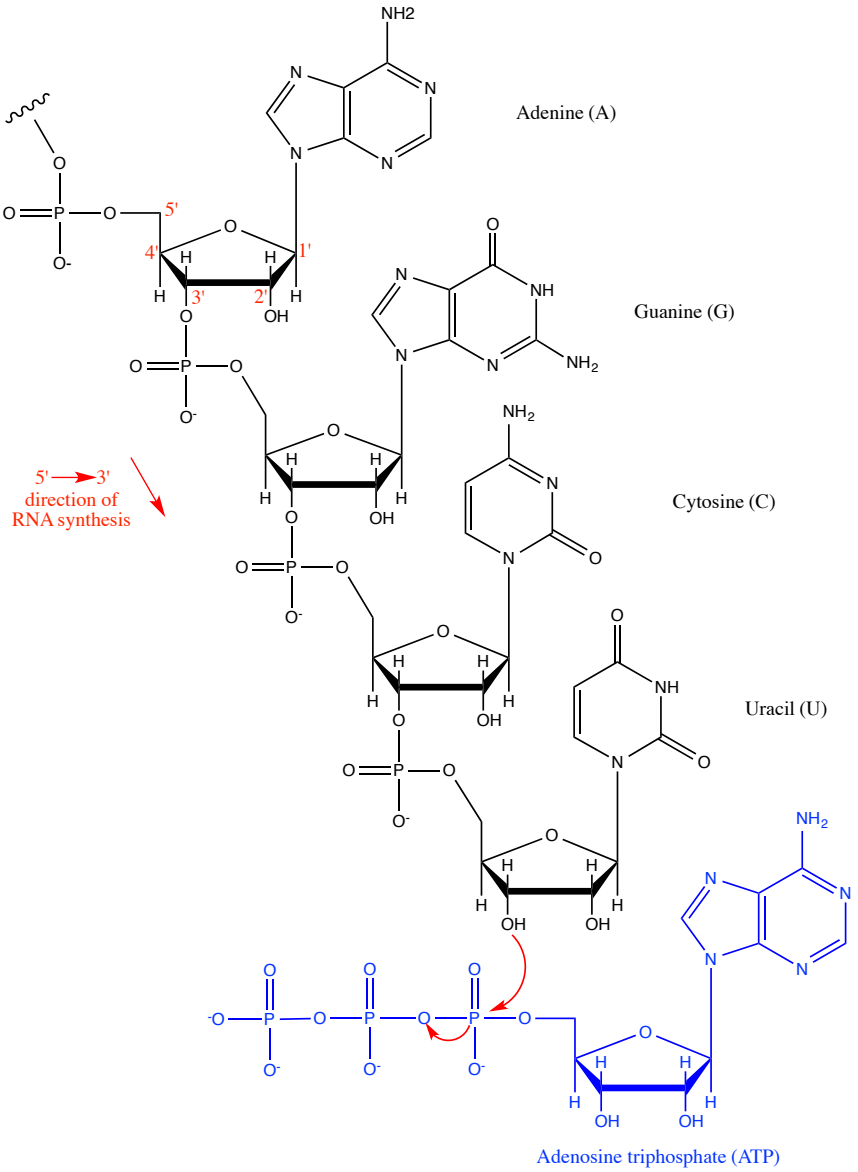

Figure 1.5: **Schematic of RNA synthesis.** A chemical structure of a growing RNA chain is shown with four RNA bases: A, G, C, U (black). Carbon atoms in the first ribose are labeled to illustrate the direction of RNA synthesis. A molecule of ATP (blue) is being incorporated into the chain. RNA polymerase catalyzes the breaking of the phosphoanhydride bond and forming the phosphodiester bond (curved red arrows). Pyrophosphate is released as a byproduct of the reaction.

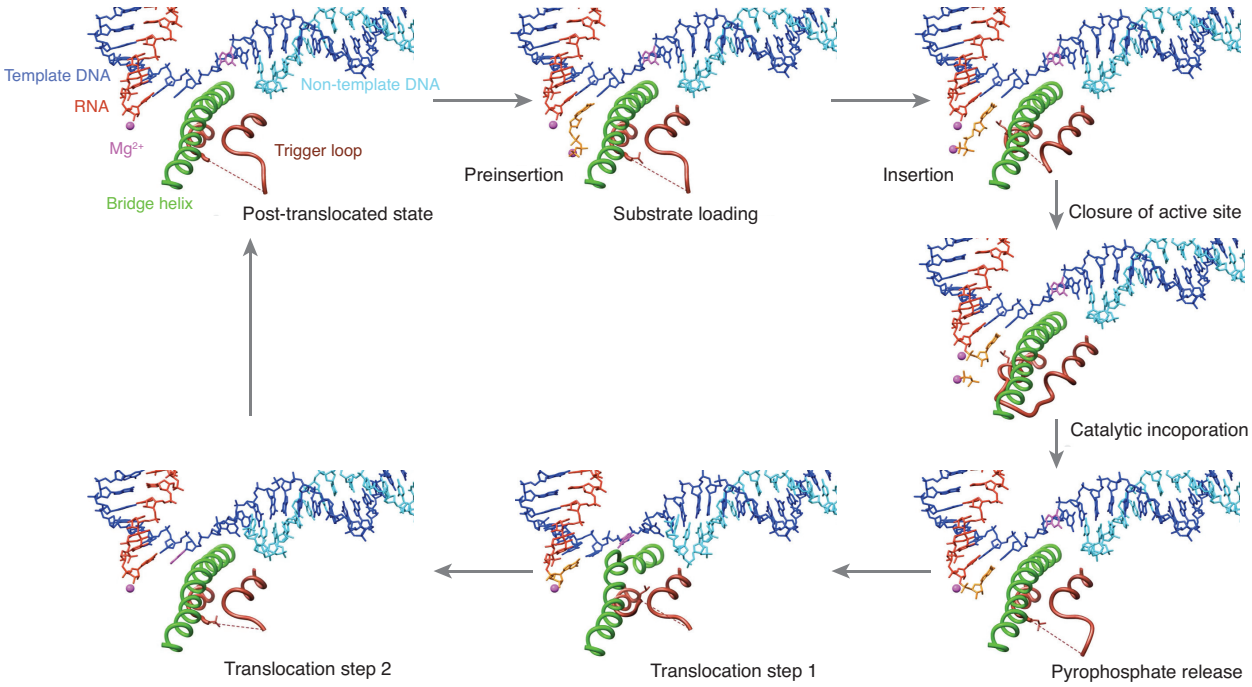

Figure 1.6: **Snapshots of the active center during the nucleotide addition cycle.** Representative images from a movie [26] show different functional states of the elongation complex: post-translocated state, substrate binding, active site closure, catalytic incorporation, pyrophosphate release, and two-step translocation.

transcription.

Structurally, the nucleosome comprises 147 bp of DNA wrapping 1.75 turns around the histone octamer, two copies of H2A, H2B dimers and a H3-H4 tetramer (Figure 1.7B) [87]. It shapes like a disc with 11-nm diameter and 5.5-nm thickness [87]. In the nucleosome, the structured core histone domain interacts with surrounded DNA [110], while the flexible histone tails extend away from nucleosomal DNA. These histone tails contain sites for post-translational modifications, such as acetylation, methylation, and phosphorylation. Consequently, the tails play important roles in inter-nucleosome interactions and serve as recognition sites for additional factors [86]. The nucleosome is a general repression to gene expression. In vitro, a single nucleosome is sufficient to stop or greatly delay transcription [58, 69, 140, 71, 14]. Different elements of the nucleosome—the histone tails, the histone core domain, and the DNA sequence—provide platforms to regulate access to the genetic material; thus, modification of these elements should diversify the dynamics of gene expression and transcription.

At the mechanistic level, the nucleosome regulates transcription of Pol II by acting as a fluctuating barrier that only permits the forward movement of Pol II, when the local nucleosomal DNA is unwrapped from the histones [54]. If the downstream nucleosomal

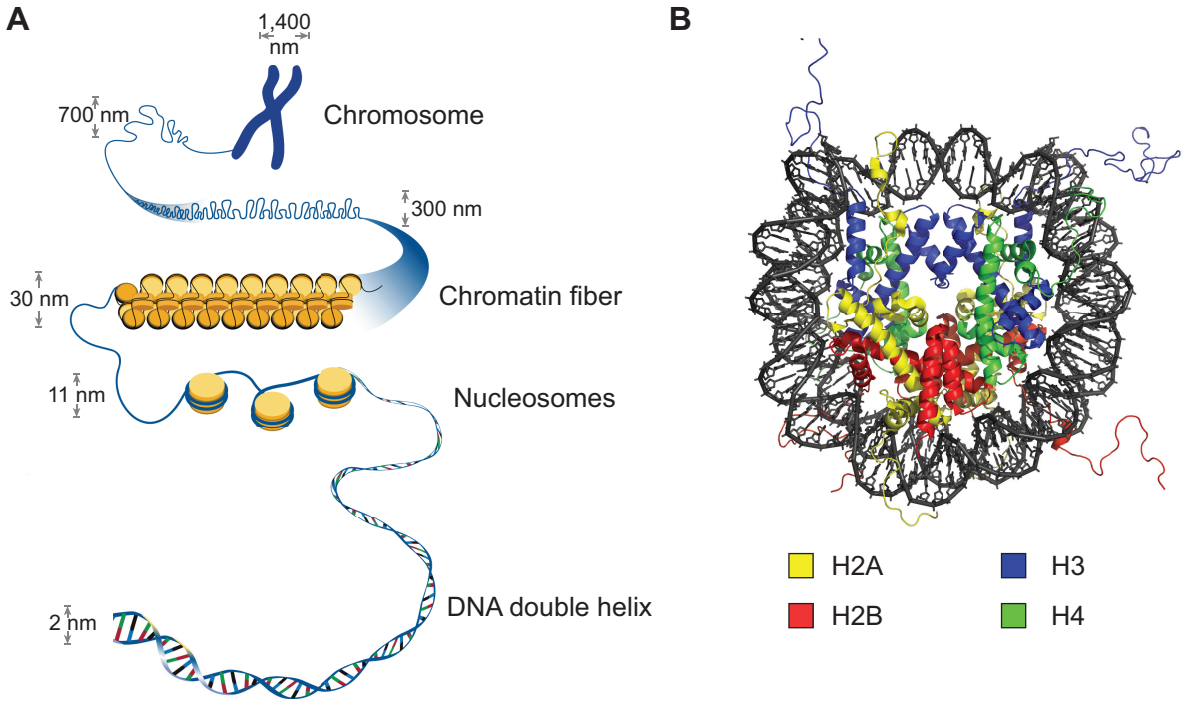

Figure 1.7: **Organization of genomic DNA into nucleosomes.** (**A**) DNA double helix is wrapped into nucleosomes, which fold into chromatin fiber. (**B**) Crystal structure of the nucleosome. PDB ID: 1KX5

DNA is wrapped the polymerase pauses, backtracks, and passively waits for the unwrapping of the nucleosome to proceed.

Furthermore, the roadblock to the transcribing RNA polymerase II represented by a nucleosome only reduces the forward translocation rate of the enzyme but does not affect its reverse motion [54], unlike an externally applied force which simultaneously affects both.

1.5 Transcription elongation factors

If a nucleosome presents a strong barrier to transcription and essentially represses gene expression, how does transcription occur so efficiently in the cell? It turns out that the cell expresses other factors to regulate the process of gene expression, such as nucleosome remodelers, and transcription factors.

As the name suggests, transcription factors are proteins that influence the process of transcription. During the elongation phases, transcription factors play various important roles in regulating the rates and fidelity of transcription. Various elongation factors have been identified: TFIIS, TFIIF, ELL, Elongin and FACT [104, 113].

In particular, TFIIS, studied extensively via biochemical and structural studies [149, 63], has been shown to rescue backtracked elongation complexes by stimulating the endonuclease

activity of polymerase to generate a new 3′-end of the transcript at the enzyme's active site. Although much is known about TFIIS, its function on the dynamics of transcription through a nucleosome context is of great interest, especially from a quantitative point of view. Although TFIIF has an established role in transcription initiation, it can also interact indirectly with Pol II in the elongation phase and is thought to affect transcription by decreasing the likelihood of the enzyme entering a pause. In vivo, various transcription factors can be available simultaneously, and the effects of their interactions in transcription dynamics is an active area of research.

1.6 Molecular mechanisms of transcription: Questions

Regulation of transcription and gene expression is clearly an important biological process that underlies numerous diseases, including Huntington's disease, Alzheimer's disease, and many cancers, which account for millions of deaths annually. Therefore, understanding the molecular basis of transcription and its regulation is essential for both diagnosis and development of new therapies.

This dissertation aims to address the following questions:

(1) What are the roles of various nucleosomal elements in establishing the magnitude and the spatial distribution of the barrier to transcription?

In chapter 2, we used optical tweezers to follow real-time trajectories of individual Pol II complexes as they transcribed through nucleosomes containing modifications either in the histone tails or at specific histone-DNA contacts. Specifically, we asked: how is the stability of nucleosomes affected by these modifications? How are the wrapping/unwrapping rates of the DNA around the histone core altered? How do they, in turn, affect the polymerase dynamics? What is the role of the enzyme?s pausing in this modified behavior? And what is the spatial extent and distribution of these effects?

(2) What is the mechanism by which Pol II converts the chemical energy into mechanical energy? What are the rates associated with each step?

In chapter 3, we used an optical tweezers based assay to follow the transcription trajectories of single yeast RNA polymerase II molecules under a variety of conditions including varying NTP concentrations, assisting and opposing applied forces, and different tracks (bare and nucleosomal DNA). The physical barrier of the nucleosome only hinders the forward translocation of the polymerase without affecting the backward movement. This unique property inspired us to use the nucleosomal barrier as a tool to separately measure the forward and reverse translocation rates of a transcribing enzyme.

(3) What are the roles of the trigger loop (TL) in regulating transcription dynamics and fidelity?

Since the TL is an important regulator of transcription for both the dynamics and the fidelity, characterizing a TL mutant RNA polymerase will not only illustrate the functions of this loop, but also provide additional information on the kinetics of transcription by multi-subunit RNA polymerases. A TL mutant Pol II from *S. cerevisiae*, *rpb1*-E1103G (glutamate to glycine), has been used to investigate the detail of transcription kinetics. This mutant has been shown to transcribe DNA at a faster overall velocity at the expense of its fidelity, and its translocation equilibrium is biased toward the pre-translocated state [89, 72]. A recently characterized force-velocity relationship of this mutant showed that it has a faster pause-free velocity and is more sensitive to force than the wild-type Pol II [82]. In chapter 4, we examined the dynamics of the E1103G Pol II and quantitatively described the role of the TL in regulating transcription dynamics, thus complementing the static snapshots obtained from structural studies with dynamical information.

(4) What are the roles of transcription factors IIS, IIF and their combination in transcription elongation through a nucleosomal template?

In chapter 5, we investigated the effects of yeast TFIIS and TFIIF on transcription elongation by Pol II, using a single-molecule optical-tweezers assay in both assisting and opposing force experimental geometries. The experiments were carried out on both bare and nucleosomal DNA. We also applied the kinetic model obtained in chapter 3 to quantify the effects of these factors.

Chapter 2

Nucleosomal elements that control the topography of the barrier to transcription

The elements that control the nucleosome barrier: the histone tails, the specific histone-DNA contacts and the underlying DNA sequence, harbor various possibilities to fine tune gene regulation. The histone tails are subjected to many post-translational modifications, such as phosphorylation on serines or threonines, methylation on lysines or arginines, acetylation and deacetylation of lysines, and ubiquitination of lysines. These modifications could be markers to recruit additional remodeling complexes and/or directly affect the stability of the nucleosomes, hence, regulating the dynamics of the transcribing Pol II.

In this chapter, we sought to separate and quantify the roles played by the various nucleosomal elements in establishing the magnitude and the spatial distribution of the barrier to transcription.

2.1 Components of the nucleosome

It has been shown that the nucleosomal barrier to transcription in vitro varies both across the extent of one nucleosome and from one type of nucleosomes to another [14, 56, 71, 130]. This variability arises from modifications in elements that control the barrier: the histone tails, the specific histone-DNA contacts, and the underlying DNA sequence. Each of these elements differentially controls the local stability of the nucleosome and, as such, can be a target of gene regulation in vivo.

Portions of this chapter were published in L. Bintu, T. Ishibashi, M. Dangkulwanich, and C. Bustamante. "Nucleosomal Elements that Control the Topography of the Barrier to Transcription." *Cell*; 151: 738-749. 2012. DOI:10.1016/j.cell.2012.10.009. Used with permission.

Specifically, the histone tails are subjected to many post-translational modifications. For example, acetylation of lysine side chains, "loosens" the DNA wrapped into the nucleosome, as demonstrated by increased DNA accessibility to nucleases [2, 51, 117] and by sensitivity to force application [15]. Indeed, removal of the histone tails decreases the barrier to transcription [130], and nucleosomes containing hyperacetylated histones are more easily transcribed by a bacteriophage polymerase than native nucleosomes [107]. However, it is not known how the transcription barrier is affected by nucleosomes acetylated only at the lysines targeted in vivo.

In contrast to the effect of the histone tails on transcription, the role of the contacts between the histone core domains and DNA is less understood. Yet, the strongest histone-DNA interactions are mediated by the histone core domains, especially those of H3 and H4, making these contacts likely candidates for gene regulation. It has been shown that point mutations in the core domains of histones H3 and H4 can partially relieve loss-of-function mutations of the chromatin remodeling complex SWI/SNF (switch/sucrose nonfermentable) in vivo [79]. Although these single amino acid mutations result in minimal structural changes to the nucleosome, they increase the mobility of nucleosomal DNA, suggesting a reduced affinity between the DNA and the mutated histones [97]. Accordingly, Sin (SWI/SNF independent) mutations are thought to lower the nucleosomal barrier to transcription [56].

In addition to histones, the DNA sequence wrapped around the octamer is also known to influence both the arrest probability and the pattern of Pol II pausing [14, 71]. The mechanism through which the DNA sequence affects nucleosomal transcription is, however, unclear. It could arise from different affinities of various sequences for the histones [85], and/or the proclivity of certain sequences to induce Pol II pausing.

2.2 Histone modifications alter passage probabilities and crossing times

We performed single molecule experiments using a dual-trap optical tweezers under assisting-force geometry (Figure 2.1A) [54]. We collected data at 300 mM KCl, an ionic strength that is slightly above physiological values (150–200 mM KCl). At this salt concentration, enough polymerases manage to pass through the nucleosomal barrier, allowing us to gather enough statistics for robust conclusions.

To understand how much of the tails' contribution is mediated through their positive charges, we "mock-acetylated" the histones by substituting all lysine residues known to be acetylated in vivo with glutamines (Materials and methods, Table 2.5). In order to examine the importance of direct histone-DNA contacts, we reconstituted core nucleosomes using the Sin mutant histones H4 R45A (Sin H4) and H3 T118H (Sin H3) (Materials and methods, Table 2.5). Representative traces presented in Figure 2.1B–D show the general trends of transcription for each construct, together with the nucleosome passage probability in each experiment. Transcription on bare DNA (Figure 2.1B) has portions of fast translocation,

Table 2.1: **General characteristics of transcription**

Template	**Passed NPS (%)**	**Time spent at the NPS (s)**	**Pause density (pauses/kbp)**	**Mean pause duration (s)**
Bare DNA	87	19.5 ± 3.5	4 ± 1	4.4 ± 0.7
Unmodified nucleosome	58	46.9 ± 5.6	14 ± 2	10.2 ± 1.1
Tailless nucleosome	71	39.5 ± 5.7	12 ± 2	7.9 ± 1.2
Acetylated nucleosome	63	45.3 ± 7.6	11 ± 2	9.6 ± 1.5
Sin H4 nucleosome	74	26.1 ± 5.4	8 ± 2	6.5 ± 0.8
Sin H3 nucleosome	78	25.7 ± 5.9	9 ± 2	5.5 ± 0.6
Unmodified nucleosome with RNase A	34	74.2 ± 19.6	18 ± 5	13.3 ± 2.8

Errors in total time spent at the NPS, pause density, and mean pause duration represent standard errors of the mean (SEM). A histogram showing the probability of passage through the NPS is shown in Figure 2.2.

punctuated by short pauses [54]. Most Pol II elongation complexes (87%) transcribed to the end of the template, crossing the nucleosome positioning sequence (NPS) in 19.5 ± 3.5 s on average. In contrast, transcription through unmodified nucleosomes (Figure 2.1B) is interrupted by very long pauses, and the total NPS crossing time varies from tens of seconds to a few minutes, with an average of 46.9 ± 5.6 s. Often these pauses turn into arrests, so only 58% of polymerases overcome the nucleosomal barrier (Figures 2.1B and 2.1). Overall transcription through tailless and acetylated nucleosomes is slightly faster than through unmodified nucleosomes (Figure 2.1C), with crossing times that are generally under 1 min (39.5 ± 5.7 and 45.3 ± 7.6 s, respectively). Both the removal and acetylation of the tails increase efficiency of NPS passage: 71% for tailless nucleosomes and 63% for acetylated nucleosomes (Figures 2.1C and 2.2), in agreement with results obtained using bulk assays of transcription [130].

Significantly, the effect of the Sin mutations on nucleosomal transcription is the largest, decreasing even further the time Pol II takes to cross the NPS (Figure 2.1D; Table 2.1), with means much closer to that on bare DNA: 26.1 ± 5.4 s for Sin H4, and 25.7 ± 5.9 s for Sin H3. Correspondingly, these mutations increase the probability of Pol II passage through the nucleosome to 74% for Sin H4 and 78% for Sin H3 (Figures 2.1D and 2.2), consistent with recently published reports [56]. Note that the effects of these single-residue Sin mutations are much larger than those of the tailless or acetylated nucleosomes, even though the tails represent 25% of the histone mass. These results point to the importance of the specific contacts that the histone-core domains make with the DNA for shaping the magnitude and spatial extent of the nucleosomal barrier.

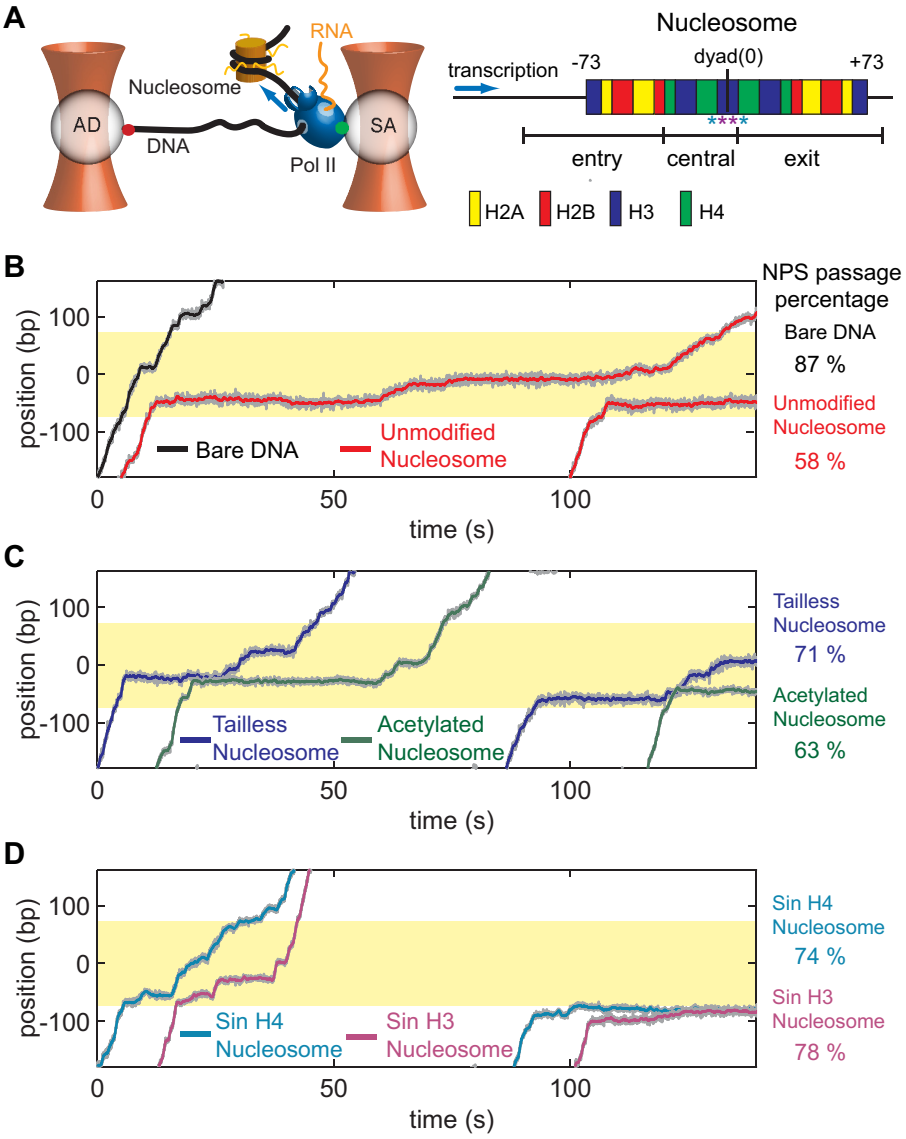

Figure 2.1: **Transcription through modified nucleosomes.** (**A**) (Left) Experimental setup for single-molecule transcription experiments. Two laser beams (red) are used to trap antidigoxigenin (AD) and streptavidin (SA) coated beads. A DNA tether is formed between a Pol II and the upstream DNA. The blue arrow shows the direction of transcription. (Right) Cartoon schematic of the histone-DNA contacts on nucleosome is shown as color coded rectangle. Asterisks are the positions of Sin H4 (cyan) and Sin H3 (purple) mutations. The position of Pol II as a function of time during single-molecule transcription of bare DNA (black) and unmodified nucleosomes (red) (**B**), tailless (blue) and mock-acetylated (green) nucleosomes (**C**), and Sin H4 (cyan) and H3 (purple) mutant nucleosomes (**D**). Traces where Pol II passed the nucleosome positioning sequence (NPS, shaded yellow) are shown on the left and traces that arrested at the nucleosome are on the right. Insets show the percentages of Pol II molecules that transcribed the entire NPS.

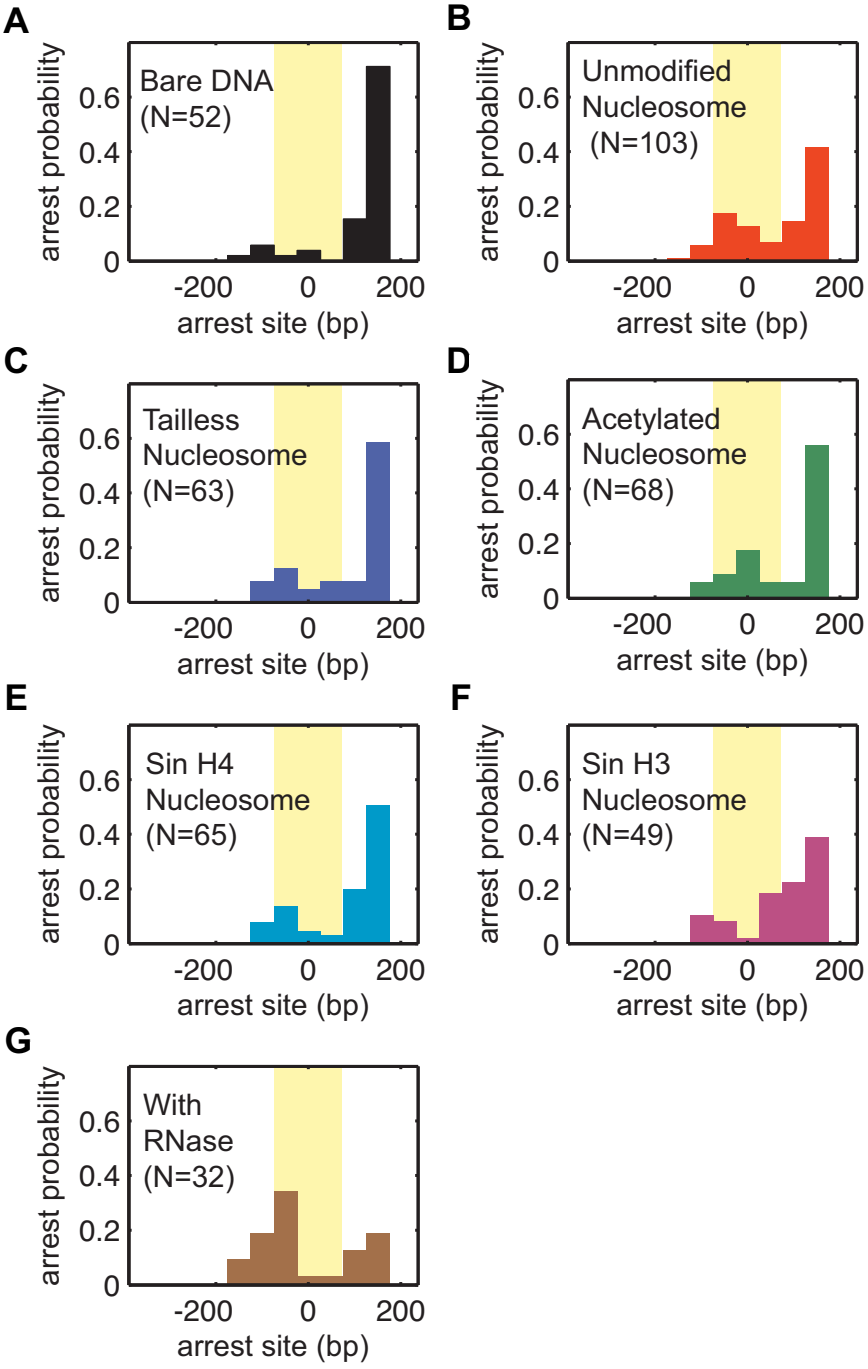

Figure 2.2: **Arrest probabilities for modified nucleosomes.** Position of terminal arrest as a function of position on the DNA template, as determined from single-molecule trajectories at 300 mM KCl for: (**A**) bare DNA, (**B**) unmodified nucleosome, (**C**) tailless nucleosome, (**D**) mock-acetylated nucleosome, (**E**) Sin H4 nucleosome, (**F**) Sin H3 nucleosome and (**G**) nucleosome with RNase A.

2.3 The histone tails gate the nucleosome entry region

For each trace where Pol II has completed transcription of the NPS, we identified the regions of pausing and active elongation, and quantified the pause durations and pause density (i.e., the number of pauses per base pair). The NPS for the core nucleosomal particle includes 147 base pairs (bp) of DNA between positions –73 and +73 bp with respect to the dyad. However, we observe increased pausing for unmodified nucleosomes compared with bare DNA earlier than position –73 (Figure 2.3), so we extend our pause analysis to the entire region between –115 and +85 (extended NPS). The inclusion of additional DNA is justified because: (1) we follow the position of the active center of the polymerase; however, its leading edge reaches the nucleosome ∼15–20 bp ahead [112], and (2) the histone tails bind additional DNA outside the region spanned by the core nucleosomal particle [4].

Pause density as a function of position on the template reveals that the effect of the nucleosomal modifications is not global, but circumscribed to certain regions along the DNA wrapped around the histone octamer. For tailless nucleosomes, most of the changes in transcription dynamics are concentrated at the entry region of the nucleosome, defined here as –115 to –35 bp with respect to the nucleosome dyad (Figure 2.3A). Moreover, compared to unmodified nucleosomes, the pauses in the entry region are significantly shorter and fewer for tailless nucleosomes (Figure 2.3B). In the central (–35 to +5 bp) and exit (+5 to +85 bp) regions, both pause densities and pause durations for tailless nucleosomes are statistically indistinguishable from those of unmodified nucleosomes (Figures 2.3A and B).

The effect of mock acetylation of the tails is smaller but similar to their removal, characterized by a reduction in both pause densities and durations in the entry region (Figures 2.3C and 2.3D), indicating that the acetylation of lysine charges constitutes only a small part of the nucleosomal barrier. As expected, the pausing in the central region is indistinguishable from unmodified nucleosomes.

The asymmetry between the results in the entry and exit regions may appear surprising at first, given the dyad symmetry of the nucleosome. However, if the tails bridge the entry and exit DNA, once this connection is broken, it cannot reform because of the physical bulkiness of the polymerase.

2.4 Histone-DNA contacts at the dyad control the nucleosomal barrier height

In contrast to the effects observed during transcription through tailless and acetylated nucleosomes, the major effect of the Sin mutants is in the central region (–35 to +5 bp), which constitutes the major barrier to transcription in the unmodified nucleosome (Figures 2.4A and C). Because we map the position of the active site of the polymerase on DNA, this major change occurs when the leading edge of the polymerases reaches the nucleosome dyad.

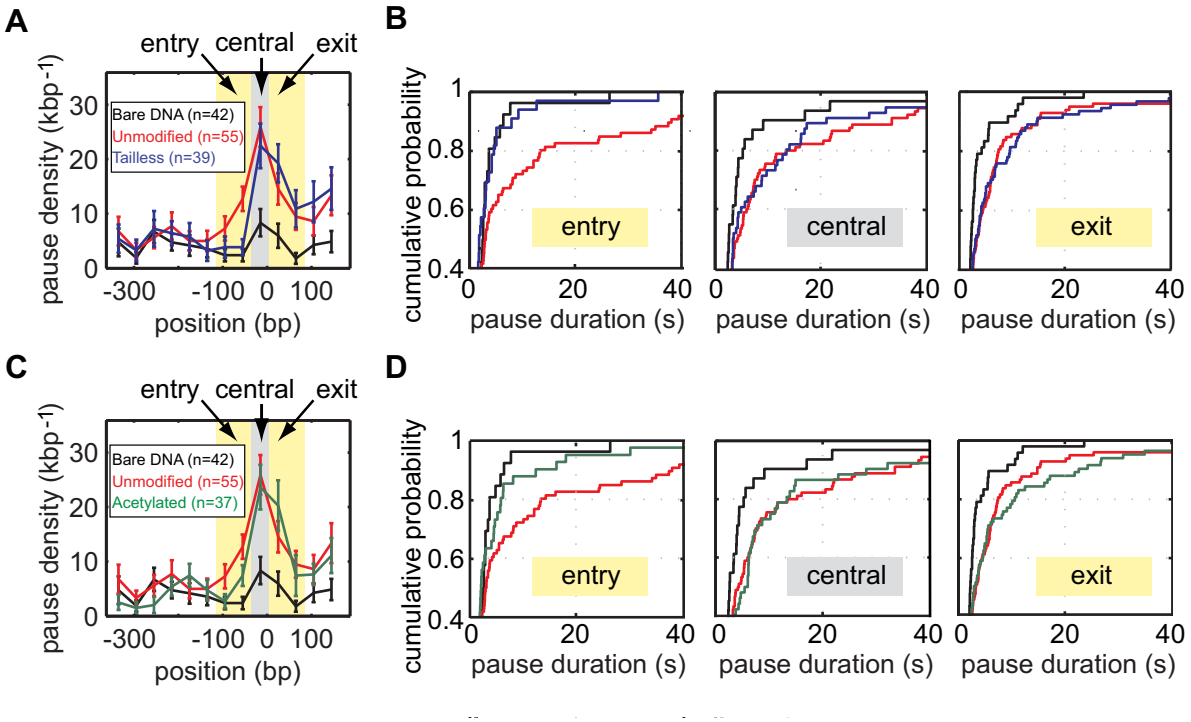

Figure 2.3: **Tails affect pausing in the nucleosome entry region.** (**A** and **C**) Pause density as a function the position of the active center of Pol II on the template for tailless and acetylated nucleosomes. The nucleosome entry/exit regions are shaded yellow, and the central region is shaded gray. Error bars represent SEM. (**B** and **D**) Pause durations in the entry, central, and exit regions of these nucleosomes.

The pause durations in the central region were significantly shorter for the Sin H4 mutant than for the unmodified nucleosome (Figure 2.4B; Table 2.1). Although both pause densities and durations in the entry are shorter for Sin H4, neither was significantly different from unmodified nucleosomes.

The strongest effects of the Sin H4 mutation are localized around the region containing the mutated amino acid. This observation agrees with the crystal structure of the Sin H4 R45A nucleosome, which shows that the change from an arginine to an alanine results in an empty minor groove of the DNA contacting this point [97].

We observe a similar pattern of pausing for the Sin H3 mutant: a strong effect on pause number and duration in the central region (Figures 2.4C and D; Table 2.1). However, in this case, the pause durations in the entry are significantly shorter than those observed for unmodified nucleosomes. This observation indicates that the effect of the mutation on pause recovery extends beyond the dyad region into the entry region of the barrier as has been suggested [97].

Similar to the tail-modified nucleosomes, the Sin mutants do not induce significant

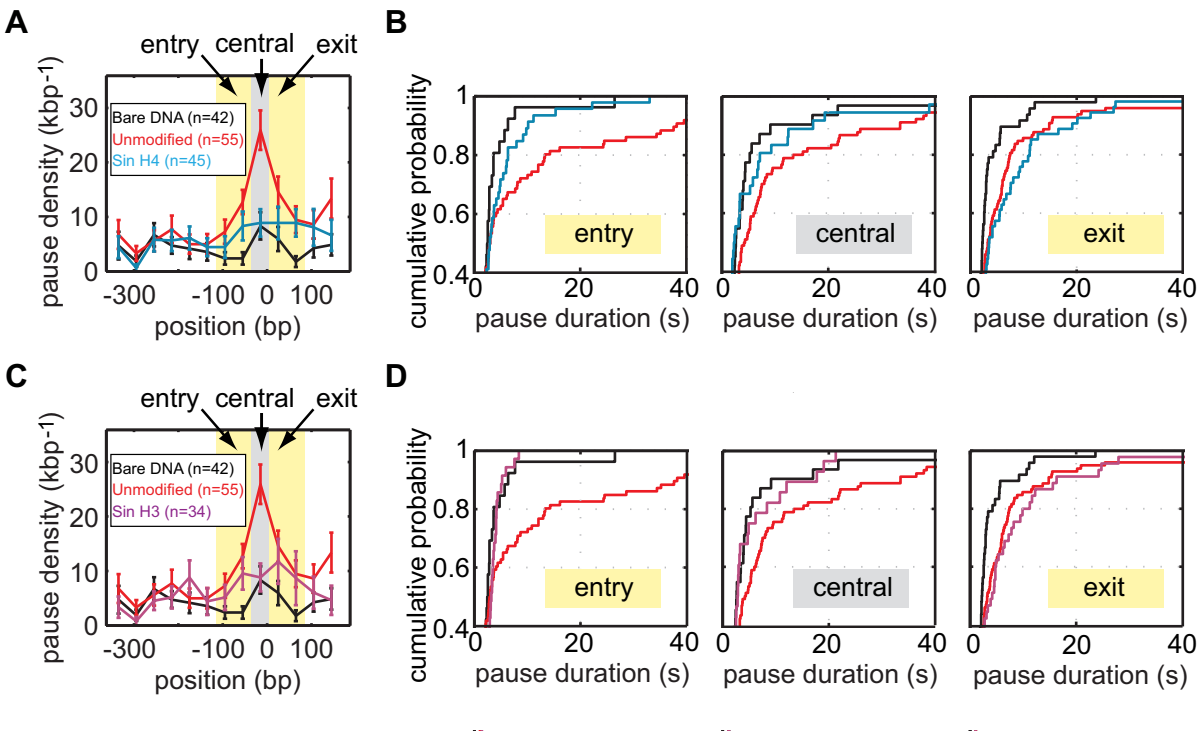

Figure 2.4: **Sin mutants destabilized at the dyad.** (**A** and **C**) Pause density as a function the position of the active center of Pol II on the template for Sin H4 and SinH3 nucleosomes. Error bars represent SEM. (**B** and **D**) Pause durations in the entry, central and exit regions of these nucleosomes.

changes in Pol II pausing in the exit. This result could be explained by the fact that when Pol II reaches the exit region, its leading edge has already passed the DNA that is in the vicinity of the mutated amino acids.

2.5 Direct measurements of nucleosomal wrapping/unwrapping dynamics

Pol II acts as a Brownian ratchet that rectifies the fluctuations of the nucleosome to gain access to the template DNA [54]; thus, it is of interest to establish what changes in nucleosomal dynamics ensue from the nucleosome modifications investigated here. Specifically, we sought to determine how the various modifications alter the nucleosomal wrapping and unwrapping rates and, therefore, the nucleosomal residence in these states. We used the experimental setup shown in Figure 2.5A [93] to monitor the dynamics of nucleosomes under force in the absence of Pol II. As the force is increased, the DNA unwraps from the nucleosome in two steps (Figure 2.5B). The first step, which occurs at low forces, is associated

with the unwrapping of the outer supercoil, and corresponds to the release of DNA from the H2A/H2B dimers. The second step, which takes place at a higher force and is associated with the inner wrap, corresponds to the central DNA coiled around the H3/H4 tetramer. In other words, the outer wrap is related to the entry/exit regions of the nucleosome, whereas the inner wrap is associated with the central region. The unwrapping forces of both steps decrease with increasing ionic strength (Figure 2.5B).

When we performed nucleosome stability experiments in the same buffer as our transcription experiments, the inner region of the nucleosome unwraps and rewraps reversibly at forces between 5 and 8 pN (Figure 2.5C). By maintaining the applied force in this range, we can calculate the rates of nucleosome wrapping and unwrapping (Figure 2.5C). Both Sin mutations decrease the wrapping rates of the nucleosome in the central region (Figure 2.5D; Table 2.2), which explains the reduced overall pausing observed in our transcriptional data. This result also helps explain the observed decrease in the efficiency of upstream histone transfer during transcription through these mutants [56]. As expected, acetylation of histone tails does not lead to significant changes in the wrapping or unwrapping rates of the central region of nucleosomes.

The results obtained with the tailless nucleosomes are perhaps more surprising. We detect an increase of both the unwrapping and the wrapping rates in the central region (Figure 2.5D; Table 2.2), but their ratio, which determines the equilibrium constant of the nucleosome between the two states, is very similar to that of unmodified nucleosomes. This result indicates that the tails affect fluctuations of the nucleosome near the dyad, but do not affect the overall stability of this region, and therefore do not significantly affect transcription.

At the ionic strength used in our transcription experiments (300 mM KCl), we do not observe a clear cooperative transition of the outer wrap (Figure 2.5B). We interpret these observations as indicative that for tailless nucleosomes, at this higher ionic strength, the outer region unwraps readily and irreversibly under the application of force. Therefore, we cannot measure wrapping and unwrapping rates at this ionic strength.

To test the trends in stability of the entry/exit region, we performed nucleosome pulling experiments at 40 mM KCl. Although 91 ± 6% of unmodified nucleosomes show a cooperative unwrapping of the outer wrap, only 56 ± 10% of the acetylated nucleosomes and as little as 13 ± 8% of the tailless nucleosomes display this transition. These results match our transcription observations that the entry region is highly destabilized for tailless and moderately so for acetylated nucleosomes. Only 56 ± 25% of Sin H3 nucleosomes showed an outer wrap, indicating that the effects of this mutation extend to the entry region. In contrast, 70 ± 19% of the Sin H4 nucleosomes showed the outer wrap, which is not significantly different from unmodified nucleosomes.

For the acetylated nucleosomes that showed reversible transitions of the outer wrap, we observed a decrease in the wrapping rate relative to unmodified nucleosomes, in agreement with our transcription results (Table 2.2; Figure 2.5D). For the Sin H3 mutant, the decrease in equilibrium constant observed in the entry region results from an increase in the unwrapping rate. We do not observe any significant changes for the wrapping or unwrapping rates of the outer wrap in the Sin H4 nucleosomes. Even at this low ionic strength we did not see

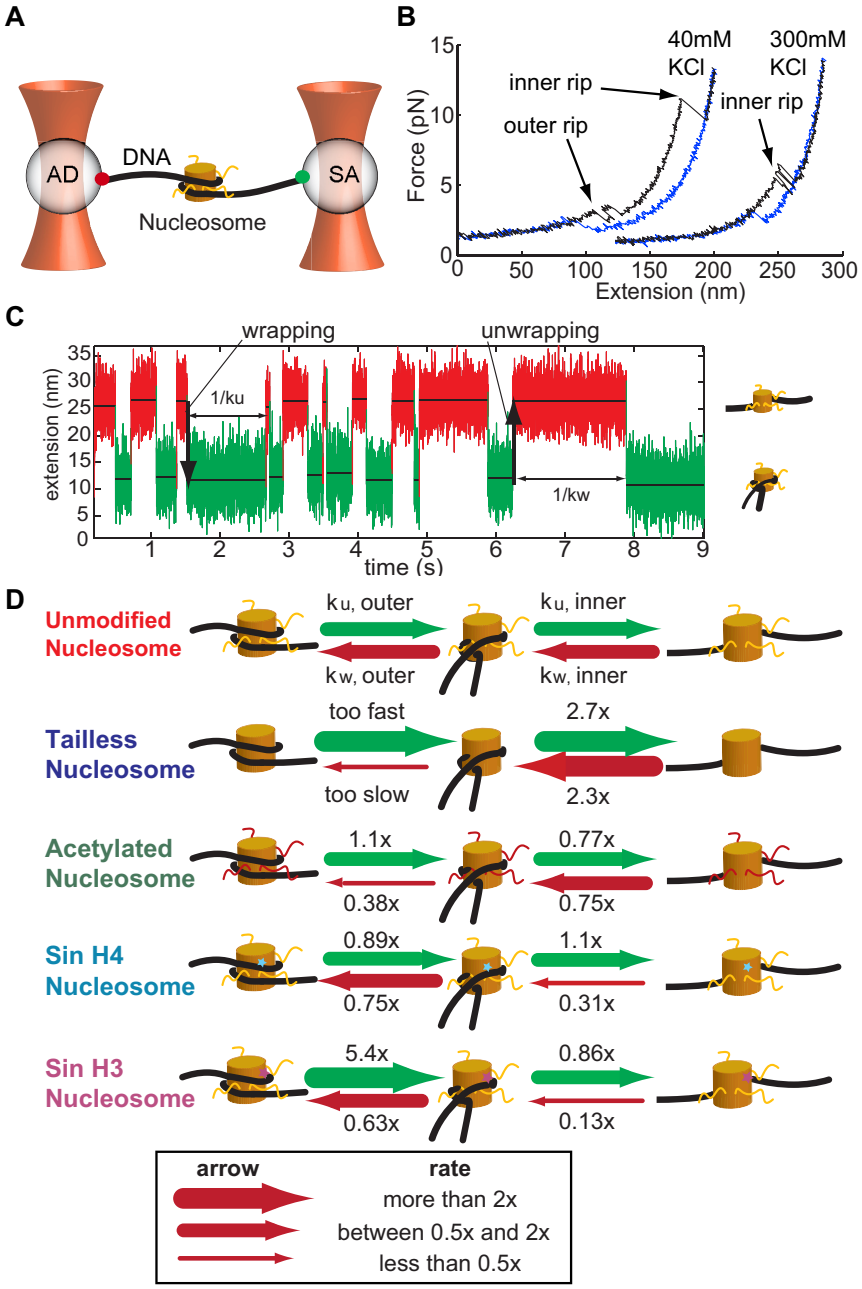

Figure 2.5: **Nucleosome dynamics in the absence of Pol II.** (**A**) Experimental setup for single-molecule nucleosome dynamics experiments. The symbols are the same as in Figure 2.1. (**B**) Force-extension curves of the nucleosome in 40 mM KCl (left) and 300 mM KCl (right). The pulling curve is in black, and the relaxation one is in blue. The unwrapping events corresponding to outer and inner rips are indicated with arrows. (**C**) An example of nucleosome transition events between the wrapped state and the unwrapped one. (**D**) Summary of changes for the nucleosome wrapping and unwrapping rates. For more information about the unwrapping and rewrapping rates of the nucleosome.

any wrapping events of the entry region in the absence of tails, suggesting that although unwrapping happens very fast, wrapping is very slow.

2.6 The template sequence modulates the strength of the nucleosomal barrier

In addition to the histones, the DNA sequence wrapped around the nucleosome greatly influences the barrier [14]. The DNA sequence can affect transcription in two different ways: by increasing the affinity of the DNA for the histones and by directly modulating the tendency of Pol II to pause [71]. We elucidate the importance of the DNA sequence in shaping the nucleosomal barrier by comparing transcription dynamics in different regions on bare and nucleosomal DNA. There is an increase in pause density on bare DNA in the central region of the NPS used in these studies (8 ± 3 per kilobase pairs [kbp^{-1}]) compared with the entry (2 ± 1 kbp^{-1}) and exit regions (4 ± 1 kbp^{-1}) (Figure 2.3A). This increase of pause density also correlates with an increase of pause duration in the central region (6.1 ± 2.2 s) compared to the entry (3.8 ± 1.0 s) and exit (3.5 ± 0.6 s) regions. The pause density at the nucleosome follows the same trend as pausing on bare DNA, displaying a peak in the central region (26 ± 4 kbp^{-1}) versus the entry and exit regions (10 ± 1 kbp^{-1} and 12 ± 2 kbp^{-1} respectively, Figure 2.3A). Pause durations at the nucleosome are also the longest (12.3 ± 2.3 s) in the central region compared with the entry (11.5 ± 2.0 s) and exit (7.0 ± 1.1 s) regions. These data reveal that at certain positions on the DNA template, the transcribing polymerase experiences an increased tendency to pause, accompanied by a slow recovery from the pause. Moreover, the presence of the nucleosome amplifies these trends.

In the backtracking model of transcription [40, 54, 154], a pause involves the forward and backward diffusion of the polymerase on DNA, and it ends when Pol II realigns its active center with the 3′-end of the RNA. If the nascent RNA forms a stable secondary structure outside of the RNA exit channel, it can prevent the polymerase from backtracking (Figure 2.6A). Indeed, we have shown recently that the presence of RNA structure decreases the number of pauses by placing a barrier to backtracking excursions [158].

To test the importance of the nascent RNA structure as a modulator of pausing at the nucleosome, we performed transcription in the presence of RNase A. This enzyme digests single-stranded RNA after U and C residues, and double-stranded RNA; thus, it should inhibit the formation of RNA secondary structure. We observe a large reduction in the probability to pass the nucleosomal barrier, from 58% in the absence of RNase to 34% in its presence (Figure 2.2). In addition, even for polymerases that pass the nucleosome, the frequency and duration of pauses increase in the presence of RNase (Table 2.1). These results support the notion that pausing at the nucleosome is mediated through polymerase backtracking, and suggest that the nascent RNA structure can play a role in preventing nucleosome-induced backtracks and aiding recovery from them.

The only regions of the NPS where we observe changes in the presence of RNase are

Table 2.2: **Nucleosome unwrapping and rewrapping rates in the absence of transcription**

entry region						
nucleosome type	k_u (s^{-1})	N Force (pN)	k_w (s^{-1})	N Force (pN)	K_w	K_w^{Rel}
unmodified	0.46±0.09	221 (F=4.1±0.5)	1.6±0.2	189 (F=3.8±0.4)	3.6±0.9	1.00
tailless	too fast	N/A	too slow	N/A		too small
acetylated	0.52±0.08	406 (F=4.1±1.0)	0.61±0.05	386 (F=3.8±1.0)	1.2±0.2	0.37±0.15
Sin H4	0.41±0.07	164 (F=4.7±1.0)	1.2±0.2	157 (F=4.2±1.0)	3.0±0.7	0.94±0.43
Sin H3	2.5±0.3	344 (F=3.4±0.2)	1.0±0.1	76 (F=3.2±0.4)	0.4±0.1	0.13±0.06
central region						
nucleosome type	k_u (s^{-1})	N Force (pN)	k_w (s^{-1})	N Force (pN)	K_w	K_w^{Rel}
unmodified	0.56±0.1	162 (F=7.4±0.7)	5.2±1	131 (F=6.8±0.4)	9.2±2.5	1.00
tailless	1.5±0.2	547 (F=7.5±0.5)	12±2	254 (F=7.2±0.4)	8.4±1.6	1.07±0.49
acetylated	0.43±0.04	433 (F=7.3±0.8)	3.9±0.8	193 (F=7.1±0.4)	10.0±2.1	1.24±0.56
Sin H4	0.61±0.1	195 (F=6.9±0.9)	1.6±0.6	74 (F=6.9±0.6)	2.6±1.1	0.34±0.22
Sin H3	0.48±0.2	61 (F=7.6±1.1)	0.65±0.5	52 (F=6.8±1.0)	1.4±1.1	0.19±0.18

These results were obtained from hopping experiments performed in the absence of Pol II (as shown in Figure 2.5B). Data were collected in 40 mM KCl for the entry region (top panel) and in 300 mM KCl for the the central region (bottom panel). Values that changed significantly for the modified nucleosomes compared to the unmodified one are shown in red.

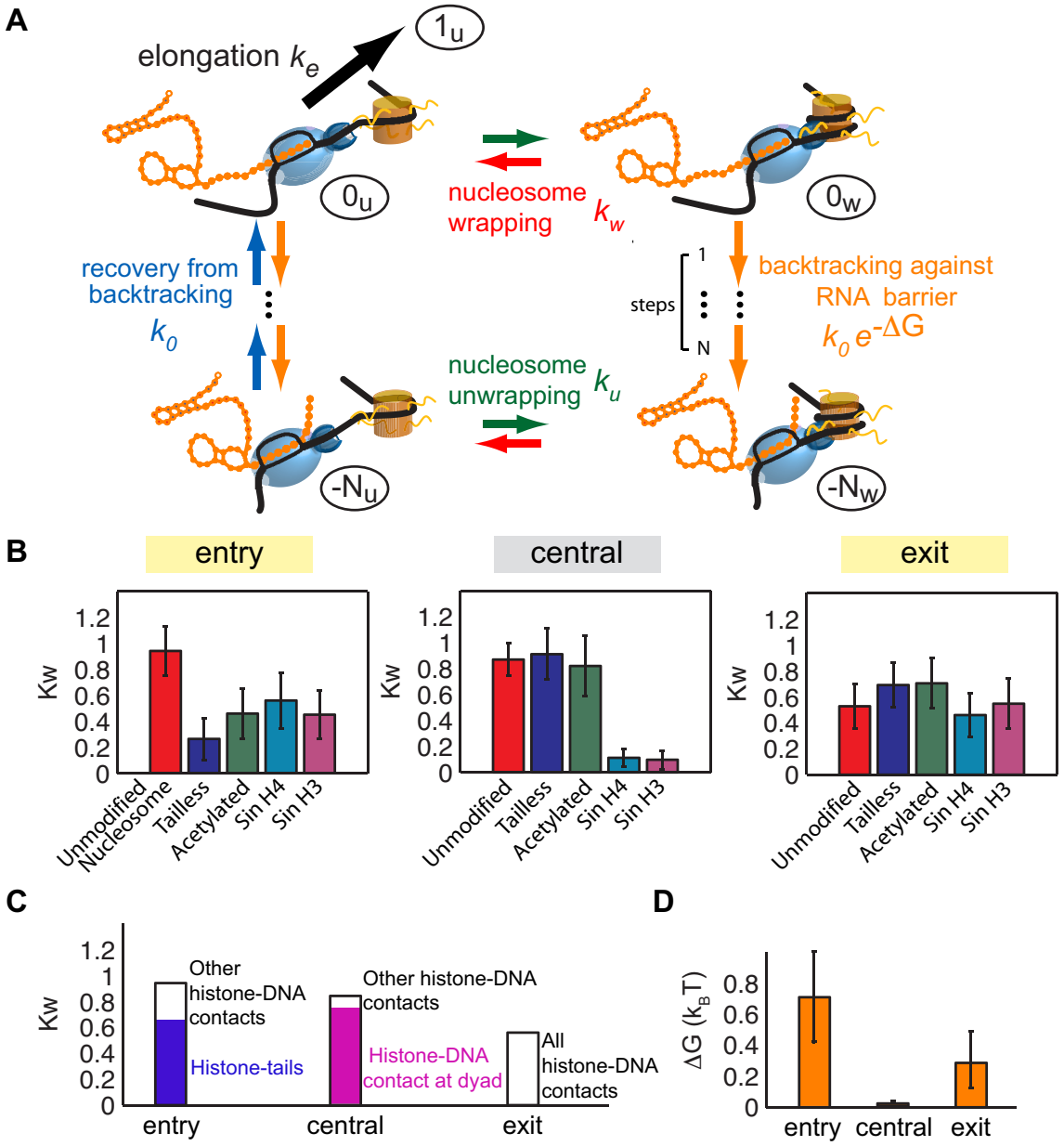

Figure 2.6: **Nucleosome wrapping equilibrium during transcription.** (**A**) Kinetic scheme of transcription through the nucleosome. The parameters involved are: the elongation rate (k_e), intrinsic diffusion rate of the polymerase during a pause (k_0), the barrier to backtracking (ΔG), and the rates of nucleosome unwrapping (k_u), and wrapping (k_w). The labels of the states indicate the number of base pairs Pol II backtracked and the state of the nucleosome: unwrapped (u) or wrapped (w). (**B**) Fitted values of local wrapping equilibrium constant of the nucleosome ($K_w = k_w/k_u$) for the three regions. (**C**) Elements that control the wrapping equilibrium for the three regions. (**D**) Magnitude of the nascent RNA barrier to backtracking in the three regions. Error bars in (B) and (D) represent 95% confidence intervals for the fit of the model parameters to the experimental data.

the entry and exit regions. More polymerases arrest in the entry region (Figure 2.2), and the mean pause duration in the exit region increases from 7.0 ± 1.1 s to 11.5 ± 3.4 s (Figure 2.7). In contrast, pause durations and densities in the central region do not change significantly in the presence of RNase (Figure 2.7). We reason that the increased pausing observed in the central region in the absence of RNase arises from lack of RNA structure behind Pol II. Because elongation competes kinetically with backtracking, the nucleosome amplifies backtracking by preventing access of Pol II to downstream DNA [54].

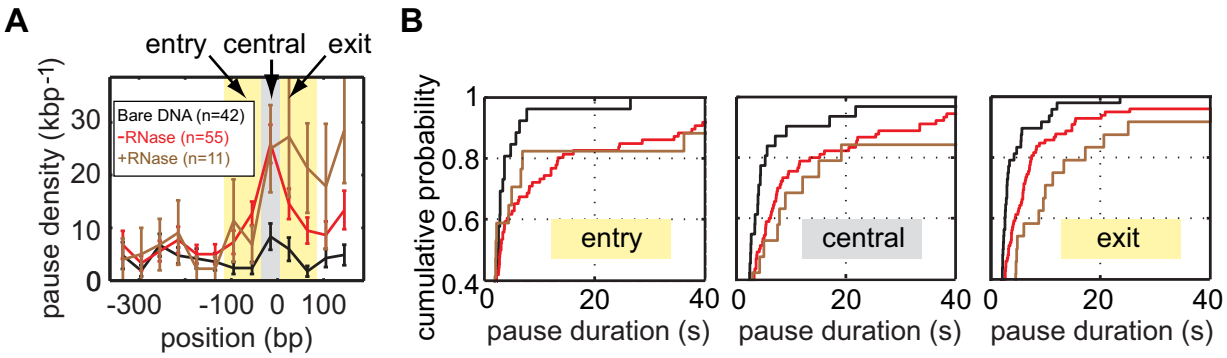

Figure 2.7: **RNase addition affects Pol II pausing.** (**A**) Pause density as a function the position of Pol II's active center on the template for bare DNA, nucleosome with/without RNase. The nucleosome entry/exit regions are shaded yellow, and the central region is shaded gray. (**B**) Pause durations in the entry, central and exit regions of the nucleosome.

2.7 A kinetic model that integrates histone-DNA interactions and sequence effects

We use our experimental data to test and extend our previously developed model of transcription through the nucleosome [54]. Note that the polymerase can only elongate when the nucleosomal DNA immediately in front of it is unwrapped. If nucleosome unwrapping fluctuations were slow, on the same time scale as backtracking, we would have to add these pauses to our pause distribution. However, because the nucleosome fluctuations are fast compared to backtracking [74, 135], pauses due to the nucleosome directly blocking the polymerase are very short (under 0.5 s), and their effect is to reduce the apparent elongation velocity instead of contributing to the measured pause distribution [54].

We extended our previous model to include the effects of different histone-DNA interactions and of the DNA sequence on transcription. The distinct behavior observed in the entry, exit and central regions of the barrier requires us to treat these regions separately. Pausing in each of these regions is affected to different extents and manners by each of the

elements that constitute the nucleosomal barrier: the DNA sequence affects Pol II backtracking through the organization of RNA secondary structure, while the histone tails and histone-DNA contacts modify the nucleosome wrapping/unwrapping equilibrium.

In both models, Pol II is either in an elongation competent state, advancing on DNA at a rate k_e, or it is in a paused backtracked state where it diffuses back-and-forth on DNA with a rate k_0 (Figure 2.6A). The previous model assumes that the landscape over which the elongation complex diffuses is uniform along DNA, and that the polymerase can backtrack unimpeded arbitrarily far. We now modify this model to include the contribution of the template sequence on pausing. Our results indicate that the effect of the sequence is to impede or facilitate backtracking through the organization of more or less RNA secondary structure behind the polymerase, respectively. We model this contribution as an average energy barrier to backtracking, ΔG, associated with each region transcribed [158]. In the absence of any external force, and assuming no RNA structure, the backward and forward diffusion rates of Pol II on DNA during backtracking are assumed equal: $k_b = k_0$ and $k_f = k_0$. The presence of RNA structure behind Pol II only modifies the backward rate (k_b), in a way that reflects the barrier height to breaking this structure: $k_b = k_0 \cdot \exp(\Delta \mathrm{G}/k_B T)$, whereas the forward rate of recovery from backtracks remains the same: $k_f = k_0$. In our experiment, we apply a forward force (F) on the polymerase, so the stepping rates during a pause become:

$$k_b = k_0 \cdot e^{(F(1-d)+\Delta G)/k_B T},$$
$$k_f = k_0 \cdot e^{(Fd/k_B T)},$$

where d is the distance to the transition state for a step (taken here to be 0.5 bp).

The effect of nucleosome fluctuations on transcription can be summarized by a single parameter: the local wrapping equilibrium constant, $K_w = k_w/k_u$, where k_u and k_w are the unwrapping and wrapping rates respectively. Because the polymerase can only recover from backtracks when the nucleosome is unwrapped in front of it, the nucleosome reduces the apparent rate of pause recovery as follows:

$$k_f = \frac{1}{(1 + K_w)} \cdot k_0 \cdot e^{Fd/k_B T},$$

but does not change the backtracking rate k_b. We account for the effects of different modifications through changes in the wrapping/unwrapping equilibrium constant K_w. Using this model, we can compute the predicted mean pause durations and densities for different values of k_0, ΔG, and K_w (Materials and methods), and vary these parameters until all predicted pause durations and densities simultaneously match their measured counterparts. Note that pause durations should only depend on the parameters describing Pol II backtracking (k_0, ΔG) and nucleosome stability (K_w). Pause densities, however, are also affected by the elongation rate of Pol II (k_e), because entry into a pause competes kinetically with elongation. We use the experimentally determined pause-free velocities to estimate the value of k_e (Figures 2.8 and 2.9).

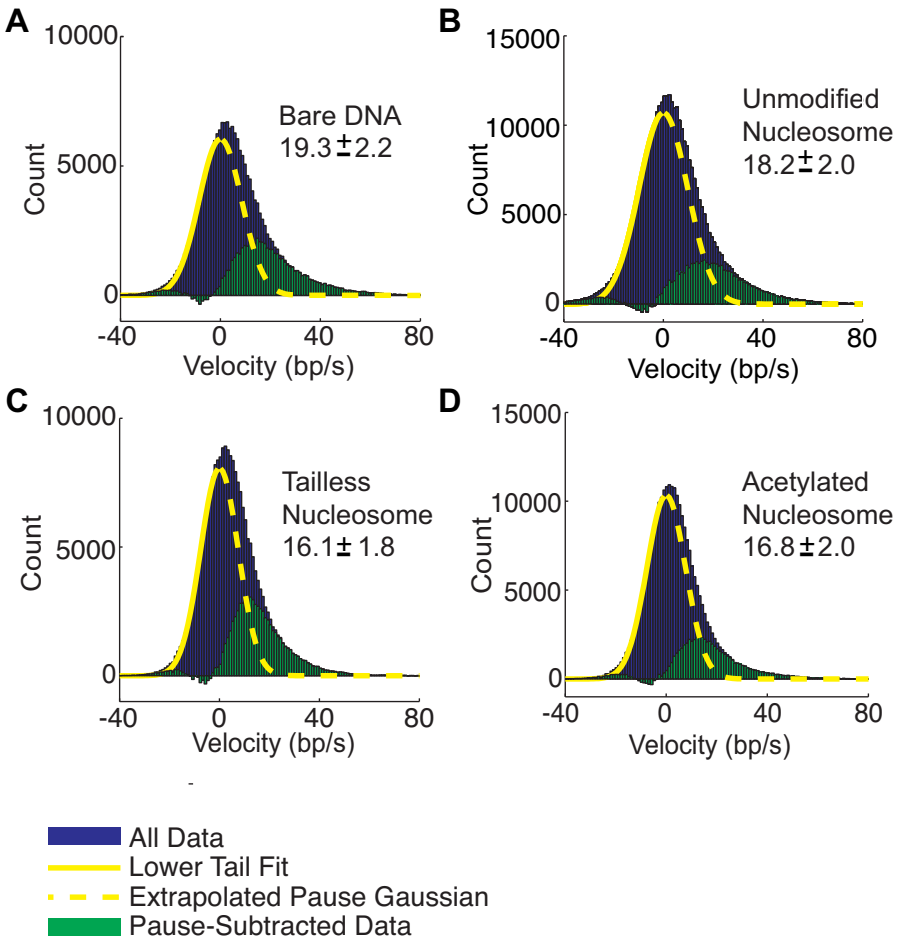

Figure 2.8: **Determination of the pause-free velocity.** (**A**–**D**) Due to limitations in picking very short pauses, we used a pause-independent algorithm to determine the pause-free velocity. For each condition, we first plotted all instantaneous velocities (slopes of the Savitzky-Golay filtered data, blue); these velocities include pauses. Since we reason that pauses represent a peak of velocities centered at zero, we fit the negative velocities with the left side of a Gaussian distribution with mean zero (Lower Tail Fit; solid yellow line). We extrapolate this distribution to the right (Extrapolated Pause Gaussian; dashed yellow line), and subtract the entire Gaussian from the overall data to obtain the pause-free velocities (green). The mean of the pause-subtracted data is shown in each panel and reported in the main text. The error is the standard error of the mean.

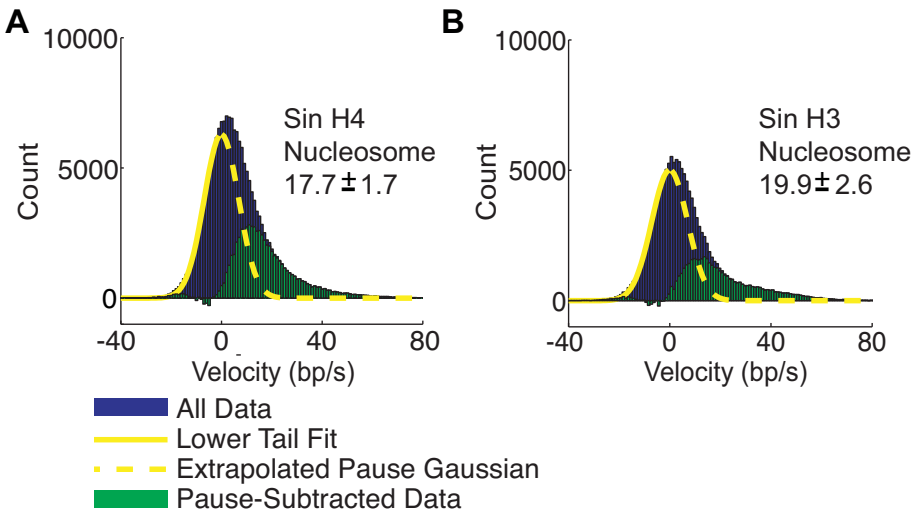

Figure 2.9: **Determination of the pause-free velocity in Sin mutants.** (**A, B**) Due to limitations in picking very short pauses, we used a pause-independent algorithm to determine the pause-free velocity. For each condition, we first plotted all instantaneous velocities (slopes of the Savitzky-Golay filtered data, blue); these velocities include pauses. Since we reason that pauses represent a peak of velocities centered at zero, we fit the negative velocities with the left side of a Gaussian distribution with mean zero (Lower Tail Fit; solid yellow line). We extrapolate this distribution to the right (Extrapolated Pause Gaussian; dashed yellow line), and subtract the entire Gaussian from the overall data to obtain the pause-free velocities (green). The mean of the pause-subtracted data is shown in each panel and reported in the main text. The error is the standard error of the mean.

For each region, we first fit the mean pause durations and densities on bare DNA to obtain a range of possible values for k_0 and ΔG. We extract $k_0 = 0.7 \pm 0.3$ s^{-1} and ΔG $= -0.7 \pm 0.3$ k_BT at the entry, $k_0 = 0.9 \pm 0.5$ s^{-1} and ΔG $= -0.05 \pm 0.05$ k_BT in the central region, and $k_0 = 1.0 \pm 0.5$ s^{-1} and ΔG $= -0.3 \pm 0.2$ k_BT at the exit. As may be expected, the intrinsic value of Pol II diffusion on DNA (k_0) is similar for the three regions of DNA, and the main difference between them is the energy barrier to backtracking (ΔG; Figure 2.6D). In the central region, ΔG is the lowest, as can be expected from the high propensity to pause. This result matches our observation that mean pause densities and durations in the presence of RNase do not change significantly in the central region (Figurefig:ModNucFigureS5), indicating that the absence of RNA structure in this region leads to increased pausing. The average RNA barrier to backtracking in the entry and exit regions is of the same order as the thermal energy, and corresponds to a decrease in the backtracking rate k_b by roughly a factor of 1.5 and 3, respectively, relative to the central region.

Ideally, we would like to directly calculate the energy necessary to unfold the RNA

Table 2.3: **Nucleosome local equilibrium during transcription.**

nucleosome type	entry region		central region		exit region	
	K_w	K_w^{Rel}	K_w	K_w^{Rel}	K_w	K_w^{Rel}
unmodified	0.94±0.19	1.0	0.87±0.12	1.0	0.53±0.17	1.0
tailless	0.26±0.16	0.3±018	0.91±0.20	1.1±0.29	0.69±0.17	1.3±0.53
acetylated	0.46±0.19	0.5±0.23	0.82±0.23	0.9±0.29	0.70±0.19	1.3±0.55
Sin H4	0.56±0.22	0.6±0.26	0.11±0.07	0.1±0.08	0.46±0.17	0.9±0.42
Sin H3	0.45±0.19	0.5±0.22	0.10±0.07	0.1±0.08	0.55±0.20	1.0±0.5

The relative wrapping equilibrium constant, K_w^{Rel}, is computed with respect to the unmodified nucleosome. Values that changed significantly for the modified nucleosomes compared to the unmodified one are in red.

secondary structure formed behind Pol II at each position on DNA. However, simulating the dynamics of folding for RNA sequences longer than 400 bases cotranscriptionally is a difficult computational problem [152]. The difficulties arise because the RNA starts folding as it is being synthesized, allowing for the formation of intermediates that are only locally and not globally stable. Moreover, weak RNA structures previously synthesized can unfold and form stronger structures as new RNA is produced. Aside from these computational difficulties, we note that previous experiments have shown that AT-rich templates lead to more polymerase pausing compared with GC-rich templates, and this difference was attributed to the fact that AU-rich RNAs form weaker secondary structures than GC-rich ones [158]. In agreement with these published results, we find that the RNA that is available for folding while Pol II is transcribing the beginning of the central region is more AU-rich than average. Taking into consideration that there are 29 bases between the active center and the point where the RNA dissociates from the surface of the polymerase [3], the RNA sequence that can fold when the active center of Pol II is in the central region (–35 to +5) corresponds to the DNA transcribed previously, between DNA positions –65 and –25. We find that the beginning of this region (–65 to –55 of the RNA transcript, corresponding to position –35 to –25 of the Pol II's active site) is 82% AU-rich, thus it can only form weak RNA structures behind Pol II. On the other hand, we cannot exclude the possibility that there may be other features in the sequence that contribute to the peak in pause density observed in the central region.

For pausing in the presence of the nucleosome, we keep k_0 and ΔG in the range determined from bare DNA, and fit for the nucleosome wrapping/unwrapping equilibrium constant (K_w), for each region and for each nucleosome modification (Table 2.3).

Table 2.4: **Changes in nucleosome wrapping equilibrium**

	K_w^{Rel} **Entry Region**		K_w^{Rel} **Central Region**	
[KCl] (mM)	**Transcription 300**	**Hopping 40**	**Transcription 300**	**Hopping 300**
Tailless	*0.3 ± 0.18*	*too small*	1.1 ± 0.29	1.1 ± 0.49
Acetylated	*0.5 ± 0.23*	*0.4 ± 0.15*	0.9 ± 0.29	1.2 ± 0.56
Sin H4	0.6 ± 0.26	0.9 ± 0.43	*0.1 ± 0.08*	*0.3 ± 0.22*
Sin H3	*0.5 ± 0.22*	*0.1 ± 0.06*	*0.1 ± 0.08*	*0.2 ± 0.18*

The relative wrapping equilibrium constant, K_w^{Rel}, is computed with respect to unmodified nucleosomes. Values that changed significantly for the modified nucleosomes compared to unmodified ones are in italics. Errors are SEM.

For the unmodified nucleosomes, the local wrapping equilibrium (K_w) does not change significantly between the entry and central regions, indicating that the histone-DNA interactions are uniform along the NPS (Figure 2.6B). The slight decrease in the wrapping equilibrium in the exit region might reflect the fact that in some cases, the histones are removed from the DNA after Pol II passes the nucleosome dyad [54, 80].

The absence of histone tails decreases the local wrapping equilibrium constant by a factor of 3 in the entry region, but does not significantly affect the central region. Mock acetylation only reduces the wrapping equilibrium by a factor of 2 in the entry region. The wrapping equilibrium does not change in the central region (Table 2.4).

The Sin H4 and Sin H3 mutants lead to a dramatic decrease of the wrapping equilibrium in the central region, by approximately a factor of 10. In addition, the Sin H3 mutant decreases the equilibrium in the entry region by a factor of 2. The Sin H4 also has a slight effect on the entry, with a destabilization just under a factor of 2 (Table 2.4). Note that the transcription and mechanical unwrapping values obtained for the entry region of the Sin H3 mutant do not match perfectly (Table 2.4). Although both are decreased compared to unmodified nucleosomes, we see a bigger decrease in the mechanical unwrapping measurements. This discrepancy may reflect the fact that the Sin H3 nucleosome does not readily rewrap once it has been mechanically unwrapped (as can be seen from the low number of rewrapping events, Table 2.2). We hypothesize that the Sin H3 nucleosome falls apart more easily when unwrapped from both sides: once the dimers are unwrapped, the central region contacts (weakest for this mutant) cannot maintain the integrity of the nucleosome. During transcription, because unwrapping takes place only from one side, the contacts between DNA and the distal dimer retain the nucleosome integrity.

Overall, the wrapping equilibrium constants extracted from the transcription data with our kinetic model are in good agreement with those we obtained by mechanical unwrapping of nucleosomes directly in the absence of Pol II (Table 2.4). This agreement shows that our model captures accurately the effect that each nucleosome element has on the different

regions of the nucleosomal barrier (Figure 2.6C).

2.8 Discussion

The single-molecule studies presented here allowed us to quantify the effect of different nucleosomal element—the histone tails, the histone core domains, and the DNA sequence—on transcription. They also revealed the existence of independently controlled spatial domains of the nucleosome: entry, central, and exit regions. The entry region corresponds to the DNA associated with the first H2A/H2B dimer encountered by the transcribing polymerase and ∼20–30 bp of additional DNA ahead of the NPS that are most likely bound by the histone tails. The central region is associated with the H3/H4 tetramer, whereas the exit region spans the last H2A/H2B dimer. These regions are affected differently by the three components of the nucleosomal barrier.

The histone tails mainly gate access into the nucleosomal region. Mock acetylation of biologically important tail lysines also decreases the barrier to transcription in the entry region, although to a smaller extent than tail removal. The small effect of mock acetylation on transcription suggests that in vivo acetylation modulates internucleosomal interactions [148] or creates targets for binding of chromatin remodeling factors to the nucleosome [83] rather than acting as an attenuator of the nucleosomal barrier.

Although removal or acetylation of the tails has a small effect on the overall efficiency of transcription, the state of the tails could be important in regulating access of chromatin remodelers to the nucleosome. For instance, histone chaperones or specific domains of chromatin remodelers could bind and sequester the tails away from the nucleosome core particle, thus opening the gate for others ATP-dependent remodelers. Once bound to the nucleosome, these factors could perturb its structure further. Indeed, this process might be important for the remodeling mechanism of nucleosome remodeling factor (NURF) and SWI/ SNF complexes [134].

Sin mutations greatly affect transcription dynamics in the central region. This destabilization corresponds to a decrease in the wrapping rate between these histones and the surrounding DNA. Surprisingly, we also observe a destabilization of the entry region of the nucleosome, especially for the Sin H3 mutant, both in transcription and nucleosome pulling experiments. These results suggest that this mutation changes the packing of the histones, and affects their ability to organize the DNA into a nucleosome. Indeed, the crystal structure [97] and molecular dynamics simulations for Sin H3 T118H [153] show that the change from threonine to the bulkier histidine leads to a rearrangement of two a helices—belonging to H3 and H4, respectively. The effect of the Sin H3 mutation could be transmitted via these helices throughout the nucleosome to produce the observed changes in transcription kinetics in the entry region in addition to the central one. However, once the polymerase passes the dyad, and disrupts the interactions of the mutated amino acids with the DNA, the opposite ends of the H3 and H4 helices can snap back into place and start interacting with the exit DNA in the same manner as in unmodified nucleosomes. The Sin H3 mutation is also known

to cause loss of hydrogen bonds around H3 N-terminal a helices [97], so it could disrupt the interaction of this tail with DNA, thus borrowing from the tailless phenotype.

Our results with the Sin mutants reveal that DNA contacts with the histones are very important for the stability of the nucleosome, and thus for the barrier to transcription elongation. Disruption of as little as one contact adjacent to the dyad can greatly weaken the barrier, suggesting that these contacts act as central control points for transcription. We speculate that there must exist factors that bind to the nucleosome and disrupt one of these contact points to make the nucleosomal DNA accessible to Pol II and other DNA-translocating motors. The histone chaperone Asf1 (Anti-silencing function 1), which can mediate chromatin disassembly during transcriptional elongation, is a good candidate for employing such a mechanism [36].

We also explore how the sequence of the local DNA, a cis-acting component, contributes to the shape of the nucleosomal barrier. On the template used for these studies (the 601 NPS), we notice an increased tendency to pause and a slow pause recovery in the central region. This trend is mirrored and amplified by the presence of the nucleosome. It has been shown that the stability of the RNA-DNA hybrid, the sequence of the downstream DNA, and the structure of the nascent RNA are important factors in determining sequence-dependent pausing [49, 62, 105]. However, despite important progress in kinetic modeling of transcriptional pausing [9, 123, 44], a consensus has not been reached on how important each of these factors is in determining Pol II pausing. In addition, it has been proposed that transcriptional pauses are associated to backtracking of the polymerase [40, 54, 71, 136]. Using a kinetic model of transcription that incorporates nucleosome fluctuations and polymerase backtracking, we show that the modulation imposed by the sequence can be taken into account as a sequence-dependent barrier to polymerase backtracking. Consistent with these results and recent work on transcription [158], we find that the action of RNase A dramatically increases the probability of arrest at the nucleosome, indicating that the RNA secondary structure acts as a barrier to polymerase backtracking, reducing the enzyme's probability of entering a pause and its average pause time. Because the values of the nucleosome wrapping equilibrium constant (K_w) extracted from the transcription data are similar in the entry and central regions, (Figure 2.6C), we predict that the central barrier would be greatly decreased if the contribution of the sequence (ΔG) were removed.

Note that although here we develop a kinetic model with the minimal number of parameters necessary to explain our observations (backtracking-k_0, RNA barrier to backtracking-ΔG, and nucleosome fluctuations-K_w), our results on the topography of the nucleosomal barrier could also be readily incorporated in other general formalisms of transcription kinetics, such as the one recently developed by Greive *et al.* [44, 45].

By monitoring the dynamics of RNA polymerase II across selectively modified nucleosomes, we have dissected the topography (height and spatial distribution) of the barrier. Three spatially distinct domains arise from this analysis: the entry, the central, and the exit regions. Each of these regions is differentially controlled by the three nucleosomal elements: the histone tails, the histone-DNA contacts, and the local DNA sequence. These results, and the kinetic model derived from these observations, suggest alternative and complemen-

tary mechanisms of control of gene expression in vivo by chromatin remodeling and other transcription factors.

2.9 Materials and methods

2.9.1 Purification and assembly of nucleosomes

Yeast histone proteins containing the deletions or substitutions indicated Table 2.5 were expressed in *Escherichia coli* BL21-codonplus (DE3), purified individually, and assembled into octamers [150]. The octamers were loaded onto a 574 bp DNA containing the 601 NPS [85, 127]. The sequences of the histones used in this study are shown in Table 2.5. (See Appendix A for more details.)

2.9.2 Single-molecule transcription

Biotinylated yeast RNA polymerase II was purified as previously reported [70]. Pol II elongation complexes (ECs) were prepared by sequential annealing of oligos, as previously published [70]. The ECs were then walked to a stall site by uridine triphosphate starvation. The complexes were produced by ligating the upstream end of the ECs to a digoxigenin containing 3-kbp DNA and the downstream end to a nucleosome-containing 574 bp fragment [54]. Single-molecule transcription assays were performed as previously described [54]. Briefly, the stalled ECs were incubated with 2.1 μm streptavidin-coated polystyrene beads (SA beads) (Spherotech), which were trapped using a dual-trap optical tweezers [94]. The upstream end of DNA was attached to a 2.1 μm antidioxigenin IgG-coated polystyrene bead (AD bead). Transcription was resumed by flowing transcription buffer (20 mM Tris-HCl [pH 7.9], 5 mM $MgCl_2$, 5 mM $ZnCl_2$, 300 mM KCl, 1 mM β-mercaptoethanol, 1 mM NTPs, and 1 mM pyrophosphate) into the chamber. Single molecule experiments in the presence of RNase were performed in the same manner as in its absence, except the transcription buffer was supplemented with 6.5 mM of RNase A (Fermentas).

2.9.3 Alignment of single-molecule traces

Even though our accuracy in measuring position changes of Pol II on DNA was high ($\sim$3 base pairs at a bandwidth of 1 Hz), our precision in determining the absolute position of Pol II on DNA was not as good. We noticed for example that our estimate of the final length of the tether before Pol II released the DNA was often slightly lower than the expected run-off length, and in some cases even higher. We attribute this to calibration errors in our experimental setup, since in a biochemical assay where we assay the length of the RNA product by running it in a gel, we obtain a very sharp distribution of run-off lengths. For this reason, we can be confident that for traces that crossed the NPS, the final position should match the run-off position of Pol II from DNA. We use this information to improve

Table 2.5: **Histone sequences used in this study.**

Nucleosome	**Sequence**
Tailless H2A	**MSGGKGGKAGSAAKASQSR**SAKAGLTFPVGRVHRLLRRGNYAQ RIGSGAPVYLTAVLEYLAAEILELAGNAARDNKKTRIIPRHLQLAIRN DDELNKLLGNVTIAQGGVLPNIHQNLL**PKKSAKATKASQEL**
Tailless H2B	**MSAKAEKKPASKAPAEKKPA**AKKTSTSTDGKKRSKARKETYSS YIYKVLKQTHPDTGISQKSMSILNSFVNDIFERIATEASKLAAYNKKST ISAREIQTAVRLILPGELAKHAVSEGTRAVTKY**SSSTQA**
Tailless H3	**MARTKQTARKSTGGKAPRKQLASKAAR**KSAPSTGGVKKPHR YKPGTVALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAI GALQESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLAR**RLRGERS**
Tailless H4	**MSGRGKGGKGLGKGGAKRHR**KILRDNIQGITKPAIRRLARRGG VKRISGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYAL KRQGRTLYGFGG
Acetylated H2A	MSGG**Q**GG**Q**AGSAAKASQSRSAKAGLTFPVGRVHRLLRRGNYAQRIG SGAPVYLTAVLEYLAAEILELAGNAARDNKKTRIIPRHLQLAIRNDDEL NKLLGNVTIAQGGVLPNIHQNLLPKKSAKATKASQEL
Acetylated H2B	MSAKAEKKPAS**Q**APAE**Q**KPAAKKTSTSTDGKKRSKARKETYSSYIY KVLKQTHPDTGISQKSMSILNSFVNDIFERIATEASKLAAYNKKSTISA REIQTAVRLILPGELAKHAVSEGTRAVTKYSSSTQA
Acetylated H3	MARTKQTAR**Q**STGG**Q**APR**Q**QLAS**Q**AARKSAPSTGGVKKPHRYKPG TVALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQ ESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLARRLRGERS
Acetylated H4	MSGRG**Q**GG**Q**GLG**Q**GGA**Q**RHR**Q**ILRDNIQGITKPAIRRLARRGGVKR ISGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYALKRQ GRTLYGFGG
Sin H4	MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGGVK**A**I SGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYALKRQ GRTLYGFGG
Sin H3	MARTKQTARQSTGGQAPRQQLASQAARKSAPSTGGVKKPHRYKPG TVALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQ ESVEAYLVSLFEDTNLAAIHAKRV**H**IQKKDIKLARRLRGERS

Deleted or modified histone residues are shown in boldface. Tailless histones were obtained by expressing truncated sequences [108, 144]. In mock-acetylated histones, all lysine residues known to be acetylated in vivo were substituted with glutamines [106, 119].

our positional accuracy. We perform a two-point calibration of all traces that transcribed past the NPS by aligning them both to the stall site and to the expected run-off length, as shown in Figure 2.10. This improved alignment of traces allows us to see position-dependent trends of transcriptional pausing, such as sequence dependent pausing much better than in our previously published results [54].

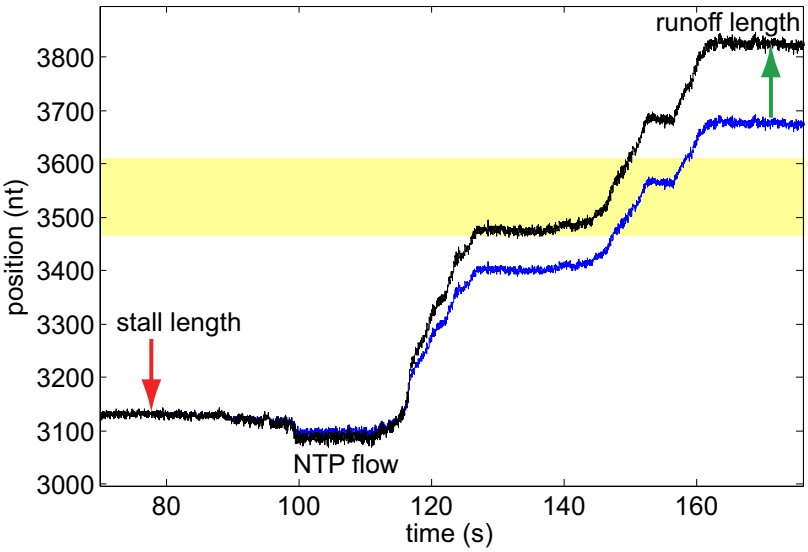

Figure 2.10: **Alignment of traces for improved precision.** An example trace is shown before (blue) and after (black) alignment. The yellow shading between 3464 and 3611 corresponds to the NPS region. Position 3538 bp corresponds to the nucleosome dyad. All trajectories were aligned such that their start position before the flow of NTPs is equal to the position of the stall site (3131 bp, red arrow). Only traces that transcribed past position 3650 bp were in addition scaled linearly such that the end position of Pol II on DNA corresponds to the expected run-off length (3824 bp, green arrow). The discrepancy between the pre and post-alignment data presented in this figure is the highest we have detected; in general the difference between the end of the trace and the expected run-off is 15-20 bp.

2.9.4 Pause analysis

Changes in position of Pol II on the DNA template were recorded at 2 kHz, averaged by decimation to 50 Hz, and then smoothed using a second-order Savitzky-Golay filter to 1 Hz. To identify pauses, we first divided the position versus time data into 3 bp bins, which is the limit of our resolution, and compute the dwell time Pol II spends in each of these bins. Pauses were defined as dwell times that were at least 1.5 times longer than the average dwell time for each trace. The pause threshold varied from trace to trace, but it was lower than 0.5 s in the majority of the traces.

2.9.5 Analysis of nucleosome wrapping/unwrapping events

For nucleosome stability measurement, we loaded histone octamers on a 2,964 bp DNA fragment containing the 601 NPS that was obtained by PCR from a modified pUC19 plasmid [162] using primers containing biotin and digoxigenin, respectively (IDT). These modifications allowed formation of a DNA tether containing a single nucleosome between the SA and AD beads held in optical traps.

In order to study the inner and outer wraps unfolding under force, once a tether was formed, the force was increased by stepping one of the traps by 2 nm every 60 ms [93].

To measure the nucleosome wrapping and unwrapping rates, the two beads were held at constant positions for 1 min. The force was approximately constant at 5 pN for the inner wrap at 40 mM KCl and 7 pN for the outer wrap at 300 mM KCl (Table 2.2). The nucleosome fluctuations at data were collected at 2 kHz, and averaged by decimation to 1 kHz (inner wrap) or 100 Hz (outer wrap). We found the wrapping and unwrapping transitions (Figure 2.5C) by running a t test analysis on these data, as described previously [22, 96].

As specified earlier, the tailless nucleosomes did not display hopping of the entry/exit region. We probed the mechanical stability of this region using the force-extension curves shown in Figure 2.5B instead. We computed the probability of seeing a distinct unwrapping event of the entry/exit region for each type of nucleosome by counting the number of times we observed an unwrapping event for the entry/exit region divided by the total number of trials. The errors reported in the main text for the probability of observing the entry/exit region reflect the width of the 95% interval, as estimated from the binomial distribution.

2.9.6 Kinetic analysis of pausing

Using the kinetic scheme shown in Figure 2.6A, we can compute the expected mean pause duration and pause densities as a function of the parameters involve [54]. Note that for computing the distribution of pause durations from this model, we assume that local nucleosome wrapping and unwrapping is much faster than polymerase backtracking (0.3–2 s^{-1}). The assumption that local nucleosome fluctuations are fast and reach equilibrium between consecutive backtracking steps is consistent with nucleosome fluctuations measured using fluorescence correlation spectroscopy [84] and FRET [74]: $\sim$4 s^{-1} for unwrapping and $\sim$20–90 s^{-1} for rewrapping. In addition, recent simulation results predicted local fluctuation between DNA and histones to be in the micro seconds range [135]. Because of this different time scales, short nucleosome fluctuations do not contribute to the distribution of pauses longer than 0.5 s, but instead they reduce the apparent pause-free velocity of the polymerase.

Therefore, both on bare DNA and at the nucleosome, the probability of observing a pause of duration t is that of a one-dimensional random walk starting at position –1 and returning to position 0:

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \frac{exp[-(k_f + k_b)t]}{t} I_1(2t\sqrt{k_f k_b}) \tag{2.1}$$

where I_1 is the modified Bessel function of the first kind, and k_b and k_f are the rates of backtracking and recovering from a backtrack respectively. However, these two rates associated with the movement of the polymerase during backtracks are different during transcription on bare DNA compared to transcription at the nucleosome.

On bare DNA $k_f = k_0 \cdot e^{F \cdot d/k_B T}$ and $k_b = k_0 \cdot e^{-F \cdot ((1-d)+\Delta G)/k_B T}$, where k_0 is the intrinsic diffusion of a backtracked Pol II on DNA, ΔG is the barrier to backtracking, F is the force we apply in our experiment (6.5 $\pm$ 1.5 pN), and d is the distance to transition (taken here to be 0.5 bp).

At the nucleosome $k_f = (k_u/k_u + k_w)k_0 \cdot e^{F \cdot d/k_B T}$ and $k_b = k_0 \cdot e^{-F \cdot ((1-d)+\Delta G)/k_B T}$, where k_u and k_w are the nucleosome unwrapping and rewrapping rates respectively. The forward rate of recovery from backtracks only depends on the ratio of the two nucleosomal rates, $K_w = k_w/k_u$, as follows: $k_f = (1/1 + K_w)k_0 \cdot e^{F \cdot d/k_B T}$.

Substituting the backtrack (k_b) and recovery (k_f) rates into Equation 5.3, we can obtain the pause durations distributions on bare DNA, $\psi_{bare}(t, k_0, \Delta G)$, and at the nucleosome $\psi_{nucl}(t, k_0, \Delta G, K_w)$.

Once we know the distribution of pauses $\psi(t)$, the mean pause duration of pauses between t_1 and t_2 can be computed:

$$< t >= \frac{\int_{t_1}^{t_2} t \cdot \psi(t) dt}{\int_{t_1}^{t_2} \psi(t) dt}$$

For fitting the average pause durations, we took $t_1 = 1$ s and $t_2 = 300$ s. Pause densities (number of pauses per base pair) can be theoretically computed as the probability of entering a pause multiplied with the probability that the pause is in our observation window (i.e., between t_1 and t_2). On bare DNA, the mean pause density is:

$$PD_{bareDNA} = \frac{k_b}{k_b + k_e} \int_{t_1}^{t_2} \psi_{bare}(t) dt$$

where k_e is the elongation rate, taken to be equal to the pause free-velocity. At the nucleosome, since Pol II can only elongate when the nucleosome is unwrapped, the elongation rate k_e gets multiplied with the probability of finding the next base pair unwrapped:

$$PD_{nucleosome} = \frac{k_b}{k_b + \frac{1}{1+K_w} k_e} \int_{t_1}^{t_2} \psi_{nucl}(t) dt$$

For fitting the average pause densities, we took $t_1 = 3$ s and $t_2 = 300$ s. For each region of the NPS (entry, central and exit), we first determine values of k_0 and ΔG for bare DNA by imposing that both the predicted mean pause durations and densities are within error of the experimental ones. These values of k_0 and ΔG are reported in the main text.

In order to fit the nucleosome rewrapping equilibrium constant for each type of nucleosome in a given region of the NPS, we vary K_w (while keeping k_0 and ΔG in the interval determined from bare DNA in that region) until the predicted pause durations and densities match the experimentally determined values These values of K_w are reported in Table 2.3.

Chapter 3

Dissection of Transcription Elongation Kinetics

During the elongation stage of transcription, RNA polymerase translocates along the DNA as it incorporates template-complementary nucleotides into a growing RNA chain, one nucleotide at a time. Many ensemble and single-molecule studies have shown that transcription elongation has two phases: active translocation and pausing. This chapter describes the characterization of molecular mechanisms of the elongation using single-molecule optical tweezers assays.

3.1 Mechanochemical coupling of the nucleotide addition cycle

Two classes of mechanisms have been offered to describe the mechanochemical coupling during the nucleotide addition cycle phase of transcription elongation. The first class, known as the power stroke mechanism, gains support from studies of single-subunit T7 RNA polymerase [157]. This model suggests that the forward translocation of RNAP is directly driven by a chemical step such as the release of the pyrophosphate (PPi). However, this mechanism cannot explain the backtracking behavior, typically observed in multi-subunit RNA polymerases. In the second class, the Brownian ratchet mechanism (Figure 3.1), the motion of the motor is directly driven by thermal fluctuations and rectified by chemical steps [143]. This mechanism postulates that at the beginning of each nucleotide addition cycle, the transcription elongation complex (TEC) oscillates back and forth on the DNA template between two translocated states [a pre-translocated, ($TEC_{n,0}$), and a post-translocated state

Portions of this chapter were published in Dangkulwanich M., Ishibashi T., Liu, S., Kireeva M.L., Lubkowska L., Kashlev M. and Bustamante C.J. "Complete dissection of transcription elongation reveals slow translocation of RNA polymerase II in a linear ratchet mechanism." *eLife*; 2:e00971. 2013. DOI:10.7554/eLife.00971. Used with permission.

($TEC_{n,1}$)] and that such thermally-driven motions are rectified forward by the incorporation of the incoming NTP [46]. Extensive structural and biochemical investigations have supported this Brownian ratchet mechanism for the translocation of multi-subunit RNAPs, including bacterial and eukaryotic enzymes [9, 11, 16, 73].

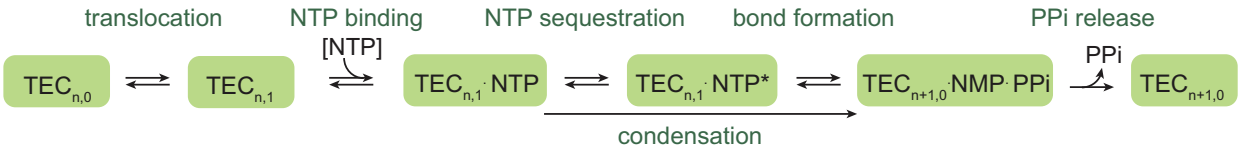

Figure 3.1: **Generalized scheme of the nucleotide addition cycle.** At the beginning of a nucleotide addition cycle, the transcription elongation complex (TEC) with a transcript length of n thermally fluctuates between the pre-translocated state ($TEC_{n,0}$) and the post-translocated state ($TEC_{n,1}$). NTP binding is followed by NTP sequestration, bond formation, and PPi release. Upon the release of the PPi, TEC is reset to the pre-translocated state ($TEC_{n+1,0}$) and ready for the next nucleotide addition cycle.

3.2 Evidence for the Brownian ratchet mechanism

As mentioned earlier, single-molecule experiments allow for the measurement and separation of the active elongation phase from the pausing phase. The pause-free velocity of the enzyme (v) is given by the distance the polymerase has to translocate during one cycle (d), divided by the total time it takes to complete the nucleotide addition cycle. This time is the sum of the times necessary to translocate (τ_{trans}), bind the incoming NTP (τ_{NTP}), complete the condensation reaction that incorporates the NTP to the RNA chain (τ_{cond}), and release the pyrophosphate before starting a new cycle (τ_{PPi}):

$$v = \frac{d}{\tau_{trans} + \tau_{NTP} + \tau_{cond} + \tau_{PPi}} \tag{3.1}$$

If translocation is driven by thermal noise and biased forward by NTP binding followed by the irreversible condensation reaction (Brownian ratchet), τ_{NTP} should be sensitive to force, since that is the step associated with net movement on DNA. Instead, if pyrophosphate release induces —or coincides with—a change in conformation of the elongation complex, triggering translocation (power stroke), τ_{PPi} would be sensitive to force instead. At limiting NTP concentrations, the time it takes to bind NTPs (τ_{NTP}) becomes dominant over the time of pyrophosphate release (τ_{PPi}). In these conditions, if elongation follows a Brownian ratchet mechanism, the velocity should be sensitive to force, while if it follows a power stroke, the velocity should not depend on force. Single-molecule data have shown that at low NTPs, the pause-free velocity does depend on force [1]. This result is, thus, inconsistent with the power stroke mechanism, and supports the Brownian ratchet mechanism for transcription elongation.

The simplest reaction pathway for a Brownian ratchet model involves only one NTP binding site and two translocation states of the polymerase [pre-translocated state ($TEC_{n,0}$) and the post-translocated state ($TEC_{n,1}$), Figure 3.1]. In order to simplify the fittings and analyses, one common assumption that most studies up to date have made is that the rates of translocation and nucleotide binding are much faster than the rate of NTP catalysis. When using this simplified assumption and the linear model (Figure 3.1) to fit the relationship between the elongation velocity and external force applied obtained from single-molecule experiments, the classical linear ratchet mechanism (Figure 3.1) had to be modified such that the incoming NTP must also bind to the pre-translocated TEC (Figure 3.2) [1, 82]. While in the linear model the incoming NTP can only bind after the polymerase has translocated, in the branched model the NTP can bind to both the pre- and the post-translocated states. However, in the pre-translocated TEC, the primary nucleotide binding site is occupied by the 3′-end of the nascent transcript [42]. Thus, the branched Brownian ratchet scheme necessarily requires a secondary NTP binding site, whose precise location and the mechanism by which the NTP is transferred to the primary site remain poorly defined.

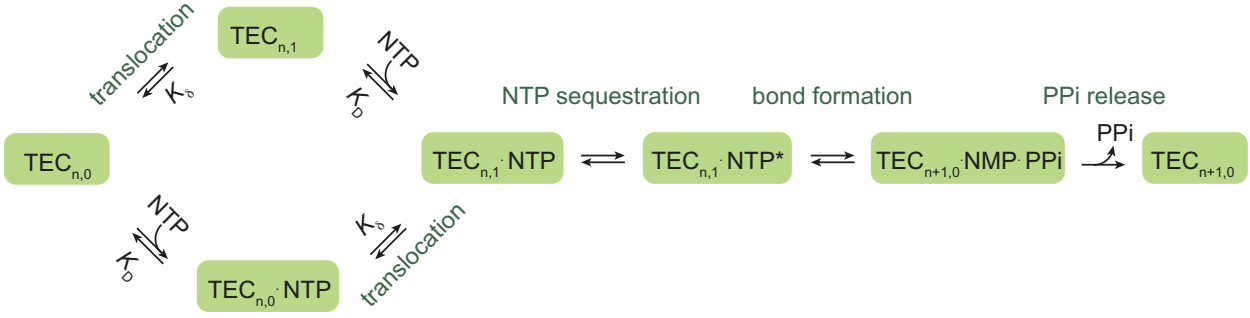

Figure 3.2: **A branched Brownian ratchet model for the nucleotide addition cycle.**

In order to understand the mechanism of transcription and its regulation, it is important to achieve a detailed description of both on- and off-pathway kinetics of the elongation reaction. Previous efforts to dissect the kinetic scheme of transcription elongation have assumed that the forward and reverse translocation steps of the Brownian ratchet occur in rapid equilibrium relative to the chemical steps in the nucleotide addition cycle [1, 9, 46, 123]. However, the assumption of fast translocation equilibrium has never been experimentally validated. In fact, recent studies suggested that the translocation step may be partially rate-limiting for the nucleotide addition cycle, which gives rise to the heterogeneous elongation rates at different template positions [57, 67, 90, 91, 98, 99].

Rather than assuming that the translocation rate of the polymerase is fast, we questioned the validity of the common assumption that the transition rates from the pre- to post-translocated states are much larger than those of other kinetic steps. Whereas the application of external force affects both the forward and the reverse translocation rates of the enzyme, a mechanical barrier such as a nucleosome only affects the forward rate, making it possible to separate it from the reverse rate. (See section 3.8 for details).

3.3 Transcriptional pausing

Another important component of transcription elongation is the pausing phase, which is an off-pathway process that plays crucial roles in the regulation of transcription elongation (Figure 3.3) [81, 102]. The mechanism of transcriptional pausing is still under debate. In one view, RNAP first enters an elemental pause state [52, 121, 129]. These elemental pauses can be subsequently stabilized into longer-lived pauses by the formation of a hairpin structure in the nascent RNA transcript or by RNAP backtracking [6, 53]. The backtracking process is caused by upstream movements of the polymerase, displacing the 3′-end of the nascent RNA away from the active site into the secondary channel of the enzyme [73, 103]. An alternative view poses that most pauses are attributed to backtracking, which can be described as a one-dimensional random walk of the enzyme along the DNA template [33, 40, 54, 92]. RNA synthesis resumes when the polymerase diffusively realigns its active site with the 3′-end of the transcript. Structural evidence for the existence of elemental pauses was recently presented in bacterial RNAP [146], but similar evidence is lacking for the eukaryotic polymerases.

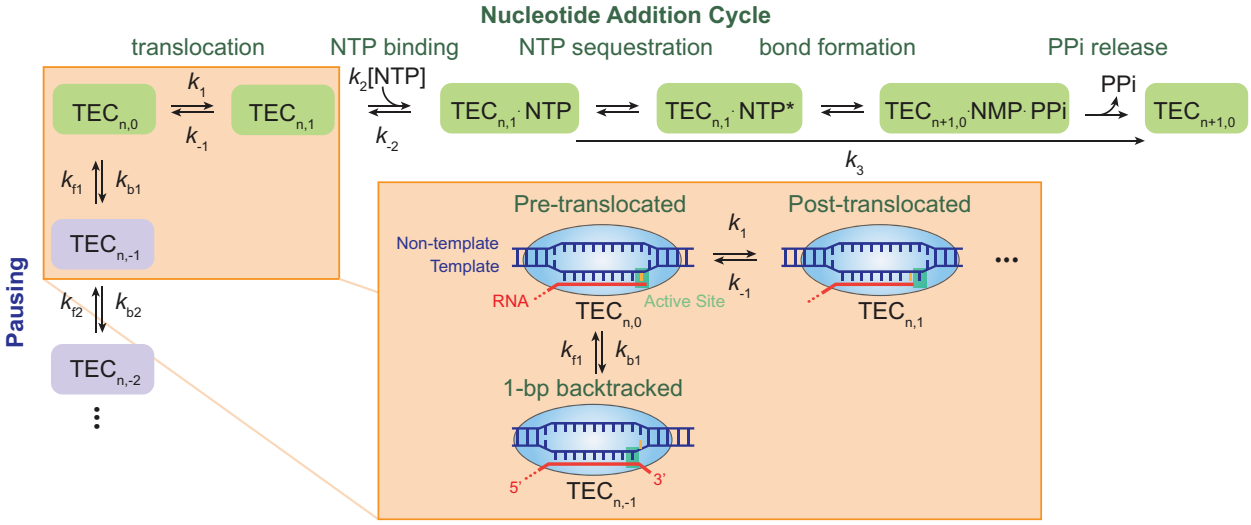

Figure 3.3: **Nucleotide addition cycle and off-pathway pausing of transcription elongation.** The nucleotide addition phase and the pausing phase are colored in green and blue, respectively. From the pre-translocated state, the polymerase can also enter the off-pathway pausing phase by backtracking. The pausing kinetics are determined by the backward stepping rate constants k_{bn} and forward stepping rate constants k_{fn}, where n denotes the number of bp backtracked. The inset shows cartoon configurations of the TEC in a pre-, a post-translocated, and a 1-bp backtracked state.

3.4 Derivation of elongation velocity equations

The Brownian ratchet kinetic scheme for the nucleotide addition cycle of transcription elongation (Figure 3.1) can be simplified to:

$$\mathrm{TEC_{n,0}} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{TEC_{n,1}} \underset{k_{-2}}{\overset{k_2[\mathrm{NTP}]}{\rightleftharpoons}} \mathrm{TEC_{n,1\text{–}NTP}} \xrightarrow{k_3} \mathrm{TEC_{n+1,0}}$$

where k_1 and k_{-1} are the forward and reverse translocation rate constants, k_2 and k_{-2} are the NTP binding and dissociation rate constants, and k_3 is the combined catalysis rate constant that includes NTP sequestration, bond formation, and PPi release. Because of the large equilibrium constant of transcription elongation and the very low PPi concentration (1 μM) in the buffer, k_3 was considered essentially irreversible. Using the concept of net rate constants [29], we can replace the reversible rate constants between two adjacent states with a single net rate constant and re-write the above scheme as:

$$\mathrm{TEC_{n,0}} \xrightarrow{k_1^{\mathrm{net}}} \mathrm{TEC_{n,1}} \xrightarrow{k_2^{\mathrm{net}}} \mathrm{TEC_{n,1\text{–}NTP}} \xrightarrow{k_3} \mathrm{TEC_{n+1,0}}$$

k_1^{net} and k_2^{net} are the net rate constants for translocation and NTP binding, respectively, which are given by:

$$k_2^{net} = k_2[NTP] \cdot \frac{k_3}{(k_{-2} + k_3)} \tag{3.2}$$

$$k_1^{net} = k_1 \cdot \frac{k_2^{net}}{(k_{-1} + k_2^{net})} = \frac{k_1 k_2 k_3 [NTP]}{k_{-1}(k_{-2} + k_3) + k_2 k_3 [NTP]} \tag{3.3}$$

The time the enzyme takes to finish one nucleotide addition cycle (τ) equals the step size of the polymerase ($d = 1$ nt) divided by the pause-free velocity (v), and also equals the sum of the inverse of each net rate:

$$\tau = \frac{d}{v} = \frac{1}{k_1^{net}} + \frac{1}{k_2^{net}} + \frac{1}{k_3} \tag{3.4}$$

Plugging Equations 3.2 and 3.3 into Equation 3.4 yields the following expression for the pause-free velocity:

$$v = \frac{k_1 k_3}{k_1 + k_3} \cdot \frac{[NTP]}{\frac{(k_1 + k_{-1}) \cdot (k_{-2} + k_3)}{(k_1 + k_3) \cdot k_2} + [NTP]} \cdot d \tag{3.5}$$

We note that this expression is more general than those shown in previous studies [1, 10], as it is derived without assuming local equilibration of translocation and NTP binding. In particular, we describe the kinetics of the translocation step with k_1 and k_{-1}, instead of a single equilibrium constant $K_\delta = k_{-1}/k_1$. Such treatment is a prerequisite to explicitly determine the forward and reverse translocation rates. Equation 3.5 can be simplified to the Michaelis-Menten equation form:

$$v = \frac{V_{max}[NTP]}{K_M + [NTP]} \tag{3.6}$$

where $V_{max} = \frac{k_1 k_3}{k_1 + k_3} \cdot d$, and $K_M = \frac{k_1 + k_{-1}}{k_1 + k_3} \cdot \frac{k_{-2} + k_3}{k_2}$.

3.5 Single-molecule transcription assay

We followed the transcriptional dynamics of individual Pol II molecules with a dual-trap optical tweezers instrument. One laser trap holds a polystyrene bead attached to a stalled Pol II molecule, while the other trap holds another bead attached to the upstream DNA template (assisting force geometry; Figure 3.4A). The experiment was carried out in a semi-passive mode under 5-8 pN of applied tension. The experiment began by holding the complex at 8 pN. Upon addition of NTP, Pol II restarts transcription, lengthening the DNA tether length and thereby a decreasing the force applied to Pol II. When the force reaches 5 pN, the trap was moved further apart to increase the tension to 8 pN. The position of the RNA polymerase along the template was calculated from the extension and the force using the worm-like-chain model of DNA elasticity [19]. As shown in the example traces (Figure 3.4B), the trajectories can be divided into two parts: the active elongation phase, where the RNAP moves rapidly along the DNA, and the pausing phase, where the RNAP appears to pause and even translocate backward.

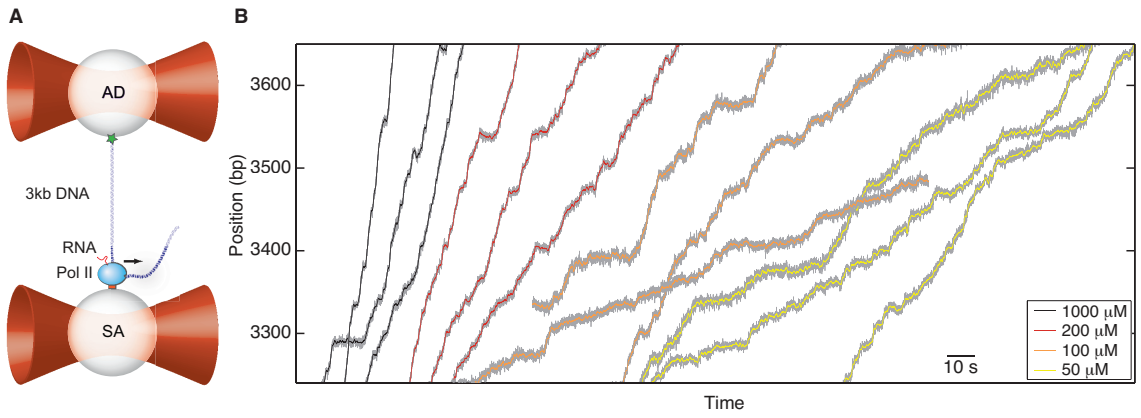

Figure 3.4: **Single-molecule transcription assay.** (**A**) Experimental setup for the single-molecule transcription assay. Each of the two optical traps holds a 2.1-μm polystyrene bead. Biotinylated Pol II is attached to the streptavidin (SA) bead. The upstream DNA is attached to the antibody (AD) bead via the digoxigenin-antidigoxigenin linkage. The black arrow indicates the direction of transcription. (**B**) Example transcription trajectories of the wild-type Pol II at four different NTP concentrations.

3.6 NTP dependence of elongation dynamics

Single-molecule transcription trajectories were collected at a range of NTP concentrations (35 μM to 2 mM) (Figure 3.4B). The relationship between pause-free velocity (v) and [NTP] for the wild-type enzyme fits well to Equation 3.6, with $V_{max} = 25 \pm 3$ nt/s and $K_M = 39 \pm 12$ μM (errors are SEM) (Figure 3.5A, gray line).

As shown in the example trajectories (Figures 3.4B), transcription elongation is punctuated by pauses of various durations. Pause density, ρ_{pause}, is defined as the average number of pauses per bp of template transcribed. As the concentration of NTP goes up, the pause-free velocity increases and the apparent ρ_{pause}, which counts pauses lasting longer than 1 s, decreases (Figure 3.5B). The inverse relationship between v and (ρ_{pause}) indicates that elongation and pausing are in kinetic competition and that pausing occurs prior to NTP binding [6, 32, 38, 81, 92]. Note that pausing has also been observed to occur after NTP binding at certain sequences for *E. coli* RNAP; however, yeast Pol II does not seem to employ such mechanism [68]. The pause-free velocities and apparent pause densities at various NTP concentrations are summarized in Table 3.1.

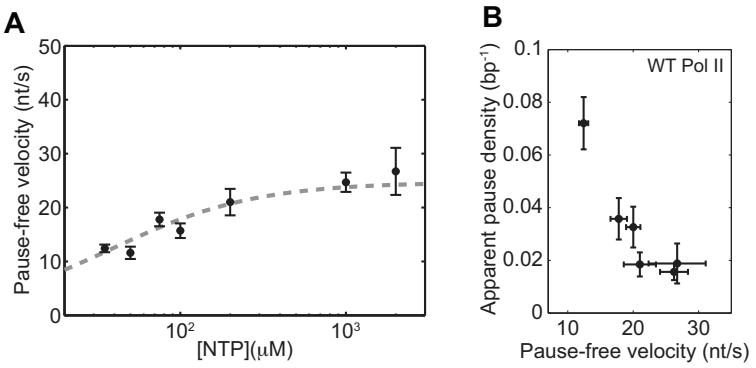

Figure 3.5: **Pause-free velocities and apparent pause densities.** (**A**) Pause-free velocities of the wild-type Pol II at various NTP concentrations. Dashed lines are fits to the Michaelis-Menten equation (Equation 3.6; $R^2 = 0.80$). (**B**) The apparent pause densities (ρ_{pause}) of the wild-type Pol II at different NTP concentrations are plotted against the corresponding pause-free velocities (v). Error bars in both figures represent standard error of the mean (SEM).

3.7 Stepping rates during a backtracked pause

Backtracking is a major mechanism for transcriptional pauses. We have previously modeled backtracking as a one-dimensional random walk of the enzyme along the DNA template [54]. In this model, Pol II diffuses back and forth on DNA with a forward stepping rate constant k_f and a backward stepping rate constant k_b during a backtracked pause. These

Table 3.1: **Summary of pause-free velocities and apparent pause densities measured at various NTP concentrations.**

[NTP] (μM)	N	Pause-free velocity (nt/s)	Apparent pause density (bp^{-1})
35	10	12.4 ± 0.7	0.0721 ± 0.0099
50	11	11.6 ± 1.1	0.0526 ± 0.0086
75	9	17.8 ± 1.2	0.0358 ± 0.0079
100	13	15.7 ± 1.4	0.0326 ± 0.0077
200	17	21.0 ± 2.4	0.0184 ± 0.0046
1000	44	24.7 ± 1.8	0.0156 ± 0.0031
2000	9	26.7 ± 4.3	0.0188 ± 0.0076

Data are shown as mean ± SEM The apparent pause densities are determined by counting pauses that last between 1 s and 120 s. N is the number of single-molecule transcription trajectories in each condition.

rate constants are dependent on the applied force (F, which is positive for assisting forces and negative for opposing forces) according to:

$$k_f = k_0 \cdot e^{F \cdot \Delta / k_B T} \tag{3.7}$$

$$k_b = k_0 \cdot e^{-F \cdot (1-\Delta) / k_B T} \tag{3.8}$$

where k_0 is the intrinsic zero-force stepping rate constant of Pol II diffusing along DNA during backtracking, Δ is the distance to the transition state for each backtracking step (taken to be 0.5 bp, or 0.17 nm), k_B is the Boltzmann constant, and T is the temperature ($k_B T$ = 4.11 pN·nm). The probability density of pause durations, $\psi(t)$, is equivalent to the distribution of first-passage times for a particle diffusing on a one-dimensional lattice to return to the origin [33], and is given by:

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \frac{exp[-(k_f + k_b)t]}{t} I_1(2t\sqrt{k_f k_b}) \tag{3.9}$$

where I_1 is the modified Bessel function of the first kind. We fit the distribution of pause durations for the wild-type enzyme on bare DNA to this model and extracted a characteristic k_0 of 1.3 ± 0.3 s^{-1} (Figure 3.6, gray dashed line). Using the values of k_0 and the applied force in our experiment (6.5 pN on average), we calculated k_f and k_b to be 1.7 ± 0.4 s^{-1} and 1.0 ± 0.3 s^{-1}, respectively (Equations 3.7 and 3.8).

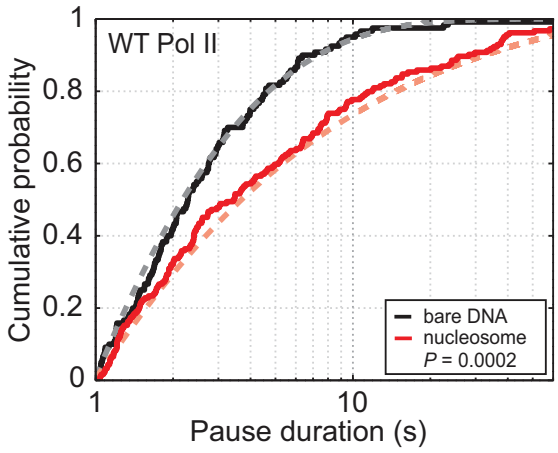

Figure 3.6: **Cumulative distribution of the pause durations.** Cumulative distribution of the durations of pauses longer than 1 s on bare DNA (black solid line) and nucleosomal DNA (red solid line). Dashed lines are theoretical fits of the experimental data to the one-dimensional diffusion model for backtracking P-value is derived from Kolmogorov-Smirnov (KS) test.

3.8 Nucleosome as a tool to perturb the forward translocation

Next, we investigated the transcriptional dynamics of Pol II through the nucleosome as a barrier to the forward translocation of the enzyme. We loaded a histone octamer on the 601 nucleosome positioning sequence (NPS) and ligated it to the downstream template [85].

3.8.1 Dynamics during nucleosomal transcription

As shown previously by Hodges and Bintu *et al.* [54], the nucleosome increases the frequency and durations pauses of the polymerase (Figures 3.7 and 3.8). The wild-type enzyme displays a two-fold increase in the apparent pause density upon encountering the nucleosome (Table 3.2). The mean pause duration on nucleosomal DNA is significantly longer than that on bare DNA (Table 3.2; Figure 3.6).

It has been shown that the nucleosomal DNA can spontaneously unwrap and rewrap around the histones [75, 84, 135]. The increased pause duration of Pol II on nucleosomal DNA can be explained by rewrapping of the DNA downstream of a backtracked Pol II, which prevents the polymerase from diffusing back to the 3′-end of the nascent RNA to resume transcription [54, 13]. Because one bp of nucleosomal DNA fluctuates much faster (>1000 s^{-1}; see Materials and methods for the derivation) than Pol II stepping (1 s^{-1}), the nucleosomal DNA in front of the polymerase reaches wrapping/unwrapping equilibrium between each backtracking step. It follows that the pause durations on nucleosomal DNA can be drawn

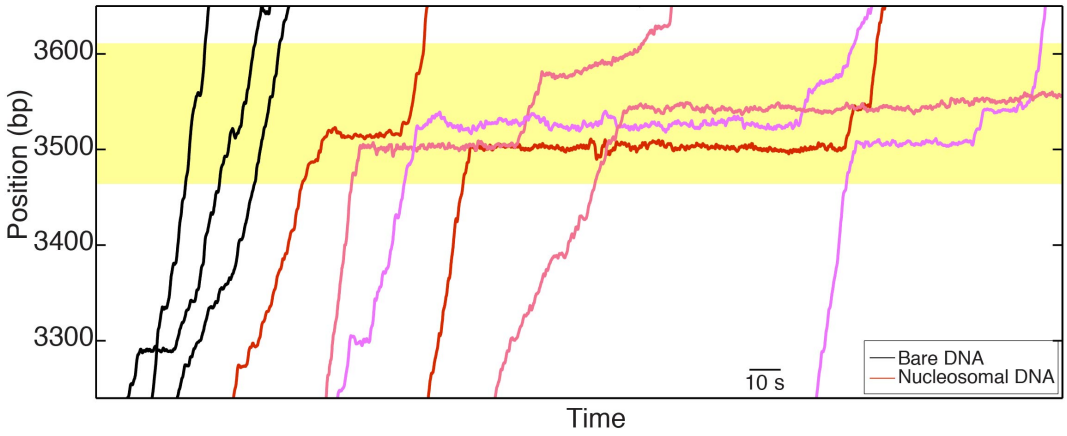

Figure 3.7: **Example transcription trajectories on bare and nucleosomal DNA.** Black traces are bare DNA transcription. Trajectories through the nucleosome are shown in red-tone colors. The data were collected at 1 mM NTP. The yellow shaded region represents the nucleosome positioning sequence (NPS).

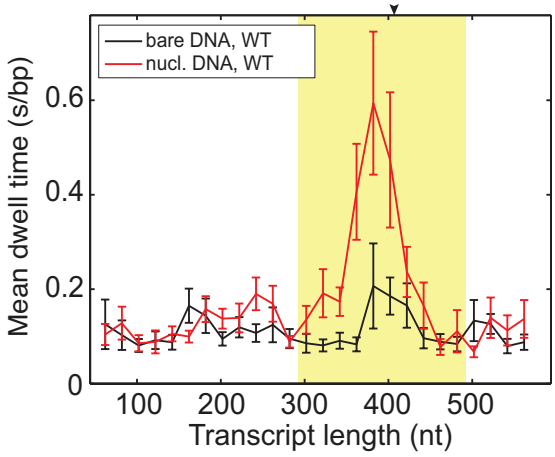

Figure 3.8: **Mean dwell times of the wild-type Pol II along the template.** Bare and nucleosomal DNA transcription are shown in black and red, respectively. The position of the nucleosomal dyad is marked with an arrow.

Table 3.2: **Apparent pause densities and mean pause durations on bare DNA and nucleosomal DNA in the extended NPS region.**

DNA template	*N*	Apparent pause density (bp^{-1})	Mean pause duration (s)
Bare	38	0.0153 ± 0.0041	3.9 ± 0.6
Nucleosomal	94	0.0280 ± 0.0036	9.4 ± 0.8

Data are shown as mean ± SEM The extended NPS region spans –115 nt to +85 nt relative to the nucleosomal dyad.

from the same distribution as on bare DNA, except that the effective forward stepping rate is reduced by a factor γ_u corresponding to the fraction of time the local nucleosomal DNA is unwrapped [54], i.e. $k_{f(nuc)} = \gamma_u \cdot k_f$. The backward stepping rate k_b is not affected by the nucleosome, because little histone transfer occurs in our experimental geometry where the DNA template is under tension [54, 12] and therefore the polymerase does not encounter any roadblock when it diffuses backward. The distribution of pause durations for wild-type Pol II on nucleosomal DNA can be correctly fit by this model with a γ_u value of 0.6 ± 0.2 (Figure 3.6, red dashed line).

3.8.2 Rates of forward translocation and catalysis

Having understood the effect of the nucleosomal barrier on the pausing dynamics, we then turned our attention to its effect on the on-pathway elongation kinetics. Interestingly, we found that the nucleosome also delays the transcribing enzyme by modulating its pause-free velocity. As the wild-type Pol II transcribes through nucleosomal DNA at saturating [NTP], its mean pause-free velocity decreases by 14% from 26.9 ± 0.8 nt/s to 23.2 ± 0.6 nt/s (Figure 3.9).

Previous results have shown that a transcribing Pol II cannot actively open a wrapped nucleosome; instead, it passively waits for the DNA immediately in front of the enzyme to spontaneously unwrap and then translocates forward through a locally unwrapped nucleosome [54]. Since the fluctuations of local nucleosomal DNA occur orders of magnitude faster than the translocations of Pol II during backtracking, we assume that they are also much faster than the on-pathway translocation steps of Pol II. Under this assumption, local DNA reaches wrapping/unwrapping equilibrium before Pol II makes a translocation step and the forward translocation rate (k_1) is effectively reduced by the fraction of time the local nucleosomal DNA is unwrapped (γ_u). The reverse translocation rate (k_{-1}) is unlikely to be affected, again due to the lack of a roadblock against reverse translocation. Thus, according to Equation 3.6, the maximum pause-free velocity for nucleosomal DNA transcription is:

$$V_{max(nucl)} = \frac{\gamma_u \cdot k_1 \cdot k_3}{(\gamma_u \cdot k_1) + k_3} \cdot d \tag{3.10}$$

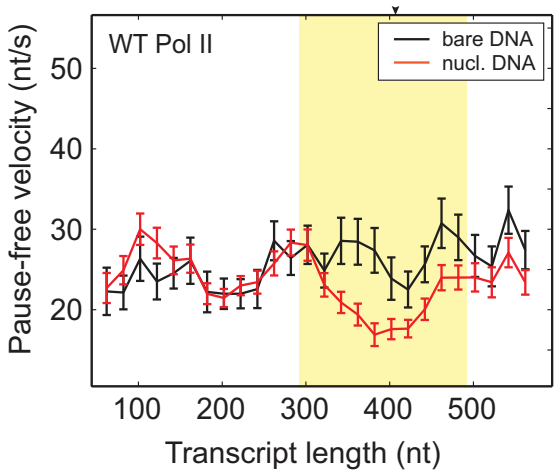

Figure 3.9: **Comparison of pause-free velocities on bare DNA and nucleosomal DNA.** Pause-free velocities of the wild-type Pol II on bare DNA (black) and nucleosomal DNA (red) are plotted as a function of the transcript length. The nucleosomal dyad position corresponds to a transcript length of 407 nt. The extended NPS region (–115 nt to +85 nt relative to the nucleosomal dyad) is highlighted in yellow. The arrow on the top axis marks the position of the dyad.

In comparison, the maximum pause-free velocity for bare DNA transcription is:

$$V_{max} = \frac{k_1 \cdot k_3}{k_1 + k_3} \cdot d \tag{3.11}$$

Using an optimal γ_u value of 0.6, we solved Equations 3.10 and 3.11 and obtained $k_1 = 112 \pm 30$ s^{-1} (indeed much slower than local DNA wrapping/unwrapping) and $k_3 = 35 \pm 3$ s^{-1} for the wild-type. Importantly, these values show that the forward translocation rate is only three times faster than the catalysis rate and, therefore, has a significant contribution to the overall elongation velocity. These numbers were extracted by using the average values of the pause-free velocity and γ_u over the whole nucleosomal region. Such a simplifying treatment is based on the observations that both the pause-free velocity (Figure 3.9) and the local DNA wrapping equilibrium [13] do not change substantially along the NPS.

3.8.3 The first backtracking step is distinct from subsequent steps

The pause density, ρ_{pause}, is governed by the kinetic competition between pause entry and elongation. Previously, an overall elongation rate, which includes translocation, NTP binding, and catalysis, was used in the expression for ρ_{pause} [52, 54, 163]. A more accurate treatment is to use the elementary rate constant in the elongation pathway directly con-

nected to pausing, which is the net rate constant for forward translocation, (Figure 3.3; Equation 3.3):

$$\rho_{pause} = \frac{k_{b1}}{k_{b1} + k_1^{net}} = \frac{k_{b1}}{k_{b1} + \frac{[NTP]}{\frac{k_{-1}(k_{-2}+k_3)}{k_2 k_3} + [NTP]} \cdot k_1} \tag{3.12}$$

where k_{b1} is the rate constant of entering the 1-bp backtracked pausing state. At saturating NTP concentrations ([NTP]$\gg k_{-1}(k_{-2}+k_3)/(k_2k_3)$), k_1^{net} becomes equivalent to k_1. Hence,

$$\rho_{pause(sat)} = \frac{k_{b1}}{k_{b1} + k_1} \tag{3.13}$$

where $\rho_{pause(sat)}$ is the pause density at saturating NTP concentration. In order to obtain a true pause density, the apparent ρ_{pause} needs to be corrected to include pauses shorter than 1 s that are missed by our pause detection algorithm. After such a correction (Materials and methods), the total $\rho_{pause(sat)}$ is 0.045 ± 0.012 bp^{-1}. Solving Equation 3.13 yields $k_{b1} = 5.3 \pm 2.0$ s^{-1}. This value is approximately five times larger than subsequent backward stepping rates, which are force-biased stepping rates obtained from Equation 3.8 ($k_{bn} = 1.0 \pm 0.3$ s$^{-1}, n \geq 2$). The difference between k_{b1} and k_{bn} indicates that the first backtracking transition is easier to make than subsequent backtracking transitions. Using this value of k_{b1}, along with the value of γ_u obtained above, we can predict a nucleosomal pause density of 0.035 ± 0.015 bp^{-1} for pauses longer than 1 s, which agrees with the experimental measurement (Table 3.2).

3.8.4 Reverse translocation rate

We have determined the rates of forward translocation (k_1) and catalysis (k_3) in the nucleotide addition cycle and shown that they are comparable. What remains unknown is the reverse translocation rate k_{-1}, which may also affect the elongation velocity under subsaturating NTP conditions (Equation 3.5). To determine k_{-1}, we examined the pause densities measured at various NTP concentrations. Equation 3.12 can be re-written as:

$$\rho_{pause} = \frac{k_{b1}}{k_{b1} + \frac{[NTP]}{\frac{k_{-1}K}{k_3} + [NTP]} \cdot k_1} \tag{3.14}$$

where $K = (k_{-2} + k_3)/k_2$. The total ρ_{pause} as a function of [NTP] fits well to Equation 3.14 (Figure 3.10). Using the values of k_1, k_3, and k_{b1} determined above, we obtained $k_{-1}K$ equal to $(4.7 \pm 0.5) \times 10^3 \mu$M·s^{-1}. We then revisited the relationship between the pause-free velocity and [NTP] (Figure 3.5A), which follows Michaelis-Menten kinetics. According to Equation 3.5, the Michaelis constant K_M is expressed as:

$$K_M = \frac{k_1 + k_{-1}}{k_1 + k_3} \cdot \frac{k_{-2} + k_3}{k_2} = \frac{k_1 + k_{-1}}{k_1 + k_3} \cdot K \tag{3.15}$$

Plugging the values of K_M, k_1, k_3, and $k_{-1}K$ into Equation 4.2 yields the values of K and k_{-1} for the wild-type Pol II: $K = 9.2$ μM and $k_{-1} = 510$ s^{-1}. We could further calculate the translocation equilibrium constant, K_δ = [pre-translocated]/[post-translocated] $= k_{-1}/k_1 = 4.6$. This result indicates that the enzyme favors the pre-translocated state to the post-translocated one, in agreement with most previous reports [11, 10, 72, 91].

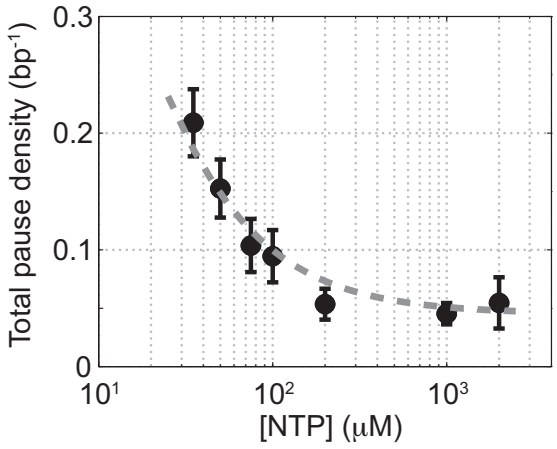

Figure 3.10: **Relationship between pause density and NTP concentration.** The total pause density for the wild-type Pol II (black circles) is plotted against the NTP concentration. The gray dashed line is the fit to Equation 3.14 ($R^2 = 0.93$).

3.9 Force-velocity relationship

A central piece of evidence previously used to favor a branched kinetic scheme (Figure 3.2) over a simpler linear scheme (Figure 3.1) for the nucleotide addition cycle is the relationship between the pause-free velocity (v) and the applied force (F) [1, 82]. However, in those studies, translocation was assumed to be in rapid equilibrium relative to catalysis. Having explicitly determined the translocation rates ($k_{\pm 1}$) and found that the forward translocation rate (k_1) is comparable to the catalysis rate (k_3), we went on to examine whether a linear kinetic scheme (Figure 3.1) is sufficient to explain the F-v relationship, which for such scheme can be expressed as:

$$v(F) = \frac{k_1(F) \cdot k_3}{k_1(F) + k_3} \cdot \frac{[NTP]}{\frac{[k_1(F)+k_{-1}(F)] \cdot K}{k_1(F)+k_3} + [NTP]} \cdot d \tag{3.16}$$

We assume that only the translocation transitions in the nucleotide addition cycle are force-sensitive and that the translocation rates depend on force according to the Boltzmann-type equation: $k_1(F) = k_1(0) \cdot e^{F\delta/k_BT}$, and $k_{-1}(F) = k_{-1}(0) \cdot e^{-F\cdot(1-\delta)/k_BT}$, where δ is the

distance to the transition state for forward translocation, the only unknown variable left in the equation. We measured the pause-free velocity at different applied forces and obtained values in good agreement with previously published single-molecule data [82] (Figure 3.11).

The velocity of the wild-type enzyme shows a weak but detectable dependence on force. The F-v plots can be fit well to Equation 3.16 with δ of 0.46 ± 0.09 bp. Therefore, it is indeed possible to explain the observed force-velocity relationship of transcription elongation with a classic, non-branched Brownian ratchet mechanism, in which NTP binding occurs after translocation.

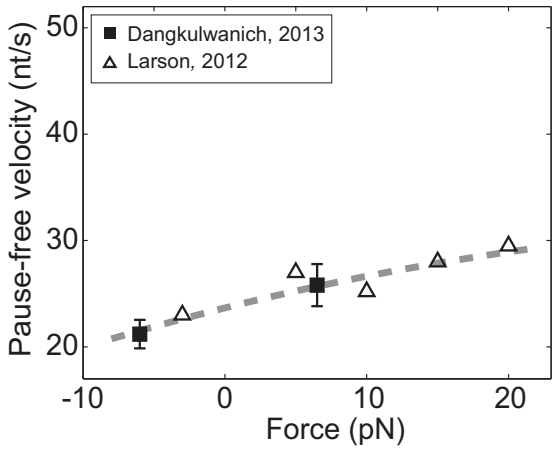

Figure 3.11: **Relationship between transcription velocity and applied force.** The pause-free velocity of the wild-type Pol II is plotted against the applied force. Experimental data in the present study are shown in solid squares (error bars indicate SEM). Open triangles represent data from a previously published single-molecule study [82]. The combined data are fit to the force-velocity relationship predicted by a linear Brownian ratchet model (dashed line), yielding a characteristic distance to the transition state $\delta = 0.46 \pm 0.09$ bp (error is SEM, $R^2 = 0.88$). Positive and negative force values indicate assisting and opposing forces, respectively.

3.10 Discussion

3.10.1 Rate-limiting steps in the Brownian ratchet mechanism

RNAP transcribes DNA through a multi-step kinetic pathway. The rate-limiting nature of the various steps in the nucleotide addition cycle has so far remained largely conjectural. Almost all the existing kinetic studies of transcription elongation relied on the major assumption that translocation and NTP binding follow rapid equilibrium kinetics. As a result,

the catalytic step occurring after NTP binding has been assigned to be rate-limiting of the overall elongation reaction.

The linear Brownian ratchet mechanism that assumes fast translocation equilibrium predicts that, as the NTP concentration increases, the force-sensitivity of the elongation velocity decreases and eventually vanishes, because the enzyme spends less time in the load-sensitive translocation steps. However, the F-v relationships of the enzyme obtained from optical tweezers studies have shown significant dependence of elongation velocity on external force even at saturating NTP concentrations [1, 10, 82], in contradiction to the above prediction. To account for this discrepancy, a modified, branched ratchet model was proposed in which the NTP must also bind to a secondary site on the polymerase in the pre-translocated configuration. Although the existence of such additional binding site may be rationalized by the downstream allosteric site [43, 55], the "E" site or pre-insertion site [126, 147], or the tilted hybrid structure [25], whether it constitutes a significant pathway in the elongation reaction and how it is related to the primary nucleotide binding pathway remain obscure. More importantly, the branched model neglects the possibility that the translocation steps may not be as fast as assumed.

In this study, we tested this possibility of slow translocation by placing a nucleosome in the path of the transcribing polymerase and directly determining the rates of forward and reverse translocation. Our analyses show that the forward translocation rate is in fact within the same order of magnitude as the catalysis rate. For the wild-type Pol II, k_1 is only 2.5 times higher than k_3 (Figure 3.12; Table 3.3). Our results demonstrate that a linear ratchet model can explain the transcriptional kinetics of Pol II and that it is not necessary to invoke a conceptually more complicated branched model, as long as the constraint of fast translocation equilibrium is relieved. Note that although our data argue against rapid oscillation of the ratchet, they still support the notion that the enzyme is able to spontaneously diffuse along the DNA between the pre- and post-translocated states, as suggested by the Brownian ratchet mechanism.

We extracted the values of k_1 and k_3 by comparing the maximum pause-free velocities on bare DNA and nucleosomal DNA (Equations 3.10 and 3.11). In principle, k_1 and k_3 can also be determined by examining V_{max} as a function of applied force:

$$V_{max}(F) = \frac{k_1(F) \cdot k_3}{k_1(F) + k_3} \cdot d \tag{3.17}$$

where $k_1(F) = k_1(0) \cdot e^{F\delta/k_BT}$. Using our data and the previously published data [82] collected at saturating [NTP] (1 mM) and various forces (Figure 3.11), we fit the V_{max}-F dependence to Equation 3.17 and obtained the values of $k_1 = 87 \pm 61\ \mathrm{s}^{-1}$, $k_3 = 33 \pm 8\ \mathrm{s}^{-1}$, and $\delta = 0.64 \pm 0.58$ bp for the wild-type Pol II. Thus, the same qualitative conclusion that both translocation and catalysis are rate-limiting for the elongation reaction can be drawn from this alternative approach. Compared to the approach of using the nucleosomal barrier as a tool to determine k_1 and k_3, fitting the V_{max}-F relationship involves one additional free parameter (δ) and the values are less constrained (larger errors). In the future, it is worthwhile to use either of

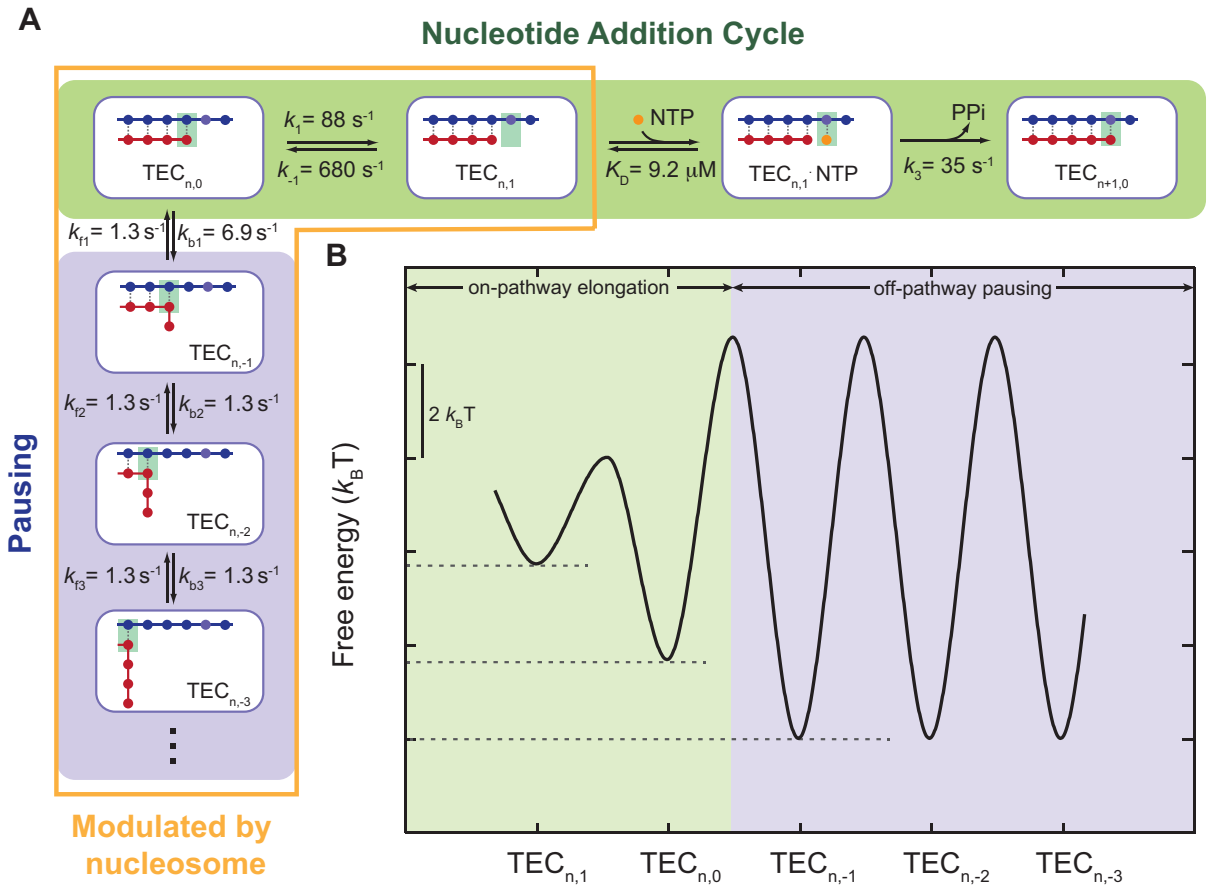

Figure 3.12: **A quantitative kinetic model for transcription elongation.** (**A**) A comprehensive kinetic characterization of the nucleotide addition phase (highlighted in green) and the pausing phase (highlighted in blue) for transcription by the wild-type Pol II. Inside the yellow box are the transitions affected by the nucleosomal barrier. (**B**) The schematic translocation free energy landscape at a given RNA length (solid black). The on-pathway elongation is highlighted in green and the off-pathway pausing is highlighted in blue.

these two approaches or both to test whether prokaryotic transcription also employs a linear ratchet mechanism.

3.10.2 The energy landscape for transcription elongation

With the same transcript length, RNAP is able to move back and forth on the DNA template, forming different TEC configurations (Figure 3.12). Each translocation state corresponds to a local energy minimum [9, 123, 154]. Transitions between the pre- and post-translocated states, together with NTP binding and catalysis, constitute the active elongation pathway (Figure 3.13, green). The enzyme can also enter the pausing pathway by transiting from the pre-translocated state to the backtracked states (Figure 3.13, blue). The hyper-translocated states, in which the enzyme undergoes further forward translocation beyond 1 bp, are energetically unfavorable. The rate constants extracted from our single-molecule experiments translate into a free energy landscape for Pol II's mechanical translocations and chemical transitions (Figure 3.13; Materials and methods), which reveals many detailed features of the kinetics of Pol II transcription.

First, the staircase shape formed by the energy minima of post-translocated, pre-translocated, and 1-bp backtracked states shows that the off-pathway backtracked states are thermodynamically more stable than the on-pathway states (Figure 3.12). This feature confers the enzyme its propensity to enter the pausing pathway, which is the central mechanism for various types of transcriptional control, such as arrest, proofreading, co-transcriptional RNA folding, and recruitment of regulators.

Second, the energy barrier from the pre-translocated to the 1-bp backtracked state is 2.5 k_BT higher than the barrier from the pre-translocated to the post-translocated state, causing k_1 to be more than 10 times faster than k_{b1}. Thus, at the beginning of each nucleotide addition cycle, the pre-translocated TEC favors the catalysis-competent post-translocated

Table 3.3: **Summary of kinetic parameters measured in this study.**

Parameters	**Wild-type Pol II**
k_1 (s^{-1})	88 ± 23
k_{-1} (s^{-1})	$\sim$680
$K_\delta = k_{-1}/k_1$	$\sim$7.7
$K = (k_{-2} + k_3)/k_2(\mu M)$	$\sim$9.2
k_3 (s^{-1})	35 ± 3
k_{b1} (s^{-1})	6.9 ± 2.6
k_{f1}(s^{-1})	1.3 ± 0.3
k_{bn} (s^{-1}), $n \geq 2$	1.3 ± 0.3
k_{fn}(s^{-1}), $n \geq 2$	1.3 ± 0.3

The values reported in the text were measured at 6.5 pN of applied assisting force and are normalized to zero force here.

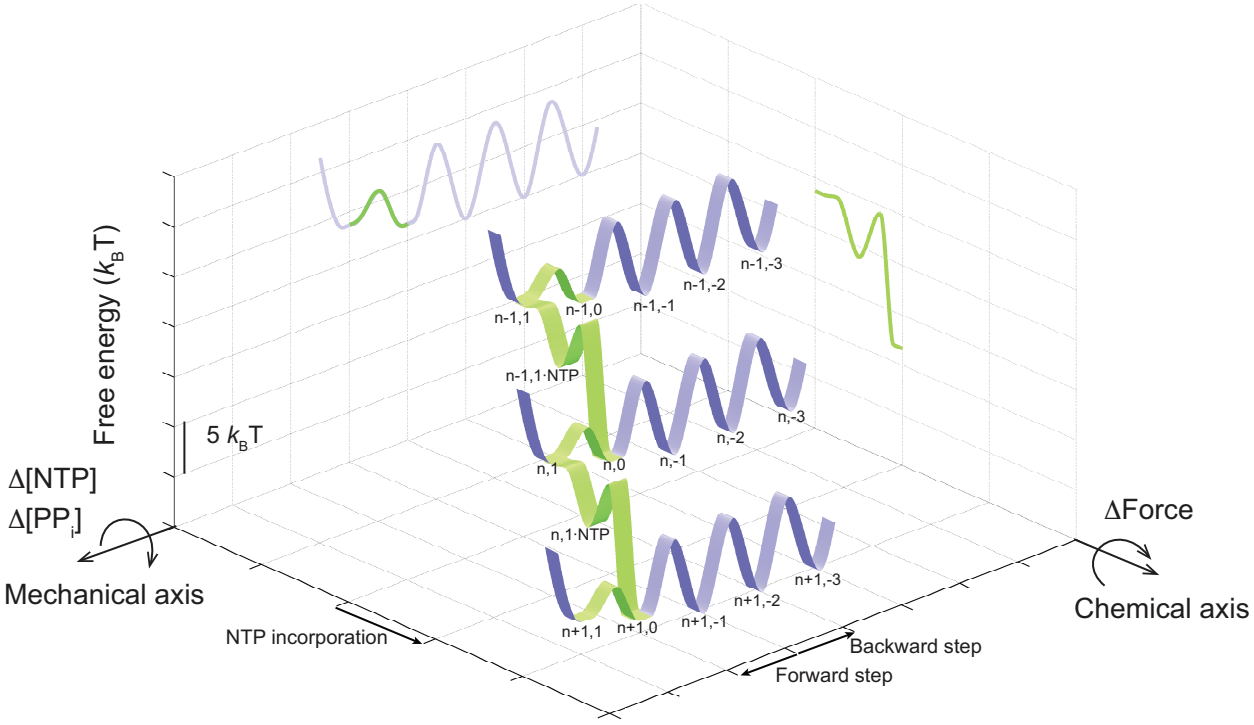

Figure 3.13: **The schematic three-dimensional free energy landscape for transcription elongation by the wild-type Pol II.** The ribbons represent the minimal energy paths of the nucleotide addition cycle (green) and the off-pathway processes (blue). The nomenclature for the TECs (e.g. n,1) is the same as that used in Figure 3.12. Chemical and mechanical transitions are shown in two orthogonal axes. Mechanical perturbations, such as force, affect the mechanical transitions of the enzyme by tilting the landscape around the chemical axis to a first approximation, while chemical perturbations, such as [NTP] and [PPi], rotate the energy levels of chemical transitions around the mechanical axis, again to a first approximation. Two-dimensional projections on the grids highlight the relative free energy of each state. This diagram is calculated at zero force, 1 mM NTP and 1 μM PPi.

state kinetically over the 1-bp backtracked state, even though it is thermodynamically more favorable to move in the opposite direction. This property ensures that pausing only occurs sporadically so that the transcript can be synthesized within a reasonable amount of time. In addition, the barriers between neighboring backtracked states are also relatively high, preventing the enzyme from backtracking too far, which could lead to transcriptional arrest.

Third, the first backtracking step appears to be unique from further backtracking steps in two aspects. Kinetically, entering the 1-bp backtracked state is easier than entering subsequent backtracked states, as reflected by the difference between k_{b1} and $k_{bn}(n \geq 2)$. Such a difference is supported by structural data: the structure of an arrested Pol II complex suggests that backtracking beyond 1 bp is disfavored as it is sterically hindered by a "gating" tyrosine (Rpb2-Y769) [25]. Thermodynamically, transiting from the pre-translocated state to the 1-bp backtracked state is favorable, while backtracking for more steps yields no additional energetic benefit. This result can also find structural support: the first backtracked nucleotide is stabilized by a binding pocket formed by several Pol II residues, whereas the second or third backtracked nucleotide makes no additional contact to the enzyme [142].

Thus, our model depicts an enzyme with a delicate balance between active elongation and inactive pausing [137]. This model can serve as a framework to study the effects of DNA sequence and nascent RNA structure on transcriptional dynamics [9, 123, 158]. Moreover, this model may improve our understanding of the control of transcription fidelity. The 1-bp backtracked state is closely associated with the proofreading process of Pol II, as the enzyme in this location preferentially cleaves the $3'$ dinucleotide of the RNA containing the mismatched base, emptying the active site for NTP binding [142]. It is possible that nucleotide misincorporation slows down forward translocation, thereby promoting the entry to the pausing pathway and the removal of the dinucleotide.

It is worth noting that we cannot definitively rule out the alternative scenario in which the first unique pausing state corresponds to a non-backtracked intermediate. Nonetheless, no evidence has been found for the universal occurrence of such an intermediate in Pol II transcription. The interpretation that most pauses in Pol II transcription are caused by enzyme backtracking is more parsimonious, especially given the corroborating structural data mentioned above. The elementary rate constants extracted from our analyses should provide a reference frame for future computational studies aiming to fully describe the molecular trajectory of a transcribing polymerase.

3.11 Materials and methods

3.11.1 Proteins and DNA preparation

Biotinylated wild-type *S. cerevisiae* Pol II (unphosphorylated C-terminal domain) were purified as previously described [71]. The 3-kb DNA handle was prepared by PCR from Lambda DNA (NEB) using a digoxigenin-labeled primer, and restriction digestion with AvaI (NEB) to create a complementary overhang to the upstream end of the complex.The 574-bp DNA

template was prepared by PCR from a modified pUC19-N1 [162] containing the 601 nucleosome positioning sequence (NPS) [85], and restriction digestion with BstAPI (NEB) to generate a ligatable overhang. Each histone protein was recombinantly expressed and purified from *E. coli*, reconstituted to octamers [150], and loaded on the NPS-containing DNA using salt gradient dialysis [127]. The sequences of DNA primers used to amplify the constructs are provided in Table 3.4. Note that the 601 NPS used here differs from the Widom 601 by several base pairs. Zhang *et al.* mutated the sequence to add SaII recognition site around the first third of the NPS [162]. The alignment of the two sequences is shown in Figure A.2 in the Appendix.

Table 3.4: **Sequences of oligonucleotides used in transcription assays.**

Geometry	**Name**	**Sequence (5′ to 3′)**
assisting force	3-kb Dig handle for	/5DigN/GGGAGTGATTTCCGTCTTACGGT
	3-kb Lambda AvaI rev	AATTATCTCGGGCATACAGCAACAACATGG
	N1 BstAPI for	CAGATTGTACTGAGAGTGCACCA
	N1 598bp rev	ATGACCATGATTACGCCAAG
	TDS	/5Phos/AGCATAATCCTGAATATGGCAAGTTACATAGATAAGTTGGTCGGT TGGGGTTTGTGTGGCTTCGTCGGGCGTCTTCTACATACTACTCCTACC
	NDS	GGTAGGAGTAGTATGTAGAAGACGCCCGACGAAGCCACACAAACCCCAA CCGACCAACTTATCTATGTAACTTGCCATATTCAGGATTATGCTCAT
opposing force	DraIII-phage-rev	AATATTCACCATGTGTTAGAAAACGATAACACCGTG
	dig-phage-fwd	/5DigN/AAGCTGCATGTGCTGGAACTTCAC
	TDS	/5Phos/GGTGTCGCTTGGGTTCTCTTTTCGCCTTGTCTCGGGCGTCGGCT GTAAGTATCCTATACC
	NDS	/5Phos/CCGACGGTATAGGATACTTACAGCCGACGCCCGAGACAAGGCGAA AAGAGAACCCAAGCGACACCCAT
both	RNA9	rGrArCrGrCrCrCrGrA

3.11.2 Assembly of transcription elongation complexes

The transcription elongation complexes (TECs) were assembled by annealing a 9-nt RNA primer (IDT) to a 93-nt template DNA, incubating the hybrid with a biotinylated Pol II, and subsequently annealing a 96-nt complementary DNA using previously published sequences and procedures (Figure 3.14)[54]. The sequences of the template DNA (TDS), non-template DNA (NDS), and RNA primer (RNA9) are provided in Table 3.4. The TEC was walked to a stall site by addition of ATP, CTP and GTP. In the assisting force geometry, the downstream end of the stalled TEC was ligated to the 574-bp DNA containing the 601 NPS (with or without a preloaded nucleosome), while its upstream end was ligated to the 3-kb DNA handle. In the opposing force geometry, the downstream end of the TEC was ligated to a 4-kb DNA amplified from Lambda DNA [158]. The complexes were incubated with 2.1-μm streptavidin-coated beads (Spherotech), and DNA tethers were formed in a dual-trap optical tweezers instrument by attaching the digoxigenin-labeled DNA handle to a 2.1-μm anti-digoxigenin IgG-coated bead. In the assisting force geometry, Pol II and its upstream DNA were under tension, while no external force was applied to the downstream

nucleosome. The tension in the upstream DNA prevented intra-nucleosomal loop transfer and thus ensured that the nucleosome was always ahead of the transcribing polymerase. Transcription was restarted in optical tweezers by addition of NTPs (Thermo Scientific). The transcription buffer contains 20 mM Tris-HCl (pH 7.9), 5 mM $MgCl_2$, 10 μM $ZnCl_2$, 1 mM β-mercaptoethanol, 1 μM pyrophosphate, 300 mM KCl, and NTPs ranging from 35 μM to 2 mM each. All the nucleosomal transcription were done in 1 mM NTPs. The assay was carried out at salt concentration higher than physiological condition in order to facilitate the transcription through the nucleosome, thus improving the throughput of the experiment.

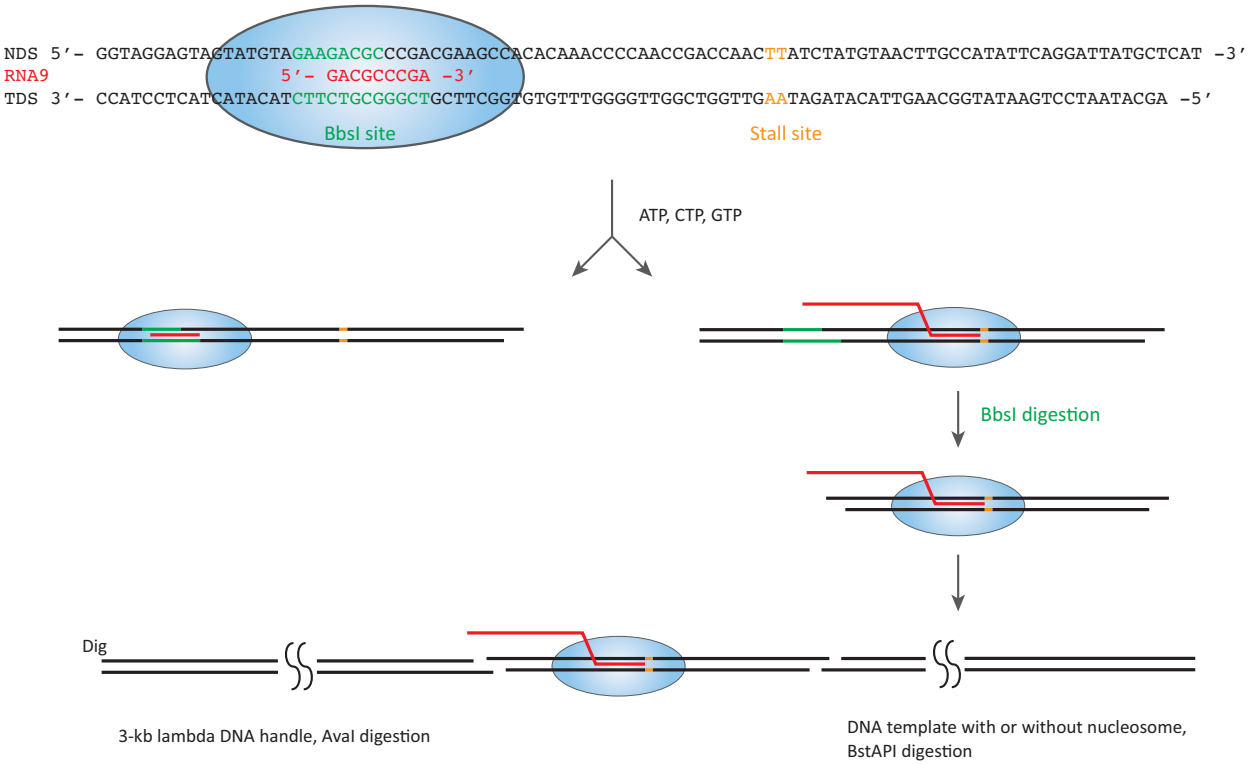

Figure 3.14: **Assembly of Pol II elongation complexes.** The elongation complex were prepared by annealing the RNA9 (red) to the TDS (black). Addition of Pol II and the NDS yielded functional TECs that overlapped a BbsI restriction site (green). The sequence of the oligo nucleotides are shown. TECs were walked to a stall site by the addition of 10 μM each of three nucleotide triphosphates (ATP, GTP, and CTP). As active polymerases moved to the starvation site (orange), they revealed a BbsI restriction site upstream. These TECs were then digested with BbsI and ligated to upstream DNA containing an AvaI-digested overhang on one end and a digoxigenin at the other end. The downstream end was ligated to a BstAPI-digested overhang DNA, which can be preloaded with a nucleosome.

3.11.3 Single-molecule transcription assay

The optical-tweezers based single-molecule transcription elongation assays were performed in passive force mode as previously reported [54, 13, 40, 31, 158]. In assisting force geometry, the stall complex was held at 8 pN for 30 s to established a starting tether length prior to inducing NTPs into the fluidic cell. As Pol II transcribes, the tether lengthens and the force drops. When the force reached 5 pN, one of the traps was moved in to increase the distance between the traps as well as the force, thus maintaining the force in the range of 5-8 pN. Opposing force experiments were started below 3 pN and the force increases as transcription proceeds.

3.11.4 Data collection and analysis

Position data were recorded at 2,000 Hz, averaged and decimated to 50 Hz, and filtered using a second-order Savitzky-Golay filter with a time constant of 1 s. The contour length of the DNA was calculated from the extension and force using the worm-like-chain formula of DNA elasticity [19] with a persistent length of 30 nm. This value of persistent length was obtained from pulling 3-kb DNA in our transcription buffer (data not shown). Data between each trap/force jump were stitched together. Missed data during the jump were extrapolated from the trajectories immediately before the jump. To alleviate calibration error and improve positional accuracy, single-molecule transcription traces that passed 85% of the template were aligned using both the stall site and the expected run-off length [13]. Shorter traces were also proportionally extended based on the average error from the run-off traces. To identify pauses, we computed the dwell time of Pol II at each nucleotide position. Pauses were identified from dwell times that were longer the average dwell time by at least two standard deviations. Due to the limited spatial resolution, we joined pauses that were separated by 3 bp or fewer into a single continuous pause. Pauses longer than 1 s are most likely caused by backtracking [91] and were counted. Pause-free velocities were calculated from time derivatives of the filtered position data, with a lower threshold of 2 nt/s to remove pauses. All curve fittings were performed by non-linear regression of the means weighted by the inverse of the variances.

3.11.5 Estimation of the timescale of local nucleosomal DNA fluctuations

Fluorescence correlation spectroscopy and fluorescence resonance energy transfer experiments showed that the first 20-30 bp of DNA at the nucleosome ends spontaneously unwrap and rewrap on the histone surface every $10-250$ ms [75, 84]. The timescale of the 1-bp DNA fluctuations has not been directly reported but can be estimated from the experimental results above for longer DNA fluctuations. Assuming the wrapping/unwrapping kinetics is uniform along the DNA, we can model the unwrapping of a 25-bp DNA segment as:

$$0 \underset{k_w}{\overset{k_u}{\rightleftarrows}} 1 \underset{k_w}{\overset{k_u}{\rightleftarrows}} 2 \underset{k_w}{\overset{k_u}{\rightleftarrows}} 3 \cdots \underset{k_w}{\overset{k_u}{\rightleftarrows}} 24 \xrightarrow{k_u} 25$$

where k_u and k_w are the local unwrapping and wrapping rate constants of each base pair, respectively. Since the local wrapping equilibrium constant has been shown to be close to 1 [54, 13], we further approximate k_u and k_w with a single value k:

$$0 \underset{k}{\overset{k}{\rightleftarrows}} 1 \underset{k}{\overset{k}{\rightleftarrows}} 2 \underset{k}{\overset{k}{\rightleftarrows}} 3 \cdots \underset{k}{\overset{k}{\rightleftarrows}} 24 \xrightarrow{k} 25$$

A net rate constant can substitute for each pair of forward and reverse rate constants [29]:

$$0 \xrightarrow{k^{net}_{0\to1}} 1 \xrightarrow{k^{net}_{1\to2}} 2 \xrightarrow{k^{net}_{2\to3}} 3 \cdots \xrightarrow{k^{net}_{23\to24}} 24 \xrightarrow{k^{net}_{24\to25}} 25$$

The net rate constants are given by:

$$\begin{aligned}
k^{net}_{24\to25} &= k \\
k^{net}_{23\to24} &= k \cdot \frac{k^{net}_{24\to25}}{k^{net}_{24\to25} + k} = \frac{k}{2} \\
k^{net}_{22\to23} &= k \cdot \frac{k^{net}_{23\to24}}{k^{net}_{23\to24} + k} = \frac{k}{3} \\
&\vdots \\
k^{net}_{1\to2} &= k \cdot \frac{k^{net}_{2\to3}}{k^{net}_{2\to3} + k} = \frac{k}{24} \\
k^{net}_{0\to1} &= k \cdot \frac{k^{net}_{1\to2}}{k^{net}_{1\to2} + k} = \frac{k}{25}
\end{aligned}$$

The time required for unwrapping 25 bp of DNA equals to the total time of unwrapping each bp of DNA:

$$\tau_{0\to25} = \frac{1}{k^{net}_{0\to1}} + \frac{1}{k^{net}_{1\to2}} + \cdots + \frac{1}{k^{net}_{24\to25}} = \frac{325}{k} \approx (10 - 250)ms$$

Thus, the time for 1-bp DNA to unwrap from the nucleosome is expected to be less than 1 ms:

$$\tau_{0\to1} = \frac{1}{k} < 1ms$$

In the same way, we can also show that the 1-bp DNA rewrapping occurs on a similar timescale ($\tau_{1\to0} < 1$ ms). In addition, molecular dynamics simulations also suggested that

the local nucleosomal DNA fluctuates very fast (ns–μs timescale) [135]. In fact, we expected that in our buffer conditions (300 mM KCl), the nucleosomal fluctuations are even faster than the value estimate from the fluorescence experiments, which were performed at lower ionic strengths. Therefore, we assume that the 1-bp of DNA in front of the polymerase unwraps and rewraps much faster than the translocation of the enzyme.

3.11.6 Correction for undercounted short pauses

Experimentally we only counted pauses with lifetimes between 1 s and 120 s. The total pause density $\rho_{pause,total}$ is given by:

$$\rho_{pause,total} = \frac{k_b}{k_b + k_1^{net}} = \frac{\rho_{pause,1<t<120}}{\int_1^{120} \psi(t)dt}$$

The correction factor can be solved analytically to be 2.9 for the wild-type Pol II.

Chapter 4

Roles of the Trigger Loop Element

The trigger loop (TL) is a highly conserved, mobile structural motif near the active center of the polymerase. It folds into α-helical hairpins that make direct contacts with the cognate NTP in the active site to aid substrate selection [141]. It also plays regulatory roles in the pausing phase [161]. Furthermore, the TL also interacts with a nearby element called the bridge helix. These two structural elements have been shown to function together to drive translocation [11, 16, 124, 132]. A mushroom toxin α-amanitin specifically inhibits the activity of Pol II by preventing or altering the TL movement [18, 141].

Several mutations in the TL have been identified to significantly alter the activity and fidelity of yeast Pol II. A conserved histidine residue in the TL has been proposed to interact directly with the substrate and mutations of His 1085 to alanine or phenylalanine are lethal to the cells [61]. Interestingly, a substitution of a glutamate residue at the end of the TL to glycine (E1103G) causes the cells to be sensitive to nucleotide-depleting drug 6-azauracil and become dependent on TFIIS for viability [89]. From a thorough biochemical characterization, the E1103G Pol II exhibits a faster overall elongation rate with a reduction in its fidelity, namely the mutation promotes nucleotide misincorporation [72].

In this chapter, we characterized the transcription dynamics of the E1103G mutant Pol II at the single-molecule level using optical tweezers in order to further elucidate the function of the TL.

Portions of this chapter were published in 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*; 2:e00971. 2013. DOI:10.7554/eLife.00971. Used with permission.

4.1 NTP-dependent elongation dynamics of the E1103G mutant Pol II

We employed the previously described single-molecule optical tweezers assay to track the trajectories of the mutant, and compare its transcription dynamics to that of the wild-type Pol II. We performed our assay at a range of 50 μM to 2 mM of NTP to characterize the detailed kinetic parameters of transcription elongation. Representative trajectories of the mutant and the wild-type Pol II at saturating and sub-saturating concentrations of NTP illustrate that the mutant transcribes DNA at a faster speed and pauses less often than the wild-type at each NTP concentration (Figure 4.1). The relationship between pause-free velocity (v) and [NTP] for both enzymes fit well to the Michaelis-Mention equation, with $V_{max} = 25 \pm 3$ nt/s and $K_M = 39 \pm 12$ μM (errors are SEM) for the WT Pol II. For the E1103G Pol II, we found that its maximum pause-free velocity is ∼1.5-fold higher than that of the wild-type, with $V_{max} = 38 \pm 5$ nt/s and $K_M = 62 \pm 15$ μM (Figure 4.2A). Having higher pause-free velocities than the WT, the mutant enzyme also has fewer pauses (Figure 4.2B). The pause-free velocities and apparent pause densities at various NTP concentrations are summarized in Table 4.1.

Table 4.1: **Summary of pause-free velocities and apparent pause densities of the E1103G Pol II.**

[NTP] (μM)	N	Pause-free velocity (nt/s)	Apparent pause density (bp^{-1})
35	10	16.1 ± 0.9	0.0374 ± 0.0055
50	13	15.2 ± 1.0	0.0266 ± 0.0055
75	13	18.9 ± 1.5	0.0290 ± 0.0069
100	13	24.2 ± 1.7	0.0106 ± 0.0036
200	13	27.4 ± 3.9	0.0100 ± 0.0027
400	10	35.6 ± 1.9	0.0094 ± 0.0062
1000	96	37.6 ± 4.9	0.0051 ± 0.0008
2000	15	42.1 ± 4.9	0.0083 ± 0.0011

Data are shown as mean ± SEM The apparent pause densities are determined by counting pauses that last between 1 s and 120 s. N is the number of single-molecule transcription trajectories in each condition.

4.2 Nucleosomal transcription

As described in the previous chapter, transcription through a nucleosomal barrier can be exploited to further dissect the elongation kinetics of the RNA polymerase. We followed transcription of the E1103G Pol II as it transcribes through the nucleosome. Even though the mutant has a higher pause-free velocity than the wild-type, the nucleosome still presents

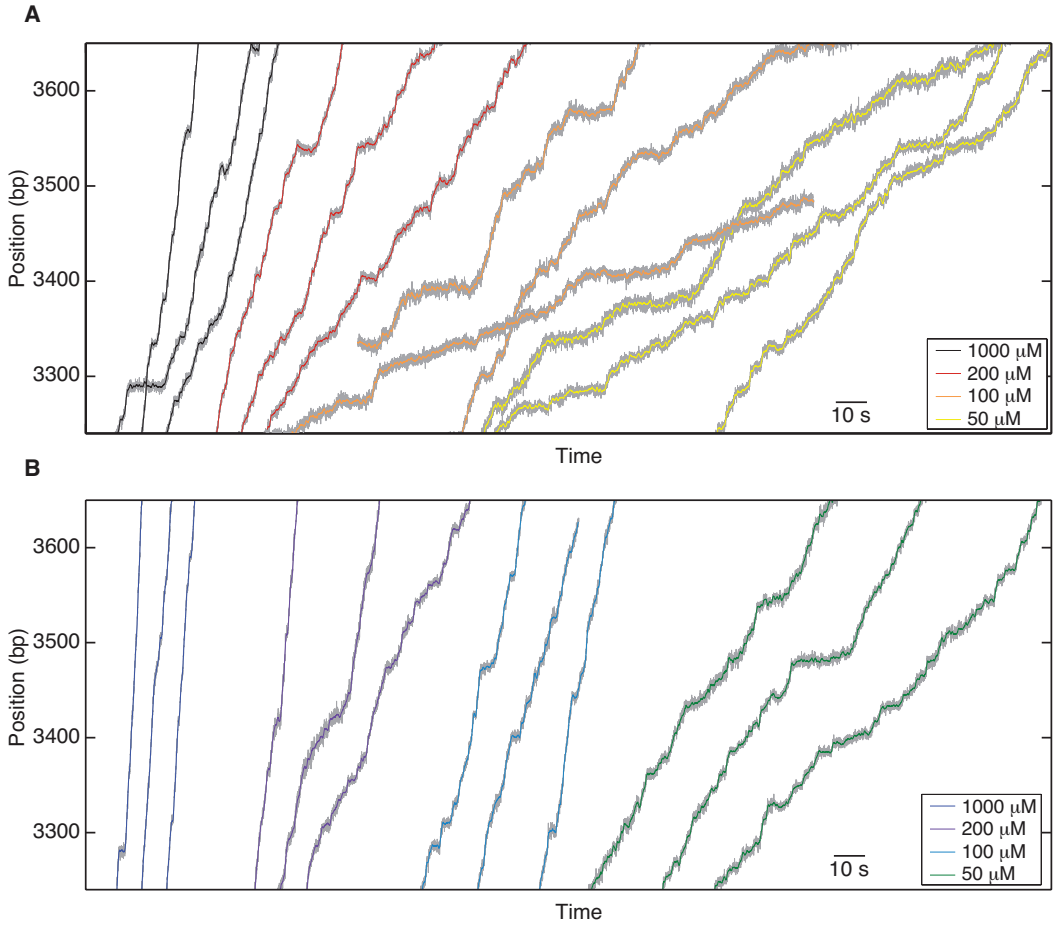

Figure 4.1: **Transcription trajectories of the WT and the E1103G Pol II.** (**A**) Example trajectories of the WT Pol II at various NTP concentrations. (**B**) Example trajectories of the E1103G Pol II at the same concentrations of NTP.

a strong barrier for the enzyme (Figure 4.3). Similar to the wild-type, the enzyme spends longer time crossing the nucleosomal barrier (Figure 4.4). Notably, the nucleosome slows down the mutant enzyme much more than the wild-type. Its mean pause-free velocity reduces by 35% from 39.8 ± 0.6 nt/s to 26.0 ± 0.7 nt/s (Figure 4.5), whereas that of the wild-type decreases by 14% from 26.9 ± 0.8 nt/s to 23.2 ± 0.6 nt/s.

In the same manner with the calculation shown in the previous chapter, the presence of the nucleosome slows down the forward translocation of the enzyme by a fraction of the time that the nucleosomal DNA unwrapped from the surface. By comparing the velocity on bare and nucleosomal DNA, we can extract the translocation rate (k_1) of 50 ± 4 s^{-1} and the catalysis rate (k_3) of 195 ± 65 s^{-1}. In comparison, k_1 of the wild-type is 112 ± 30 s^{-1} and k_3 = 35 ± 3 s^{-1}. The mutation decreases the forward translocation rate, but increases the catalysis rate. The mutant's higher k_3 compensates for its lower k_1, rendering its overall

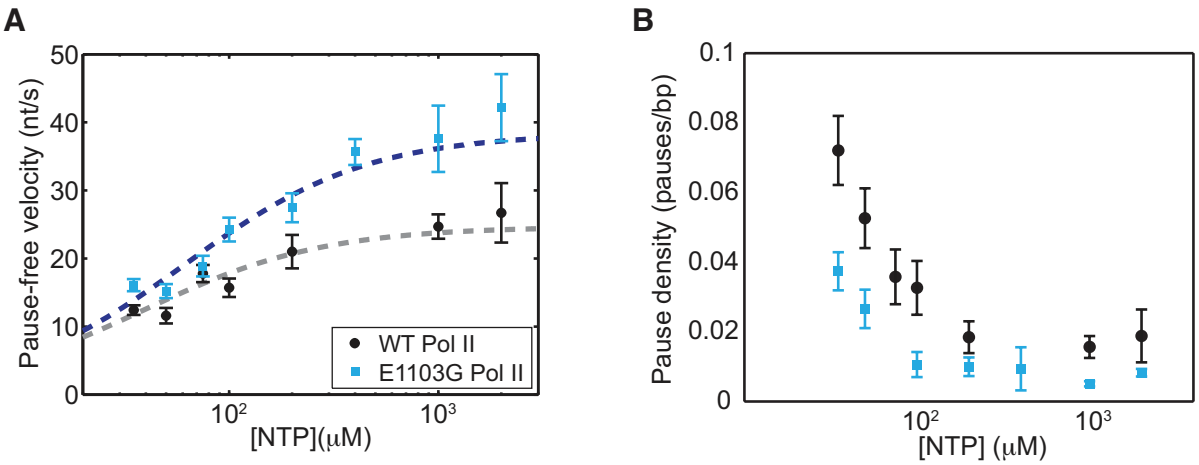

Figure 4.2: **Pause-free velocities and pause density at various NTP concentrations.** (**A**)Pause-free velocities of the wild-type (black) and mutant Pol II (blue). Dashed lines are fits to the Michaelis-Menten equation ($R^2 = 0.80$ for the wild-type; $R^2 = 0.85$ for the mutant). (**B**) The mutant (blue) has fewer pauses than the WT (black) at each NTP concentration.

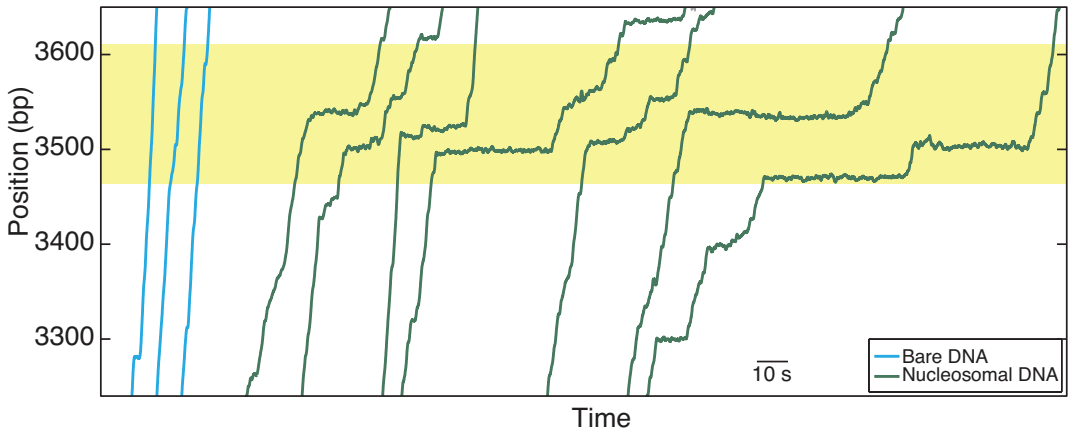

Figure 4.3: **Example transcription trajectories of the E1103G Pol II on bare and nucleosomal DNA.** Transcription on bare and nucleosomal DNA are shown in blue and green, respectively. The data were collected at 1 mM NTP. The yellow shaded region marks the location of the 601 NPS.

velocity faster than the wild-type.

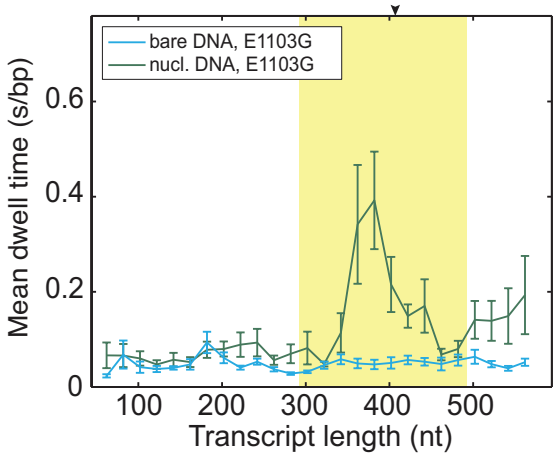

Figure 4.4: **Mean dwell times of the E1103G Pol II along the template.** Bare and nucleosomal DNA transcription are shown in blue and green, respectively. The position of the nucleosomal dyad is marked with an arrow.

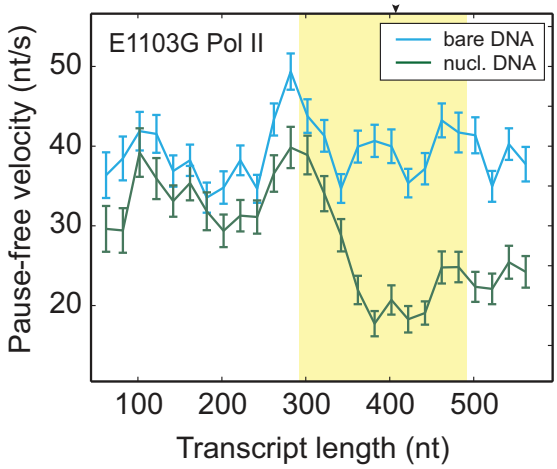

Figure 4.5: **Comparison of the pause-free velocities for the E1103G Pol II on bare DNA and nucleosomal DNA.** Pause-free velocities of the E1103G mutant Pol II on bare DNA (blue) and nucleosomal DNA (green) are plotted as a function of the transcript length. These experiments were conducted at 1 mM NTP. Note that after the polymerase exits the nucleosome, the velocity does not return to the same level of that on bare DNA. This observation could be rationalized if the nucleosome rolls along the DNA and remains ahead of the transcribing polymerase in a fraction of the traces. Error bars are SEM

4.3 Pausing kinetics

We also compared the pausing kinetics between the wild-type and the mutant enzymes. Interestingly, on bare DNA, the mutation only affects the distribution of pauses that are

shorter than 2 s (Figure 4.6A, $P = 0.003$, Kolmogorov-Smirnov test). In contrast, the distributions of longer pauses are indistinguishable between the mutant and the wild-type Pol II (Figure 4.7, $P = 0.9$). It is possible to rationalize this observation if the mutation selectively influences the kinetics of the first backtracking step (k_{b1} and/or k_{f1}) without affecting subsequent backtracking steps, given that pauses of short durations involve small backtracking excursions and that entering the 1-bp backtracked state is distinct from entering longer backtracked ones (k_{b1} is different from k_{bn}, $n \geq 2$). The first backward stepping rate (k_{b1}) only influences the pause density but not the pause duration, while the first forward stepping rate (k_{f1}) does affect the pause duration. Specifically, the increase in short pauses can be explained if the mutation increases k_{f1} and accelerates the return from a pause to active elongation. Indeed, Monte Carlo kinetic simulations show that setting k_{f1} to be larger than 4 s^{-1} —2.4 fold higher than the wild-type value (1.7 ± 0.4 s^{-1})— can reproduce the experimentally observed pause duration distributions for the mutant Pol II on bare DNA (Figure 4.6A, blue dashed line) and nucleosomal DNA (Figure 4.6B, green dashed line). Moreover, by comparing the experimentally measured and simulated pause densities using different k_{b1} values, we can set a lower bound for the mutant's k_{b1} to be 2.8 s^{-1} (Figure 4.9).

Taken together, we have shown that the rate of entering the 1-bp backtracked state is higher than those of entering further backtracked states, and that the E1103G mutation modulates the transition kinetics between the 1-bp backtracked state and the pre-translocated state. Until now, k_{f1} and k_{b1} have been assumed to be identical with the other stepping rates during backtracking (k_{fn} and $k_{bn}, n \geq 2$)[40, 54, 13]. Our data here suggest that the first backtracking step should be treated differently, consistent with published structural data [25, 142] (See Discussion).

4.3.1 The rate of reverse translocation

We have determined the rates of forward translocation (k_1) and catalysis (k_3) in the nucleotide addition cycle, and shown that the k_1 is the rate-limiting step for the mutant enzyme. The reverse translocation k_{-1}, which may affect the velocity under sub-saturating NTP conditions, still remains unknown. To determine k_{-1}, we examined the pause densities measured at various NTP concentrations. Equation 3.12 from the previous chapter is repeated here:

$$\rho_{pause} = \frac{k_{b1}}{k_{b1} + \frac{[NTP]}{\frac{k_{-1}K}{k_3} + [NTP]} \cdot k_1} \tag{4.1}$$

where $K = (k_{-2} + k_3)/k_2$. The total ρ_{pause} as a function of [NTP] fits well to Equation 4.1 (Figure 4.2A). Using the values of k_1, k_3, and k_{b1} determined above, we obtained $k_{-1}K$ equal to $(2.5 \pm 0.4) \times 10^4 \mu\text{M}\cdot\text{s}^{-1}$. We then revisited the relationship between the pause-free velocity and [NTP] (Figure 4.2A), which follows Michaelis-Menten kinetics. The Michaelis constant K_M can be expressed as:

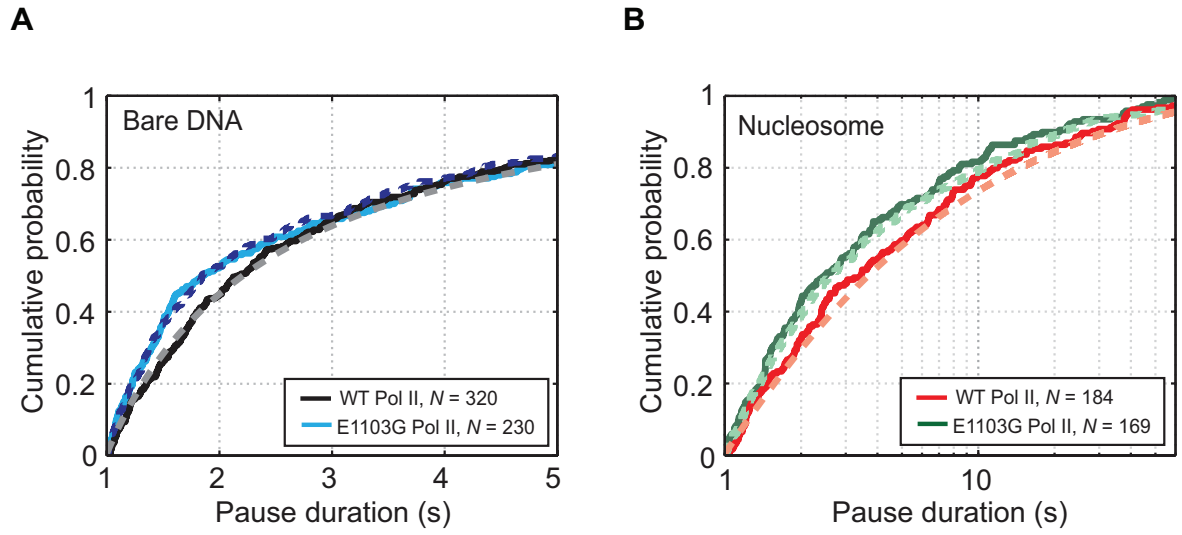

Figure 4.6: **Pause durations on bare DNA and nucleosomal DNA.** (**A**) Cumulative distributions of the pause durations on bare DNA for the wild-type Pol II (black solid line) and the mutant enzyme (blue solid line). The wild-type curve is fit to the one-dimensional random walk model for backtracked pausing (gray dashed line). The blue dashed line represents the simulated pause duration distribution for the mutant enzyme, using a k_{f1} value of 4 s^{-1}. (**B**) Cumulative distributions of the pause durations in the nucleosome region for the wild-type enzyme (red solid line) and the mutant enzyme (green solid line). The wild-type curve is fit to the one-dimensional diffusion model for backtracked pausing, using a γ_u value of 0.6 (red dashed line). The green dashed line is the simulated pause duration distribution for nucleosomal DNA transcription by the mutant enzyme, using a k_{f1} value of 4 s^{-1}.

$$K_M = \frac{k_1 + k_{-1}}{k_1 + k_3} \cdot \frac{k_{-2} + k_3}{k_2} = \frac{k_1 + k_{-1}}{k_1 + k_3} \cdot K \tag{4.2}$$

For the wild-type enzyme, we plugged in known values of K_M, k_1, k_3, and $k_{-1}K$ into Equation 4.2 to find K and k_{-1}. For the mutant enzyme, the value of K cannot be constrained from Equation 4.2 due to the relatively large experimental error. We took a different approach to constrain K_{mutant} by simulating the ρ_{pause}-[NTP] relationship with varying K_{mutant} values and then comparing it to the experimental data (Figure 4.9). We found that the simulated curve significantly deviates from the experimental curve when K_{mutant} becomes larger than 100 μM. Hence, we could set the upper bound of K_{mutant} to be 100 μM. Using the $k_{-1}K$ value of $(2.5 \pm 0.4) \times 10^4 \mu\text{M}\cdot\text{s}^{-1}$ obtained from Equation 4.1, we could set the lower bound of k_{-1} for the mutant to be 210 s^{-1}. The notion that the NTP dissociation rate is much faster than the catalysis rate ($k_{-2} \gg k_3$) has been widely used in the kinetic studies of RNAP and DNA polymerase, and is supported by biochemical evidence [9, 39, 91, 109]. It follows from this notion that $K = (k_{-2} + k_3)/k_2 \approx k_{-2}/k_2$. Thus, K becomes virtually identical to K_D,

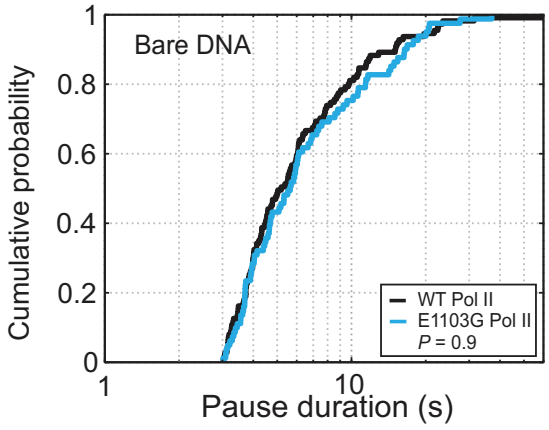

Figure 4.7: **Cumulative pause duration distributions of pauses longer than 3 s.** The curves for the wild-type and the E1103G mutant Pol II are statistically indistinguishable (P = 0.9, Kolmogorov-Smirnov test).

the NTP dissociation constant. Because the mutated residue (Glu1103) is located distal from the NTP interacting part of the TL [141] and the E1103G mutation affects TL closure and NTP sequestration after the initial docking step [72], the NTP binding/dissociation kinetics are unlikely to be significantly affected by the mutation. Therefore, it is reasonable to assume that the wild-type and the mutant enzymes share similar K_D values ($\sim$9.2 μM). Under this assumption, we could estimate the k_{-1} value for the mutant to be $\sim$2700 s^{-1}. The translocation equilibrium constant, K_δ = [pre-translocated]/[post-translocated] = k_{-1}/k_1 is approximately 54. In contrast, the K_δ of the wild-type is only 4.6. This result indicates that the mutation shifts the translocation equilibrium to favor the pre-translocated state more than the wild-type.

4.4 The mutation increases the force sensitivity of the polymerase

While the velocity of the wild-type Pol II weakly depends on force, the velocity of the mutant displays a much stronger force dependence (Figure 4.10). This observation arises from the mutant's lower forward translocation rate and its translocation equilibrium bias. The F-v plots can be fit well to Equation 3.16 with δ of 0.24 $\pm$ 0.05 bp for the mutant (Figure 4.10).

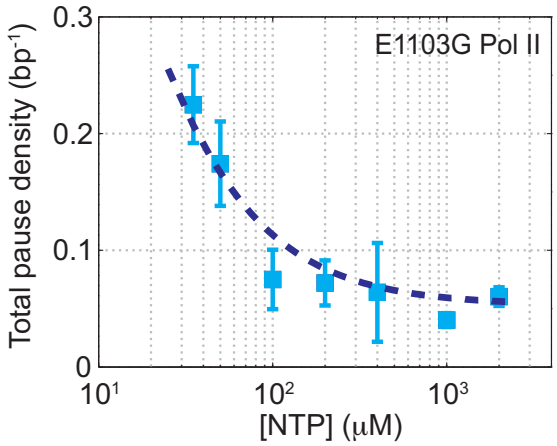

Figure 4.8: **Relationship between pause density and NTP concentration.** The total pause density for the E1103G Pol II (blue squares) is plotted against the NTP concentration. The gray dashed line is the fit to Equation 4.1 ($R^2 = 0.93$).

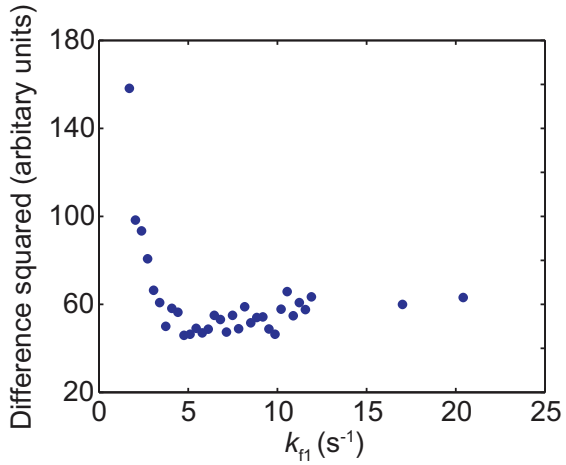

Figure 4.9: **Comparison between the simulated and experimentally obtained distribution of pause durations for nucleosomal DNA transcription by the mutant Pol II.** The square of the difference between the experimental and simulated data is plotted as a function of k_{f1} used in the simulation.

4.5 Discussion

4.5.1 The TL is the global regulator of transcription dynamics

From our experiments and analyses, we found that the kinetics of E1103G Pol II differ from that of the wild-type in various steps of the transcription elongation, both in the nucleotide

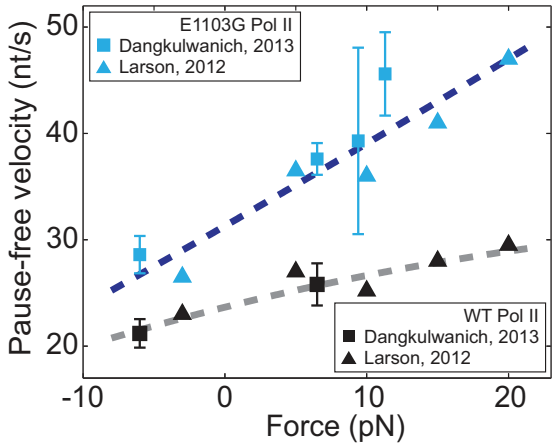

Figure 4.10: **Relationship between transcription velocity and applied force of the E1103G Pol II.** Open triangles represent data from a previously published single-molecule study [82]. The combined data are fit to the force-velocity relationship predicted by a linear Brownian ratchet model (dashed line), yielding $\delta = 0.24 \pm 0.05$ bp (error is SEM, $R^2 = 0.85$). Positive and negative force values indicate assisting and opposing forces, respectively.

addition cycle and the pausing phase. The kinetic parameters of the E1103G Pol II are summarized in Table 4.2.

Table 4.2: **Summary of kinetic parameters of the E1103G Pol II.**

Parameters	Wild-type Pol II	E1103G Pol II
k_1 (s^{-1})	88 ± 23	44 ± 4
k_{-1} (s^{-1})	~680	$\sim 4.1 \times 10^3$
$K_\delta = k_{-1}/k_1$	~7.7	~ 92
$K = (k_{-2} + k_3)/k_2 (\mu M)$	~9.2	~ 9.2
k_3 (s^{-1})	35 ± 3	195 ± 65
k_{b1} (s^{-1})	6.9 ± 2.6	$\sim 3.7^*$
k_{f1}(s^{-1})	1.3 ± 0.3	$\sim 3.1^*$
k_{bn} (s^{-1}), $n \geq 2$	1.3 ± 0.3	1.3 ± 0.3
k_{fn}(s^{-1}), $n \geq 2$	1.3 ± 0.3	1.3 ± 0.3

The values reported in the text were measured at 6.5 pN of applied assisting force and are normalized to zero force here.

The kinetic characterization of the E1103G mutant Pol II reveals that this TL mutation results in many modifications to the enzyme dynamics (Table 4.2). Between the pre-translocated state and the post-translocated state, the mutant is significantly more bi-

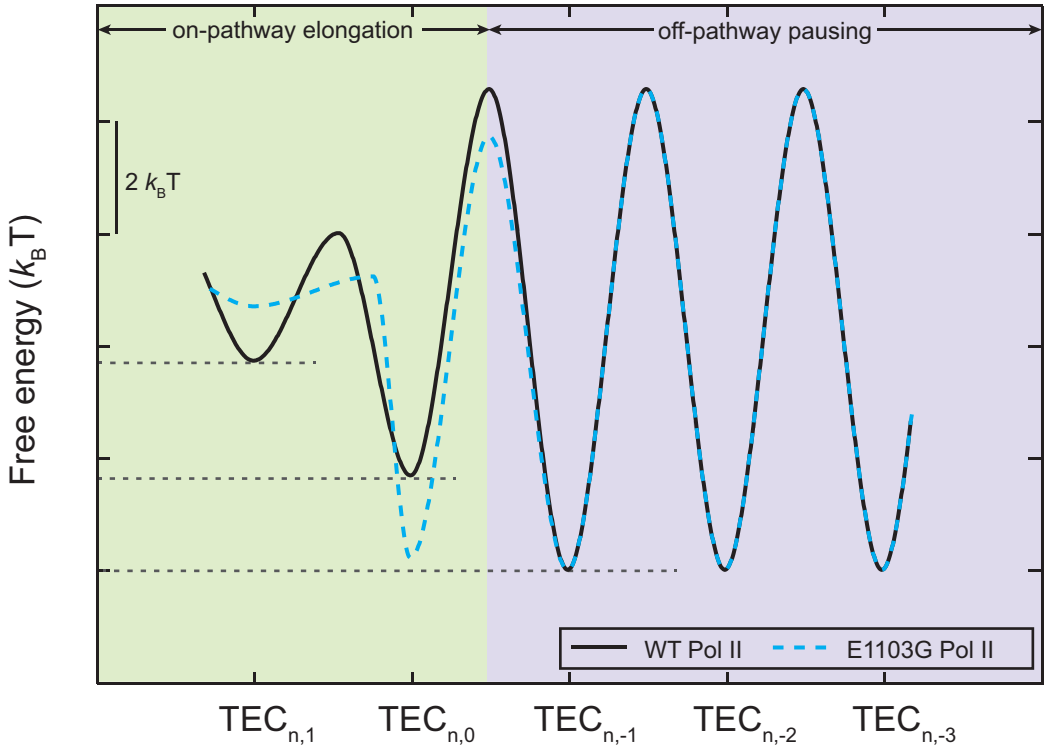

Figure 4.11: **The schematic translocation free energy landscape at a given RNA length** for the wild-type Pol II (solid black) and the E1103G Pol II (dashed cyan). The on-pathway elongation is highlighted in green and the off-pathway pausing is highlighted in blue.

ased toward the former than the wild-type. This property, together with its lower forward translocation rate, renders the mutant's elongation velocity more sensitive to perturbations of its forward translocation, such as an externally applied force (Figure 4.10) or the presence of a nucleosomal barrier (Figure 4.5). It has been shown that the inter-conversion between pre- and post-translocated states involves the transitions of the TL between an open conformation and a wedged conformation [16]. The mutation appears to modulate the enzyme's translocation kinetics by altering the rates of transition between these two conformations.

Furthermore, our analyses lead to the conclusion that the faster overall elongation velocity of the mutant is due to its much greater catalysis rate despite a slower translocation step. The increase of the catalysis rate is most likely due to a faster NTP sequestration step induced by the closure of the TL [72]. The lack of hydrogen bonding between T1095 and the mutated E1103 residue may destabilize the open state of the TL and speed up its closure [139].

The E1103G mutation also affects the pausing kinetics. Specifically, a decrease in the activation energy required to return from the first backtracked state to the pre-translocated

state accelerates the recovery from a pause (Figure 4.11). Consequently, the mutant populates the 1-bp backtracked state less than the wild-type. This property might affect the overall fidelity of transcription. It has been previously shown that E1103G mutation strongly promotes incorporation of non-cognate NMP and mismatch extension [61, 72]. The destabilization of the 1-bp backtracked state relative to the pre-translocated state in the E1103G mutant, established in this work, is consistent with its efficient mismatch extension and suggests that this mutation might also confer a defect in proofreading activity.

Together, our results suggest that the dynamics of TL are involved in multiple phases of transcription elongation, including translocation, catalysis, and pausing. In vivo, various transcription factors and small molecules can directly manipulate the TL dynamics and regulate transcription elongation. For example, transcription factor IIS (TFIIS) stimulates the endonuclease activity of Pol II by replacing the TL with its zinc finger domain, and thus, rescues transcription elongation by creating a new 3′-end of the transcript at Pol II's active site [63]. In fact, the viability of yeast cells expressing only the E1103G mutant Pol II is strictly dependent on TFIIS [89]. It is interesting to investigate how these trans-acting factors modify the rate-limiting mechanism and detailed kinetics of the elongation reaction either via the TL or other elements in the enzyme.

4.6 Materials and methods

4.6.1 Monte Carlo simulation

From an elongation-competent state, Pol II can either elongate by 1 nt with the net forward translocation rate and incorporate an NMP to the RNA transcript, or enter a backtracked pause by 1 nt. During a pause, Pol II diffuses forward and backward with force-biased rate constants k_f and k_b, respectively. For each condition, we simulated 100 trajectories and extracted the pause durations and densities to compare with the experimentally measured values. We also use this simulation to correct for undercounted short pauses. Using the lower bounds of k_{f1} and k_{b1}, we obtained a correction factor of $\sim$7 for the mutant Pol II.

Chapter 5

Regulation of Elongation Dynamics by Transcription Factors IIS and IIF

In cells, transcription is not carried out by RNA polymerase alone. Many factors exist to regulate the process. Some factors directly interact with the polymerase; some influence the process through an interaction with a mediator. In particular, transcription factors IIS and IIF are known to regulate transcription elongation by directly interacting with the polymerase [118]. In this chapter, we studied the effects of TFIIS, TFIIF or the combination of both on Pol II transcription dynamics at the single molecule level. We applied our linear Brownian ratchet kinetic model of slow Pol II translocation, described in previous chapters, to quantify the effects of these factors.

5.1 Transcription factors IIS and IIF

TFIIS inserts its hairpin loop into the pore to position the active site for RNA cleavage [63]; thus, rescuing backtracked Pol II molecules by stimulating the intrinsic endonucleolytic activity of the enzyme [25]. An internal scission of the RNA backbone removes 2-nucleotide (nt) or longer fragments of the nascent RNA and returns the enzyme to a post-translocated state, from which it then resumes transcription elongation [59]. Misincorporated nucleotides favor the backward movements of the enzyme; thus, TFIIS-induced cleavage promotes transcription fidelity both in vitro and in vivo [76, 77, 128]. Although TFIIS is not necessary for the wild-type cells, the ability to cleave the nascent RNA transcript is crucial for cell viability [116]. At the single molecule level, TFIIS can assist Pol II to transcribe against higher external loads [40].

Portions of this chapter were published in T. Ishibashi, M. Dangkulwanich, Y. Coello, T.A. Lionberger, L. Lubkowska, A.S. Ponticelli, M. Kashlev, and C. Bustamante. "Transcription factors IIS and IIF enhance transcription efficiency by differentially modifying RNA polymerase pausing dynamics." *Proc Natl Acad Sci U S A*. 111(9):3419-24. 2014. DOI:10.1073/pnas.1401611111. Used with permission

By contrast, little is known about TFIIF in the elongation phase of transcription. It has an established role in transcription initiation, where it associates with Pol II and five other general transcription factors to form the preinitiation complex. TFIIF is necessary for the recruitment of Pol II to the preinitiation complex, and it either recruits or retains TFIIB during transcription initiation [21, 41, 65, 30]. A recent cryo-EM reconstruction of the preinitiation complex of human Pol II suggests that TFIIF can stabilize the downstream DNA along the cleft of the enzyme [50], in addition to stabilizing the RNA–DNA hybrid within the polymerase [65]. TFIIF is also involved in the elongation phase of transcription in vivo in both yeast [78] and mammals [30]. It has been shown to contribute to the efficiency of the early stages of the elongation [155, 24, 131], and stimulate the overall elongation rate in mammalian systems [58, 125, 160]. However, the detailed mechanism by which TFIIF affects the elongation phase is still unknown. Moreover, although TFIIF is known to bind elongating Pol II in yeast [78], its effects on the elongation process have not been demonstrated.

As discussed above, the presence of histones organizes the DNA in the form of nucleosomes, which prevents the progression of Pol II along the DNA template. Gel-based biochemical assays have shown that TFIIS strongly stimulates in vitro transcription through a single nucleosome [71] and even chromatin templates containing multiple nucleosomes [66, 47]. In a mammalian transcription system, TFIIF has been shown to enhance nucleosomal passage of Pol II, and the presence of both TFIIS and TFIIF significantly improves the efficiency of passage [88]. It remains unclear whether similar effects would occur in the yeast enzyme.

5.2 Optical tweezers elongation assay in the presence of factors

To investigate the effects of transcription factors during Pol II elongation, we used optical tweezers to apply force and monitor the position of Pol II along the DNA template in real time. In an opposing force configuration, a DNA tether was created by attaching a stalled biotinylated Pol II elongation complex to a streptavidin (SA) bead and the digoxigenin-labeled downstream end of the DNA template to an anti-digoxigenin (AD) bead (Figure 5.1A) [40]. Alternatively, we also switched the direction of the applied force to assist transcription elongation by labeling the upstream end of the DNA with a digoxigenin molecule and attaching it to an AD bead (assisting force configuration). Transcription factors (TFIIS, TFIIF, or both) were introduced to the sample chamber at the same time with NTP substrates.

The overall elongation of Pol II in the presence of TFIIF (7.7 ± 0.8 nt/s) or TFIIS (7.3 ± 1.3 nt/s) is faster than that in the absence of the factors (4.9 ± 0.8 nt/s) (Figure 5.1B). The shown rates were measured in passive mode under 4 to 7 pN of opposing loads, where we have sufficient resolution and number of traces; errors are SEM unless otherwise specified. At higher force of 10 pN (Figure 5.1C), we zoomed in to illustrate the effects of the TFIIS and TFIIF (discussed later). As mentioned previously, transcription elongation is punctuated by pauses, which can be separated from active elongation to obtain pause-free velocities. The

mean pause-free velocities are 19 ± 2 nt/s for Pol II in the absence of transcription factors (Figure 5.2), in good agreement with previously published results [82, 31]; 21 ± 2 nt/s with TFIIS; and 23 ± 2 nt/s with TFIIF (opposing force range of 4 to 7 pN).

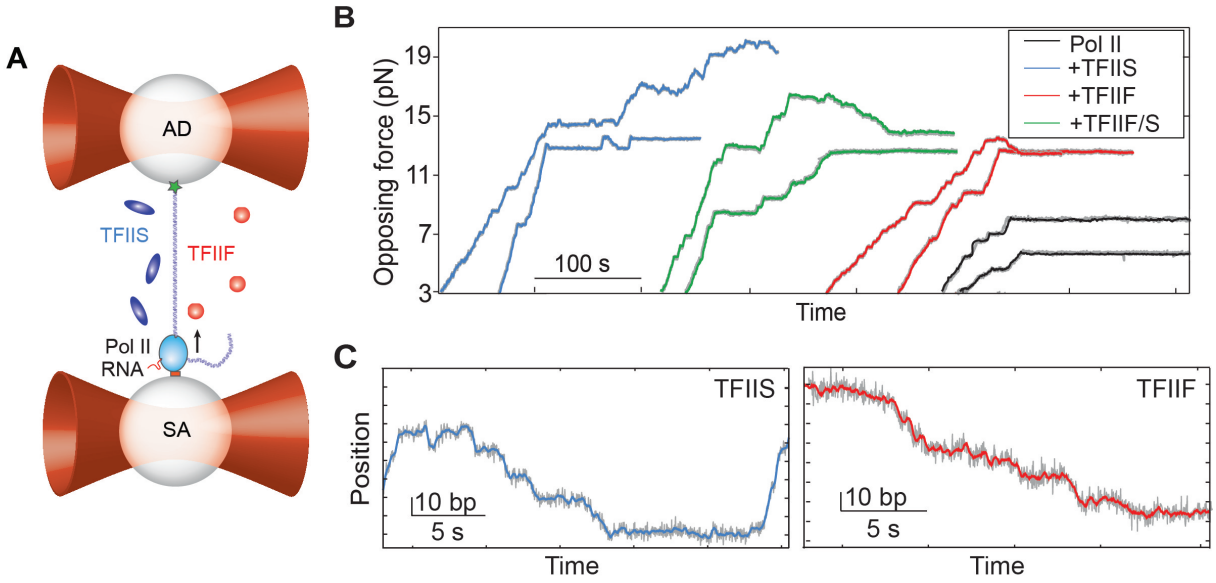

Figure 5.1: **Single-molecule transcription elongation in the presence of transcription factors TFIIS and TFIIF.** (**A**) Experimental geometry for the single-molecule transcription assay in opposing force configuration. (**B**) Example traces of Pol II transcription in the absence of factors (black), with TFIIS (blue), with TFIIF (red), and with both TFIIF and TFIIS (green). (**C**) An enlarged typical example trace of Pol II transcription with TFIIS (Left) and TFIIF (Right) at high opposing force (∼10 pN).

The pause-free velocities at different force ranges studied were also not significantly different from that in the absence of transcription factors (Figure 5.2). This observation suggests that another part of the kinetic pathway, the off-pathway pausing, should be affected by these transcription factors. Indeed, the number of pauses detected (pauses lasting between 1 s and 120 s) decreases (Figure 5.3). The two transcription factors affect the pause durations differently (Figure 5.4). In the presence of TFIIS, pauses become shorter in both assisting and opposing force configurations (Figures 5.4A and B). On the other hand, TFIIF only shortened the pauses in the assisting force geometry (Figures 5.4B and C).

The transcriptional pauses also become significantly shorter in the presence of TFIIS in both opposing and assisting force configurations [Kolmogorov-Smirnov (K-S) test: opposing force in Figure 5.4A, $P = 0.096$; and assisting force in Figure 5.4C, $P \sim 0.01$]. Moreover, TFIIS also increased the stall force of Pol II from 6.7 ± 0.4 pN to 9.0 ± 0.8 pN (Figure 5.5). As a fraction of backtracked Pol II molecules are rescued in the presence of TFIIS, Pol II is more likely to transcribe against higher opposing loads [40]. Although these forces are

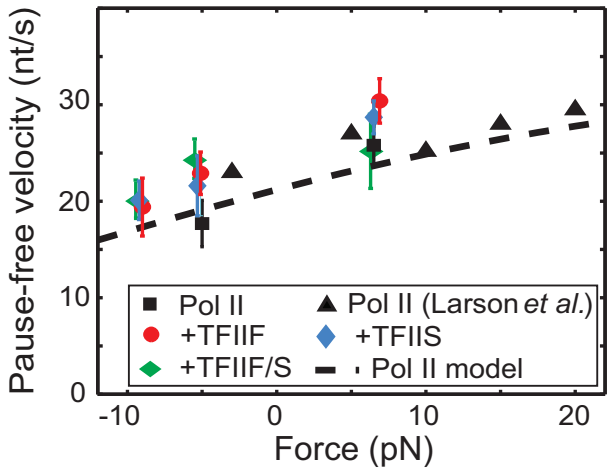

Figure 5.2: **Pause-free velocity against applied forces.** The pause-free velocity at various force ranges in both assisting (+) and opposing (–) force configurations. The presence of TFIIS (blue), TFIIF (red) or both (green) did not strongly affect the pause-velocity of the enzyme. Error bars represent SEMs.

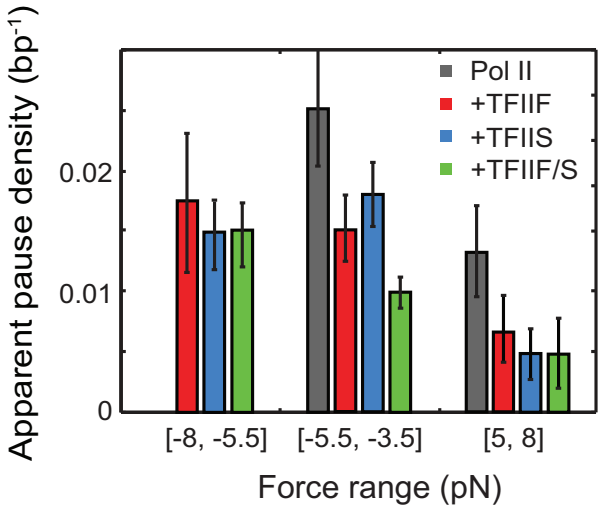

Figure 5.3: **Mean pause densities.** Mean pause densities of Pol II in the absence and presence of transcription factors over the force range studied. Note that we excluded the apparent pause density of Pol II in the 7-pN to 10-pN range in the plot because the mean stall force of Pol II is about 7 pN. Vertical lines represent SEMs.

somewhat different from those reported previously [40], it is known that the stall forces will change with the DNA templates because of the differences in the stability of the nascent RNA secondary structures, determined in part by the GC content [158]. The template used in this study has a uniformly lower GC content than that used by Galburt *et al.* (Figure 5.6)

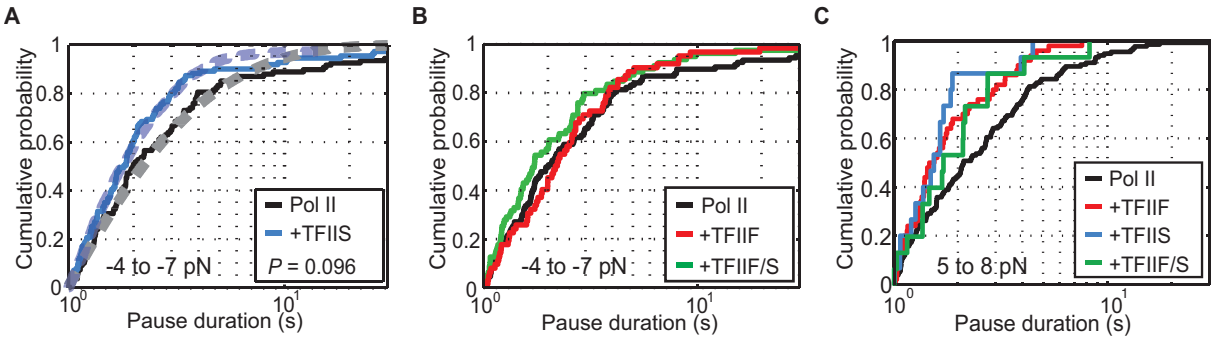

Figure 5.4: **The cumulative pause duration distributions.** (**A**) The cumulative pause duration distributions of Pol II (black) and Pol II with TFIIS (blue) in the 4 to 7-pN opposing force range. Dashed lines represent fits from the models. (**B**) The cumulative pause duration distributions of Pol II (black), Pol II with TFIIF (red), and Pol II with TFIIF/TFIIS (green) in the 4 to 7-pN opposing force range. (**C**) The cumulative distributions of pause durations on bare DNA under assisting loads of 5 to 8 pN for Pol II alone (black), Pol II with TFIIF (red), Pol II with TFIIS (blue), and Pol II with TFIIF/TFIIS (green).

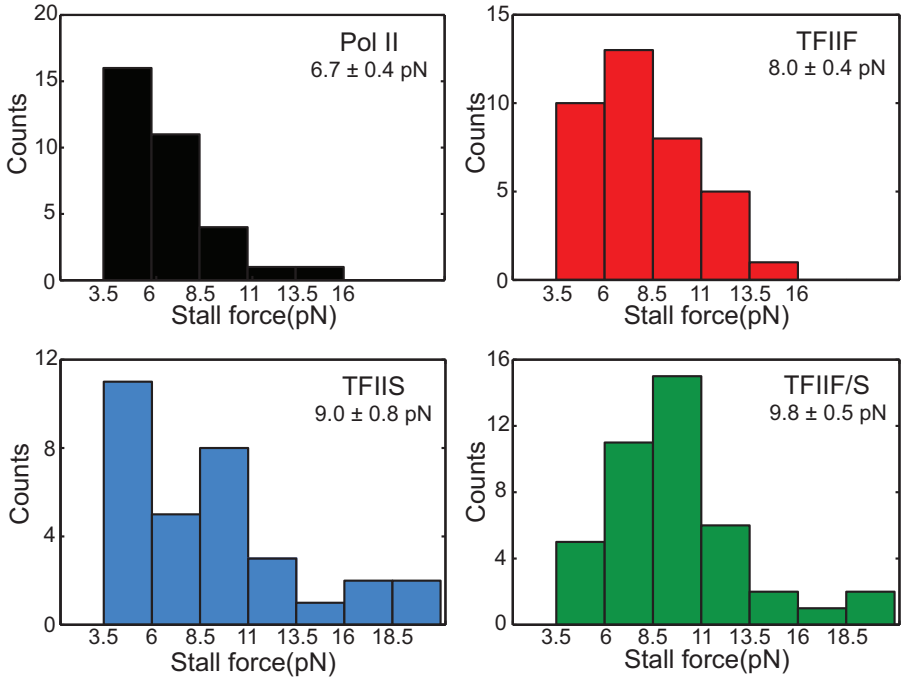

Figure 5.5: **The distributions of the Pol II stall force in the presence of transcription factors.** Both TFIIS and TFIIF increase the stall force of Pol II. Numbers indicate means ± SEM.

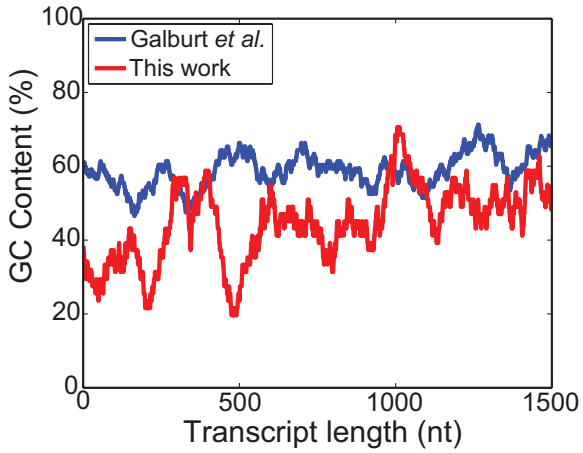

Figure 5.6: **The GC content in the templates.** The percentage of GC residues along the template for the DNA sequence used here (red, [158] and that used by Galburt *et al.* (blue, [40]).

[40]; thus, we attribute differences in observed stall forces in part to RNA structures. In addition, the stall force in the presence of TFIIS reported here likely represents a lower bound estimate of the effect of this factor. Despite a saturating concentration of TFIIS (2 μM, 20 times greater than the K_d [151]), it is possible that TFIIS does not bind to all Pol II molecules throughout the elongation process, as noted by Galburt *et al.* [40]. At forces much higher than the mean stall force, Pol II typically backtracks over a large distance and cannot resume transcription from these backtracks. However, in the presence of TFIIS, we observed that the factor can rescue Pol II molecules that have backtracked more than 10 bp (Figure 5.1C). In addition, the enzyme seems to backtrack in a stretch of 10 or 20 bp in the presence of TFIIS (Figure 5.1C). Perhaps, this observation suggests a more stable backtracked conformation of Pol II with interactions of the backtracked RNA-binding site on the enzyme and 9-nt RNA in Cheung *et al.* [25]. After backtracking for ∼10 nt, the polymerase pauses with favorable interactions in the backtracked RNA-binding site. Subsequently, TFIIS stimulates the RNA cleavage. At this point, the polymerase can extend the RNA or backtrack with certain probabilities. In the example trace shown (Figure 5.1C), the enzyme backtracked in stretches of ∼10 nt for ∼60 nt and paused for ∼10 s before it could resume active elongation. These real-time observations agree with previous experiments suggesting that TFIIS enhances Pol II recovery from backtracked pauses by stimulating its endonucleolytic activity [25, 40].

As with TFIIS, TFIIF similarly decreased the number of pauses detected in both opposing and assisting force geometries (Figure 5.4). TFIIF shortened the duration of pauses in the assisting force configuration ($P = 0.014$, Figure 5.4C). However, this effect in the distribution of pause durations was not detected in the opposing force configuration (P ∼0.5, Figure 5.4C). This observation suggests that the effects of TFIIF was weak and the presence of force, which increases the pause durations, overwhelmed the effects of the factor. Taken

together, these results reveal that TFIIF regulates transcription elongation by modifying the pausing dynamics of Pol II. As would be expected for a transcription factor that decreases the probability of Poll II entering a paused state (thus making Pol II less susceptible to transcriptional arrest), the stall force of Pol II also increases in the presence of TFIIF: 8.0 $\pm$ 0.4 pN, compared with 6.7 $\pm$ 0.4 pN for Pol II alone (Figure 5.5).

Having established the effects on transcription elongation of each individual factor, we sought to characterize Pol II dynamics in the presence of both TFIIS and TFIIF, as may occur in vivo. We found that the distribution of pause durations in the presence of both transcription factors does not differ significantly from that observed in the presence of either TFIIF or TFIIS alone [(K-S test, $P = 0.05$), Figures 5.4B and C, opposing and assisting force configurations, respectively]. Furthermore, the apparent pause density in the presence of both transcription factors is lower than that observed in the presence of either transcription factor alone only in the lower opposing force range studied (Figure 5.3). Finally, the stall force in the presence of both TFIIS and TFIIF was 9.8 $\pm$ 0.5 pN compared with 9.0 $\pm$ 0.8 pN and 8.0 $\pm$ 0.4 pN in the presence of only TFIIS and only TFIIF, respectively (Figure 5.5). Our results indicate that there is a weak enhancement of transcription elongation when both transcription factors are present simultaneously relative to the enhancement observed when either factor is present alone.

5.3 Effects of transcription factors during nucleosomal transcription

In eukaryotic cells, genomic DNA is wrapped into nucleosomes, which regulate transcription by acting as a barrier to Pol II elongation [71, 54, 14, 28]. Therefore, defining the mechanisms by which transcription factors assist Pol II elongation through nucleosomes lies at the heart of understanding transcription regulation in the cell.

To this end, we used assisting force geometry with the downstream DNA template harboring a nucleosome whose position was defined by a 601 nucleosomal positioning sequence (NPS). Nucleosomal transcription experiments were carried out under assisting force geometry to avoid applying tension to the nucleosome and destabilizing it. In the presence of either transcription factor, we observed that Pol II spends less time at the entry of the nucleosome (–115 bp to –35 bp with respect to the nucleosome dyad) and that Pol II transcribes nucleosomal DNA more efficiently, as reflected by the probabilities of nucleosomal passage (Figure 5.7). In 300 mM KCl, 63% of Pol II molecules (59 of 93 molecules in total) were found to pass through the nucleosome in the absence of any transcription factor, 74% in the presence of TFIIS (42 of 57 molecules), 72% in the presence of TFIIF (23 of 32 molecules), and 77% when both TFIIS and TFIIF (18 of 23 molecules) were present (Figure 5.7C). Our results indicate that one factor may interfere with the other's function; hence, we did not observe a quantitative addition of their effects.

Mechanistically, the nucleosome acts as a rapidly fluctuating barrier that allows the

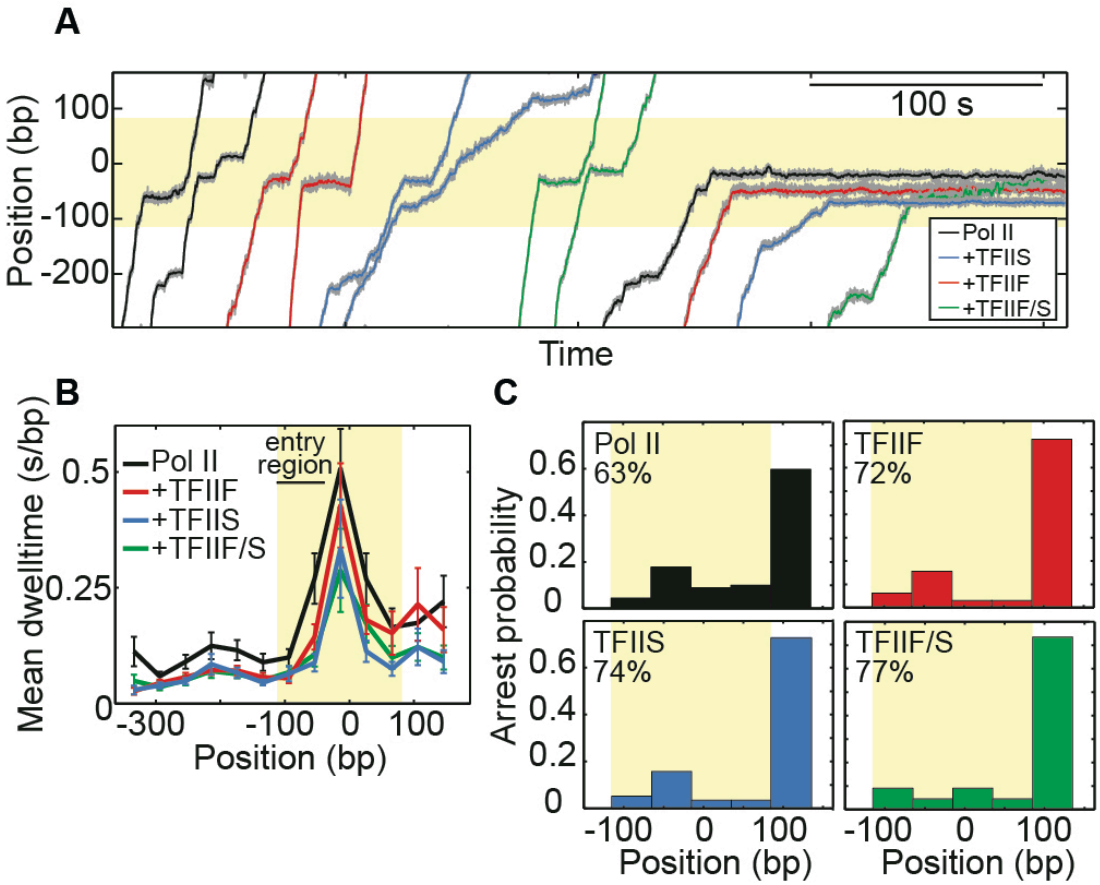

Figure 5.7: **Transcription factors TFIIF and TFIIS enhance Pol II elongation through the nucleosome.** (**A**) Example traces of Pol II transcription on nucleosomal DNA in the absence of factors (black), with TFIIF (red), with TFIIS (blue), and with TFIIF/TFIIS (green). Some molecules stop in the nucleosome region (right side). (**B**) Mean dwell times along transcription distance in relative to the nucleosomal dyad. Error bars represent SEMs. (**C**) Histograms of transcription arrest sites in the presence of different factors. Numbers are the percentage of Pol II molecules that passed through the nucleosome. The extended NPS region (–115 nt to +85 nt) is highlighted in yellow. All nucleosomal transcription experiments were done under an assisting load of 5-8 pN.

polymerase to progress only when it is unwrapped in front of the enzyme. In the absence of TFIIS, a backtracked Pol II must wait for DNA unwrapping before it can diffuse forward and recover from a backtracked pause [54]. In the presence of TFIIS, however, RNA cleavage places Pol II in the elongation-competent state, rescuing the enzyme from a backtracked pause and thus facilitating transcription through a nucleosome. The effect of TFIIF, namely preventing pause entering and reducing pause durations of Pol II, can similarly explain how this factor facilitates transcription through a nucleosome [159, 57].

We observed that the presence of transcription factors does not substantially affect the pause-free velocities of Pol II even in the nucleosome region (Figure 5.8). This result is consistent with the pausing dynamics being responsible for Pol II's enhanced processivity through the nucleosome. Although TFIIS and TFIIF function via different mechanisms (Discussion), both factors favor the on-pathway phase, reducing the probability of arrest and, consequently, increasing nucleosomal passage.

5.4 A kinetic model that explains the effects of transcription factors

To quantitatively describe the observed effects established here for the transcription factors, we modified the linear Brownian ratchet model for Pol II elongation, discussed in previous chapters [31] to account for the effects of transcription factors. Briefly, at each nucleotide position along the DNA template, the pretranslocated Pol II can either transit to the post-translocated state and incorporate a nucleotide (the "on-pathway" mechanism, green in Figure 5.9) or enter a pause ("off-pathway" mechanism, purple in Figure 5.9). In the on-pathway mechanism, Pol II thermally fluctuates between the pretranslocated state (denoted $\mathrm{TEC}_{n,0}$) and the posttranslocated state ($\mathrm{TEC}_{n,1}$). By convention, the first subindex (n) corresponds to the RNA transcript length and the second subindex indicates the translocation state (0 for "pre" or 1 for "post"). Translocation by Pol II occurs with a forward rate, k_1, and a backward rate, k_{-1}. Once in the posttranslocated state, NTP can bind to the active site and rectify the forward translocation with the NTP binding (k_2) and dissociation rates (k_{-2}). After NTP binding, the enzyme catalyzes the phosphodiester bond formation with nascent RNA and releases PPi; we represent the combined catalysis rate constant that includes bond formation and PPi release by the rate k_3. Pol II then completes a cycle of nucleotide addition, moves forward on the DNA by 1 bp, and returns to the pretranslocated state with one additional nucleotide in the RNA transcript ($\mathrm{TEC}_{n+1,0}$). At each position along the template, Pol II may also enter into a pause with the rate k_{b1}, thus kinetically competing with a forward translocation (with rate k_1). If Pol II backtracks, it enters the off-pathway pausing states (purple in Figure 5.9). There, Pol II diffuses along the DNA template with force-biased diffusion rate constants in the forward (k_f) and backward (k_b) movements (downstream and upstream in Figure 5.9, respectively), given by:

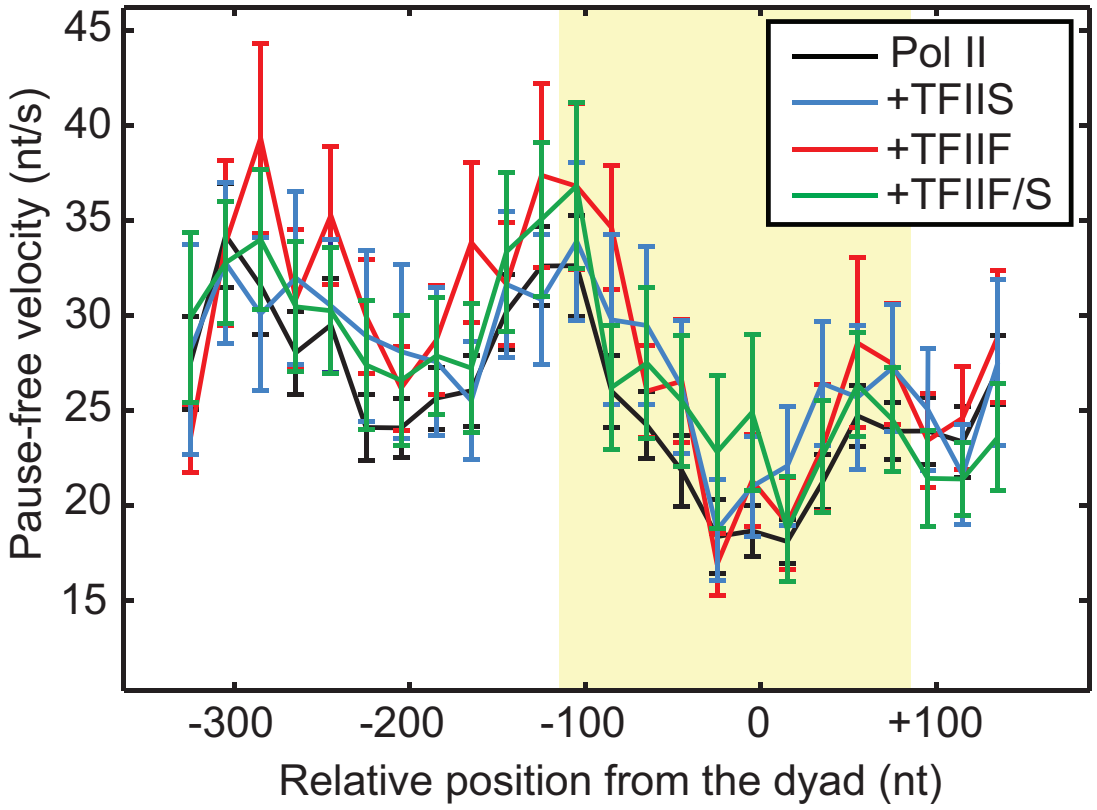

Figure 5.8: **Pause-free velocities of Pol II along the DNA template.** Pause-free velocities of Pol II as a function of template position in the presence of transcription factors TFIIS and TFIIF. Pause-free velocities were measured in the assisting force range 5-8 pN. Template positions are relative to the nucleosome dyad. The extended nucleosome positioning sequence (–115 nt to +85 nt) is highlighted in yellow.

$$k_f = k_0 \cdot e^{F \cdot \Delta / k_B T} \tag{5.1}$$

$$k_b = k_0 \cdot e^{-(F \cdot \Delta + \delta G_{RNA})/k_B T} \tag{5.2}$$

Here, k_0 is the intrinsic rate constant describing Pol II diffusion along DNA during backtracking at zero force, Δ is the distance to the transition state for each backtracking step (taken here to be 0.5 bp), F is the applied external force, and δG_{RNA} is an energy barrier to backtracking due to the nascent RNA secondary structure behind Pol II [158]. In a paused state, Pol II performs a random walk back and forth along the DNA until the 3′-end of the RNA restores registration with the active site of Pol II [40]. Therefore, the

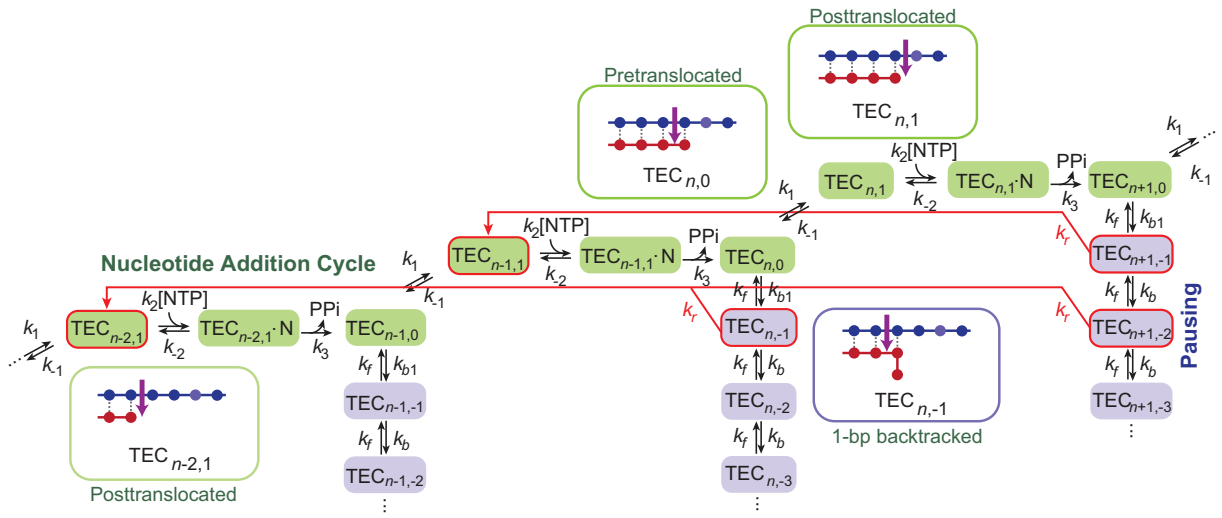

Figure 5.9: **Kinetic model of transcription elongation by Pol II in the presence of TFIIS.** Transcription elongation by Pol II is composed of the on-pathway elongation (green) and the off-pathway pausing (purple). Forward translocation (k_1) competes with entry into backtracked pauses (k_{b1}). In the absence of TFIIS, pause recovery requires forward diffusion of the enzyme (k_f) to a pretranslocated state such as $TEC_{n,0}$. TFIIS introduces a new pause recovery mechanism (k_r, red arrows) that takes a backtracked Pol II in the state $TEC_{i,-j}$ to the on-pathway posttranslocated state $TEC_{i-j-1,1}$. Cartoon configurations of Pol II TECs in the pre- and posttranslocated and 1-bp backtracked states show that TFIIS-stimulated transcript cleavage rescues the 1-bp backtracked Pol II complex ($TEC_{n,-1}$), transferring it to the elongation-competent posttranslocated state $TEC_{n-2,1}$. The magenta arrow represents the active site of the enzyme. The RNA transcript and template DNA are shown in red and blue, respectively.

pause durations can be modeled as the first passage times of a 1D random walker and the probability density of pause durations, $\psi(t)$, is then given by:

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \frac{exp[-(k_f + k_b)t]}{t} I_1(2t\sqrt{k_f k_b}) \tag{5.3}$$

where I_1 is the modified Bessel function of the first kind.The pause density ρ_{pause} (in bp^{-1}) reflects the probability of entering a pause, which results from the competition between the first backtracking step (k_{b1}) and the net forward translocation (k_1). At saturating NTP concentrations, such as in our experimental condition (1 mM), k_1 can adequately describe the net rate of forward translocation (Figure 5.9, [31]); thus the pause density can be written as:

$$\rho_{pause(sat)} = \frac{k_{b1}}{k_{b1} + k_1} \tag{5.4}$$

We have only counted pauses with lifetimes between 1 s and 120 s in our experiments. In order to relate the theoretical absolute pause density given by Equation 5.4 with the experimental result, the theoretical density must be multiplied by the fraction of pauses that are within the pause detection limits (1-120 s). The theoretical apparent pause density in the range of 1-120 s can then be calculated by

$$\rho_{pause(sat),1<t<120} = \rho_{pause(sat)} \cdot \int_{1}^{120} \psi(t)d(t) \tag{5.5}$$

We fitted our opposing pause density and pause duration data in the absence of transcription factors based on this model. Using previously reported rates for transcription by Pol II at zero force, $k_{b1} = 8.0$ s^{-1} and $k_1 = 88$ s^{-1} [31], we extracted the rate $k_0 = 1.6 \pm 0.3$ s^{-1} and $\delta G_{RNA} = 1.1 \pm 0.2$ k_BT using the expressions for the pause duration distribution and apparent pause density (Equations 5.3 and 5.5, respectively) [158].

In the presence of TFIIS, we observed that pause durations became shorter due to the TFIIS-stimulated endonucleolytic activity of Pol II [59, 40]. To quantitatively address this effect, we introduced an additional backtracked pause recovery pathway to the kinetic scheme described above (Figure 5.9, red arrows). This new recovery pathway takes Pol II from any off-pathway, backtracked state ($\mathrm{TEC}_{i,-j}$, where the second negative subindex indicates a backtracked state and j indicates the number of base pairs backtracked) to the on-pathway posttranslocated state ($\mathrm{TEC}_{i-j-1,1}$) with a recovery rate k_r (Figure 5.9) as a result of the j+1 nt RNA transcript cleavage. For instance, in Figure 5.9, this pause recovery pathway takes a 1-bp backtracked Pol II molecule ($\mathrm{TEC}_{n+1,-1}$) to the posttranslocated state ($\mathrm{TEC}_{n-1,1}$) after a 2-nt RNA transcript cleavage (red arrow). The corresponding configurations of Pol II transcription elongation complexes are highlighted in Figure 5.9. In the presence of TFIIS, two possible mechanisms exist to rescue the enzyme from a backtracked pause state: the return of Pol II to transcription competent on-pathway state by diffusion, or the TFIIS-stimulated endonucleolytic cleavage. While Pol II is in the backtracked state, the forward and backward diffusive movements compete with TFIIS-stimulated backtrack recovery.

Note that the introduction of the rate k_r effectively shortens pause durations as compared with the original model (i.e., where $k_r = 0$ s^{-1}). Accordingly, a pause can no longer be described as a random walk of the enzyme along DNA with return to the pre-translocated state; consequently, the probability density of pause durations can no longer be described by Equation 5.3. To determine the rate k_r, we performed a kinetic Monte Carlo simulation of the kinetic mechanism shown in Figure 5.9. Fitting the experimental pause durations and densities to our model, we extracted a recovery rate $k_r = 0.4 \pm 0.2$ s^{-1}. Figure 5.4A compares the experimental pause duration distribution in the presence of TFIIS with the theoretical prediction obtained with this k_r value. This simulation shows that the decreased pause density observed in the presence of TFIIS can be rationalized by the increased fraction of pauses that become shorter than our experimental detection limit (1 s). The extracted value of k_r rate is on the same order of magnitude with the endonucleolytic rate $k_{clv} = 0.1$ s^{-1} recently estimated [34].

In the case of TFIIF, we observed both reductions in the pause density for Pol II under opposing and assisting forces, and also reductions in pause duration only under assisting force conditions. The reduced pause density can in principle be explained either by an increased forward translocation rate (k_1), a decreased rate of pause entry (k_{b1}), or both (see Equation 5.4). From the pause density observed in the presence of TFIIF, we estimate that if this factor only affected k_{b1}, then this rate would have to decrease from 8.0 s^{-1} to 5.7 s^{-1}. If, on the other hand, TFIIF only increased k_1, then k_1 would have to rise from 88 s^{-1} to 124 s^{-1}. Note that these values are force-dependent, and reported here are average rates extrapolated to zero force. Also note that the changes in k_1 required to explain the reduced pause density would not lead to an observable change in pause-free velocity, given the variance of our data. Moreover, changes to either or both of these rates would not explain the decrease in pause duration observed under assisting loads. For this observation, our model suggests various possible scenarios. With respect to the shorter pause durations observed in the presence of TFIIF under assisting forces, our model suggests various possible scenarios. Mechanistically, shorter pause durations could arise from a higher rate of forward diffusion (k_f), a reduced rate of backward diffusion (k_b), a combination of the two, or a higher rate of intrinsic diffusion (k_0). At present, we cannot distinguish among these possibilities.

5.5 Discussion

5.5.1 The quantitive effects of TFIIS

Biochemical characterizations of TFIIS have shown that it enhances the intrinsic endonucleolytic activity of Pol II. Here, we observed that TFIIS stimulates transcription elongation by shortening the durations of transcriptional pauses, without affecting the pause-free velocity. From these observations, we have measured the TFIIS-stimulated rate of recovery from a paused state (k_r) to be 0.4 ± 0.2 s^{-1}. The TFIIS-stimulated recovery positions the enzyme in the posttranslocated state, from which Pol II can resume active elongation, eliminating the required diffusional search of the enzyme for the 3′-end of the nascent RNA. Hence, by modulating the pausing dynamics, TFIIS strongly increases the efficiency of transcription elongation. Note that our assay does not measure RNA transcript cleavage directly, but instead it monitors transcription elongation resumption. Therefore, the TFIIS-induced recovery described by the rate k_r does not necessarily equal the rate of cleavage of the nascent RNA at Pol II's active site. The pause recovery process may also include other rates such as diffusion of the cleaved RNA transcript out of the enzyme and conformational changes of Pol II required to resume active transcription elongation. Notice also that the kinetic model presented here considers an average rate of recovery k_r without accounting for its possible dependence on the backtracked distance, the DNA sequence, or the external applied force. In addition, the model supposes that every TFIIS molecule in the studied TFIIS-Pol II complexes is active. Thus, the extracted rate k_r is likely a lower bound estimate for the TFIIS-stimulated rate of recovery.

Table 5.1: **General characteristics of transcription of Pol II in the presence of TFIIS and TFIIF.**

Observable	***F* (pN)**	**Pol II**	**+TFIIS**	**+TFIIF**	**+TFIIF/TFIIS**
Dwell times (ms/bp)	-3.5 to -8	176 ± 23	129 ± 17	122 ± 10	106 ± 16
Pause density (pauses/bp)	-4 to -7	0.0237 ± 0.0049	0.0164 ± 0.0031	0.0176 ± 0.0036	0.0114 ± 0.0019
Dwell times before the nucleosome (ms/bp)	5 to 8	98 ± 22	51 ± 10	47 ± 11	74 ± 26
Dwell times in the nucleosome (ms/bp)	5 to 8	253 ± 57	172 ± 70	195 ± 56	164 ± 54
Nucleosomal passage (%, [N passed/N total])	N/A	63 [59/93]	74 [42/57]	72 [23/32]	77 [18/23]

Data are shown as mean ± SEM N is the number of single-molecule transcription trajectories in each condition.

5.5.2 The effects of TFIIF in Pol II pausing regulation

The single-molecule experiments presented here allowed us to separate pauses from active elongation of Pol II. This separation reveals that the effect of TFIIF is primarily in regulating the pausing phase by decreasing the frequency at which the enzyme enters a pause and by shortening their durations in a force dependent manner. In the presence of TFIIF, the pauses become shorter under assisting loads (Figure 5.4C), but they remain unchanged under opposing loads (Figure 5.4A). It is possible that the high propensity of the enzyme to backtrack under opposing forces partly masks the effect of TFIIF. Using the previously proposed backtracking kinetic model for transcriptional pauses [54, 13, 40, 31, 158], we can express the reduced pause durations in the presence of TFIIF in terms of a reduced backward diffusion rate (k_b), an increased forward diffusion rate (k_f), a combination of both, or an increased in the intrinsic diffusion rate (k_0). Our results do not allow us to discern between these possible scenarios. TFIIF has been shown to help stabilize the short RNA-DNA hybrid during the initiation phase [50]. A similar stabilization of RNA-DNA hybrid during the elongation phase could constitute the mechanism by which TFIIF prevents pause entering and shortens pause durations. Specifically, the TFIIF-enhanced stability of the RNA-DNA hybrid disfavors fraying of the 3′-end from the active site of the enzyme; thus, forward translocation is favored over pausing. As Pol II begins to backtrack, the opening of such hybrid is difficult, which should lead to a decrease in k_{b1} and k_b. Similarly, if the closure of the RNA-DNA hybrid near the active site during backtracking is favored by TFIIF, the k_f would increase. Perhaps, the effect of TFIIF could be equally distributed between all of these processes.

Alternatively, the observation that TFIIF shortens the durations of pauses only under assisting loads could indicate the existence of pauses that are not associated with enzyme backtracking. It has been shown that the bacterial polymerase can exist in a non-backtracked paused state characterized by an open clamp conformation [146]. Since TFIIF has been proposed to stabilize the closed clamp conformation of Pol II [50], it is possible that under assisting load this factor could both decrease the probability of pausing by preventing entering into an analogous (non-backtracked) paused state for the eukaryotic enzyme, and shorten its lifetime in this state.

5.5.3 Mechanistic interference of the two factors

As each factor enhances transcription elongation in a different manner, combining both TFIIS and TFIIF could in principle significantly improve the efficiency of transcription elongation both on bare and nucleosomal DNA as has been described for the human system [88]. However, we did not observe significant enhancement in the presence of both TFIIS and TFIIF simultaneously relative to each factor alone. Since TFIIF reduces the probability of pausing, its presence would be expected to weaken the effect of TFIIS, whose activity necessarily depends on a backtracked complex. This mechanistic "interference" may explain why the presence of both factors does not result in a quantitatively additive effect. When both

factors are present with nucleosomal DNA, we observed a weak enhancement of nucleosomal passage relative to when either factor is present in isolation. Thus, with the yeast system, we do not observe the quantitatively additive effect described in a previous bulk study of human nucleosomal transcription elongation when both TFIIS and TFIIF were present in the medium [88].

In summary, our studies provide a quantitative kinetic model of the mechanism through which transcription factors TFIIS and TFIIF affect the elongation process and highlight the importance of pausing for transcription regulatory in the cell. The molecular mechanisms we describe for these transcription factors show that they have evolved to limit the time Pol II remains in a catalytically inactive, paused state. Elucidating the function, at the single molecule level, of other transcription factors known to play a role in transcription elongation will be of central importance to ultimately understand the detailed mechanisms through which Pol II activity is regulated.

5.6 Materials and methods

5.6.1 Preparation of proteins and assay

Yeast transcription factor TFIIS (Δ1-113) with a 6x-His tag was expressed in *E.coli* BL21 (DE3), and purified using a HisTrap HP column (GE healthcare). The purified protein was dialyzed against 50 mM NaCl, 50 mM Tris (pH 8.0), and 10% glycerol for storage. The expression plasmid was a gift from the C. Kane laboratory (Department of Molecular and Cell Biology, University of California, Berkeley, CA). Two recombinant subunits of yeast TFIIF (Tfg1-Tfg2) were expressed and purified as previously reported [156].

For experiments in the presence of transcription factors, 2 μM of TFIIF and/or 2 μM of TFIIS was mixed with 1 mM NTP and introduced to the enzyme in the fluidic chamber to restart transcription.

5.6.2 Monte Carlo simulation

The kinetic mechanism shown in Figure 5.9 was simulated for saturating NTP conditions. In such a case, the on-pathway kinetic mechanism (green boxes in Figure 5.9) reduces to two irreversible steps with rates k_1 and k_3, corresponding to forward elongation and catalysis, respectively [31]. Thus, when Pol II is in the elongation-competent state ($\mathrm{TEC}_{n,0}$), it can either translocate with rate k_1 to the posttranslocated state $\mathrm{TEC}_{n,1}$ (and subsequently to state $\mathrm{TEC}_{n+1,0}$ with rate k_3) or enter a backtracked pause with rate k_{b1}. During a pause the enzyme diffuses back and forth with force-biased rate constants k_b and k_f, respectively. In the presence of TFIIS, a backtracked pause recovery step with rate k_r can take Pol II from any backtracked state to the corresponding posttranslocated state $\mathrm{TEC}_{n,1}$ (red arrows, Figure 5.9). Simulations were carried out at a force of 5 pN, which is the average force of the pauses analyzed.

Chapter 6

Conclusions and Future Directions

In this dissertation, I have presented the work that characterizes the dynamics of RNA polymerase II transcription through a single nucleosome, and dissects the contribution of the elements of the nucleosomal barrier—the histone tails, the specific histone-DNA interactions, and the DNA sequence, using a single-molecule optical tweezers based assay. Since the nucleosome selectively biases the forward translocation of the polymerase, we employed the nucleosomal barrier as a tool to specifically perturb forward translocation of a transcribing Pol II and characterize its kinetics. The resulting model has been extended to describe the effects on the dynamics of transcription elongation of a point mutation on the trigger loop element of Pol II, transcription factors IIS, IIF and the combination of the two factors. We modeled the nucleotide addition cycle of Pol II in a linear, non-branched ratchet mechanism in which translocation is one of the rate-limiting steps, and further determined the major kinetic parameters in the elongation cycle. These rates provide complementary dynamics information for snapshots of Pol II structures obtained from various x-ray crystallographic and electron microscopic studies, thus putting us one step closer to describe the full molecular trajectory of a complex molecular machine like RNA polymerase II.

Our experiments on the trigger loop mutant (E1103G) Pol II attempted to further characterize the roles of a specific element near the active site. We found that this loop affects multiple parts of both the on- and off-pathway of transcription elongation, including translocation, catalysis and pausing. On the other hand, transcription factors IIS and IIF mainly influence the pausing phase of transcription elongation. These examples illustrate the application of the slow-translocation, linear Brownian ratchet mechanism of transcription elongation. Furthermore, the slow translocation of the polymerase may explain the consensus pause-induced sequence discovered from genome-wide mapping of bacterial RNA polymerase (RNAP) occupancies with base-pair resolution [138]. With translocation rate of RNAP being one of the rate-limiting steps, the strength of DNA base pairing interactions that RNAP must break during the elongation contribute significantly to the overall rate and even induce pausing of the enzyme.

6.1 Topographical map of transcription through the nucleosome

A closer examination of the transcription trajectories reveals many interesting features. Instead of constructing histograms, I attempted to further extract the details of position-dependent information from dwell times and more accurately represent the data via the kernel density estimator [27]. Typically, we represent position-dependent data with a histogram, where each data point is sorted into a bin of certain size. However, the histogram depends on the bin size, discards uncertainty information of the data, and represents the data in a discrete and discontinuous manner. In a kernel density plot, every data point is represented as a Gaussian centered at the value of the data-point with a width given by the uncertainty of the data-point value.

To construct the kernel density estimate and extract useful information, the transcription trajectories must align very well. Our precision in determining the absolute position of Pol II on DNA was not as good as our accuracy in measuring position changes of Pol II on DNA. The final length of the tether before the enzyme released the DNA template was often slightly shorter than the expected runoff length. We attribute this discrepancy to the calibration errors in our experiment, because we obtained a sharp runoff band from our bulk biochemical assay where we measured the length of the RNA. Therefore, we aligned the traces that transcribed through 85% of the total DNA length to both the stall site and the expected runoff length (Chapter 2, Materials and methods). A collection of the aligned position-time transcription traces was analyzed with the kernel density method, using an arbitrary bandwidth of 3 (Figure 6.1). The kernel density estimates of individual traces were added. The peaks in this residence time plot show the locations on the DNA that the polymerase spent longer time transcribing, which can be interpreted at pause sites. High-resolution gel-based biochemical assays have described two strong pause sites as yeast Pol II transcribes through a nucleosome on the 601 NPS at position +15 nt and +45 nt from the beginning of the NPS, which correspond to position 3479 and 3509 in our single-molecule assay [14]. In addition to the strongest pause site at the beginning of the NPS, we also observed smaller pause sites after the nucleosomal dyad, or even after the NPS (Figure 6.1, red). These peaks could indicate additional pause sites, arising from the DNA sequence or the periodicity of the octamer-DNA interactions. Nonetheless, we cannot rule out that they result from misalignment of the transcription traces.

In addition to the histograms of mean dwell time (Figures 6.2A and C), we applied the same method of kernel density estimate to the bare DNA transcription by the wild-type Pol II (Figure 6.2B) and the bare DNA and nucleosomal DNA transcription by E1103G Pol II (Figure 6.2D). The peak locations from both the histograms and the kernel density estimates agree well with each other. Note that the magnitude of the dwell time plots cannot be compared directly, as the kernel density estimate shown here is the sum rather than the average of all the traces. Both variants of Pol II appear to pause at position ∼3300, which could indicate a specific pausing sequence in the DNA arising from strong DNA base pairing

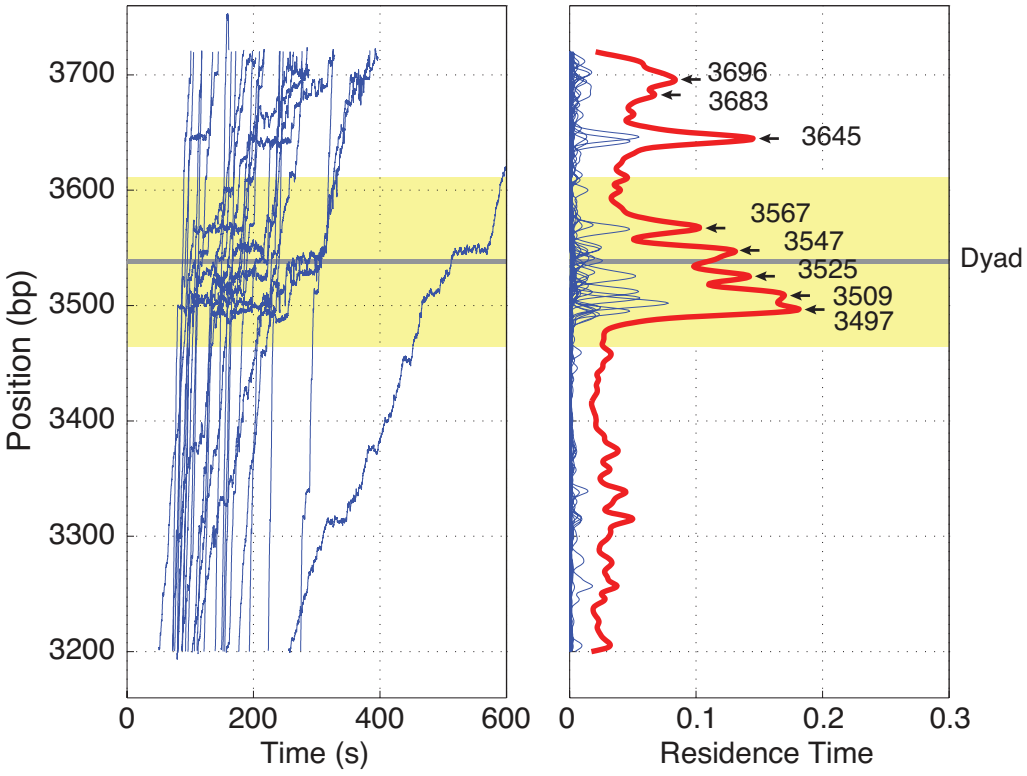

Figure 6.1: **Kernel density estimation of dwell times.** (Left) Single molecule transcription traces of the WT pol II through the nucleosome. (Right) The kernel density estimate constructed from the data on the left. The sum of the kernel density estimate (red) shows that pol II spends longer time at those positions on average. The yellow shaded region represents the 601 NPS.

interactions in front of Pol II, a lack of RNA secondary structure behind the polymerase, or a combination of both. We previously assumed that the prominent pause site in the middle of the NPS is coming from an amplification of sequence-dependent pauses by the nucleosome. However, the kernel density estimate (Figure 6.2B) shows that the pause sites for the bare and nucleosomal DNA of the wild-type are not aligned. They appear to be aligned in the histograms because of binning of the data. Moreover, the E1103G Pol II does not recognize this pause site in the middle of the NPS (Figures 6.2C and D). We learned that the ability of the E1103G pol II to backtrack is lower than that of the wild-type; thus, the pause site in the wild-type transcription could arise from backtracking of the enzyme due to a lack of RNA secondary structure. Another interesting feature that appears in the kernel density plots is the $\sim$15 bp periodicity of pause sites in the NPS of both the wild-type and the E1103G Pol II (Figures 6.2B and D). Interestingly, another investigation by Jin *et al.*, which employed

E. coli RNA polymerase as a surrogate, observed a ∼10 bp periodicity before reaching the nucleosomal dyad [60]. The results of these two experiments cannot be compared directly because: (1) they were performed at different ionic strengths, which are known to affect the fluctuations of DNA around the nucleosome strongly, (2) the prokaryotic and eukaryotic enzymes may have different intrinsic properties, and (3) our experiments used recombinantly expressed yeast histones, whereas Jin *et al.* used purified HeLa histones [60]. Nonetheless, it appears that the ∼5-bp periodicity of the histone-DNA interactions in the nucleosome [48] is strongly related to the location of RNA polymerases' pauses. An experiment at higher resolution and precision will certainly clarify this issue.

As mentioned earlier, the precision in determining the absolute position of the enzyme on the DNA template is crucial to extract meaningful information from the data. To draw accurate information from the kernel density estimates, additional improvements of location determination and trace alignment are required. One method that can be implemented is to incorporate well-defined pause sites on the template as fiducial markers, and align these locations in all the traces. A caveat to this method is that all Pol II molecules must stop at these sites long enough to be distinguished from other nonspecific locations. Such well-characterized pause sites for eukaryotic Pol II are not known. Perhaps, the identified pause sites from *E. coli* RNA polymerase, such as *his* site, *ops* site, or the consensus pause element identified from whole-genome sequencing [138] can be tested.

If the kernel density estimates of the dwell times along the template accurately represent the pausing pattern of the polymerase in the nucleosome (Figures 6.2B,D), it is clear that the pausing pattern of Pol II in the nucleosome is asymmetrically distributed. The pauses before the dyad are stronger than those at the dyad or after. The DNA-histones interaction map from a single-molecule unzipping study also shows that the strengths of the interactions are dependent on unzipping directions [48]. Therefore, it is interesting to transcribe a nucleosome in the reverse direction and map the pausing pattern of the enzyme. This experiment will further elucidate the details of nucleosomal transcription, identify location of pause sites and measure their strengths along the nucleosome.

6.2 Mechanics of nucleosomes and internucleosome interactions

Previous work from both the Bustamante lab and others have shown that when exerting force on the two DNA ends, a single nucleosome unfolds in two transitions: (1) the outer turn of the DNA unwraps at low force ∼3-5 pN, and (2) the inner turn of the DNA unwraps from the octamer surface at higher force with a very broad distribution ∼10-35 pN (Figure 6.3) [93, 115]. The force where these transitions occur depends on the species and types of histones used, as well as the buffer components. Moreover, the first transition is usually seen as a gradual transition, which implies a non-cooperative peeling of the outer-turn DNA from the histones. The presence of a clear low-force transition depends not only on the

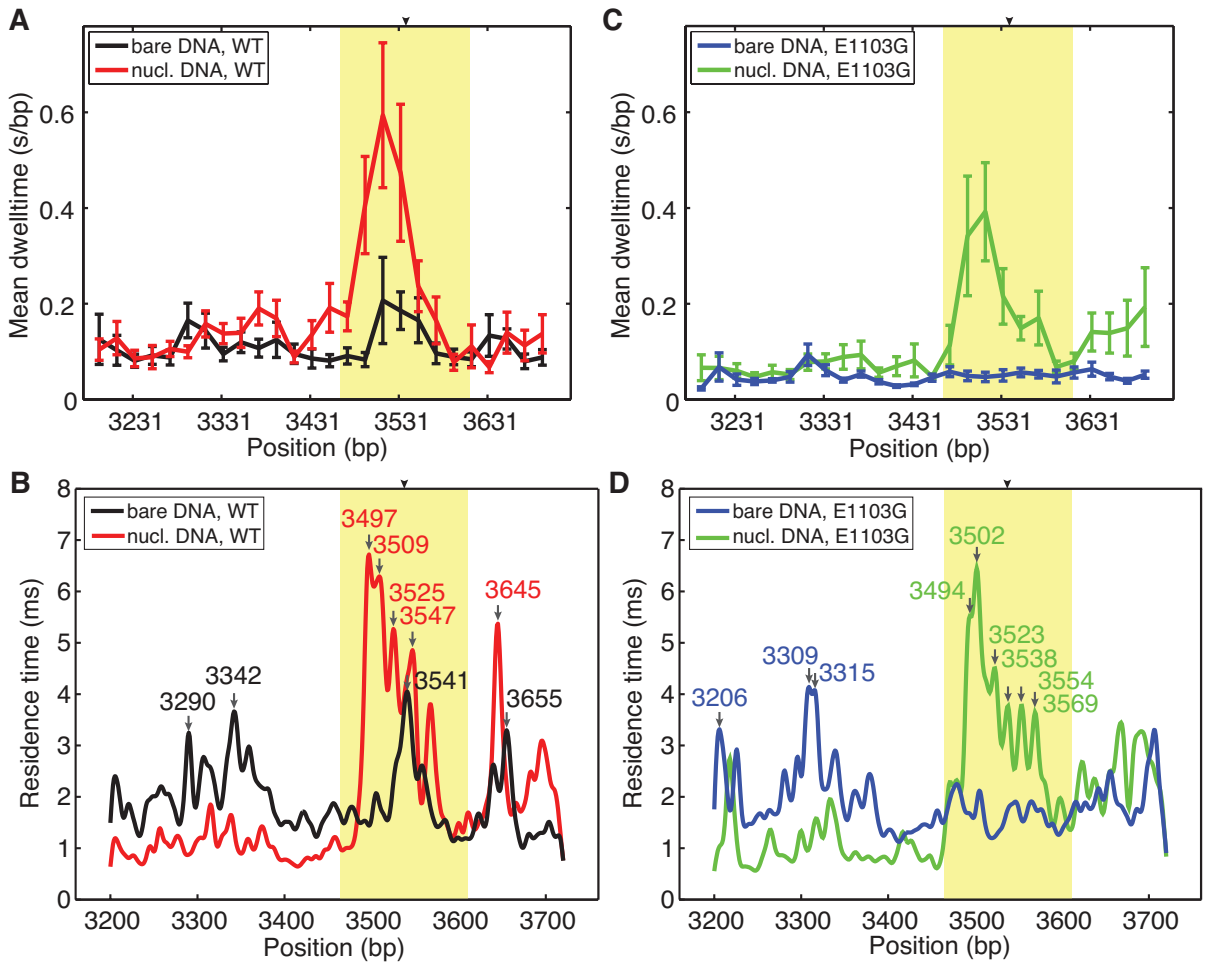

Figure 6.2: **Kernel density estimation of dwell times along the template.** (**A**) Histograms of the mean dwell times of the WT Pol II on bare DNA (black) and on nucleosomal DNA (red) with the bin size of 20 bp. (**B**) Kernel density estimates of the same data in (A) the peak positions are labeled. (**C**) Histograms of the mean dwell times of the E1103G Pol II on bare DNA (blue) and on nucleosomal DNA (green) with the bin size of 20 bp. (**D**) Kernel density estimates of the same data in (C) the peak positions are labeled. A black arrow on the top of the figures marks the position of the nucleosomal dyad. The yellow shaded region indicates the NPS (between position 3464 and 3611). The number above each peak indicates their positions.

factors mentioned above, but also the torsional state of the DNA. Under applied torque, the frequency of observing a clear low-force transition increases [115].

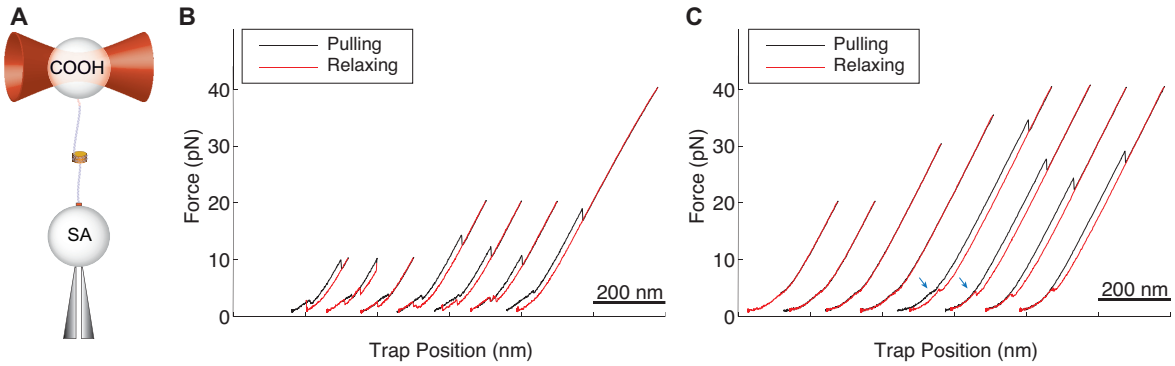

Figure 6.3: **Nucleosome pulling experiments.** (**A**) Experimental geometry for nucleosome pulling experiments. One end of the DNA is ligated to a carboxylated bead cross-linked to a short DNA with a ligatable overhang. One bead is held in an optical trap, while another is held in a micropipette by suction. (**B**) Example force-extension curves from nucleosome pulling experiments, which show a clear transition at low force. (**C**) Another example, where a transition at low force appears as a non-cooperative transition, rather than a clear rip. Note that this transition disappears in the subsequent pulling cycle after the first high-force transition appears (blue arrows). The pulling curves are colored in black and the relaxing are colored in red.

These pulling experiments on single nucleosomes will provide insights on the magnitude of the nucleosomal barrier that obstructs processes that require access to the genetic materials. Example of such processes are transcription, DNA replication, and repair. However, in cells, nucleosomes are folded into a higher-order structure of chromatin fibers, which further condense the DNA and add additional barriers. A key parameter in internucleosome interactions is the length of linker DNA between adjacent nucleosomes. In cells, the linker DNA size varies between 20-60 bp according to species and this number fluctuates from nucleosome to nucleosome [23]. A crystal structure of tetranucleosome array containing 20 bp linker indicated that a trinucleosome is the minimum structure required to attain higher order chromatin structures and nucleosomes spaced two positions apart (i.e., with one nucleosome between them) interact (Figure 6.4) [114]. The linker DNA can be considered to be a torsionally stiff spring linking nucleosomes. Since DNA has a periodicity of 10.5 bp/turn, changes of just one base pair in the linker length will change the relative angle orientation between adjacent nucleosomes by ∼34 degrees, and significantly alter the internucleosomal spatial relationships in higher-order chromatin structure (Figure 6.4). Thus, it is interesting to characterize the interactions between nucleosomes as a function of linker DNA length in a trinucleosome array.

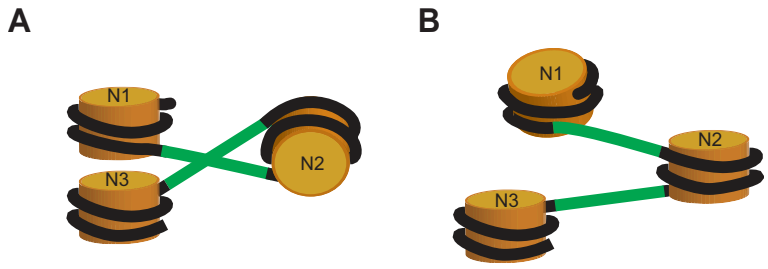

Figure 6.4: **Cartoon of trinucleosome arrays** (**A**) A trinucleosome array where the flat faces of the first (N1) and the third (N3) nucleosomes could be interacting in their parallel orientation. (**B**) Due to torsional stiffness of DNA, changing the length of the linker DNA (green) could affect the orientation of the nucleosomes and disrupt the interactions.

To this end, I have prepared DNA construct for di- and tri-nucleosomes with linker length of 20, 25, 42, 47, and 63 bp between the 601 NPS. These constructs are designed such that they would position the third nucleosome either parallel or perpendicular to the first one. We expect that internucleosome interactions would be stronger in the parallel constructs, whereas the perpendicular constructs would not exhibit interactions between nucleosomes. Characterization of these minimal arrays both structurally and dynamically will be important to understand the nature of the barrier to transcription on chromatin.

Moreover, elements of the nucleosome, especially the histone tails, could play important roles in regulating internucleosome interactions. Their effects could be quantify by characterizing the stability of nucleosomal arrays comprising of modified histones .

Another important structural protein that constitutes the higher-order structure of chromatin fibers is linker histone protein, H1. It is known that the occupancy of histone H1 affects the expression level of genes [122]. Although the affinity of H1 to DNA is not as strong as that of other histones (H2A, H2B, H3, and H4), this weaker interaction is the basis of another mechanism for control of gene expression: displacement of H1 by various transcription regulation proteins has been shown to directly affect gene expression [164]. However, the contribution of the linker histone to the stability and dynamics of the nucleosome has not been addressed. A quantitive study of H1's interaction to the nucleosome and its contribution to the structural integrity of chromatin fibers will be important to describe the mechanism and the dynamics of gene expression controls.

6.3 Concluding thoughts

Made up of distinct, strictly regulated steps, transcription is highly regulated by many cis- (e.g. DNA sequence, the TL element of Pol II) and trans-acting factors (e.g. nucleosome, transcription factors IIS and IIF). The emergence of single-molecule studies over the last two decades has provided many mechanistic insights into this process, especially for the dynamics of transcription elongation.

These studies have also begun to provide a coherent picture of the energy flow during the transcription cycle. In particular, the processes through which the energy released in the binding and hydrolysis of NTPs is converted into mechanical movement of the enzyme along the template and the generation of force. As a result, scientists are now adding rich dynamic information derived from carefully designed single-molecule experiments to the increasing number of crystal structures depicting snapshots of RNA polymerases in different states of their kinetic cycle. Single-molecule techniques, such as optical-tweezers, have proven their power to obtain a more realistic picture of the complex dynamics and control of transcription, without sacrificing the precision and quantitative description afforded by other in vitro studies. These studies have laid a solid foundation towards our understanding of DNA-directed molecular machines and brought us closer to obtaining "moving pictures" of transcription and gene expression regulation as occurred in vivo.

Since the human epigenome has been published [111], further research aiming to figure out the roles of specific modifications on both the DNA and histones will be crucial to our understanding of gene regulation in vivo. Some of these modifications may exert direct effects on biological processes, while some serve as a recognition platform to recruit additional factors. Nonetheless, the complexity of biological process is still a gigantic puzzle. These studies only describe the mechanistic insights for transcription elongation and the roles that a few of the cis- and trans-acting factors regulate transcription elongation dynamics. Yet, much of the gene expression regulation occurs at the initiation and termination stages. We are barely touching the surface to understand the complex dynamics of biological processes.

Bibliography

[1] 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*, 438:460–5, 2005. doi: 10.1038/nature04268.

[2] J. Anderson, P. Lowary, and J. Widom. Effects of histone acetylation on the equilibrium accessibility of nucleosomal DNA target sites. *J Mol Bio*, 307(4):977 – 985, 2001. doi: 10.1006/jmbi.2001.4528.

[3] J. Andrecka, R. Lewis, F. Brückner, E. Lehmann, P. Cramer, and J. Michaelis. Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc Natl Acad Sci U S A*, 105(1):135–140, 2008. doi: 10.1073/pnas.0703815105.

[4] D. Angelov, J. M. Vitolo, V. Mutskov, S. Dimitrov, and J. J. Hayes. Preferential interaction of the core histone tail domains with linker dna. *Proc Natl Acad Sci U S A*, 98(12):6599–6604, 2001. doi: 10.1073/pnas.121171498.

[5] K.-J. Armache, S. Mitterweger, A. Meinhart, and P. Cramer. Structures of complete RNA polymerase II and its subcomplex, rpb4/7. *J Bio Chem*, 280(8):7131–7134, 2005. doi: 10.1074/jbc.M413038200.

[6] I. Artsimovitch and R. Landick. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A*, 97:7090–5, 2000. doi: 10.1073/pnas.97.13.7090.

[7] A. Ashkin. Acceleration and trapping of particles by radiation pressure. *Physical Review Letters*, 24(4):156–159, 1970. doi: 10.1103/PhysRevLett.24.156.

[8] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm, and S. Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Optics letters*, 11(5):288, 1986. doi: 10.1364/OL.11.000288.

[9] L. Bai, A. Shundrovsky, and M. D. Wang. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J Mol Biol*, 344:335–49, 2004. doi: 10.1016/j.jmb.2004.08.107.

[10] L. Bai, R. M. Fulbright, and M. D. Wang. Mechanochemical kinetics of transcription elongation. *Phys Rev Lett*, 98:068103, 2007. doi: 10.1103/PhysRevLett.98.068103.

[11] 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*, 120:183–93, 2005. doi: 10.1016/j.cell.2004.11.045.

[12] L. Bintu, M. Kopaczynska, C. Hodges, L. Lubkowska, M. Kashlev, and C. Bustamante. The elongation rate of RNA polymerase determines the fate of transcribed nucleosomes. *Nat Struct Mol Biol*, 18:1394–9, 2011. doi: 10.1038/nsmb.2164.

[13] L. Bintu, T. Ishibashi, M. Dangkulwanich, Y. Y. Wu, L. Lubkowska, M. Kashlev, and C. Bustamante. Nucleosomal elements that control the topography of the barrier to transcription. *Cell*, 151:738–49, 2012. doi: 10.1016/j.cell.2012.10.009.

[14] V. A. Bondarenko, L. M. Steele, A. Ujvári, D. A. Gaykalova, O. I. Kulaeva, Y. S. Polikanov, D. S. Luse, and V. M. Studitsky. Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. *Mol Cell*, 24(3):469–479, 2006. doi: 10.1016/j.molcel.2006.09.009.

[15] B. Brower-Toland, W. D. A, R. M. Fulbright, J. T. Lis, W. L. Kraus, and M. D. Wang. Specific contributions of histone tails and their acetylation to the mechanical stability of nucleosomes. *J Mol Biol*, 346:135–146, 2005. doi: 10.1016/j.jmb.2004.11.056.

[16] F. Brueckner and P. Cramer. Structural basis of transcription inhibition by alpha-amanitin and implications for RNA polymerase II translocation. *Nat Struct Mol Biol*, 15:811–8, 2008. doi: 10.1038/nsmb.1458.

[17] F. Brueckner, J. Ortiz, and P. Cramer. A movie of the RNA polymerase nucleotide addition cycle. *Curr Opin Struct Bio*, 19(3):294–299, 2009. doi: 10.1016/j.sbi.2009.04.005.

[18] D. A. Bushnell, P. Cramer, and R. D. Kornberg. Structural basis of transcription: alpha-Amanitin-RNA polymerase II cocrystal at 2.8 Å resolution. *Proc Natl Acad Sci U S A*, 99(3):1218–22, 2002. doi: 10.1073/pnas.251664698.

[19] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265:1599–600, 1994. doi: 10.1126/science.8079175.

[20] C. Bustamante, Y. R. Chemla, and J. R. Moffitt. High-resolution dual-trap optical tweezers with differential detection: Instrument design. *Cold Spring Harbor Protocols*, 2009(10):pdb.ip73, 2009. doi: 10.1101/pdb.ip73.

[21] P. Čabart, A. Újvári, M. Pal, and D. S. Luse. Transcription factor TFIIF is not required for initiation by RNA polymerase II, but it is essential to stabilize transcription factor

TFIIB in early elongation complexes. *Proc Natl Acad Sci U S A*, 108(38):15786–15791, 2011. doi: 10.1073/pnas.1104591108.

[22] N. J. Carter and R. A. Cross. Mechanics of the kinesin step. *Nature*, 435(7040): 308–312, 2005. doi: 10.1038/nature03528.

[23] C. E. Castro. Nutrient effects on DNA and chromatin structure. *Annu Rev Nutr*, 7: 407–421, 1987. doi: 10.1146/annurev.nu.07.070187.002203.

[24] B. Cheng and D. H. Price. Properties of rna polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. *J Biol Chem*, 282(30):21901–21912, 2007. doi: 10.1074/jbc.M702936200.

[25] A. C. Cheung and P. Cramer. Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, 471:249–53, 2011. doi: 10.1038/nature09785.

[26] A. C. Cheung and P. Cramer. A movie of RNA polymerase II transcription. *Cell*, 149: 1431–7, 2012. doi: 10.1016/j.cell.2012.06.006.

[27] G. Chistol. *Dissecting the Operating Mechanism of a Biological Motor One Molecule at a Time*. PhD thesis, University of California at Berkeley, 2013.

[28] L. S. Churchman and J. S. Weissman. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330):368–373, 2011. doi: 10.1038/nature09652.

[29] W. W. Cleland. Partition analysis and the concept of net rate constants as tools in enzyme kinetics. *Biochemistry*, 14:3220–4, 1975. doi: 10.1021/bi00685a029.

[30] M. Cojocaru, C. Jeronimo, D. Forget, A. Bouchard, D. Bergeron, P. Côte, G. G. Poirier, J. Greenblatt, and B. Coulombe. Genomic location of the human RNA polymerase II general machinery: evidence for a role of TFIIF and Rpb7 at both early and late stages of transcription. *Biochem J*, 409(1):139–147, 2008. doi: 10.1042/BJ20070751.

[31] 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*, 2:e00971, 2013. doi: 10.7554/eLife.00971.

[32] R. J. Davenport, G. J. Wuite, R. Landick, and C. Bustamante. Single-molecule study of transcriptional pausing and arrest by *E. coli* RNA polymerase. *Science*, 287:2497–500, 2000. doi: 10.1126/science.287.5462.2497.

[33] M. Depken, E. A. Galburt, and S. W. Grill. The origin of short transcriptional pauses. *Biophys J*, 96:2189–93, 2009. doi: 10.1016/j.bpj.2008.12.3918.

[34] M. Depken, J. M. R. Parrondo, and S. W. Grill. Intermittent transcription dynamics for the rapid production of long transcripts of high fidelity. *Cell reports*, 5(2):521–530, 2013. doi: 10.1016/j.celrep.2013.09.007.

[35] S. Elsässer. *Dissecting mammalian replication-independent chromatin assembly: biochemical and structural studies on the H3.3-specific histone chaperones HIRA and DAXX.* PhD thesis, Rockefeller University, 2012.

[36] C. M. English, M. W. Adkins, J. J. Carson, M. E. A. Churchill, and J. K. Tyler. Structural basis for the histone chaperone activity of asf1. *Cell*, 127(3):495–508, 2006. doi: 10.1016/j.cell.2006.08.047.

[37] M. Feig and Z. F. Burton. RNA polymerase II with open and closed trigger loops: active site dynamics and nucleic acid translocation. *Biophysical journal*, 99(8):2577–2586, 2010. doi: 10.1016/j.bpj.2010.08.010.

[38] 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. *Proc Natl Acad Sci U S A*, 99:11682–7, 2002. doi: 10.1073/pnas.142417799.

[39] J. E. Foster, S. F. Holmes, and D. A. Erie. Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell*, 106:243–52, 2001. doi: 10.1016/S0092-8674(01)00420-2.

[40] 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*, 446:820–3, 2007. doi: 10.1038/nature05701.

[41] M. A. Ghazy, S. A. Brodie, M. L. Ammerman, L. M. Ziegler, and A. S. Ponticelli. Amino acid substitutions in yeast TFIIF confer upstream shifts in transcription initiation and altered interaction with RNA polymerase II. *Mol Cell Bio*, 24(24):10975–10985, 2004. doi: 10.1128/MCB.24.24.10975-10985.2004.

[42] A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science*, 292(5523):1876–1882, 2001. doi: 10.1126/science.1059495.

[43] X. Q. Gong, C. Zhang, M. Feig, and Z. F. Burton. Dynamic error correction and regulation of downstream bubble opening by human RNA polymerase II. *Mol Cell*, 18:461–70, 2005. doi: 10.1016/j.molcel.2005.04.011.

[44] S. J. Greive, B. A. Dyer, S. E. Weitzel, J. P. Goodarzi, L. J. Main, and P. H. von Hippel. Fitting experimental transcription data with a comprehensive template-dependent modular kinetic model. *Biophys J*, 101(5):1166–1174, 2011. doi: 10.1016/j.bpj.2011.07.043.

[45] S. J. Greive, J. P. Goodarzi, S. E. Weitzel, and P. H. von Hippel. Development of a "modular" scheme to describe the kinetics of transcript elongation by RNA polymerase. *Biophys J*, 101(5):1155–1165, 2011. doi: 10.1016/j.bpj.2011.07.042.

[46] R. Guajardo and R. Sousa. A model for the mechanism of polymerase translocation. *J Mol Biol*, 265:8–19, 1997. doi: 10.1006/jmbi.1996.0707.

[47] M. Guermah, V. B. Palhan, A. J. Tackett, B. T. Chait, and R. G. Roeder. Synergistic functions of SII and p300 in productive activator-dependent transcription of chromatin templates. *Cell*, 125(2):275–286, 2006. doi: 10.1016/j.cell.2006.01.055.

[48] M. A. Hall, A. Shundrovsky, L. Bai, R. M. Fulbright, J. T. Lis, and M. D. Wang. High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat Struct Mol Biol*, 16:124–9, 2009. doi: 10.1038/nsmb.1526.

[49] P. J. Hawryluk, A. Ujvári, and D. S. Luse. Characterization of a novel RNA polymerase II arrest site which lacks a weak 3' RNA-DNA hybrid. *Nucleic Acids Res*, 32(6):1904–1916, 2004. doi: 10.1093/nar/gkh505.

[50] Y. He, J. Fang, D. J. Taatjes, and E. Nogales. Structural visualization of key steps in human transcription initiation. *Nature*, 495(7442):481–486, 2013. doi: 10.1038/nature11991.

[51] T. R. Hebbes, A. L. Clayton, A. W. Thorne, and C. Crane-Robinson. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. *EMBO J*, 13(8):1823–1830, 1994.

[52] 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*, 125:1083–94, 2006. doi: 10.1016/j.cell.2006.04.032.

[53] K. M. Herbert, W. J. Greenleaf, and S. M. Block. Single-molecule studies of RNA polymerase: motoring along. *Annu Rev Biochem*, 77:149–76, 2008. doi: 10.1146/annurev.biochem.77.073106.100741.

[54] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, 325:626–8, 2009. doi: 10.1126/science.1172926.

[55] S. F. Holmes and D. A. Erie. Downstream DNA sequence effects on transcription elongation. allosteric binding of nucleoside triphosphates facilitates translocation via a ratchet motion. *J Biol Chem*, 278:35597–608, 2003. doi: 10.1074/jbc.M304496200.

[56] F.-K. Hsieh, F. Michael, A. Újvári, V. M. Studitsky, and D. S. Luse. Histone Sin mutations promote nucleosome traversal and histone displacement by RNA polymerase II. *EMBO reports*, 11:705–710, 2010. doi: 10.1038/embor.2010.113.

[57] M. Imashimizu, M. L. Kireeva, L. Lubkowska, D. Gotte, A. R. Parks, J. N. Strathern, and M. Kashlev. Intrinsic translocation barrier as an initial step in pausing by RNA polymerase II. *J Mol Biol*, 425:697–712, 2013. doi: 10.1016/j.jmb.2012.12.002.

[58] M. G. Izban and D. S. Luse. Factor-stimulated RNA polymerase II transcribes at physiological elongation rates on naked DNA but very poorly on chromatin templates. *J Bio Chem*, 267(19):13647–13655, 1992.

[59] M. G. Izban and D. S. Luse. SII-facilitated transcript cleavage in RNA polymerase II complexes stalled early after initiation occurs in primarily dinucleotide increments. *J Bio Chem*, 268(17):12864–12873, 1993.

[60] 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. *Nat Struc Mol Bio*, 17:745–752, 2010. doi: 10.1038/nsmb.1798.

[61] C. D. Kaplan, K. M. Larsson, and R. D. Kornberg. The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. *Mol Cell*, 30:547–56, 2008. doi: 10.1016/j.molcel.2008.04.023.

[62] R. G. Keene, A. Mueller, L. Robert, and L. London. Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N- and c-*myc*genes. *Nucleic Acids Res*, 27(15):3173–3182, 1999. doi: 10.1093/nar/27.15.3173.

[63] H. Kettenberger, K.-J. Armache, and P. Cramer. Architecture of the RNA polymerase II-TFIIS complex and implications for mRNA cleavage. *Cell*, 114(3):347–357, 2003. doi: 10.1016/S0092-8674(03)00598-1.

[64] H. Kettenberger, K.-J. Armache, and P. Cramer. Complete rna polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell*, 16(6):955 – 965, 2004. doi: 10.1016/j.molcel.2004.11.040.

[65] D. A. Khaperskyy, M. L. Ammerman, R. C. Majovski, and A. S. Ponticelli. Functions of *Saccharomyces cerevisiae* TFIIF during transcription start site utilization. *Mol Cell Bio*, 28(11):3757–3766, 2008. doi: 10.1128/MCB.02272-07.

[66] J. Kim, M. Guermah, and R. G. Roeder. The human PAF1 complex acts in chromatin transcription elongation both independently and cooperatively with SII/TFIIS. *Cell*, 140(4):491–503, 2010. doi: 10.1016/j.cell.2009.12.050.

[67] M. Kireeva, M. Kashlev, and Z. F. Burton. Translocation by multi-subunit RNA polymerases. *Biochim Biophys Acta*, 1799:389–401, 2010. doi: 10.1016/j.bbagrm.2010.01.007.

[68] M. L. Kireeva and M. Kashlev. Mechanism of sequence-specific pausing of bacterial RNA polymerase. *Proc Natl Acad Sci U S A*, 106:8900–5, 2009. doi: 10.1073/pnas.0900407106.

[69] M. L. Kireeva, W. Walter, V. Tchernajenko, V. Bondarenko, M. Kashlev, and V. M. Studitsky. Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. *Mol Cell*, 9(3):541–552, 2002. doi: 10.1016/S1097-2765(02)00472-0.

[70] M. L. Kireeva, L. Lubkowska, N. Komissarova, and M. Kashlev. Assays and affinity purification of biotinylated and nonbiotinylated forms of double-tagged core RNA polymerase II from saccharomyces cerevisiae. In S. Adhya and S. Garges, editors, *RNA Polymerases and Associated Factors, Part C*, volume 370 of *Methods in Enzymology*, pages 138 – 155. Academic Press, 2003. doi: 10.1016/S0076-6879(03)70012-3.

[71] M. L. Kireeva, B. Hancock, G. H. Cremona, W. Walter, V. M. Studitsky, and M. Kashlev. Nature of the nucleosomal barrier to RNA polymerase II. *Mol Cell*, 18:97–108, 2005. doi: 10.1016/j.molcel.2005.02.027.

[72] M. L. Kireeva, Y. A. Nedialkov, G. H. Cremona, Y. A. Purtov, L. Lubkowska, F. Malagon, Z. F. Burton, J. N. Strathern, and M. Kashlev. Transient reversal of RNA polymerase II active site closing controls fidelity of transcription elongation. *Mol Cell*, 30:557–66, 2008. doi: 10.1016/j.molcel.2008.04.017.

[73] N. Komissarova and M. Kashlev. RNA polymerase switches between inactivated and activated states by translocating back and forth along the DNA and the RNA. *J Biol Chem*, 272:15329–38, 1997. doi: 10.1074/jbc.272.24.15329.

[74] W. J. Koopmans, R. Buning, T. Schmidt, and J. van Noort. spFRET using alternating excitation and FCS reveals progressive DNA unwrapping in nucleosomes. *Biophys J*, 97:195–204, 2009. doi: 10.1016/j.bpj.2009.04.030.

[75] W. J. A. Koopmans, T. Schmidt, and J. van Noort. Nucleosome immobilization strategies for single-pair FRET microscopy. *Chemphyschem*, 9(14):2002–2009, 2008. doi: 10.1002/cphc.200800370.

[76] H. Koyama, T. Ito, T. Nakanishi, N. Kawamura, and K. Sekimizu. Transcription elongation factor SII maintains transcriptional fidelity and confers oxidative stress resistance. *Genes to Cells*, 8(10):779–788, 2003. doi: 10.1046/j.1365-2443.2003.00677.x.

[77] H. Koyama, T. Ito, T. Nakanishi, and K. Sekimizu. Stimulation of RNA polymerase II transcript cleavage activity contributes to maintain transcriptional fidelity in yeast. *Genes to Cells*, 12(5):547–559, 2007. doi: 10.1111/j.1365-2443.2007.01072.x.

[78] N. J. Krogan, M. Kim, S. H. Ahn, G. Zhong, M. S. Kobor, G. Cagney, A. Emili, A. Shilatifard, S. Buratowski, and J. F. Greenblatt. RNA polymerase II elongation factors of *Saccharomyces cerevisiae*: a targeted proteomics approach. *Mol Cell Bio*, 22 (20):6979–6992, 2002. doi: 10.1128/MCB.22.20.6979-6992.2002.

[79] W. Kruger, C. L. Peterson, A. Sil, C. Coburn, G. Arents, E. N. Moudrianakis, and I. Herskowitz. Amino acid substitutions in the structured domains of histones H3 and H4 partially relieve the requirement of the yeast SWI/SNF complex for transcription. *Genes Dev*, 9(22):2770–2779, 1995. doi: 10.1101/gad.9.22.2770.

[80] O. I. Kulaeva, F.-K. Hsieh, and V. M. Studitsky. RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. *Proc Natl Acad Sci U S A*, 107 (25):11325–11330, 2010. doi: 10.1073/pnas.1001148107.

[81] R. Landick. The regulatory roles and mechanism of transcriptional pausing. *Biochem Soc Trans*, 34:1062–6, 2006. doi: 10.1042/BST0341062.

[82] M. H. Larson, J. Zhou, C. D. Kaplan, M. Palangat, R. D. Kornberg, R. Landick, and S. M. Block. Trigger loop dynamics mediate the balance between the transcriptional fidelity and speed of RNA polymerase II. *Proc Natl Acad Sci U S A*, 109:6555–60, 2012. doi: 10.1073/pnas.1200939109.

[83] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, 2007. doi: 10.1016/j.cell.2007.01.015.

[84] G. Li, M. Levitus, C. Bustamante, and J. Widom. Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struc Mol Bio*, 12(1):46–53, 2005. doi: 10.1038/nsmb869.

[85] P. T. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Bio*, 276(1): 19–42, 1998. doi: 10.1006/jmbi.1997.1494.

[86] K. Luger and T. J. Richmond. The histone tails of the nucleosome. *Curr Opin in Genet Dev*, 8(2):140 – 146, 1998. ISSN 0959-437X. doi: 10.1016/S0959-437X(98)80134-2.

[87] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature*, 389(6648):251–260, 1997. doi: 10.1038/38444.

[88] D. S. Luse, L. C. Spangler, and A. Újvári. Efficient and rapid nucleosome traversal by RNA polymerase II depends on a combination of transcript elongation factors. *J Bio Chem*, 286(8):6040–6048, 2011. doi: 10.1074/jbc.M110.174722.

[89] F. Malagon, M. L. Kireeva, B. K. Shafer, L. Lubkowska, M. Kashlev, and J. N. Strathern. Mutations in the saccharomyces cerevisiae RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics*, 172:2201–9, 2006. doi: 10.1534/genetics.105.052415.

[90] A. M. Malinen, M. Turtola, M. Parthiban, L. Vainonen, M. S. Johnson, and G. A. Belogurov. Active site opening and closure control translocation of multisubunit RNA polymerase. *Nucleic Acids Res*, 40:7442–51, 2012. doi: 10.1093/nar/gks383.

[91] D. O. Maoileidigh, V. R. Tadigotla, E. Nudler, and A. E. Ruckenstein. A unified model of transcription elongation: what have we learned from single-molecule experiments? *Biophys J*, 100:1157–66, 2011. doi: 10.1016/j.bpj.2010.12.3734.

[92] Y. X. Mejia, H. Mao, N. R. Forde, and C. Bustamante. Thermal probing of e. coli RNA polymerase off-pathway mechanisms. *J Mol Biol*, 382:628–37, 2008. doi: 10.1016/j.jmb.2008.06.079.

[93] S. Mihardja, A. J. Spakowitz, Y. Zhang, and C. Bustamante. Effect of force on mononucleosomal dynamics. *Proc Natl Acad Sci U S A*, 103(43):15871–15876, 2006. doi: 10.1073/pnas.0607526103.

[94] J. R. Moffitt, Y. R. Chemla, D. Izhaky, and C. Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proc Natl Acad Sci U S A*, 103(24):9006–9011, 2006. doi: 10.1073/pnas.0603342103.

[95] J. R. Moffitt, Y. R. Chemla, S. B. Smith, and C. Bustamante. Recent advances in optical tweezers. *Annu Review of Biochem*, 77:205–228, 2008. doi: 10.1146/annurev.biochem.77.043007.090225.

[96] J. R. Moffitt, Y. R. Chemla, K. Aathavan, S. Grimes, P. J. Jardine, D. L. Anderson, and C. Bustamante. Intersubunit coordination in a homomeric ring ATPase. *Nature*, 457(7228):446–450, 2009. doi: 10.1038/nature07637.

[97] U. M. Muthurajan, Y. Bao, L. J. Forsberg, R. S. Edayathumangalam, P. N. Dyer, C. L. White, and K. Luger. Crystal structures of histone Sin mutant nucleosomes reveal altered protein–DNA interactions. *EMBO J*, 23(2):260–271, 2004. doi: 10.1038/sj.emboj.7600046.

[98] Y. A. Nedialkov, X. Q. Gong, S. L. Hovde, Y. Yamaguchi, H. Handa, J. H. Geiger, H. Yan, and Z. F. Burton. NTP-driven translocation by human RNA polymerase II. *J Biol Chem*, 278:18303–12, 2003. doi: 10.1074/jbc.M301103200.

[99] Y. A. Nedialkov, E. Nudler, and Z. F. Burton. RNA polymerase stalls in a post-translocated register and can hyper-translocate. *Transcription*, 3(5):260–269, 2012. doi: 10.4161/trns.22307.

[100] K. Neuman and S. Block. Optical trapping. *Review of Scientific Instruments*, 75:2787, 2004. doi: 10.1063/1.1785844.

[101] K. C. Neuman and A. Nagy. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Meth*, 5(6):491–505, 2008. doi: 10.1038/nmeth.1218.

[102] E. Nudler. RNA polymerase backtracking in gene regulation and genome instability. *Cell*, 149:1438–45, 2012. doi: 10.1016/j.cell.2012.06.003.

[103] E. Nudler, A. Mustaev, E. Lukhtanov, and A. Goldfarb. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell*, 89:33–41, 1997. doi: 10.1016/S0092-8674(00)80180-4.

[104] G. Orphanides, G. LeRoy, C. H. Chang, D. S. Luse, and D. Reinberg. FACT, a factor that facilitates transcript elongation through nucleosomes. *Cell*, 92(1):105–116, 1998. doi: 10.1016/S0092-8674(00)80903-4.

[105] M. Palangat and R. Landick. Roles of RNA:DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II. *J Mol Biol*, 311(2): 265–282, 2001. ISSN 0022-2836. doi: 10.1006/jmbi.2001.4842.

[106] M. A. Parra, D. Kerr, D. Fahy, D. J. Pouchnik, and J. J. Wyrick. Deciphering the roles of the histone H2B N-terminal domain in genome-wide transcription. *Mol Cell Bio*, 26(10):3842–3852, 2006. doi: 10.1128/MCB.26.10.3842-3852.2006.

[107] R. U. Protacio, G. Li, P. T. Lowary, and J. Widom. Effects of histone tail domains on the rate of transcriptional elongation through a nucleosome. *Mol Cell Bio*, 20: 8866–8878, 2000. doi: 10.1128/MCB.20.23.8866-8878.2000.

[108] Q. Ren and M. A. Gorovsky. The nonessential H2A N-terminal tail can function as an essential charge patch on the H2A.Z variant N-terminal tail. *Mol Cell Bio*, 23(8): 2778–2789, 2003. doi: 10.1128/MCB.23.8.2778-2789.2003.

[109] G. Rhodes and M. J. Chamberlin. Ribonucleic acid chain elongation by escherichia coli ribonucleic acid polymerase. i. isolation of ternary complexes and the kinetics of elongation. *J Biol Chem*, 249:6675–83, 1974.

[110] T. J. Richmond and C. A. Davey. The structure of DNA in the nucleosome core. *Nature*, 423(6936):145–150, 2003. doi: 10.1038/nature01595.

[111] Roadmap Epigenomics Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y.-C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R. Dixon,

K.-H. Farh, S. Feizi, R. Karlic, A.-R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. M. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L.-H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meoptissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. Integrative analysis of 111 reference human epigenomes. *Nature*, 518(7539): 317–330, 2015.

[112] I. Samkurashvili and D. S. Luse. Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. *J Biol Chem*, 271(38):23495–23505, 1996. doi: 10.1074/jbc.271.38.23495.

[113] A. Saunders, J. Werner, E. D. Andrulis, T. Nakayama, S. Hirose, D. Reinberg, and J. T. Lis. Tracking FACT and the RNA polymerase II elongation complex through chromatin in vivo. *Science*, 301(5636):1094–1096, 2003. doi: 10.1126/science.1085712.

[114] T. Schalch, S. Duda, D. F. Sargent, and T. J. Richmond. X-ray structure of a tetranucleosome and its implications for the chromatin fibre. *Nature*, 436(7047):138–141, 2005. doi: 10.1038/nature03686.

[115] M. Y. Sheinin, M. Li, M. Soltani, K. Luger, and M. D. Wang. Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. *Nat Commun*, 4:2579, 2013. doi: 10.1038/ncomms3579.

[116] S. Sigurdsson, A. B. Dirac-Svejstrup, and J. Q. Svejstrup. Evidence that transcript cleavage is essential for RNA polymerase II transcription and cell viability. *Mol Cell*, 38(2):202–210, 2010. doi: 10.1016/j.molcel.2010.02.026.

[117] R. T. Simpson. Structure of chromatin containing extensively acetylated H3 and H4. *Cell*, 13(4):691–699, 1978. doi: 10.1016/0092-8674(78)90219-2.

[118] R. J. Sims, R. Belotserkovskaya, and D. Reinberg. Elongation by RNA polymerase II: the short and long of it. *Genes Dev*, 18(20):2437–2468, 2004. doi: 10.1101/gad.1235904.

[119] N. Suka, Y. Suka, A. A. Carmen, J. Wu, and M. Grunstein. Highly specific antibodies determine histone acetylation site usage in yeast heterochromatin and euchromatin. *Mol Cell*, 8(2):473 – 479, 2001. ISSN 1097-2765. doi: 10.1016/S1097-2765(01)00301-X.

[120] K. Svoboda and S. M. Block. Biological applications of optical forces. *Annu Rev Biophys and Biomol Struct*, 23(1):247–285, 1994. doi: 10.1146/annurev.bb.23.060194.001335.

[121] J. F. Sydow and P. Cramer. RNA polymerase fidelity and transcriptional proofreading. *Curr Opin Struct Biol*, 19(6):732–739, 2009. doi: 10.1016/j.sbi.2009.10.009.

[122] H. J. Szerlong and J. C. Hansen. Nucleosome distribution and linker DNA: connecting nuclear function to dynamic chromatin structure. *Biochem Cell Biol*, 89(1):24–34, 2011. doi: 10.1139/O10-139.

[123] V. R. Tadigotla, D. O. Maoileidigh, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein. Thermodynamic and kinetic modeling of transcriptional pausing. *Proc Natl Acad Sci U S A*, 103:4439–44, 2006. doi: 10.1073/pnas.0600508103.

[124] L. Tan, S. Wiesler, D. Trzaska, H. C. Carney, and R. O. Weinzierl. Bridge helix and trigger loop perturbations generate superactive RNA polymerases. *J Biol*, 7:40, 2008. doi: 10.1186/jbiol98.

[125] S. Tan, T. Aso, R. C. Conaway, and J. W. Conaway. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J Bio Chem*, 269(41):25684–91, 1994.

[126] D. Temiakov, N. Zenkin, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, E. Kashkina, M. Savkina, S. Zorov, V. Nikiforov, N. Igarashi, N. Matsugaki, S. Wakatsuki, K. Severinov, and D. G. Vassylyev. Structural basis of transcription inhibition by antibiotic streptolydigin. *Mol Cell*, 19:655–66, 2005. doi: 10.1016/j.molcel.2005.07.020.

[127] A. Thastrom, P. T. Lowary, and J. Widom. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods*, 33:33–44, 2004. doi: 10.1016/j.ymeth.2003.10.018.

[128] M. J. Thomas, A. A. Platas, and D. K. Hawley. Transcriptional fidelity and proofreading by RNA polymerase II. *Cell*, 93(4):627 – 637, 1998. doi: 10.1016/S0092-8674(00)81191-5.

[129] I. Toulokhonov, J. Zhang, M. Palangat, and R. Landick. A central role of the RNA polymerase trigger loop in active-site rearrangement during transcriptional pausing. *Mol Cell*, 27:406–19, 2007. doi: 10.1016/j.molcel.2007.06.008.

[130] A. Újvári, F.-K. Hsieh, S. W. Luse, V. M. Studitsky, and D. S. Luse. Histone N-terminal tails interfere with nucleosome traversal by RNA polymerase II. *J Biol Chem*, 283(47):32236–32243, 2008. doi: 10.1074/jbc.M806636200.

[131] A. Újvári, M. Pal, and D. S. Luse. The functions of TFIIF during initiation and transcript elongation are differentially affected by phosphorylation by casein kinase 2. *J Biol Chem*, 286(26):23160–23167, 2011. doi: 10.1074/jbc.M110.205658.

[132] D. G. Vassylyev, M. N. Vassylyeva, J. Zhang, M. Palangat, I. Artsimovitch, and R. Landick. Structural basis for substrate loading in bacterial RNA polymerase. *Nature*, 448:163–8, 2007. doi: 10.1038/nature05931.

[133] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R.-R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. The sequence of the human

genome. *Science*, 291(5507):1304–1351, 2001. doi: 10.1126/science.1058040.

[134] M. Vignali, A. H. Hassan, K. E. Neely, and J. L. Workman. Atp-dependent chromatin-remodeling complexes. *Mol Cell Bio*, 20(6):1899–1910, 2000. doi: 10.1128/MCB.20.6.1899-1910.2000.

[135] K. Voltz, J. Trylska, N. Calimet, J. C. Smith, and J. Langowski. Unwrapping of nucleosomal DNA ends: a multiscale molecular dynamics study. *Biophys J*, 102:849–58, 2012. doi: 10.1016/j.bpj.2011.11.4028.

[136] P. H. von Hippel and E. Delagoutte. A general model for nucleic acid helicases and their "coupling" within macromolecular machines. *Cell*, 104(2):177–190, 2001. doi: 10.1016/S0092-8674(01)00203-3.

[137] P. H. von Hippel and Z. Pasman. Reaction pathways in transcript elongation. *Biophys Chem*, 101-102:401–23, 2002. doi: 10.1016/S0301-4622(02)00160-6.

[138] 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*, 344(6189):1285–1289, 2014. doi: 10.1126/science.1253458.

[139] C. Walmacq, A. C. Cheung, M. L. Kireeva, L. Lubkowska, C. Ye, D. Gotte, J. N. Strathern, T. Carell, P. Cramer, and M. Kashlev. Mechanism of translesion transcription by RNA polymerase II and its role in cellular resistance to DNA damage. *Mol Cell*, 46:18–29, 2012. doi: 10.1016/j.molcel.2012.02.006.

[140] W. Walter, M. L. Kireeva, V. M. Studitsky, and M. Kashlev. Bacterial polymerase and yeast polymerase II use similar mechanisms for transcription through nucleosomes. *J Bio Chem*, 278(38):36148–36156, 2003. doi: 10.1074/jbc.M305647200.

[141] D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, and R. D. Kornberg. Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell*, 127:941–54, 2006. doi: 10.1016/j.cell.2006.11.023.

[142] D. Wang, D. A. Bushnell, X. Huang, K. D. Westover, M. Levitt, and R. D. Kornberg. Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*, 324:1203–6, 2009. doi: 10.1126/science.1168729.

[143] H. Wang and G. Oster. Ratchets, power strokes, and molecular motors. *Applied Physics A*, 75(2):315–323, 2002. doi: 10.1007/s003390201340.

[144] X. Wang, S. C. Moore, M. Laszckzak, and J. Ausió. Acetylation increases the alpha-helical content of the histone tails of the nucleosome. *J Biol Chem*, 275(45):35013–35020, 2000. doi: 10.1074/jbc.M004998200.

[145] L. Wasserman. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics. Springer, 2004.

[146] A. Weixlbaumer, K. Leon, R. Landick, and S. A. Darst. Structural basis of transcriptional pausing in bacteria. *Cell*, 152:431–41, 2013. doi: 10.1016/j.cell.2012.12.020.

[147] K. D. Westover, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell*, 119:481–9, 2004. doi: 10.1016/j.cell.2004.10.016.

[148] H. R. Widlund, J. M. Vitolo, C. Thiriet, and J. J. Hayes. DNA sequence-dependent contributions of core histone tails to nucleosome stability: differential effects of acetylation and proteolytic tail removal. *Biochemistry*, 39(13):3835–3841, 2000. doi: 10.1021/bi991957l.

[149] M. Wind and D. Reines. Transcription elongation factor SII. *Bioessays*, 22(4):327–36, 2000.

[150] J. Wittmeyer, A. Saha, and B. Cairns. DNA translocation and nucleosome remodeling assays by the RSC chromatin remodeling complex. *Methods Enzymol*, 377:322–43, 2004. doi: 10.1016/S0076-6879(03)77020-7.

[151] J. Wu, D. E. Awrey, A. M. Edwards, J. Archambault, and J. D. Friesen. In vitro characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. *Proc Natl Acad Sci U S A*, 93(21):11552–11557, 1996.

[152] A. Xayaphoummine, T. Bucher, and H. Isambert. Kinefold web server for rna/dna folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res*, 33(suppl 2):W605–W610, 2005. doi: 10.1093/nar/gki447.

[153] F. Xu, A. V. Colasanti, Y. Li, and W. K. Olson. Long-range effects of histone point mutations on DNA remodeling revealed from computational analyses of sin-mutant nucleosome structures. *Nucleic Acids Res*, 38(20):6872–6882, 2010. doi: 10.1093/nar/gkq506.

[154] T. D. Yager and P. H. von Hippel. A thermodynamic analysis of RNA transcript elongation and termination in escherichia coli. *Biochemistry*, 30:1097–118, 1991. doi: 10.1021/bi00218a032.

[155] Q. Yan, R. J. Moreland, J. W. Conaway, and R. C. Conaway. Dual roles for transcription factor IIF in promoter escape by RNA polymerase II. *J Biol Chem*, 274(50): 35668–35675, 1999.

[156] C. Yang, D. A. Khaperskyy, M. Hou, and A. S. Ponticelli. Improved methods for expression and purification of *Saccharomyces cerevisiae* TFIIF and TFIIH; identification of a functional *Escherichia coli* promoter and internal translation initiation within

the n-terminal coding region of the TFIIF TFG1 subunit. *Protein Expr Purif*, 70(2): 172–178, 2010. doi: 10.1016/j.pep.2009.09.021.

[157] Y. W. Yin and T. A. Steitz. The structural mechanism of translocation and helicase activity in t7 RNA polymerase. *Cell*, 116:393–404, 2004. doi: 10.1016/S0092-8674(04) 00120-5.

[158] B. Zamft, L. Bintu, T. Ishibashi, and C. Bustamante. Nascent RNA structure modulates the transcriptional dynamics of RNA polymerases. *Proc Natl Acad Sci U S A*, 109:8948–53, 2012. doi: 10.1073/pnas.1205063109.

[159] C. Zhang and Z. F. Burton. Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J Mol Bio*, 342(4):1085–1099, 2004. doi: 10.1016/j.jmb.2004.07.070.

[160] C. Zhang, H. Yan, and Z. F. Burton. Combinatorial control of human RNA polymerase II (RNAP II) pausing and transcript cleavage by transcription factor IIF, hepatitis delta antigen, and stimulatory factor II. *J Bio Chem*, 278(50):50101–50111, 2003. doi: 10.1074/jbc.M307590200.

[161] J. Zhang, M. Palangat, and R. Landick. Role of the RNA polymerase trigger loop in catalysis and pausing. *Nat Struc Mol Bio*, 17(1):99–104, 2010. doi: 10.1038/nsmb.1732.

[162] Y. Zhang, C. L. Smith, A. Saha, S. W. Grill, S. Mihardja, S. B. Smith, B. R. Cairns, C. L. Peterson, and C. Bustamante. DNA translocation and loop formation mechanism of chromatin remodeling by SWI/SNF and RSC. *Mol Cell*, 24:559–68, 2006. doi: 10.1016/j.molcel.2006.10.025.

[163] 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. *Mol Cell*, 44:635–46, 2011. doi: 10.1016/j.molcel.2011.09.018.

[164] J. Zlatanova, C. Seebart, and M. Tomschik. The linker-protein network: control of nucleosomal DNA accessibility. *Trends Biochem Sci*, 33(6):247–253, 2008. doi: 10. 1016/j.tibs.2008.04.001.

Appendix A

Protocols

A.1 Preparation of nucleosomes

We expressed *S. Cerevisiae* and *X. Laevis* histones recombinantly in *E. coli* BL21(DE3). The expression plasmids for 6x-His tagged *X. Laevis* histones were gifts from J. Chin laboratory in MRC laboratory of molecular biology, Cambridge, UK. The details protocol for expression and purification of these histones were explained in details elsewhere [150, 35]. It is crucial to check the amount of bacterial DNA in the histone preparation by measuring the ratio of absorbance at 260/280. The contaminated DNA will shield the interaction of the histones with target DNA, thus resulting in poor nucleosome formation. The ratio should be lower than 0.5, an indication that the DNA contamination is minimal. For *S. Cerevisiae* histones, our A260:280 ratio is around 1, which indicates that we have DNA contamination from the expression. With *X. Laevis* histones, the preparation contains less DNA contamination. Histidine tags on *X. Laevis* histones can be removed with TEV protease, divided into 3-4 mg aliquots, and lyophilized for storage.

A.1.1 Octamer reconstitution

To reconstitute nucleosomes, each histone: H2A, H2B, H3, and H4 were solubilized in 6M Guanidinium HCl, 20 mM Tris pH 7.5 and 5 mM DTT for 2 hours. We mixed the histones in equimolar ratio, using 10% excess of H2A and H2B, and adjusted the final protein concentration to 1 mg/ml. We dialyzed the mixture using 3500 MWCO dialysis membrane against 1 L of 2 M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, and 5 mM β-mercaptoethanol for 3 hours and changed the buffer for an overnight dialysis. The octamer was purified using a Superdex 200 HR (10/30) size exclusion column. and collect 0.3 mL fractions. The peak fractions were analyzed by 15% SDS-PAGE. The octamer fractions were combined and dialyze against 10 mM Tris pH 7.5, 2 M NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, and 50% glycerol for storage at -20 °C.

A.1.2 Nucleosome reconstitution

We reconstituted nucleosomes by salt gradient dialysis. Histone octamers and 601-NPS containing DNA were mixed at equimolar ratio in 2M NaCl and slowly dialyzed into low salt buffer (see below). Since the majority of the interactions between the histones and DNA are electrostatic, the charges shield the interactions at high salt. When the salt is removed slowly, the DNA and histones interact and the octamer loads on to the 601 NPS region in the DNA as it is the most thermodynamically stable position.

High salt buffer (400 mL)

Component	**Concentration (mM)**
NaCl	2000
Tris-HCl (pH 7.9)	10
EDTA	1
DTT	0.5
PMSF	0.1

Low salt buffer (2L)

Component	**Concentration (mM)**
Tris-HCl (pH 7.9)	10
EDTA	1
DTT	0.5
PMSF	0.1

In the final volume of 50 μL, we used 0.2 μM of the 574-bp DNA used in transcription experiments, or 0.1 μM of 1.6 or 3-kbp DNA used in pulling experiments. Small scale dialysis buttons can be created from PCR caps. Previously, we added BSA in our reconstitution to prevent histone loss from sticking to the surface, but it is not crucial. We also found that BSA can form aggregation at high salt and interfere with some other experiments, such as AFM imaging.

After mixing the DNA and histones octamers in appropriate ratio in dialysis buttons, place the buttons in 400 mL of high salt buffer and slowly flow in 1.6 L of the low salt buffer at the rate of 1 mL/min using a peristaltic pump. We kept the volume of the buffer constant by piercing a hole on the side of the container at the starting volume of the dialysis. This step takes about 27 hours to complete. The next step is change the buffer and dialyze against 400 mL of the low salt buffer for 6 hours or overnight.

When the dialysis is finished, collect the sample and centrifuge at 15,000 rcf for 10 mins to remove any aggregates, measure the DNA concentration and check the loading quality

of the nucleosomes by restriction enzyme digestion and run the digested samples on a 5% polyacrylamide-TBE native gel (Figure A.1). The DNA sequence is shown in the next section.

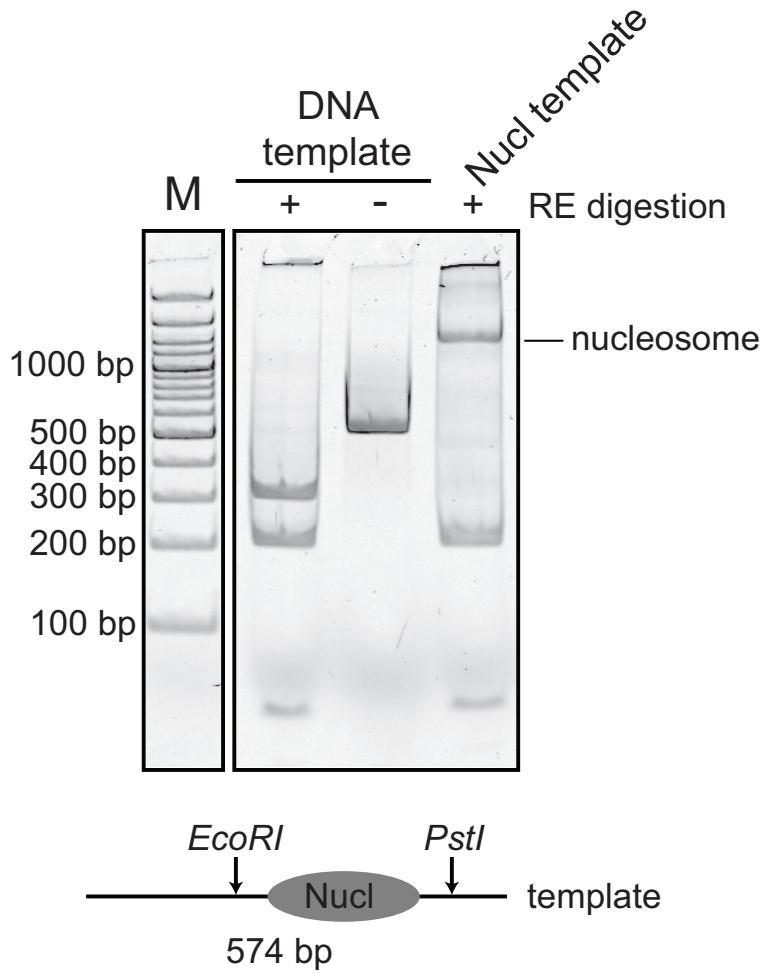

Figure A.1: **Gel electrophoresis assay for nucleosome loading quantification.** For a typical nucleosomal template used for single-molecule transcription experiments, we want close to 100% loading efficiency. The DNA flanking both sides of the DNA were cleaved with restriction enzyme EcoRI and PstI. The first lane is the digest bare DNA. The second lane is the DNA without digestion. The third lane is the same DNA with loaded nucleosomes after restriction enzyme digestion. The observed upward shifted band in the last lane and the disappearance of the 300-bp band indicates that the histones are loaded on the NPS. M is the 100-bp DNA ladder.

A.2 DNA sequences

The sequence of the 601 NPS used used in these experiments differ slightly from the original 601 sequence reported by Lowary and Widom (Figure A.2)[85]. Several residues in the NPS were mutated for restriction enzyme digestion purposes. The sequence of the NPS containing 598-bp DNA template is also provided (Figure A.3).

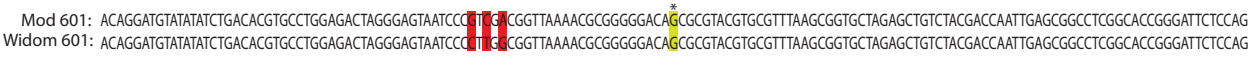

Figure A.2: **Alignment of the NPS sequences.** The mutated bases are highlighted in red. The nucleosomal dyad is highlighted in yellow with the asterisk.

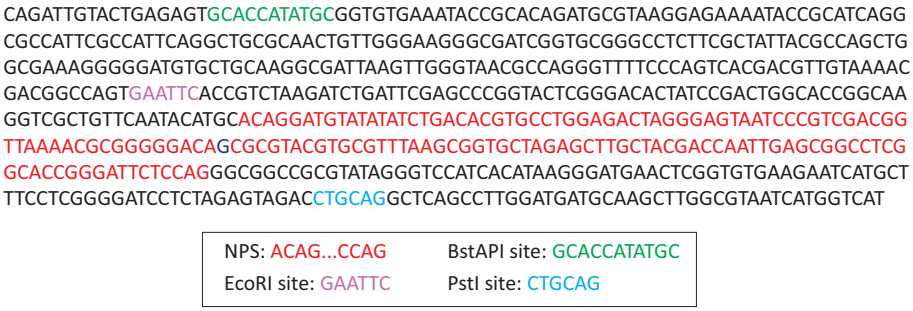

Figure A.3: **Nucleosomal DNA template.** The DNA template amplified from pUC19-N1. The recognition sequences of restriction enzymes used in the experiments are highlighted.

A.3 Additional data analysis

A.3.1 Maximum likelihood estimate for pause duration analysis

We modeled the pause durations of the RNA polymerase as the first-passage times for return to the origin of a Poisson stepper on a one-dimensional lattice [33]. Thus, the probability density function of pause durations is given by:

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \frac{exp[-(k_f + k_b)t]}{t} I_1(2t\sqrt{k_f k_b})$$

where I_1 is the modified Bessel function of the first kind, and k_f and k_b are the force biased diffusion rate constants. We extracted the intrinsic zero-force stepping rate (k_0) by comparing the cumulative distribution function (CDF) of the observed pause durations to the theoretical CDF with various k_0 values, and using the Kolmogorov-Smirnov (KS) test to find the P-value. The k_0 values that yield the P-value greater than 0.05, an indication that

the two distribution are not statistically different with 95% confidence interval, are averaged, and reported as a range of the extracted k_0 values (Figure A.4). For the wild-type Pol II, the k_0 extracted from this method is $1.25 \pm 0.35\ s^{-1}$.

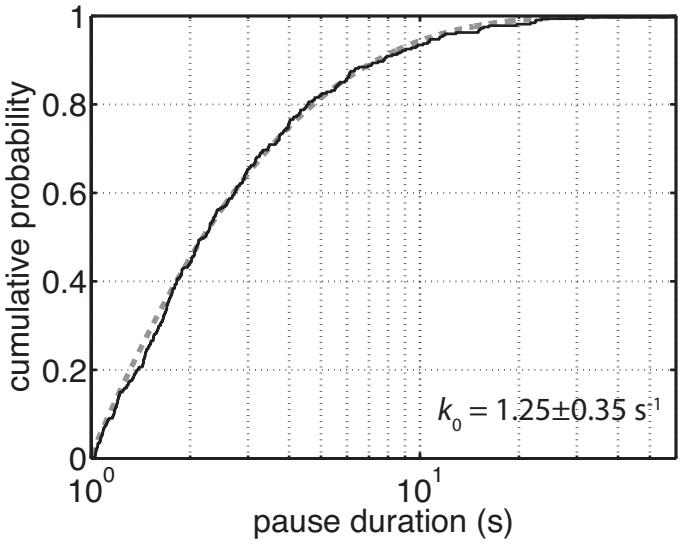

Figure A.4: **Fitting the CDF of the pause durations.** The CDF of the experimental pause durations of the wild-type Pol II is shown in black. The gray dashed line is the theoretical CDF of the k_0 of $1.25 \pm 0.35\ s^{-1}$.

Alternatively, we can extract the k_0 more accurately by fitting the probability density of pause durations through a method of maximum likelihood. By experimental constraints, we only observed pauses (t) between 1-120 s. Therefore, we conditioned the probability distribution as followed:

$$P(t|1 \leq t \leq 120) = \frac{P(t)}{P(1 \leq t \leq 120)} = \frac{\psi(t)}{\int_1^{120} \psi(t)dt}$$

Pause durations data $(t_1, t_2, t_3, ...)$ are independent and identically distributed observations. The unknown k_0 is the parameter for this distribution. To find the value of k_0 through the method of maximum likelihood (L), one first specifies the likelihood function (L):

$$
\begin{aligned}
L(k_0; t_1, ..., t_n) &= P(k_0|t_1, ..., t_n) \\
&= \frac{P(t_1, ..., t_n|k_0)P(k_0)}{P(t_1, ..., t_n)} \\
&\propto P(t_1, ..., t_n|k_0) \\
&= P(t_1|k_0) \times P(t_2|k_0) \times ... \times P(t_n|k_0) \\
&= \prod_{i=1}^{n} P(t_i|k_0)
\end{aligned}
$$

On the third line, we used an argument that there is no prior knowledge on k_0, and the denominator is independent of k_0. On the fourth line, we used the fact that all observations are independent of each other. To simplify the computation, we find a value of k_0 that maximizes the logarithm of the likelihood function (Figure A.5). This method gives the best-fit k_0 of 1.36 s^{-1}, whereas fitting the CDF yields k_0 of 1.25 $\pm$ 0.35 s^{-1}. The 95% confidence bounds of the k_0 estimator [1.36, 1.95] was estimated from bootstrapping method (Figure A.6) [145].

$$
\ln L(k_0; t_1, ..., t_n) = \sum_{i=1}^{n} P(t_i|k_0)
$$

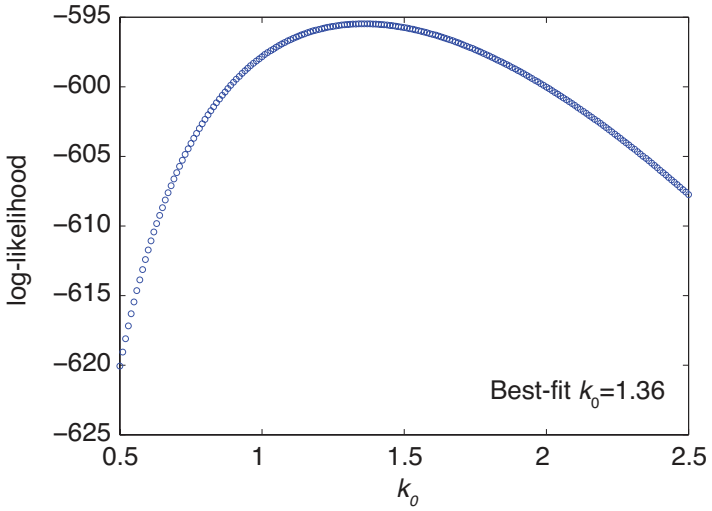

Figure A.5: **Maximum log-likelihood estimator.** The log-likelihood function is plotted as a function k_0. The k_0 value, which maximizes the log-likelihood function is shown.

To illustrate the influence of k_0 on the probability density function, we plotted the theoretical distributions using k_0 values of 0.1 and 10 (Figure A.7). With a higher k_0 value, we

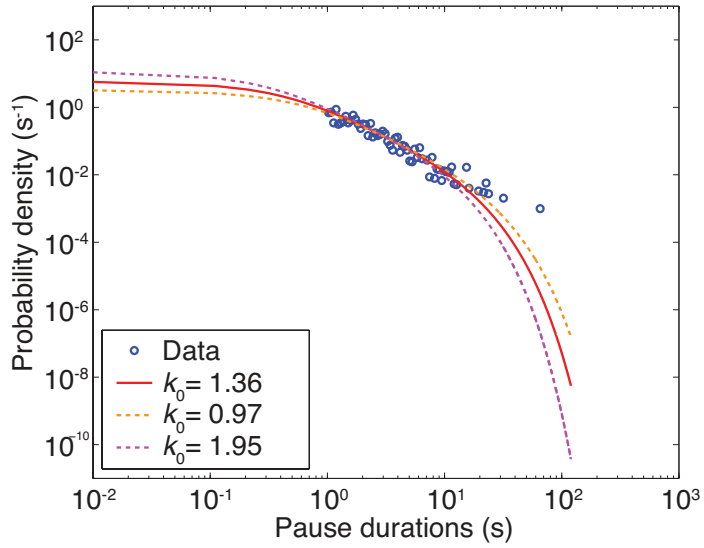

Figure A.6: **Probability density distribution of pause durations.** The experimentally observed pauses is shown in blue circles and the theoretical distribution (ψ(t)) using k_0 of 1.36 as estimated from the maximum likelihood method is shown in red. The 95% confidence interval of the k_0 estimator is [0.97, 1.95], and the distributions from the lower and upper confidence bounds are shown in yellow and pink dashed lines, respectively.

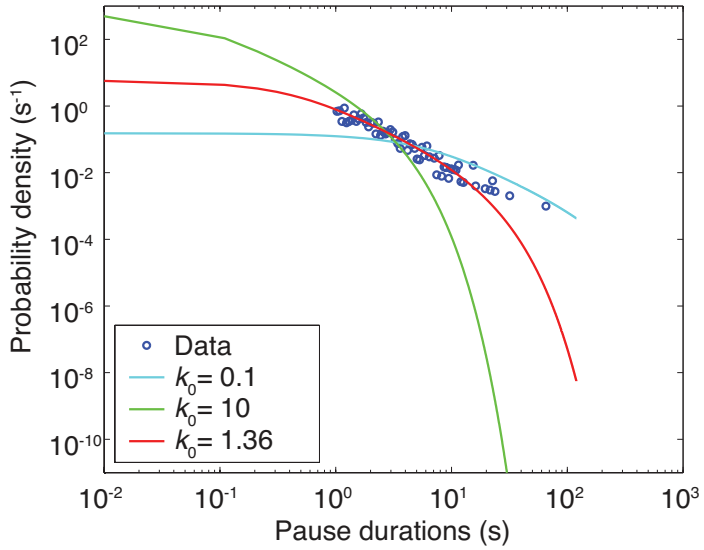

Figure A.7: **Probability density distribution of pause durations with various k_0.** The theoretical distributions of pause durations using the k_0 of 0.1 (blue) and 10 (green) are shown to illustrate the expected shifts of the plot with different values of k_0.

expected a higher fraction of short pauses than observed, while a smaller k_0 of predicts that there should be fewer short pauses.