UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Dynamic Interactions and Molecular Rearrangements Occurring when RNA Polymerase II Meets the Nucleosome

Permalink
https://escholarship.org/uc/item/3kx3312n

Author
Bintu, Lacramioara

Publication Date
2010

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Dynamic Interactions and Molecular Rearrangements Occurring when RNA Polymerase II Meets the Nucleosome

by

Lacramioara Bintu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carlos Bustamante, Chair
Assistant Professor Ahmet Yildiz
Professor Evangelina Nogales de la Morena

Fall 2010

Dynamic Interactions and Molecular Rearrangements Occurring when RNA Polymerase II Meets the Nucleosome

Copyright 2010
by
Lacramioara Bintu

Abstract

Dynamic Interactions and Molecular Rearrangements Occurring when RNA Polymerase II Meets the Nucleosome

by

Lacramioara Bintu

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Carlos Bustamante, Chair

Transcription by RNA Polymerase II (Pol II) represents a major control point for regulation of eukaryotic gene expression. Yet, the mechanistic details and dynamics of a large number of transcriptional regulatory processes are currently unknown. Many of these processes, such as chromatin remodeling and epigenetic silencing, are mediated through nucleosomes, which comprise the repeating units of chromatin. Thus, it is of great interest to investigate the real-time dynamics of Pol II when it encounters a nucleosome, and to determine what happens to nucleosomes upon the passage of a transcribing Pol II.

It has been shown that a single nucleosome is sufficient to halt or greatly slow transcription by Pol II *in vitro*, and factors that restrict transcriptional backtracking by Pol II also relieve nucleosome-induced pauses and arrests. These observations suggest that the influence of the nucleosome is mediated through polymerase backtracking. Unfortunately, the temporal resolution provided by these biochemical studies was not sufficient to provide mechanistic information about the nucleosomal barrier. Using optical tweezers, we studied nucleosomal transcription of single Pol II complexes in real time, and obtained direct evidence for the first time that a nucleosome acts as mechanical fluctuating barrier that both increases the tendency of the polymerase to enter a backtracked pause and slows its recovery from these pauses. These changes in pause durations quantitatively agree with a model where temporary rewrapping of the nucleosomal DNA immediately downstream of the backtracked Pol II prevents enzyme recovery from its paused state. Furthermore, our studies revealed that Pol II does not actively separate the nucleosomal DNA from the surface of the histones, but, instead, acts as a ratchet that rectifies local nucleosomal unwrapping events to gain access to downstream DNA and overcome the nucleosomal barrier.

In vivo, the histone tails are essential for the regulation of gene expression, so we investigated their direct effect on the dynamics of transcription elongation. We found that removal of the tails favors progression of Pol II into the entry region of the nucleosome, by increasing the DNA fluctuations in this region. However, since our data shows that the magnitude of

the barrier to transcription is highest in the central region of the nucleosome, and the tails only affect the entry region of the nucleosome, we investigated what interactions control the strength of the barrier near the nucleosome dyad. To this end, we used nucleosomes with point mutations in the histone-fold domains of H3 and H4, mutations that affect histone-DNA contacts at the dyad and that have been shown to partially relieve the requirement of the chromatin remodeling factor SWI/SNF *in vivo*. We found that these mutations abolish the barrier to transcription in the central region by increasing the local unwrapping rate of the DNA from the surface of the histones near the dyad. We speculate that factors that could bind to the nucleosome and specifically disrupt even a single DNA-histone contact in this region would have a profound effect on transcription.

Nucleosomes are disrupted to varying degrees by transcription elongation, with outcomes ranging from partial loss to complete removal and exchange of histones. *In vitro* studies with the phage SP6 RNA polymerase and RNA Polymerase III have shown that upon transcription the histone octamer moves upstream, while Pol II leads to the formation of a hexamer whose position on DNA is unchanged. The histone transfer process is believed to involve looping of the DNA template, but claims of template looping have so far relied on indirect evidence. Using atomic force microscopy, we obtained direct evidence of looping in the form of polymerase-nucleosome complexes in which the histones bridge the DNA upstream and downstream of the polymerase simultaneously. We also showed that a small fraction of the transcribed nucleosomes moved upstream of their original position. Significantly, we found that the fraction of the transcribed nucleosomes that are remodeled to hexasomes versus the ones that are transferred as intact octamers depends on the speed of elongation. A simple model involving the kinetic competition between the rates of transcription elongation, histone transfer, and histone-histone dissociation quantitatively rationalizes our observations and unifies results obtained with other polymerases.

In conclusion, we show how the finely tuned interplay between polymerase dynamics and nucleosome fluctuations determines the outcome of transcription, and we propose that factors affecting the relative magnitude of these processes provide the physical basis for the regulation of gene expression.

To my parents, Elena and Gheorghe Bintu, and to my teacher Dana Cornea,
who have given me wings to fly.

Contents

List of Figures

List of Tables

List of Symbols

Pol II	RNA Polymerase II
NTPs	nucleotide triphosphates
bp	base pairs
AFM	Atomic Force Microscopy
SEM	Standard Error of the Mean
F	force
T	temperature
k_B	Boltzman constant
k_BT	thermal energy
k_e	rate of Pol II elongation
k_0	rate of Pol II diffusion on DNA while paused
k_b	rate of Pol II backtracking
k_f	rate of Pol II recovery from backtracks
k_u	rate of nucleosome unwrapping
k_w	rate of nucleosome rewrapping
γ_u	probability of finding the nucleosome locally unwrapped
K_w	nucleosome wrapping equilibrium constant
k_{ue}	combined rate of nucleosome unwrapping and net elongation
k_t	rate of histone transfer from downstream to upstream DNA
k_d	rate of dimer dissociation from the histone octamer

Acknowledgments

I would like to thank my advisor, Professor Carlos Bustamante, who has been extremely supportive of me and my work throughout my graduate career. He taught me how to ask the right questions, he constantly coached me on how to express my thoughts clearly, and he always knew how to put this work in the bigger frame. However, the most important teaching he passed onto me is best summarized in the words of William Turner: "You may have heard the world is made up of atoms and molecules, but it's really made up of stories."

The experiments I performed as a graduate student built on work done by previous members of the lab: Jason Choy, who assembled the first incarnation of the optical tweezers instrument I used for my research, Stephan Grill and Eric Glaburt, who shared with me their assays and models for single-molecule transcription of Pol II on bare DNA, and Yongli Zhang who taught me how to work with nucleosomes.

The research presented here was made possible by of a set of very fruitful collaborations. In particular, the work presented in Chapter 2 was a project shared equally with Courtney Hodges, the work in Chapter 3 was performed together with Toyotaka Ishibashi, and the work presented in Chapter 4 resulted from a joint endeavor with Marta Kopaczynska. In addition, this research was performed in collaboration with the laboratory of Dr. Mikhail Kashlev, from the National Cancer Institute. Dr. Kashlev offered constructive criticism on this work throughout the years, Dr. Lucyna Lubovska purified the polymerase, and Dr. Maria Kireeva provided training and advice on the biochemistry of transcription, especially in the initial phases of the project.

Past and current members of the Bustamante Lab have created and maintained an amazing environment that allows both science and friendships to flourish. In addition to the people mentioned above, I am also grateful to Gheorghe Chistol, Manchuta Dangkulwanich, Phillip Elms, Craig Hetherington, Christian Kaiser, Ariel Kaplan, Aathavan Karankuran, Rodrigo Maillard, Yara Mejia, Jeff Moffitt, Bibiana Onoa, Jason Quinlan, Claudio Rivetti, Jae-Yen Shin, Sam Sternberg Melania Strycharska, Patrick Visperas, Anna Wiedmann, Gloria Wu, Shannon Yan, and Brad Zamft for providing help and advice with experiments, for organizing amazing gatherings and trips, and for insightful discussions about life and the universe.

I would also like to thank Professor Rob Phillips for convincing me that I was still fit to do science when I went through the usual graduate school blues, and to Professors Ahmet Yildiz, Evangelina Nogales, and Ignacio Tinocco for their comments and suggestions on this work.

Finally, I am grateful to my beloved husband, Anton Geraschenko, whose affection and support has brought out the best in me throughout the years.

Chapter 1

Introduction

1.1 Evolution and Its Connection to Gene Expression

The last 150 years constituted a revolution in biology and started providing answers to a fundamental human question: ***Where do we come from?***

This ideological revolution started in the middle of the nineteenth century, its most important seed planted in 1859 by the publication of Charles Darwin's book *On the Origin of Species* [1]. In this work, Darwin proposed the theory of evolution based on inheritance and natural selection.

At about the same time, Gregor Mendel was performing experiments on plant hybridization [2, 3], and coming up with quantitative rules for inheritance and even a rudimentary form of the "gene" concept as the unit for inheritance, though under a different name [4]. Unfortunately Mendel's work remained largely ignored, and was rediscovered independently in the 1890s by Carl Correns and Hugo de Vries [4]. Correns also discovered cytoplasmic inheritance transmitted through organelles that are independent of the nucleus (such as chloroplasts or mitochondria), while de Vries introduced the term "mutation," and developed a mutation theory of evolution.

However, it was still not clear at the time exactly how features were passed from parent to progeny. Even though the concept of proteins had been already described in 1838 [5], and nucleic acids were discovered by Friedrich Miescher as early as 1869 [4], it took another century and the work of many scientists to flesh out the connection between nucleic acids, proteins, and inheritance. A major milestone was achieved in 1953, when James Watson and Francis Crick proposed the double-helix model of DNA [6]. In fact, it was Crick who first stated what is now known as the central dogma of molecular biology [7, 8], shown in Figure 1.1, which explicitly connects nucleic acids and proteins through clear biological processes: DNA replication, DNA to RNA transcription and RNA to protein translation. Starting with the 1940's, the scientific community has witnessed (and often awarded Nobel Prizes for) a myriad of discoveries of the enzymes involved in replication, transcription, translation, and many other related processes.

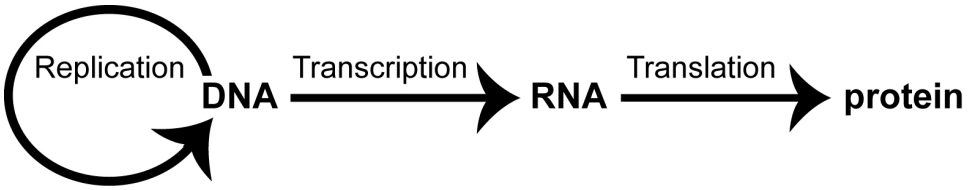

Figure 1.1: **The central dogma of molecular biology.**

With this renewed biochemical understanding of the workings of cell in general and of inheritance in particular, views about evolution also started shifting from a "survival of the species" view, to a gene-centered one. In this description of evolution, the unit of selection is a gene (as represented in the entire population), rather than the individual or the species. This approach elegantly explains the evolutionary pressure for kin altruism, and was highly popularized through the 1976 book *The Selfish Gene* by Richard Dawkins [9].

If indeed natural selection is working on genes, then we would better understand what genes are, and how they are inherited and expressed to give rise to different phenotypes. While we can describe a gene simply as "the unit of inheritance," in order to fully understand its magic, we should be able to define a gene biochemically as well. One could naively describe a gene as the DNA sequence that encodes for a functional protein or RNA. The problem with this definition is that it conflicts with the "unit of inheritance" one. The view that the gene is only a DNA sequence treats the RNA and the proteins as mute actors in the inheritance play, while it is very clear that they are not. Both RNA and proteins are inherited along with DNA and they greatly modify how the DNA blueprint is implemented. This is obvious when we think about development: all the cells in our bodies have the same DNA, but skin cells are structurally and functionally very differently from neurons for example. In fact even in 1909, when Wilhelm Johannsen coined the word "gene", he had the amazing foresight to warn against the "conception of the gene as a material, morphologically characterized structure which is very dangerous for the smooth advance of genetics" [4].

There is recent evidence to suggest that the DNA sequence is not necessarily the unit of inheritance. For instance, a cell can change DNA mutations back to the normal sequence by using the RNA inherited from past generations [10]. Moreover, changes in the levels of RNA or proteins inherited from the parents can change the phenotype of the offspring, and can be in turn passed to their children [11, 12].

This process of passing information to the next generation through means other than the DNA sequence is called "epigenetic inheritance". Because epigenetics has important implications on how we think about evolution [13], it is essential that we understand how it affects gene expression.

1.2 The Nucleosome: a Regulator of Gene Expression

In eukaryotes the DNA is packaged into chromosomes, which are a well-organized and tightly controlled mix of DNA and proteins called chromatin. An important way of controlling gene expression is by modulating chromatin structure and accessibility [14]. Chromatin has as its basic unit the nucleosome, which was discovered by Roger Kornberg in 1974 [15] while he was working as a postdoctoral researcher with Aaron Klug and Francis Crick. The high-resolution crystal structure of a nucleosome was solved in 1997 [16], and is shown Figure 1.2a. The nucleosome core particle consists of 147 base pairs of DNA wrapped around a histone octamer, which in turn is made up of a tetramer containing histones H3 and H4 and two H2A-H2B dimers (Figure 1.2b).

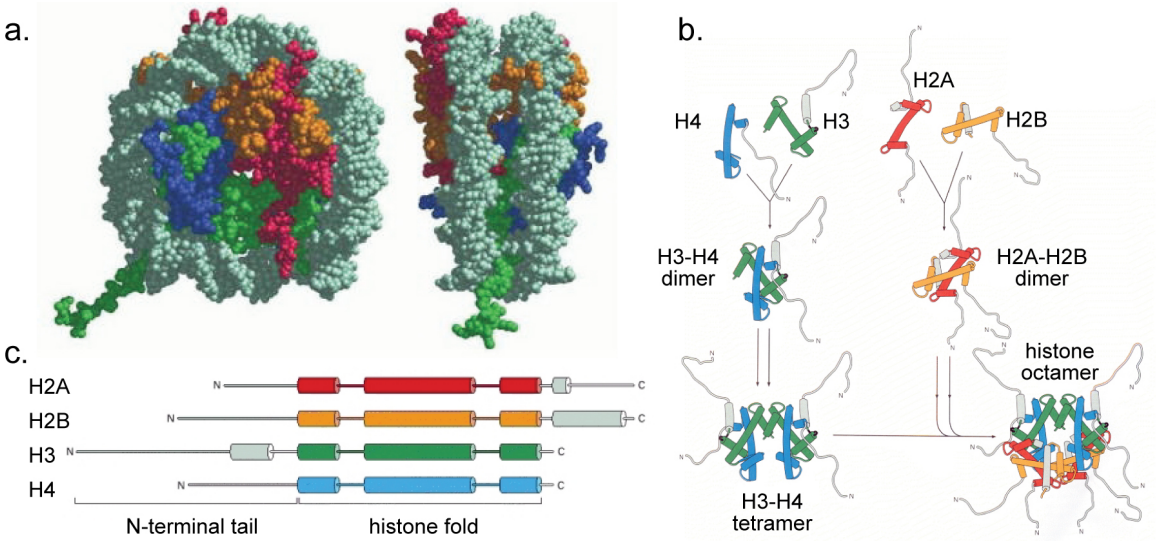

Figure 1.2: **The structure of the nucleosome core particle. (a)** The DNA double helix (light gray) is wrapped 1.65 times in a left-handed coil around a disc-shaped histone octamer. **(b)** The histone H3H4 dimer and the H2A-H2B dimer are formed from the handshake interaction. An H3-H4 tetramer forms the scaffold of the octamer onto which two H2A-H2B dimers are added, to complete the assembly. **(c)** Each of the core histones contains an N-terminal tail and a histone fold region. Adapted from *Molecular Biology of the Gene* by Alberts et al. [17].

As shown Figure 1.2c, the histones consists of two structural and functional units: the histone folds and the tails. The histone fold regions are the ones involved in direct histone-histone and histone-DNA interactions. The histone tails protrude from the disc-shaped core structure, and most of their lengths weren't visible in the crystal structure [16], suggesting that their conformation is flexible. In the cell, the tails could either bind DNA outside the

147 base pairs that are included in the nucleosome core particle, or they could bind to other proteins, such as chromatin remodeling factors.

Interestingly, *in vivo*, the histone tails are subject to several forms of post-translational covalent modifications that have strong effects on gene expression. This form of epigenetic control of gene expression gave rise to the histone code hypothesis [18], which states that specific patterns of histone modifications allow specific chromatin remodeling factors and RNA polymerases to recognize regions they are supposed to remodel or transcribe. Testing out this hypothesis is currently an area of intensive research.

1.3 Transcription Elongation by RNA Polymerase II

Transcription of DNA into messenger RNAs, the precursors to proteins, sits at the heart of the central dogma of molecular biology. In eukaryotes, this process is performed by RNA polymerase II (Pol II), a large molecular complex (550 kDa, 12 subunits) that has been extensively studied.

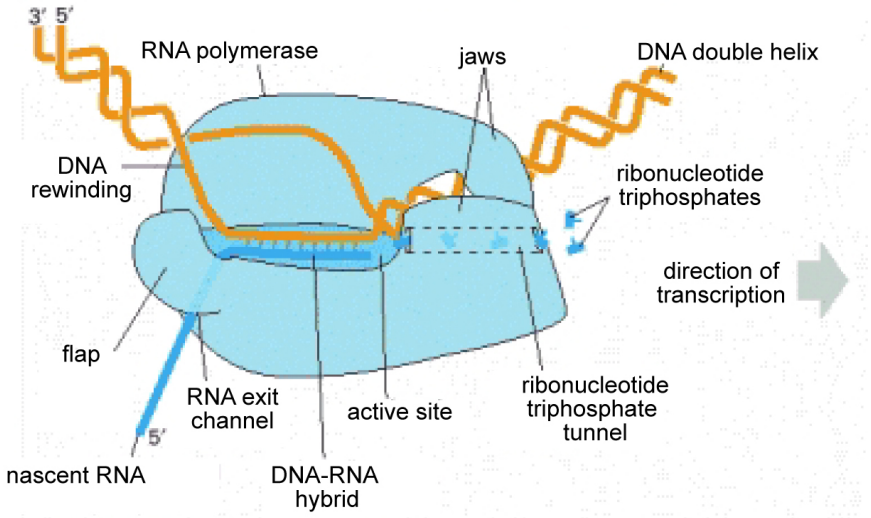

Figure 1.3: **The basic mechanism of transcription elongation.** During transcription elongation, Pol II unwinds the double helix, forming "the elongation bubble," which shifts one base pair at a time on DNA. Ribonucleotide triphosphates (NTPs: ATP, UTP, CTP, and GTP) enter through the nucleotide tunnel and are added to the growing RNA chain at the active center of the enzyme. The newly synthesized RNA chain is a complementary copy of the template DNA strand, and forms a 9 base pairs DNA-RNA hybrid with this DNA. Adapted from *Molecular Biology of the Gene* by Alberts et al. [17].

Since the discovery of the RNA polymerase in the 1960's [19], a lot of progress was made

towards understanding how this enzyme works. Crystal structures of the Pol II holoenzyme [20] and of the elongation complex consisting of Pol II, DNA and RNA [21], along with clever biochemical assays performed on Pol II (reviewed in [22]) or its bacterial counterpart (reviewed in [23]) revealed a lot about the basic mechanism of transcription, briefly described in Figure 1.3. Pol II uses nucleotide triphosphates (NTPs) as building blocks for the RNA nascent chain. The hydrolysis of the NTPs accompanying polymer synthesis also serves as the source of energy for the movement of the polymerase on DNA.

In parallel with the characterization of Pol II, a large number of factors that modify transcription have been identified (reviewed in [24]). The effects of these factors are often mediated through chromatin [25–28], pointing to the need of understanding the mechanism of nucleosomal transcription.

1.4 Nucleosomal Transcription: Questions

As described in the previous two sections, an important step towards the bigger goal of unlocking the mechanism of gene expression consists of understanding transcription regulation in the context of chromatin. Chromatin-mediated regulation can take place at all three stages of transcription—initiation, elongation and termination (see [29] and [30] for reviews). We focus on uncovering how Pol II elongation is modulated by the nucleosome, and how this encounter affects in turn the structure of the nucleosome. More specifically, we would like to find out:

- **What exactly is happening at the molecular level when Pol II attempts to transcribe the DNA wrapped into a nucleosome?** It is clear that the histone octamer and the polymerase cannot both occupy the DNA at the same time, so there must exist a competition between them for this physical space. It is known that Pol II pauses and sometimes arrests upon reaching the nucleosome. How does the nucleosome induce this transcriptional behavior and what factors control the barrier that the nucleosome represents for Pol II? We address these questions in Chapters 2 and 3 respectively.

- **What happens with the histones during and after transcription?** There is evidence that during transcription the histones can be either removed from DNA or transferred behind the polymerase. Different answers arise, depending on the experimental methods and the type of polymerase used for transcription. Can we rationalize these results and predict when and why one scenario dominates over the other? Moreover, in the case when the histones are transferred on the template upstream of the polymerase, how does this process take place? We address these issues in Chapter 4.

Through the experiments and analysis presented in this work, we hope to confirm certain claims about nucleosomal transcription and bring new insight into our current understanding of the dynamics interactions and molecular rearrangements that accompany this process.

Chapter 2

The Mechanism of Nucleosomal Transcription

During transcription elongation in eukaryotes, Pol II must overcome the transcriptional barriers imposed by nucleosomes in chromatin. It has long been known that histones generally inhibit transcription [31], yet well-positioned nucleosomes are ubiquitous throughout eukaryotic chromosomes [32, 33] and are generally believed to be important scaffolds for the regulation of gene expression in eukaryotes.

Biochemical assays of nucleosomal transcription show that a single nucleosome can arrest or greatly slow the rate of transcription by Pol II *in vitro* [34–38]. Polymerases arrested at nucleosomes are often backtracked many bases [37]. In addition, TFIIS—a transcription elongation factor that stimulates the polymerase's inherent transcript endonucleolytic cleavage activity to allow backtracked enzymes to recover from arrest [39]—has also been shown to increase the net rate of transcription through a nucleosome [37]. These experimental results suggest that the influence of the nucleosome is mediated through polymerase backtracking. Moreover, nucleosomes exhibit fast wrapping and unwrapping fluctuations, which were proposed to play a significant role on the dynamics of transcribing polymerases [40].

In order to understand the exact nature of the polymerase pausing at the nucleosome, and the mechanism of overcoming the nucleosomal barrier, we characterize the dynamics of Pol II as it transcribes nucleosomal DNA in real time. To do this, we take advantage of three important experimental advances: factor-free initiation of Pol II [41, 42], the ability of loading nucleosomes at known positions on DNA [43], and the high temporal and position resolution of single-molecule assays of transcription [44, 45].

2.1 Real-Time Transcription Through the Nucleosome

We followed individual Pol II elongation complexes in real time as they transcribe through single nucleosomes, using a dual-trap optical tweezers. Pol II ternary elongation complexes

Portions of this chapter were published in C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. **Nucleosomal Fluctuations Govern the Transcription Dynamics of RNA Polymerase II**, *Science* 325, 626628 (2009) doi:10.1126/science.1172926. Used with permission.

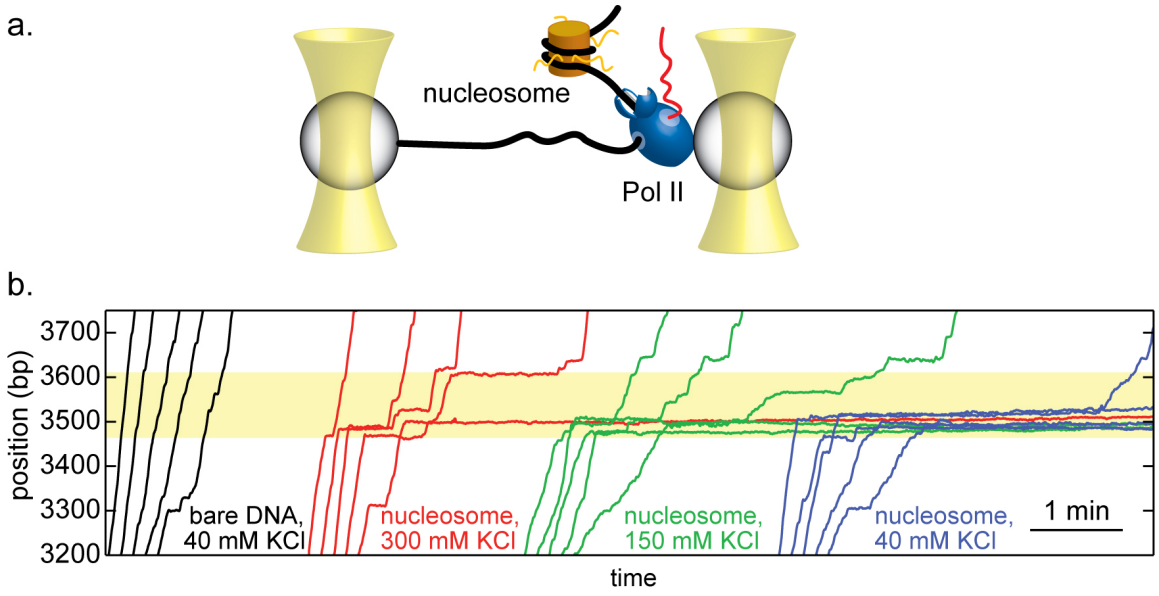

Figure 2.1: **Single-molecule transcription through the nucleosome.** **(a)** Geometry for the dual-trap optical tweezers experiments. Pol II (blue) is attached to a streptavidin coated bead (gray). The upstream DNA (black) is attached to an antidigoxigenin coated bead (gray). The optical traps are shown in yellow and the RNA in red. **(b)** Representative trajectories of individual transcribing polymerases over time are shown with or without the nucleosome at different ionic strengths.

from *S. cerevisiae* were prepared and their downstream DNA ligated to DNA containing a nucleosome pre-loaded on the 601 nucleosome positioning sequence (NPS) [43]. A tether was created between two beads—one attached via biotin/streptavidin to a stalled Pol II, the other to the DNA upstream of the stalled enzyme via digoxigenin/anti-digoxigenin pair—and a solution containing ribonucleotide triphosphates was delivered to the flow cell (Figure 2.1a). Polymerase activity resulted in the lengthening of the upstream DNA and caused the force between the two beads to decrease. The real-time position of the polymerase along the DNA template was calculated by fitting the measured force to the worm-like chain formula of DNA elasticity [46].

In the absence of a nucleosome, the polymerase generally proceeded to the end of the DNA template, interrupted only by a few short pauses (Figure 2.1b, black traces). When we pre-loaded a single core nucleosome onto the template, Pol II showed pronounced changes in its dynamics at the NPS. The behavior of Pol II at a nucleosome varied greatly, ranging from many pauses, to one or two pauses, to complete arrest (Figure 2.1b, colored traces). Arrest is defined as a pause that lasts longer than 20 minutes or until the tether breaks.

2.2 Arrests Dependence on the Ionic Strength

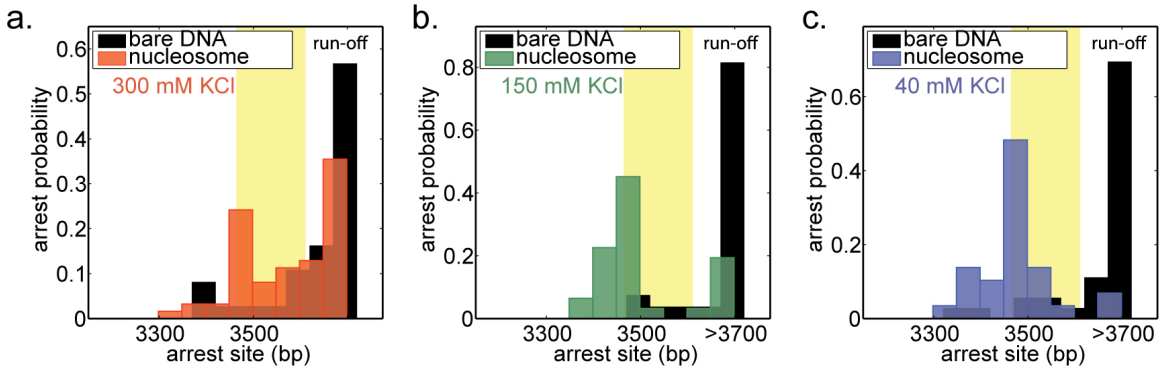

Figure 2.2: **Arrest probability depends on ionic strength.** Probability of arrest or termination as a function of Pol II's position on the template DNA at **(a)** 300 mM KCl, **(b)** 150 mM KCl, and **(c)** 40 mM KCl. The data for transcription of bare DNA is shown in solid black and the nucleosome data is shown in semi-translucent colors. The shaded region represents the NPS.

As DNA must be separated from the charged surface of the histone octamer for the polymerase to proceed, a large fraction of the nucleosomal barrier should be of electrostatic origin. The effect of ionic strength on the transcription of nucleosomal DNA by Pol II (Figure 2.2) showed that there is indeed a marked decrease in the frequency of arrest at the nucleosome with increasing salt, from 93% at 40 mM KCl, to 77% at 150 mM KCl, to 48% at 300 mM KCl, in agreement with previous biochemical ensemble studies [37, 38]. We also checked that the influence of ionic strength on arrest parallels a decrease in the mechanical stability of the nucleosome with salt, but does not correlate with changes in the dynamics of transcription on bare DNA.

In order to observe the effects of ionic strength on nucleosome stability, we analyzed force-extensions curves of untranscribed nucleosomes using previously described techniques [47]. We loaded the nucleosome on the same DNA we use for the transcription experiments, except that we introduced a biotin at the end of template using a modified primer. This allowed us to pull on nucleosomes in the absence of the biotinylated Pol II. The most reliable signature of the nucleosome is the sudden opening of the inner wrap at increased force (Figure 2.3). We quantified both the distribution of forces at which this rip occurred as well as the size of the rip, as a function of [KCl]. As expected, the force necessary to unwrap the nucleosome decreased as we increased the salt concentration, confirming that the relief in transcriptional arrests is a result of reduced histone-DNA interactions (Figure 2.4). The extension change associated with nucleosome unwrapping did not change significantly with ionic strength, suggesting that the nucleosome is still intact as the salt concentration is increased.

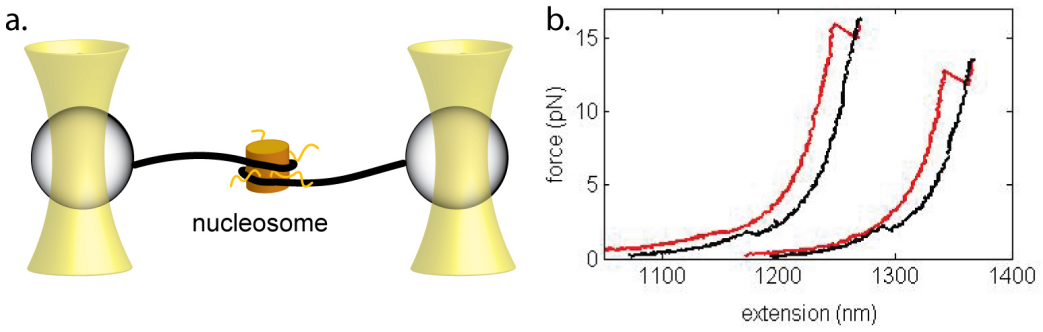

Figure 2.3: **Pulling on single nucleosomes.** **(a)** Experimental setup for nucleosome pulling experiments. **(b)** Typical force-extensions curves of the nucleosome. Pulling curves are in red, and relaxation ones are in black.

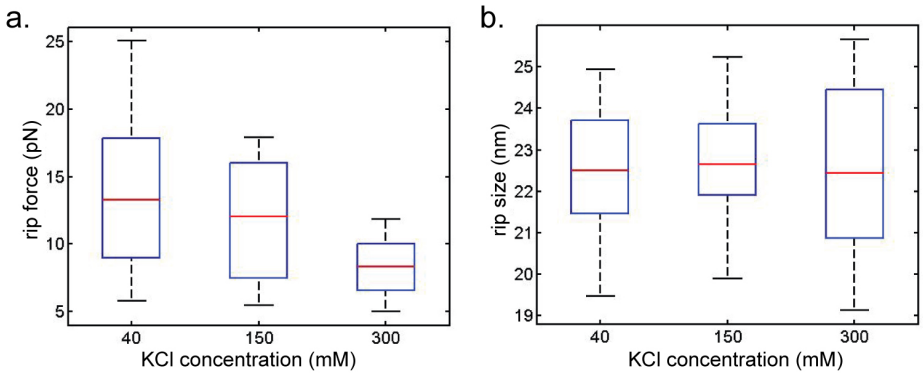

Figure 2.4: **Mechanical stability of the nucleosome with ionic strength.** **(a)** The forces necessary to unwrap the nucleosome decrease with increasing salt concentration. **(b)** The size of the inner wrap is independent of salt concentration. The red lines represent the means of the data, the blue boxes show one standard deviation, and the dotted lines show the entire range of the data.

In order to check whether the intrinsic dynamics of Pol II are affected by ionic strength, we compared transcription on bare DNA at 40, 150 and 300 mM KCl. Transcription of bare DNA by Pol II was unaffected over this salt range (Figure 2.5). Neither the mean dwell time nor the pause free velocity shows any trend with increasing KCl concentrations. Therefore, differences in nucleosomal transcription at different ionic strengths (Figure 2.2) are the direct result of changes in the nucleosome's mechanical stability.

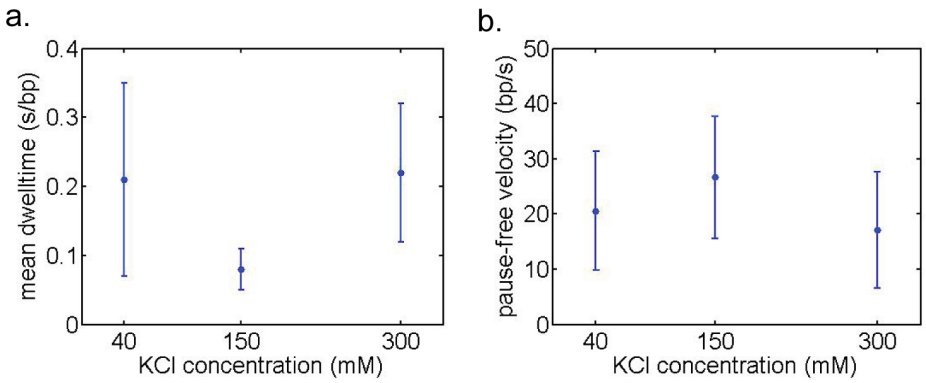

Figure 2.5: **Transcription on bare DNA at different ionic strengths.** **(a)** Pause durations as a function of ionic strength. **(b)** Pause-free velocities as a function of ionic strength. Error bars are standard deviations among traces.

2.3 Detailed Dynamics of Pausing and Elongation

Because a majority of the polymerases were able to overcome the nucleosomal barrier at 300 mM KCl, we conducted more detailed studies of nucleosomal transcription at this ionic strength.

The average time Pol II spent at the NPS was much longer in the presence of the nucleosome than on bare DNA (0.4 s/bp, compared to 0.1 s/bp), showing a significant increase even before Pol II reached the beginning of the NPS (Figure 2.6). The increase in dwell time right before the start of the NPS could have two causes. Firstly, in our single-molecule experiments we are following the position of the active center of the polymerase on the template; however, the leading edge of the polymerase reaches the nucleosome ∼20 bp before the active center does [48]. Secondly, while the nucleosome core particle only organizes 147 bp of DNA, the histone tails could bind additional DNA situated at the entry into the nucleosome [49], thus retarding the polymerase in this region. We address this possibility in Chapter 3.

The effect of the nucleosome on transcription was greater before the polymerase passed the dyad axis of the nucleosome, in agreement with previous studies [38]. The attenuation of the nucleosomal barrier that we observe towards the end of the NPS might be caused by the complete detachment of the histones from the downstream DNA. We discuss this hypothesis in detail in Chapter 4.

There are three possible explanations for the reduced overall transcription rate of the nucleosomal DNA: the presence of the nucleosome could induce Pol II to pause more frequently, slow the recovery of Pol II from pausing, and/or slow the pause-free velocity of the polymerase. We examine each of these alternatives.

To establish whether or not the nucleosome affects pause entry, we counted all pauses of at least 3 s and recorded their positions on the DNA template (Figure 2.7a). The nucleosome

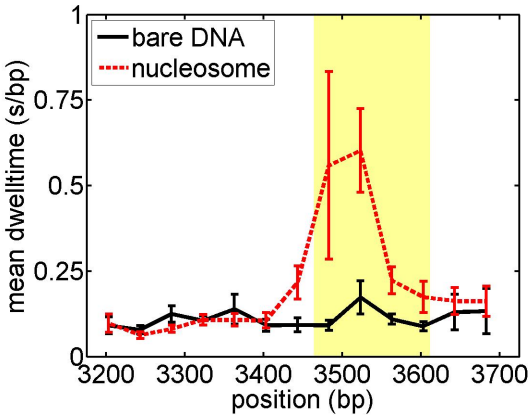

Figure 2.6: **Total time spent at the nucleosome.** The mean time the polymerase spent in each bin is plotted for traces with (dashed red line) and without (solid black line) a nucleosome. Mean dwelltime includes the time spent during both active elongation and pausing. Shaded region corresponds to the NPS, and error bars are SEM.

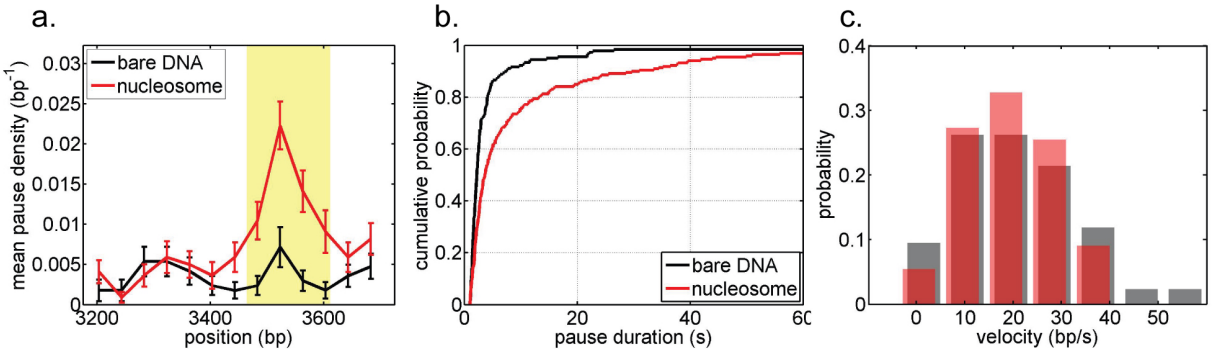

Figure 2.7: **Effect of the nucleosome on transcription kinetics.** In each subplot, only traces that passed the NPS are considered. **(a)** Pause density with a nucleosome (red) and on bare DNA (black). Error bars are SEM. **(b)** Cumulative distributions of pause durations, both with (red) and without (black) a nucleosome present. **(c)** Pause-free velocities with (pink) and without (gray) a nucleosome.

locally increased the probability of Pol II to enter a paused state more than threefold from 0.004 ± 0.001 bp^{-1} on bare DNA to an average of 0.014 ± 0.003 kbp^{-1} at the nucleosome. The effect on pause density is strongest right before the polymerase reaches the dyad axis of the nucleosome, reaching a peak of 0.022 ± 0.003 kbp^{-1}.

To determine if the nucleosome alters the kinetics of pause recovery by Pol II, we com-

pared the distributions of pause durations longer than 1 s in the nucleosomal region for traces with and without a nucleosome. Pausing behavior was highly variable among trajectories and, even within a trajectory, pausing at the nucleosome varied from short intermittent events to longer pauses. For comparison of pausing data, we applied the two-sample Kolmogorov-Smirnov test. This nonparametric test requires no knowledge of the underlying probability density functions, and compares the maximum deviation between two populations in their empirical cumulative distributions. All p-values reported for pause durations refer to the two-tailed version of this statistical test. A comparison of the cumulative distributions of the pause durations shows that the nucleosome biases the polymerase towards longer pauses at the NPS ($p < 0.001$, Figure 2.7b), with the mean pause duration on bare DNA of 4.8 ± 0.8 s increased to 10.7 ± 1.2 s at the nucleosome. Thus, the nucleosome slows the underlying pause recovery mechanism of the polymerase.

Finally, analysis of pause-free velocities obtained with an algorithm that identifies and eliminates transcriptional pauses (Section 2.8) revealed that pause-free velocity at the NPS is only slightly reduced from 21.2±2 bp/s on bare DNA to 18.9±2 bp/s in the presence of a nucleosome (Figure 2.7c). This difference is not statistically significant.

We conclude that the nucleosome only affects entry into and recovery from the paused state, but does not affect the nucleotide addition cycle of Pol II. For the remainder of this chapter, we build a model that allows us to explain our experimental observations and make further predictions on nucleosomal transcription.

2.4 Kinetic Model of Backtracking Coupled with Nucleosome Fluctuations

Many transcriptional pauses of Pol II on bare DNA are associated with backtracking of the enzyme along the DNA template [45], a process which results in misalignment of the 3' end of the transcript with respect to the enzyme's active center [50, 51]. Pauses end when the polymerase diffusively realigns the dislocated 3' end of the transcript with its active site and resumes elongation. We incorporate nucleosome fluctuations into this model of transcriptional pausing in order to rationalize our observations of the dynamics of Pol II at the nucleosome.

2.4.1 Pause Durations

According to the backtracking theory, during the paused state the polymerase performs a random walk on DNA, sliding the entire elongation bubble back and forth one base pair at a time. The distribution of transcriptional pause durations is then equivalent to the distribution of first-passage times for return to the origin of a Poisson stepper that takes integral steps along a one-dimensional lattice [52]. We re-derive the distribution of pause durations ,$\psi(t)$, in Section 2.9.2, and arrive at the following expression:

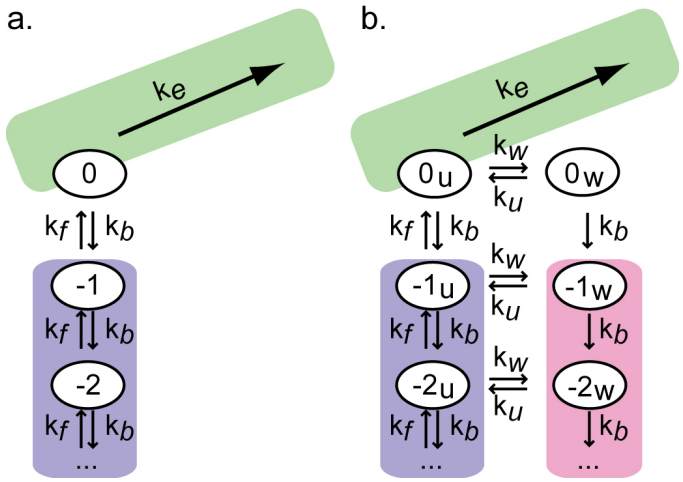

Figure 2.8: **Kinetic model of transcription.** Kinetic states and rates for **(a)** transcription on bare DNA and **(b)** transcription at the nucleosome. The green shaded region corresponds to the on-pathway elongation cycle (k_e). The pink and blue shaded regions represent off-pathway paused states where Pol II is backtracked; negative numbers indicate how many bases Pol II has backtracked from the elongation competent state, denoted by 0. The forward and backward stepping rates during a backtrack are given by k_f and k_b, respectively. The label u refers to the nucleosome being locally unwrapped (blue shaded region), while w denotes the states where the nucleosome is wrapped in front of Pol II (pink shaded region). The wrapping and unwrapping rates are k_w and ku.

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \cdot \frac{e^{-(k_f + k_b)t}}{t} \cdot I_1(2t\sqrt{k_f k_b}), \tag{2.1}$$

where I_1 is the modified Bessel function of the first kind, and k_f and k_b are the forward and backward stepping rates, respectively, during a backtrack (see Figure 2.8).

If we ignore any non-uniformities in the DNA sequence of the template, the forward and backward rates are equal, and are given by the intrinsic rate of Pol II diffusion along DNA during a backtrack, which we call k_0:

$$k_f = k_0 \text{ and } k_b = k_0. \tag{2.2}$$

In our experimental setup we apply an assisting force (F) to the polymerase, decreasing the backwards rate and increasing the forward one as follows:

$$k_f = k_0 e^{F \cdot d / k_B T} \text{ and } k_b = k_0 e^{-F \cdot d / k_B T}, \tag{2.3}$$

where d is the distance to the transition state for a step (taken here to be 0.5 bp). In the experiments described here we maintained the applied force between 4-8 pN, with an average of 6 pN. Note that for small forces and long pause durations $\psi(t)$ reduces to the $t^{-3/2}$ power-law dependence previously reported [45].

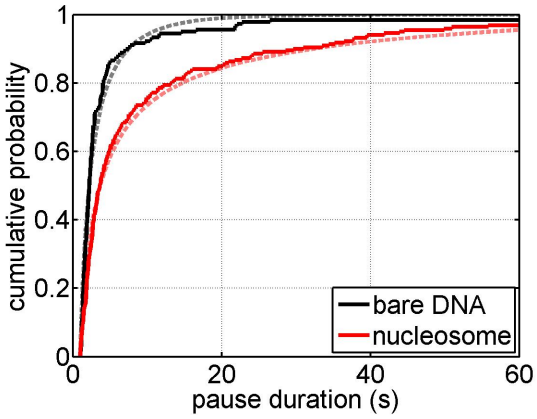

Figure 2.9: **Pause durations fits against the kinetic model.** The cumulative distribution of pause durations on bare DNA (continuous black line) is fit for k_0 using the normalized integral of $\psi(t)$ (dashed gray line). The cumulative distribution of pause durations at the nucleosome (continuous red line) is fit for γ_u using the normalized integral of $\psi_N(t)$ (dashed pink line).

The effect of the nucleosome on transcription can then be understood in light of the diffusive nature of backtracks and the dynamics of nucleosome wrapping/unwrapping fluctuations. Because backtrack pausing competes with continued elongation, any barrier that restricts elongation will cause the polymerase to backtrack more frequently. If, moreover, this barrier follows the polymerase during diffusive backtracking, as expected if nucleosomal DNA rewraps after the polymerase steps back, it will also frustrate backtrack recovery, thus increasing pause durations. Local nucleosomal fluctuations are very fast relative to the diffusive stepping rate k_0 [40], indicating that the nucleosome reaches fast local equilibrium between each backtracking step. Thus, the effect of these fluctuations on the dynamics of Pol II depends only on the fraction of time the nucleosome is locally unwrapped, γ_u, given by:

$$\gamma_u = \frac{k_u}{k_u + k_w}, \tag{2.4}$$

where k_w and k_u are the rates of local wrapping and unwrapping of the DNA template around the core histones, respectively.

A backtracked polymerase cannot actively separate downstream nucleosomal DNA from

the surface of the core histones because it possesses no energy source. Thus, we expect Pol II pause durations on nucleosomal DNA to be drawn from the same distribution as on bare DNA, but with a net forward stepping rate reduced by a factor corresponding to the fraction of time the local nucleosomal DNA is unwrapped:

$$k_{f\,(\mathrm{nucl})} \rightarrow \gamma_u \cdot k_f \tag{2.5}$$

$$\psi_N(t) = \sqrt{\frac{\gamma_u \cdot k_f}{k_b}} \cdot \frac{e^{-(\gamma_u \cdot k_f + k_b)t}}{t} \cdot I_1(2t\sqrt{\gamma_u \cdot k_f k_b}). \tag{2.6}$$

To compare the theoretical predictions of pause duration with the experimental data, we used cumulative probability distributions. The cumulative distribution of $\psi(t)$ for pauses between our lower limit of 1 second and our arrest cutoff of 20 minutes is given by the integral of the pause durations distribution:

$$C_\psi(t) = \frac{\int_1^t \psi(\tau)d\tau}{\int_1^{1200} \psi(\tau)d\tau}. \tag{2.7}$$

The normalization in the denominator represents the fraction of pauses between our observational limits. This normalization is largely independent of the particular choice for the upper cutoff because almost no pauses are expected to last longer than a few minutes. We numerically integrated $\psi(t)$ as shown above to obtain the theoretical curves shown in Figure 2.9.

We first fit the cumulative distribution corresponding to $\psi(t)$ to the pause durations on bare DNA to determine a value for k_0. This curve fits well to the observed pause durations in the absence of a nucleosome for pauses at least 1 s in duration (Figure 2.9b, gray dashed line), consistent with a diffusive mechanism of backtrack pause recovery. This analysis yielded a value of $k_0 = 1.3 \pm 0.2\ \mathrm{s}^{-1}$, indicating that diffusion of a backtracked enzyme is a relatively slow process compared to the rate of transcription elongation.

Using the value of k_0 determined above and fitting the cumulative distribution of pause durations in the presence of a nucleosome at 300 mM KCl to the diffusive backtracking model yields $\gamma_u = 0.59 \pm 0.05$, indicating nucleosomal DNA is locally unwrapped approximately half of the time immediately downstream of a polymerase. During backtracking events, the DNA downstream of a backtracked polymerase can stochastically rewrap around the core histones, restricting Pol II from diffusing back to the 3' end of the nascent RNA to resume transcription, thereby increasing pause durations.

2.4.2 Pause-Free Velocities

The model presented in Figure 2.8 predicts that the nucleosome only affects the paused state of the polymerase, but shouldn't affect its pause-free velocity. If one is able to remove

all pauses from the experimental traces, the pause-free velocity obtained this way should reflect the net irreversible elongation rate k_e. Removing all pauses is technically difficult; while it's relatively easy to identify pauses above 1 s in duration, correctly identifying pauses under 0.2 s in our system is impossible, as we filter the data to 50 Hz before we find pauses. Because of this caveat, in our previous publication of these data, we reported a decreased pause-free velocity at the nucleosome compared with that on bare DNA [53]. However, here we have improved our pause picking algorithm, and, as expected, we find that the pause-free velocity of Pol II doesn't change as it transcribes the nucleosomal DNA.

We take the value of the elongation rate to be the pause-free velocity measured on bare DNA $k_e = 21.2 \pm 1.9$ bp/s (Figure 2.7). Since it is easier to remove pauses from data on bare DNA compared with nucleosomal data, we expect the bare DNA value to be closer to the true pause-free elongation rate of Pol II.

2.4.3 Number of Pauses

During active transcription, forward elongation competes kinetically with backtrack pausing; therefore, the increased pause density at a nucleosome can be used to infer changes in the net elongation rate. Specifically, derivation of the predicted pause densities allows us to discriminate between two possible scenarios: one in which Pol II can only elongate against a locally unwrapped nucleosome (Figure 2.8), and another in which the polymerase can advance by actively unwrapping nucleosomal DNA (Figure 2.11).

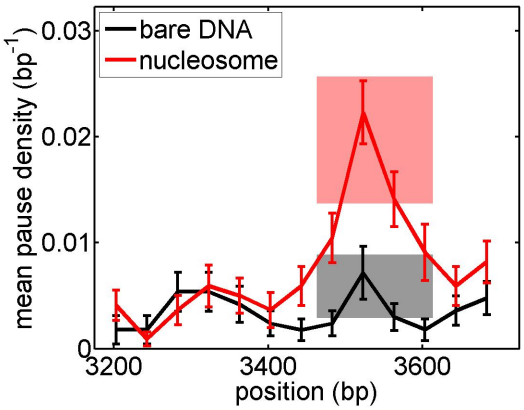

Figure 2.10: **Model predictions of pause density.** Pause density as a function of position on the template with a nucleosome (red line) and on naked DNA (black line). Error bars are SEM. The shaded regions represent the predicted confidence interval for pause density on naked DNA (gray) and at the nucleosome (pink) based on the model presented in the text.

We first examine the pause density predictions of the model where Pol II cannot actively

separate core histones from DNA (Figure 2.8). In this view, the frontal block provided by a wrapped nucleosome is sufficient to inhibit transition from the pre- to post-translocation stage of the polymerase's mechanochemical cycle. If nucleosomal wrapping/unwrapping dynamics are much faster than backtrack stepping by the polymerase, the pause density in the presence of the nucleosome has the same general form as on bare DNA, except that the elongation rate is multiplied by the probability of finding the nucleosome locally unwrapped, γ_u (see Appendix 2.9.1 for derivation):

$$P_{\text{bare DNA}} = \frac{k_b}{k_b + k_e} \tag{2.8}$$

$$P_{\text{nucleosome}} = \frac{k_b}{k_b + \gamma_u k_e} \tag{2.9}$$

The probabilities we calculated above describe the probability of entering a pause at a given position on the template and correspond experimentally to the total number of pauses per base pair. However, due to experimental noise and because of finite averaging bandwidth, we chose to avoid miscounting shorter pauses by only measuring pause durations that were three seconds or longer. When comparing theory with the experimental data, we must take into account all the very short pauses that we did not analyze.

Because we know the underlying probability density function for pause durations, $\psi(t)$, we can calculate what fraction of the pauses are between our lower cutoff pause time (3 s, for pause densities) and our upper cutoff (1200 s) by integrating $\psi(t)$ within those times. To calculate the pause density for pauses within the experimentally measurable range, we multiply the theoretical pause density by this fraction:

$$P_{\text{naked DNA}}(3 < t < 1200) = \frac{k_b}{k_b + k_e} \int_3^{1200} \psi(t)\, dt \tag{2.10}$$

Similarly at the nucleosome, the predicted pause density when using pause cutoffs is:

$$P_{\text{nucleosome}}(3 < t < 1200) = \frac{k_b}{k_b + \gamma_u k_e} \int_3^{1200} \psi_N(t)\, dt \tag{2.11}$$

The correction in pause density due to pauses outside our experimental measurements is roughly a factor of 2 in each case.

We use this model of pause density with the parameters we already determined from pause-free velocities (k_e), pause durations on bare DNA (k_0), and at the nucleosome (γ_u) to predict the number of pauses on bare DNA and at the nucleosome. These predictions, shown as shaded regions in Figure 2.10, match the experimental data very well, providing an independent, fit-free check of the model presented in Figure 2.8.

2.5 Pol II Does Not Actively Peel DNA from Histones

The alternative scenario, in which the polymerase can actively open a wrapped nucleosome and elongate through it with a rate $k_{e,w}$ (Figure 2.11), predicts a different pause density (see Section 2.9.1 for derivation):

$$P_{\text{nucleosome, active unwrapping}} = \frac{k_b}{k_b + \gamma_u k_e + (1-\gamma_u)k_{e,w}} \tag{2.12}$$

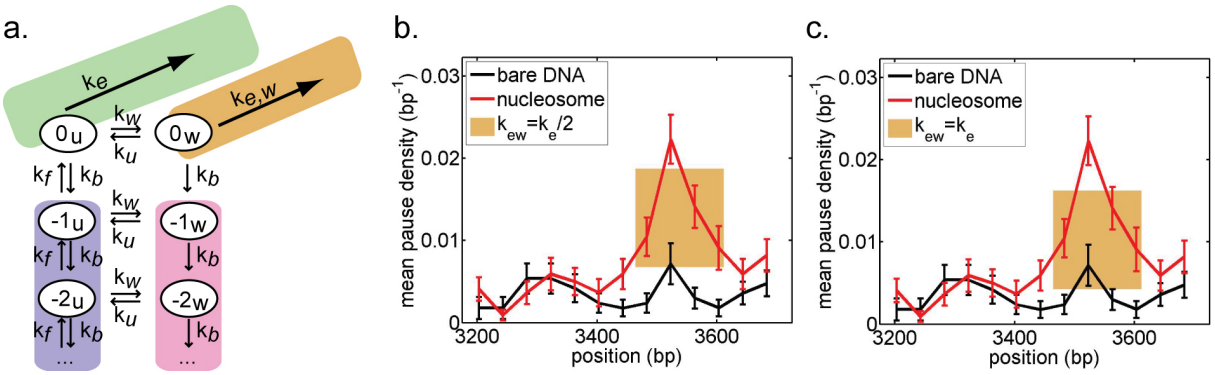

Figure 2.11: **Alternative kinetic model – active nucleosome unwrapping. (a)** Kinetic states and rates for transcription transcription at the nucleosome when Pol II can actively separate DNA from the surface of the histones at a rate $k_{e,w}$ (brown shading). The green shaded region corresponds to the on-pathway elongation cycle (k_e) on DNA already unwrapped from histones due to nucleosomal fluctuations. The pink and blue shaded regions represent off-pathway paused states where Pol II is backtracked, as before. The confidence intervals predicted by this alternative kinetic model are shown as brown shaded regions for two values of $k_{e,w}$: **(b)** $k_{e,w}$ equal to half the elongation rate on bare DNA and, **(c)** $k_{e,w}$ equal to the full elongation rate on bare DNA. Neither value of $k_{e,w}$ predicts a high enough pause density at the nucleosome to match our experimental data.

The pause density predictions of the model presented in Figure 2.11a allow us to place bounds on the rate of active elongation through a locally wrapped nucleosome ($k_{e,w}$). As shown in Figure 2.11, the active elongation model does not fit the experimentally observed peak in pause density. The only scenario when the model can accurately describe the high number of pauses that we observe is when $k_{e,w} \approx 0$. In that case, we recover the situation presented in Figure 2.8, in which the polymerase does not actively unwrap the nucleosomal DNA, but instead waits for transient nucleosomal unwrapping fluctuations to step forward and read the DNA template.

Thus, during both backtracking and elongation, Pol II waits for fluctuations that lo-

cally unwrap the nucleosome to advance, consistent with mechanisms proposed by others that invoke ratcheting during the transcription elongation cycle [54, 55]. In this way, the polymerase rectifies unwrapping fluctuations of the nucleosome as it moves along the DNA.

2.6 Relationship Between Local and Global Nucleosome Unwrapping

In the model described so far for transcription of the nucleosomal DNA, we assumed that the nucleosome and the polymerase compete for the DNA at each base pair. This assumption is very convenient because it allows us to derive a closed form solution for the parameters we measure in our single-molecule experiments: the distribution of pause durations, the probability of entering a pause and the elongation velocity. However, from the crystal structure of the nucleosome core particle, we know that the histones contact the DNA directly only every 5 or 10 base-pairs [16]. This means that the value of γ_u, which is the probability of finding the nucleosome unwrapped at each base pair, is an average value. We expect this average value (γ_u) to be smaller than the probability of nucleosome unwrapping at each specific contact point (γ_{Cu}). Intuitively, if only N_C such specific contact points exist along the entire length of the nucleosmal DNA (N_T), the relationship between the average and the contact-specific unwrapping probabilities should be of the form:

$$\gamma_u = \frac{N_C}{N_T} \cdot \gamma_{Cu} + \frac{(N_T - N_C)}{N_T} \cdot 1 \tag{2.13}$$

If we define the number of base pairs between consecutive histone-DNA contacts to be:

$$D_C = \frac{N_T}{N_C}, \tag{2.14}$$

equation 2.13 simplifies to:

$$\gamma_u = \frac{\gamma_{Cu}}{D_C} + \frac{(D_C - 1)}{D_C} \tag{2.15}$$

Note that equation 2.15 sets a bound on the average γ_u. It has to always be higher than $(D_C - 1)/D_C$. If the distance between contact points were really 5 base pairs and the intervening bases behaved as if they were bare DNA, then γ_u would always have to be greater than 0.8. This is not the case, as we measure $\gamma_u \cong 0.6$. This implies that a model where the histones make specific contacts with the DNA every 5 base pairs and don't influence much the rest of the DNA is unrealistic. An alternative that works out quantitatively with our model is that the contact points are on average every two base pairs ($D_C = 2$), and the histones have a 20% probability of being open at each contact point ($\gamma_{Cu} = 0.2$). Other models with non-uniform γ_{Cu} at each base pair (sometimes bigger than 0.5, but sometimes much smaller) could also work, and this is probably the case in reality.

It may seem strange at first that we obtain an average unwrapping probability higher than 0.5 without the nucleosomes falling apart. Equation 2.15 and the discussion above partially explain why our average value of nucleosome unwrapping is higher than 0.5, even when the contact specific unwrapping probabilities are much smaller than that. But we can do better: we can explicitly connect the macroscopic apparent dissociation constant of the entire nucleosome with the contact-specific unwrapping probabilities. Using the kinetic scheme shown in Figure 2.12, where the nucleosome sequentially unwraps one contact at a time, we can write the apparent dissociation constant of the nucleosome (K_d^{nuc}) in terms of the fractions of the nucleosomes that are completely unwrapped (U_{all}) and the fraction that are completely wrapped (W_{all}):

$$K_d^{nuc} = \frac{U_{all}}{W_{all}} \tag{2.16}$$

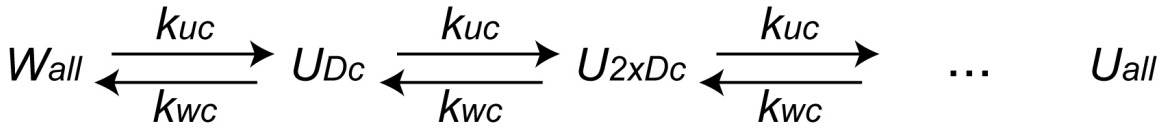

Figure 2.12: **Connection between local and global nucleosome unwrapping.** The nucleosome sequentially unwraps one contact at a time. The nucleosomes that are completely unwrapped are denoted as U_{all}) and the ones that are completely wrapped as W_{all}. Intermediates that have n contact points unwrapped have the subscript $n\text{x}D_C$. The rates of wrapping and unwrapping at each contact point are k_{wc} and k_{uc} respectively.

Following the scheme in Figure 2.12, we can write:

$$\begin{aligned} \frac{U_{D_C}}{W_{all}} &= \frac{k_{uc}}{k_{wc}} \\ \frac{U_{2\text{x}D_C}}{U_{D_C}} &= \frac{k_{uc}}{k_{wc}} \\ &\vdots \\ \frac{U_{all}}{U_{all-D_C}} &= \frac{k_{uc}}{k_{wc}} \end{aligned} \tag{2.17}$$

This set of telescoping equations simplifies to:

$$\frac{U_{all}}{W_{all}} = \left(\frac{k_{uc}}{k_{wc}}\right)^{N_C}, \tag{2.18}$$

where N_C is the number of contacts between histones and DNA, and k_{wc} and k_{uc} are the

rates of wrapping and unwrapping at each contact point.

Combining this result with the definition of K_d^{nuc} (Equation 2.16), we can write:

$$K_d^{nuc} = \left(\frac{k_{uc}}{k_{wc}}\right)^{N_C}, \tag{2.19}$$

We now use the fact that the unwrapping probability at each specific contact point (γ_{Cu}) is defined as:

$$\gamma_{Cu} = \frac{k_{uc}}{k_{wc} + k_{uc}}, \tag{2.20}$$

to write:

$$K_d^{nuc} = \left(\frac{\gamma_{Cu}}{1 - \gamma_{Cu}}\right)^{N_C}, \tag{2.21}$$

From this result, we can see that as long as the unwrapping at specific contact points is smaller than 0.5, the dissociation constant of the nucleosome is going to be very small, in agreement with our observations that the nucleosome is very stable overall.

2.7 Implications for Gene Expression Regulation

Regulation of the rate of transcription and pausing is of great importance for co-transcriptional processes such as alternative splicing [56, 57] or polyadenylation [58, 59]. Indeed, a large number of eukaryotic genes from different species are regulated during elongation by extended pausing, including heat-shock induced genes and many important developmental genes [60–66]. Because nucleosomes increase both the number and the duration of pauses, they constitute a potential control point in the regulation of transcription elongation. Our study indicates that modulation of the wrapping/unwrapping equilibrium of DNA around the histone octamer and control of the backtracking rate of the polymerase are essential ingredients in the regulation of nucleosomal transcription.

2.8 Appendix: Materials and Methods

2.8.1 Assembly and Enrichment of Active Elongation Complexes

Biotinylated RNA polymerase II (*S. cerevisiae*, unphosphorylated C-terminal domain) was prepared as previously described [67]. Yeast core histone proteins were grown recombinantly in *E. coli* [68] and analyzed through mass spectrometry to ensure the absence of post-translational modifications. Nucleosomes were loaded onto the high-affinity 601 nucleosome positioning sequence [43] using salt dialysis reconstitution [69].

Pol II ternary elongation complexes (TECs) were prepared according to previously published methods [35, 45], as shown in Figure 2.13, using the oligonucleotides listed below

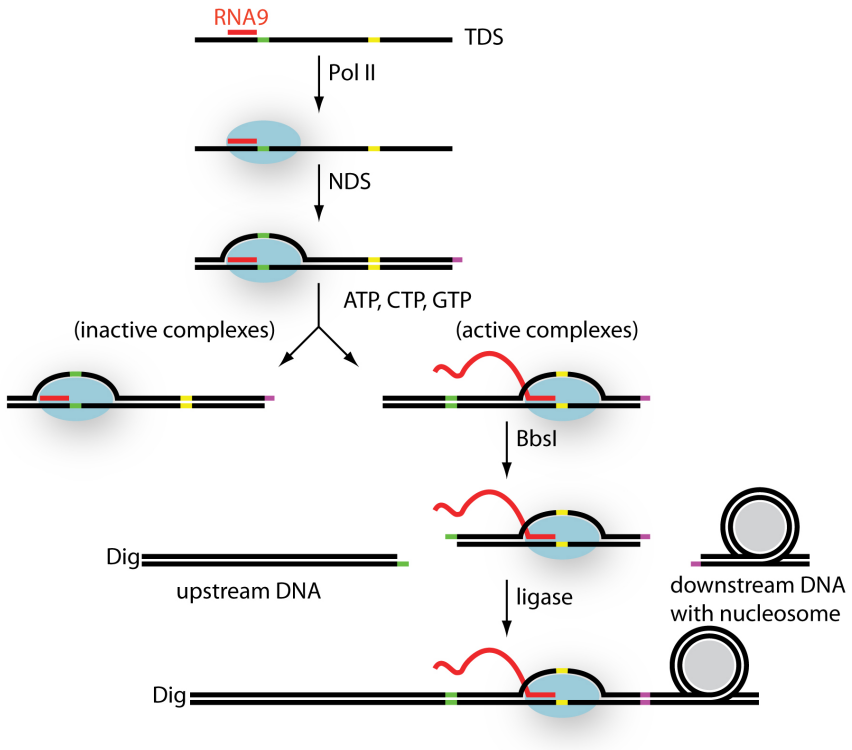

Figure 2.13: **Assembly of Pol II elongation complexes with nucleosome.** Assembly of TECs began after annealing the RNA9 (*red*) to the TDS (*black*). Addition of biotinylated Pol II and the NDS yielded functional TECs that overlapped a BbsI restriction site (*green*). TECs were walked to a nucleotide starvation stop site by the addition of 10 mM each of three nucleotide triphosphates (ATP, GTP, and CTP). As active polymerases moved to the starvation site (*yellow*), they revealed a BbsI restriction site upstream. These TECs were then digested with BbsI and ligated to upstream DNA containing a 5' AvaI overhang on one end and a digoxigenin at the other end. At the same time as the upstream ligation, we ligated downstream DNA containing a 3' BstAPI overhang and a pre-loaded nucleosome.

(IDT, Inc.). The sequences of the template DNA strand (TDS), the non-template DNA strand (NDS), and 9 bases RNA primer (RNA9) are shown in Table 2.1.

The procedure presented in Figure 2.13 allows for enrichment of active complexes, since only active TECs moved forward to reveal the BbsI restriction site. In addition, all complexes are stalled on the DNA exactly 3131 bp away from the digoxigenin; this uniform stalling allows for a calibration of the precise position of the polymerase in the optical tweezers before addition of all four nucleotide triphosphates (NTPs).

The biotinylated stalled complexes are bound to steptavidin-coated beads and a tether

Table 2.1: **Sequences used for the assembly of elongation complexes**

TDS: 5'-phosphateAGCATAATCCTGAATATGGCAAGTTACATAGATAAGTTGGTCGG TTGGGGTTTGTGTGGCTTCGTCGGGCGTCTTCTACATACTACTCCTACC-3'
NDS: 5'-GGTAGGAGTAGTATGTAGAAGACGCCCGACGAAGCCACACAAACCCCAA CCGACCAACTTATCTATGTAACTTGCCATATTCAGGATTATGCTCAT-3'
RNA9: 5'-GACGCCCGA-3'

is created using the upstream digoxigenin bound by anti-digoxigenin IgG-coated beads. The addition of all four NTPs into the flow channel allowed the polymerase to proceed with transcription. The transcription buffer contains: 20 mM Tris-HCl (pH 7.9), 5 mM MgCl2, 1 mM β-mercaptoethanol, 1 mM of each NTP, 1 μM pyrophosphate, and either 40, 150, or 300 mM KCl.

2.8.2 Spatial and Temporal Resolution

Thermal and instrumental noise both limit the resolution of our single-molecule trajectories. Although currently there is no known way to overcome thermal noise, we made several efforts to reduce the various sources of instrumental noise. The use of dual-trap optical tweezers allowed for differential detection of the movements in both traps, significantly increasing the signal-to-noise ratio of the position signal [70]. Additionally, the optical tweezers instrument was completely enclosed in a custom-made acrylic box, reducing ambient air fluctuations in the path of the laser light that contribute to noise. Finally, the instrument was located in the sub-basement floor in an acoustic isolation booth, reducing noise arising from building and acoustic vibrations [71].

All trajectories were sampled at 2000 Hz, averaged by decimation to 50 Hz, and then smoothed using a second-order Savitzky-Golay filter with a time constant of 1 s. With regard to sustained spatio-temporal resolution, the instrument is capable of reliably observing position changes as small as 1-2 nm (3-6 bp) at a bandwidth of 1 Hz, with a peak-to-peak instrumental drift of $\sim$0.5 nm ($\sim$1.5 bp) over 1 min.

Even though our accuracy in measuring position changes of Pol II on DNA was high, 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 runoff 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 runoff lengths. For this reason, we can be confident that for traces that

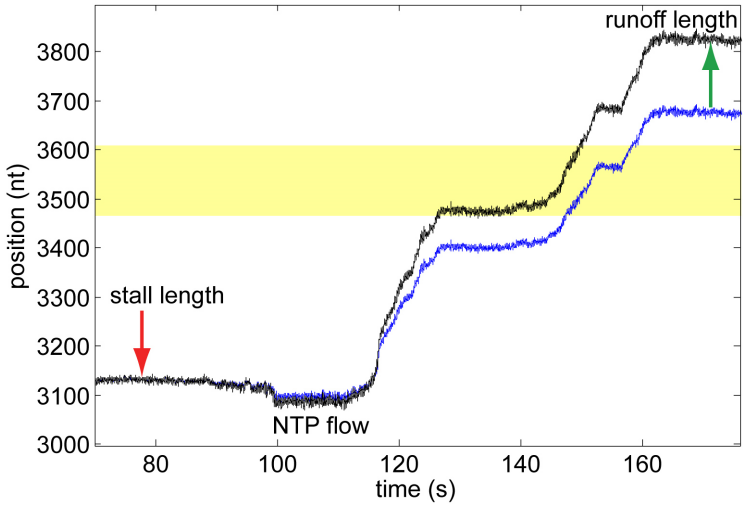

Figure 2.14: **Alignment of traces for improved precision.** All trajectories were aligned such that their start position before the flow of NTPs is equal to the position of the stall site (3131 bp, red arrow). Only traces that passed the NPS (yellow shading between 3464 and 3611) were in addition scaled linearly such that the end position of Pol II on DNA corresponds to the expected runoff length (3824 bp, green arrow). The original trace is shown in blue, and the stretched trace is in black. The discrepancy between the two traces presented in this figure is the highes we have dealt with. In general the difference between the end of the trace and the expected runoff is 15-20 bp.

crossed the NPS, the final position should match the runoff position of Pol II from DNA. We use this information to improve our positional accuracy. We perform a two-point calibration of all traces that transcribed past position 3650 (i.e. past the NPS) by aligning them both to the stall site and to the expected runoff length, as shown in Figure 2.14. 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 [53]. We will discuss these trends in more detail in Chapter 3.

2.8.3 Pause Detection and Pause-Free Velocity

We identified pauses using the position versus time data decimated to 50 Hz and smoothed to 1 Hz as described in the previous section. First, we divided the position data into 3 base pairs increments and compute the dwell time Pol II spends in each of these bins. We used bins that were 3 base pairs wide because this value represents the limit of our resolution. For each trajectory, we computed the mean and the standard deviation of the dwell times, discarded the outliers that were more than three standard deviations away from the mean,

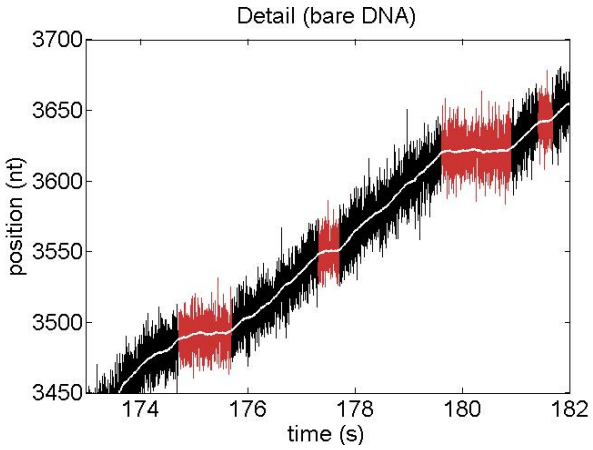

Figure 2.15: **Algorithm for picking pauses.** Trajectories were sampled at high bandwidth and averaged to lower bandwidth for further analysis. Because pause density is relatively low, our velocity resolution allowed us to identify individual pausing events (red). Pauses lasting longer than 1 s were identified with the greatest accuracy.

and recalculated the average dwell time. Finally, pauses were defined as dwells that were at least 2.5 times or longer than this average dwell time. This algorithm results in a pause threshold that varies slightly from trace to trace, but in the majority of the traces, the pause threshold is lower than 0.5 s. In addition, we joined pauses that were separated by less than three base pairs. To avoid miscounting errors caused by the effects of long-timescale drift on automated measurements, pauses longer than a few minutes were checked by eye and corrected if necessary [72].

To obtain a pause-free velocity, the times associated with the beginning and end of each pause were recorded and used to remove all identified pauses from the original (high-bandwidth unfiltered) position signal (Figure 2.16). Each pause-free trajectory was then filtered using a Gaussian convolution filter with a time constant of 1 s for further analysis. The pause-free velocities for each trace are given by the time derivatives of the filtered position signal throughout the NPS.

In our analysis of pause durations, we considered pauses with durations of 1 s or longer. However, to obtain pause-free velocities, we removed all identified pauses, including those shorter than 1 s. This difference arises because for pause durations, it is critical to have certainty regarding the identification and timing of pauses, so we have taken a more conservative approach and made measurements where we have the highest degree of certainty. For removing pauses, however, it is more important to liberally remove all identified pauses, including those whose timings are less well resolved, in order to obtain the pause-free velocity.

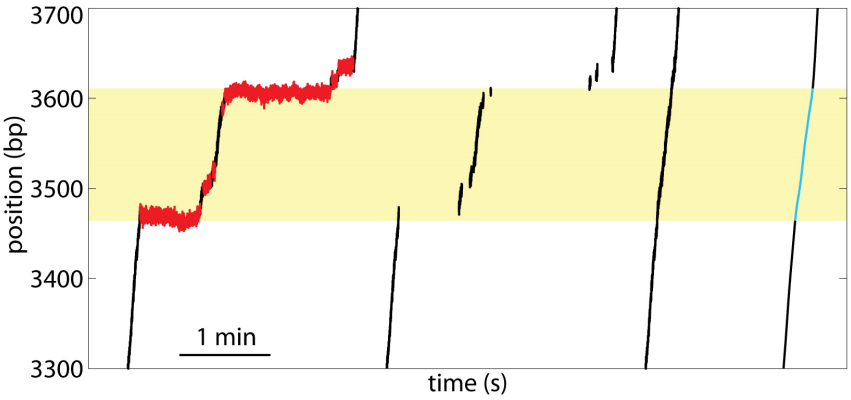

Figure 2.16: **Pause removal algorithm.** Portions of the trajectories identified as pauses (red) are removed at high bandwidth, and the remaining pause-free trajectory is rejoined then averaged to a lower bandwidth.

2.9 Appendix: Kinetic Model Details

2.9.1 Derivation of Pause Entry Probabilities

Bare DNA

In the absence of a nucleosome, the probability of entering a pause at each position on DNA is:

$$P_{\text{naked DNA}} = \frac{k_b}{k_b + k_e}$$

Nucleosome

When the nucleosome is present (Figure 2.8), if the polymerase cannot transcribe while the nucleosome is wrapped in front of it, Pol II can enter into a pause (i.e. backtrack by one base) via two pathways: it can backtrack directly when the nucleosome is unwrapped or, if the nucleosome wraps, backtracking competes with unwrapping of the nucleosome. These two paths competing with elongation correspond to the two terms in the numerator of the fraction bellow:

$$P_{\text{nucleosome}} = \frac{k_b + k_w \dfrac{k_b}{k_b + k_u}}{k_b + k_w \dfrac{k_b}{k_b + k_u} + k_e} \tag{2.22}$$

If we assume very fast kinetics of the nucleosome, in particular $k_u >> k_b$, we can rewrite:

$$P_{\text{nucleosome}} = \frac{k_b + k_w \dfrac{k_b}{k_u}}{k_b + k_w \dfrac{k_b}{k_u} + k_e} = \frac{k_b \left(1 + \dfrac{k_w}{k_u}\right)}{k_b \left(1 + \dfrac{k_w}{k_u}\right) + k_e} = \frac{k_b \left(\dfrac{k_u + k_w}{k_u}\right)}{k_b \left(\dfrac{k_u + k_w}{k_u}\right) + k_e}$$

We have already defined $\gamma_u = \dfrac{k_u}{k_u + k_w}$, so we can simplify the pause density to:

$$P_{\text{nucleosome}} = \frac{k_b \gamma_u^{-1}}{k_b \gamma_u^{-1} + k_e}, \text{ which simplifies to: } P_{\text{nucleosome}} = \frac{k_b}{k_b + \gamma_u k_e} \tag{2.23}$$

The final expression for the pause density when the nucleosome is present is intuitive since backtracking is competing with elongation, and elongation can only take place when the nucleosome is unwrapped.

Nucleosome - active DNA unwrapping

If the polymerase can transcribe while the nucleosome is wrapped in front of it, with a rate $k_{e,w}$, as shown in Figure 2.11, the probability of pausing becomes:

$$P_{\text{nucleosome, active unwrapping}} = \frac{k_b + k_w \dfrac{k_b}{k_b + k_u + k_{e,w}}}{k_b + k_w \dfrac{k_b}{k_b + k_u + k_{e,w}} + k_e + k_w \dfrac{k_{e,w}}{k_b + k_u + k_{e,w}}} \tag{2.24}$$

Again, assuming very fast kinetics of the nucleosome compared to backtracking, we can re-write the probability of entering a pause, as follows:

$$\begin{aligned} P_{\text{nucleosome, active unwrapping}} &= \frac{k_b + k_w \frac{k_b}{k_u}}{k_b + k_w \dfrac{k_b}{k_u} + k_e + k_{e,w} \dfrac{k_w}{k_u}} \\ &= \frac{k_b \left(\dfrac{k_u + k_w}{k_u}\right)}{k_b \left(\dfrac{k_u + k_w}{k_u}\right) + k_e - k_{e,w} + \left(\dfrac{k_u + k_w}{k_u}\right) k_{e,w}} \end{aligned} \tag{2.25}$$

We use the previously defined γ_u, to simplify the pause density to:

$$P_{\text{nucleosome, active unwrapping}} = \frac{k_b}{k_b + \gamma_u k_e + (1 - \gamma_u) k_{e,w}} \tag{2.26}$$

The final expression for the pause density when we allow the polymerase to actively open the nucleosome and elongate through it at a rate $k_{e,w}$ is also intuitive. In this case backtracking is competing with two different elongation rates, one when the nucleosome is unwrapped (k_e), and one through a wrapped nucleosome ($k_{e,w}$).

2.9.2 Derivation of the Pause Durations Distributions

The predicted distribution of pause durations $\psi(t)$ arises from a model in which the polymerase's dynamics during backtracking are described by a single intrinsic stepping rate of diffusion k_0 [52]. This model of polymerase backtracking has been proposed to explain pauses previously described as either "long (backtracking)" or "short (ubiquitous)" pausing [73]. Below, we derive the probability density function, $\psi(t)$, starting from first principles of counting and probability. We also briefly discuss how the backtracking dynamics give rise to two apparent pausing regimes.

Below, the DNA is treated as a one-dimensional lattice, where each lattice site corresponds a base-pair position along the DNA. We begin by considering a polymerase that arrives at a backtracked state by having moved backwards by 1 bp to position -1 on the lattice. Here, the rate-limiting step for backtrack recovery is diffusion back to the 3' end of the nascent RNA (i.e. to position 0 on the lattice), where it can re-start transcription elongation. The probability density of return times $\psi(t)$ then describes the predicted distribution of pause times.

Derivation of $\psi(t)$ begins by solving for the distribution of first-passage times for a particle at position -1, diffusing to position 0 on the lattice. A path starting from position -1 and ending at position 0 has a total of $2n+1$ steps: n backward steps, and $n+1$ forward steps. The last step in the path is always going from -1 to 0 (pause recovery), so in order to specify a path, it suffices to choose n steps (the backwards ones say) from the remaining $2n$ steps. For each value of n, the total number of trajectories starting at -1 and ending at 0 is given by the binomial coefficient:

$$\binom{2n}{n} = \frac{(2n)!}{n!\,n!} \tag{2.27}$$

However, the number of unique trajectories is smaller by a factor of $n+1$ than the total number of trajectories, as we have to discard the ones that cross 0 more than once (upon the first return to position 0, the pause ends), and is given by the Catalan numbers [74]:

$$C_n = \frac{(2n)!}{(n+1)!\,n!} \tag{2.28}$$

The probability of such a trajectory can be calculated if the probabilities of stepping forward (p) and backward (q) are known [75]. These probabilities can be calculated in terms

of Pol II's forward (k_f) and backward (k_b) stepping rates in the backtracked state:

$$p = \frac{k_f}{k_f + k_b}, \text{ and } q = \frac{k_b}{k_f + k_b}. \tag{2.29}$$

Combining equations 2.28 and 2.29, we can write the probability of having a paused trajectory with $2n + 1$ steps:

$$P(n) = \frac{(2n)!}{(n+1)!\, n!} \left(\frac{k_f}{k_f + k_b}\right)^{n+1} \left(\frac{k_b}{k_f + k_b}\right)^{n}. \tag{2.30}$$

The probability density $\psi(t)$ is calculated based on the probability of each trajectory $P(n)$ and the distribution of times associated with that trajectory $P(t \mid n)$:

$$\psi(t) = \sum_{n=0}^{\infty} P(n) \cdot P(t \mid n)\,. \tag{2.31}$$

The time durations for taking $2n+1$ Poisson steps, $P(t \mid n)$, is described by the Gamma distribution, with a time decay constant at each step depending on the sum of the forward and reverse rates, $k_f + k_b$:

$$P(t \mid n) = t^{2n} \frac{(k_f + k_b)^{2n+1}\, e^{-(k_f+k_b)t}}{(2n)!}. \tag{2.32}$$

From the definition of $\psi(t)$ (equation 2.31), and the equations describing $P(n)$ and $P(t \mid n)$ above, we can write the full expression for the pause durations distribution:

$$\psi(t) = \sum_{n=0}^{\infty} \left(\frac{(2n)!}{(n+1)!\, n!} \left(\frac{k_f}{k_f + k_b}\right)^{n+1} \left(\frac{k_b}{k_f + k_b}\right)^{n} t^{2n} \frac{(k_f + k_b)^{2n+1}\, e^{-(k_f+k_b)t}}{(2n)!} \right) \tag{2.33}$$

Simplifying the expression above and incorporating the definition of the modified Bessel function of the first kind (I_1):

$$I_1(x) = \sum_{n=0}^{\infty} \frac{1}{(n+1)!\, n!} \left(\frac{x}{2}\right)^{2n+1}, \tag{2.34}$$

allows us to rewrite this equation in the closed-form solution cited in the main text:

$$\psi(t) = \sqrt{\frac{k_f}{k_b}} \cdot \frac{e^{-(k_f + k_b)t}}{t} \cdot I_1(2t\sqrt{k_f k_b}). \tag{2.35}$$

The complete dynamics of backtracking are described as the function of the forward

and backward stepping rates k_f and k_b, which are themselves functions of the intrinsic rate of diffusing while backtracked k_0, and the applied force F according to an Arrhenius-type equation:

$$k_f = k_0 \exp(F \cdot d/k_B T) \text{ and } k_b = k_0 \exp(-F \cdot d/k_B T), \tag{2.36}$$

with the distance to transition state (d) taken here to be 0.5 bp.

While force always biases the random walk, the force dependence of any one base-pair step is relatively small. Because the distance to transition state d is very small (0.5 bp, corresponding to 0.17 nm), a large force must be exerted to yield a significant bias to the individual forward and reverse stepping rates. For instance, if even a 10-pN force is applied, the energetic bias provided by the force is $F \cdot d = 1.7$ pN $\cdot$ nm, which is small compared with thermal fluctuations ($k_B T = 4.1$ pN $\cdot$ nm).

Because short pauses consist mostly of very short backtrack excursions ($\sim$1-2 bp), they display apparent force insensitivity. In contrast, since longer pauses are comprised of many more steps, the effect of force on these pauses is compounded. Therefore, diffusive backtracking gives rise to two apparent pausing regimes that were interpreted before as resulting from different pausing mechanisms [73].

The backtracking model correctly identifies the distribution of all pauses as short as 1 second in duration (Figure 2.9), showing that all such pauses can be considered as backtrack excursions that recover through the same mechanism.

Chapter 3

Determinants of the Nucleosomal Barrier

As underlined in Chapter 1, the histone proteins contain two important regions: the histone-fold domains and the histone tails. The histone-fold domains make direct contacts with the DNA and organize it into the superhelical structure specific to the nucleosome [16]. The histone tails don't generally make such specific contacts with the DNA, but they are highly positively charged and can stabilize the nucleosome further. Both of these regions, but especially the tails, are the target of many post-translational modifications associated with gene expression [76]. In addition, the sequence of the DNA wrapped around the histones promotes pausing in certain regions of the nucleosome [34, 35], thus modulating the nucleosomal barrier.

Here, we investigate how each of these three factors—the histone tails, the specific histone-DNA contacts, and the DNA sequence—contributes directly to the barrier that the nucleosome imposes on transcription elongation by RNA polymerase II.

3.1 Transcription of Nucleosomes Containing Modified Histones

It has long been known that acetylation of histone tails leads to increased rates of transcription of chromatin-derived templates [77], and that acetylated histones are associated with transcriptionally active genes *in vivo* [78]. Acetylation and other covalent modifications such as methylation, phosphorylation and ubiquitination, are known to play diverse roles, including demarcation of heterochromatin [79–82], gene activation and silencing [81, 83, 84] and delimiting chromatin boundaries during replication [85, 86]. The proposed histone code of epigenetic regulation [18, 87] predicts that particular combinations of histone modifications serve as regulatory markers indirectly through their specific affinities for effector proteins. The identification of modular bromo- and chromodomain-containing proteins [88, 89], which

Portions of the work described in this chapter were conducted with Toyotaka Ishibashi, Yueh-Yi Wu, and Manchuta Dangkulwanich, and are used with their permission.

recognize and act on specific patterns of histone modifications, lends great support to this model; however, a variety of evidence suggests that, in addition to their role in the epigenetic code, certain histone modifications may also play an important role in the direct regulation of transcription elongation by modulating the affinity between histones and DNA.

Biochemical studies of transcription through nucleosomes have revealed that histone tails play an important role in establishing a barrier to transcription in the absence of other factors; when the histone tails are either acetylated or removed by proteolytic cleavage, the nucleosomal barrier to transcription is significantly reduced [90, 91]. In addition, DNA wrapped around acetylated histones has increased DNase I sensitivity compared to non-acetylated histones [92, 93], and acetylation also reduces the mechanical stability of chromatin templates, allowing these templates to be more easily unwrapped by the application of mechanical force [94]. Acetylation of histone tails results in the neutralization of the net positive charge of lysine side-chains. It has been speculated that this reduction in positive ionic character reduces the affinity between the core histones and DNA, resulting in a "loosening" of the DNA wrapped into a nucleoosme.

We explore the direct effect of the tails on transcription by using recombinant histones that have been expressed without the tail domains. To understand how much of the tails' effect is mediated through their positive charges, we use nucleosomes with mock-acetylated histone tails, that is tails containing point substitutions of lysines to glutamines.

In order to examine the importance of direct histone-DNA contacts on the nucleosomal barrier, we take advantage of Sin (Swi-independent) mutants of histones H3 and H4, which partially relieve loss-of-function mutations of the chromatin remodeling complex SWI/SNF (switch/sucrose nonfermentable) [95]. The structures of many of Sin-mutant nucleosomes have been obtained by X-ray crystallography [96]. Class I Sin mutants, specifically H4 R45A and H3 T118H, target specific protein-DNA contacts along the DNA minor groove. These single amino acid mutations result in minimal structural changes to the nucleosome. However, this minor reorganization of the protein structure changes the hydrogen-bond distance between the protein and many of the DNA phosphates. Each of the above mutants leads to increased mobility of the nucleosome on DNA [96], suggesting that the DNA has reduced affinity for the mutant histones, and that the mutants partially rescue SWI/SNF mutations by lowering the nucleosomal barrier to transcription.

We reconstitute core nucleosomes using the H4 R45A and H3 T118H Sin mutant histones (see Section 3.9 for details), and then directly observe transcription through these mutant particles using the experimental methods presented in Chapter 2. Throughout the rest of the text we refer to these mutant nucleosomes as Sin H4 and Sin H3 respectively.

Using the experimental setup described in Chapter 2, Figure 2.1, we followed the trajectories of single Pol II elongation complexes in real time as they transcribe through each of these modified nucleosomes. Since we were interested in investigating how the dynamics of transcription is modulated by the nucleosomal modifications, we performed the single-molecule experiments at a moderately permissive salt concentration, 300 mM KCl. Representative traces are presented in Figure 3.1, and show the general trends of transcription for each

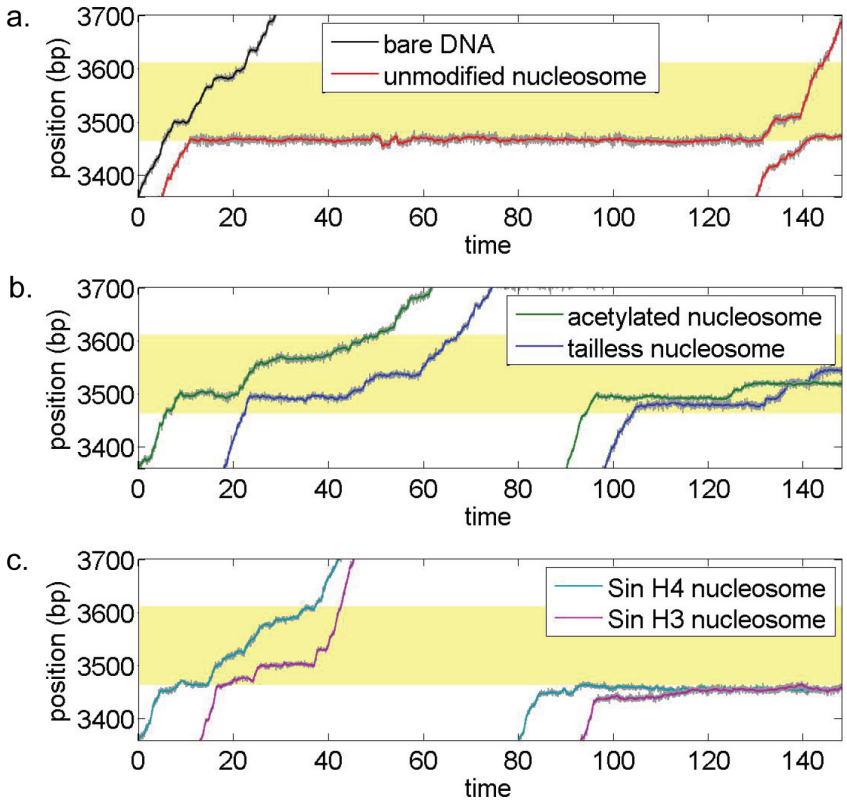

Figure 3.1: **Transcription through modified nucleosomes.** Position of Pol II during single-molecule transcription of **(a)** bare DNA (black) and unmodified nucleosomes (red), **(b)** tailless (blue) and mock-acetylated (green) nucleosomes, **(c)** Sin H3 (purple) and Sin H4 (cyan) mutant nucleosomes. Traces that passed the NPS region are shown on the left and traces that arrested at the nucleosome are on the right.

construct. For each trace where Pol II has completed transcription of the NPS, we identified the regions of pausing and active elongation, and we quantified the pause durations and the number of pauses per base pair (pause density), as described in Section 2.8.

As shown in Chapter 2, transcription on bare DNA (Figure 3.1a, black) has portions of fast translocation, punctuated by pauses which are short in duration. In this experimental setup, most Pol II elongation complexes (87%) transcribed to the end of the template (Figure 3.2a).

Transcription through unmodified nucleosomes (Figure 3.1a, red) sits at the opposite side of the spectrum. Elongation is interrupted by very long pauses, and the total time it takes to cross the nucleosome varies from tens of seconds to a few minutes. Often these pauses turn into arrests, so only 52% of polymerases manage to overcome the nucleosomal barrier (Figure 3.2b).

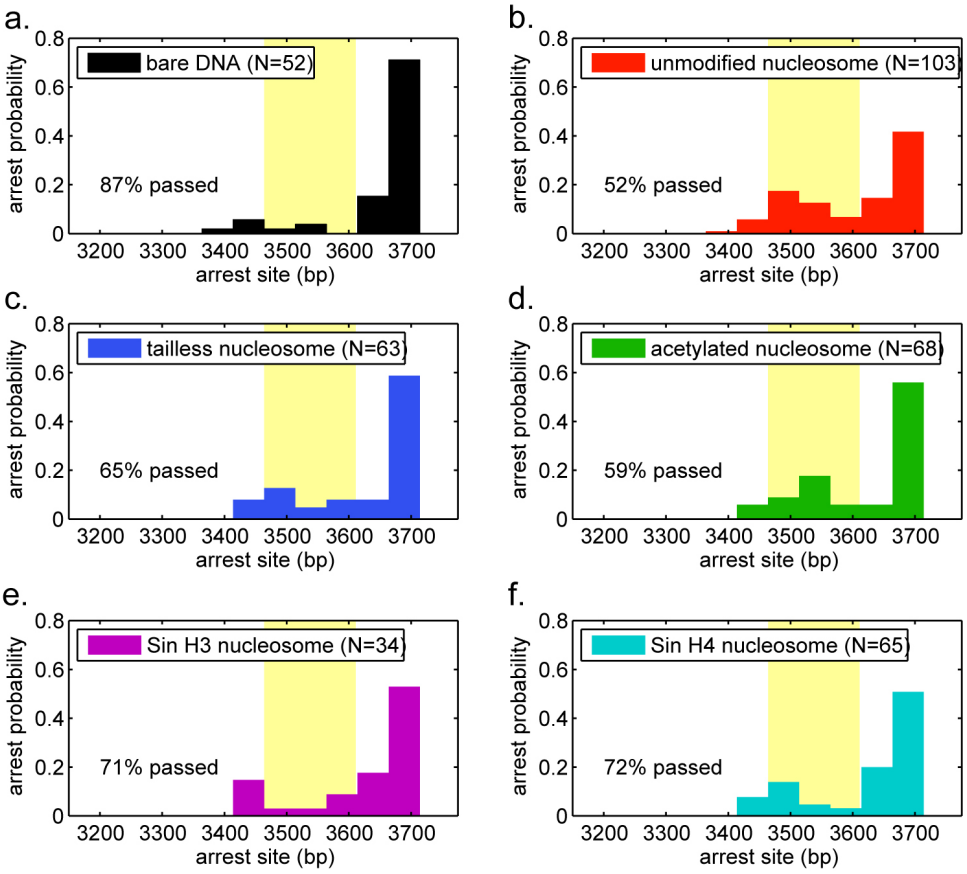

Figure 3.2: **Arrest probabilities for modified nucleosomes.** Position of terminal arrest on the DNA template, as determined from single-molecule trajectories at 300 mM KCl for: **(a)** bare DNA, and **(b)** unmodified, **(c)** tailless, **(d)** mock-acetylated, **(e)** Sin H4, **(f)** Sin H3 nucleosomes.

Overall transcription through tailless and acetylated nucleosomes is faster than through unmodified nucleosomes (Figure 3.1b, blue and green respectively), with crossing times that are generally under one minute. In fact, the average pause duration for traces that crossed the NPS for tailless nucleosomes (6.6±0.7 s) is much closer to that of bare DNA (4.8±0.8 s) than to the mean pause duration for transcription through unmodified nucleosomes (10.7±1.2 s). This observation suggests that the histone tails affect pause recovery during passage through the nucleosomal barrier. The pause durations for acetylated nucleosomes are in between tailless and unmodified, with a mean much closer to unmodified nucleosomes (8.1±1.1 s), suggesting that the lysine charges on the histone tails constitute only a small part of the interaction of the tails with DNA. Both the removal and acetylation of the tail result in an increased efficiency in crossing the NPS, : 59% for acetylated nucleosomes, and 65% for

tailless nucleosomes (Figure 3.2c and Figure 3.2d respectively), in agreement with previously published results obtained using classical biochemistry transcription assays [91]. Moreover, there is a reduction of the arrests in the entry region of nucleosomes with modified tails (Figure 3.1b and Figure 3.2c&d).

The effect of the Sin mutations on nucleosomal transcription is the greatest, decreasing even further the time it takes to cross the nucleosome (Figure 3.1c), to values that are generally under 30 seconds. Pause durations were also shorter on average for the Sin mutants compared with unmodified nucleosomes (7.7±0.9 s for Sin H4 and 6.4±1.1 s for Sin H3). Correspondingly, these mutations increase the probability of Pol II to pass the nucleosome to 71% for Sin H4 and 72% for Sin H3 (Figure 3.2e&f), in agreement with recently published reports [97].

Note that the ability of the Sin mutations to weaken the nucleosomal barrier is much larger than that of the tailless or acetylated nucleosomes. This finding is quite surprising, considering that the Sin mutants differ only by one amino acid from to the unmodified nucleosomes, while the tailless nucleosomes are missing ∼25% of the total mass of the histones. These results point to the importance of the specific contacts that the histone-fold domains make with the DNA into setting the nucleosomal barrier.

3.2 The Histone Tails Gate the Nucleosome Entry

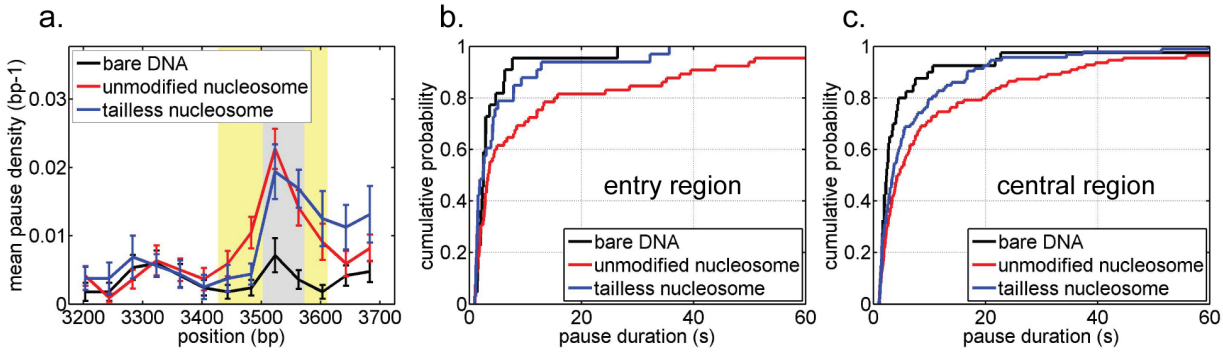

Figure 3.3: **Tails affect pausing in the nucleosome entry region. (a)** Pause density as a function the position of Pol II's active center on the template. The nucleosome entry /exit regions are shaded yellow, and the central region is shaded gray. **(b)** Pause durations in the entry region of the nucleosome. **(c)** Pause durations in the central region of the nucleosome.

A closer look at the data 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 (Figure 3.3). We define the entry region as consisting of the DNA segment bound to the H2A/H2B dimer (position -75 to -35 with respect to the dyad), as well as of an additional segment (35 base pairs) of the DNA immediately preceding it. We include this additional DNA to the definition of the entry region for two reasons: because the leading edge of the polymerase reaches the nucleosome ∼20 bp before the active center [48], and because previous experiments suggest that the tails bind the DNA outside the NPS organized around the core nucleosomal particle [49].

Accordingly, we observe that the number of pauses in this region is significantly decreased for tailless nucleosomes (Figure 3.3a). Moreover, the pauses in the entry region are significantly shorter for tailless nucleosomes (Figure 3.3b). In the central region, both pause densities and pause durations are slightly reduced by tail removal, indicating that the tails might have a small effect on the stability of the central region as well. However, this difference is not statistically significant, and it is much smaller than the large effect of tail removal on the entry.

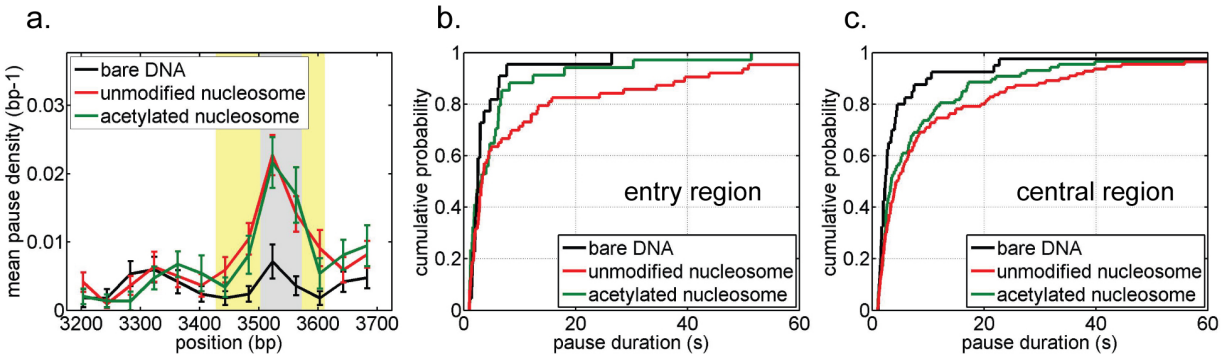

Figure 3.4: **Transcription through acetylated nucleosomes.** **(a)** Pause density as a function the position of Pol II's active center on the template. The nucleosome entry /exit regions are shaded yellow, and the central region is shaded gray. **(b)** Pause durations in the entry region of the nucleosome. **(c)** Pause durations in the central region of the nucleosome.

The effect of mock acetylation of the tail domains seems to be quite small both for pause densities and for pause durations. While we do see a reduction in the pause durations in the entry region (Figure 3.4b), pause densities in this region are not significantly different from the unmodified nucleosome (Figure 3.4a). As expected, the pausing in the central region doesn't differ from unmodified nucleosomes either. As before, we conclude that the interactions between DNA and the lysines we mutated in this study only constitute a small component of the interaction of the tails with DNA. These results suggest that, most likely, acetylation at these residues *in vivo* serves as a target for binding of specific chromatin remodeling factors to the nucleosome rather than representing a direct mechanism for

attenuation of the nucleosomal barrier.

3.3 Histone-DNA Contacts at the Dyad

3.3.1 Sin H4 Mutant: Strong Localized Effect

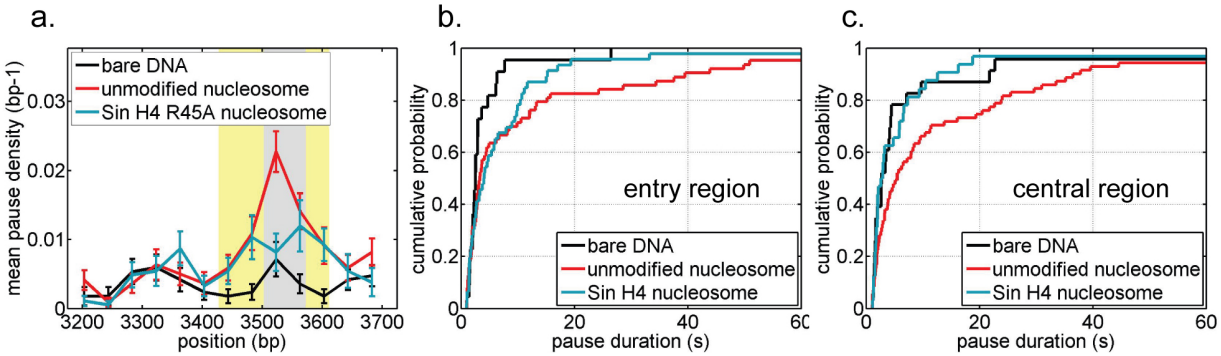

Figure 3.5: **Sin H4 nucleosome destabilized at the dyad.** **(a)** Pause density as a function the position of Pol II's active center on the template. The nucleosome entry and exit regions are shaded yellow, and the central region is shaded gray. **(b)** Pause durations in the entry region of the nucleosome. **(c)** Pause durations in the central region of the nucleosome.

In contrast to the effects observed during transcription through tailless and acetylated nucleosomes, the major effect of the Sin H4 mutant on the pause density was in the region immediately before the dyad, which in the unmodified nucleosome constitutes the major barrier to transcription. In agreement with these observations, the pause durations were significantly shorter for the Sin H4 mutant in this central region (Figure 3.5b). Since we map the position of the active center of the polymerase on DNA, we conclude that this major change occurs when the leading edge of the polymerases reaches the dyad of the nucleosome (roughly 15 bp ahead of the active center). In contrast to this strong effect at the dyad, pause densities in the entry region during elongation through Sin H4 nucleosomes were very similar with those through the unmodified nucleosomes (Figure 3.5a). The pause durations at the entry region were also not significantly different between unmodified and Sin H4 nucleosomes (Figure 3.5c).

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

3.3.2 Sin H3 mutant: Widespread Effects

We have also attempted to divide the pausing data for transcription of the Sin H3 nucleosomes into functional regions. However, since we observed a decrease in pause numbers and durations in all regions of the nucleosome (entry, central, and exit), we concluded that the Sin H3 mutation destabilizes the nucleosome overall. Therefore, we proceeded to analyzing the pausing characteristics over the entire nucleosomal region as a whole.

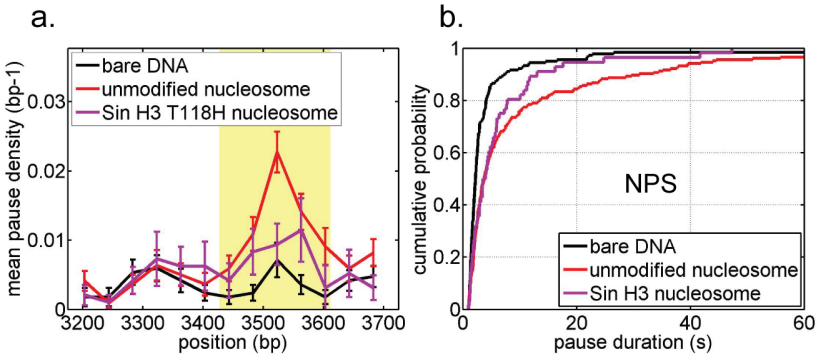

Figure 3.6: **Sin H3 destabilized overall. (a)** Pause density as a function the position of Pol II's active center on the template. **(b)** Pause durations in the entire nucleosomal region.

For the Sin H3 nucleosome, the mean pause density in the nucleosomal region was much lower than for the unmodified nucleosome—especially in the central region (Figure 3.6a). Moreover, immediately after Pol II passes the dyad region, the pause density decreased close to the value achieved on bare DNA, suggesting that the Sin H3 modification leads to an increased dissociation of the histones once the polymerase has cleared the dyad. Pause durations for the Sin H3 nucleosome were also shorter than in the unmodified nucleosome, not only in the central region, but throughout the entire NPS (Figure 3.6b).

The widespread effects that we observe for Sin H3 suggest that this mutation produces a change in the packing of the histones, and thus in their ability to organize the DNA into a nucleosome. Indeed, the crystal structure of Sin H3 T118H [96] and molecular dynamics simulations of the DNA wrapped around this mutant [98] show that the change from threonine to the bulkier histidine leads to a rearrangement of two alpha helices—belonging to H3 and H4 respectively. One end of the H3 helix containing the mutated amino-acid is in the vicinity of the dyad, and its other end contacts the DNA two double-helical turns away. The H4 helix is affected because the Sin H3 mutation disrupts hydrogen bonding between the H3 T118 and the neighboring H4 R45, pushing the H4 R45 further into the minor groove. The effect of the Sin H3 mutation could be transmitted via these helices throughout the nucleosome to produce the observed overall changes in transcription kinetics.

3.4 Changes in the Nucleosome Wrapping Equilibrium

We apply the model developed in Chapter 2 to analyze pausing in the different nucleosomal regions. This allows us to dissect in a piece-wise manner the changes in dynamics of the modified nucleosomes that give rise to the observed effects on transcription.

We have proved that pausing on bare DNA is well described by a diffusive backtracking model. In this model the polymerase enters a pause when the 3' end of the RNA becomes misaligned with the active center of the polymerase. During the paused state, Pol II diffuses back and forth on DNA with a rate k_0. Elongation resumes once the polymerase diffuses back to its original elongation-competent position on the RNA-DNA hybrid. We have also shown that a wrapped nucleosome prevents both elongation and backtrack recovery. The effect of the fluctuations on transcription can be summarized by a single parameter: the probability of finding the nucleosome unwrapped, $\gamma_u = k_u/(k_u + k_w)$.

We use the previously determined value of diffusion during backtracking on bare DNA, $k_0 = 1.3\ s^{-1}$, and extract values describing the local wrapping equilibrium of the modified nucleosomes (γ_u) by fitting the distributions of pause durations for these nucleosome constructs. Since we want to compare these results with those of the wrapping-unwrapping equilibrium measured in the absence of Pol II (presented in the next section), we convert the γ_u to a wrapped equilibrium constant, K_w, as follows:

$$K_w = \frac{k_w}{k_u} = \frac{1-\gamma_u}{\gamma_u} \tag{3.1}$$

Table 3.1: **Nucleosome local equilibrium during transcription** The relative wrapping equilibrium constant, $K_w^{Relative}$, is computed with respect to the unmodified nucleosome. Values that changed significantly for the modified nucleosomes compared to the unmodified one are in red.

	entry region		**central region**	
	K_w	$K_w^{Relative}$	K_w	$K_w^{Relative}$
unmodified	0.63 [0.2-1.5]	1	0.67 [0.53-0.83]	1
tailless	0.20 [0-0.43]	0.31	0.63 [0.38-1]	0.94
acetylated	0.25 [0-0.67]	0.40	0.63 [0.38-1]	0.94
Sin H4	0.67 [0.56-0.83]	1.1	0.28 [0-0.43]	0.41
Sin H3	0.32 [0.23-0.43]	0.51	0.32 [0.23-0.43]	0.47

As presented in Table 3.1, the absence of the tails decreases the local wrapping equilibrium constant of the nucleosome by a factor of 3 in the entry region, but does not significantly affect the dyad region. Sin H4 , by comparison, leads to a decrease in the wrapping equilibrium

by a factor of 2.5 near the dyad, but does not affect the entry into the nucleosome. Lastly, Sin H3 decreases the wrapped constant by a factor of 2 everywhere along the NPS.

3.5 Independent Measurements of Changes in Nucleosome Fluctuations

We wondered whether the effects we observe on transcription through the modified nucleosomes are a direct effect of changes in the intrinsic DNA-histone wrapping and unwrapping rates or whether Pol II contributes significantly to the destabilization of these nucleosomes. Moreover, so as to obtain a picture of the dynamic changes induced by the histone mutations in each specific scenario, we would also like to know which rate is the one that changes: the unwrapping or the re-wrapping rate of DNA to histones.

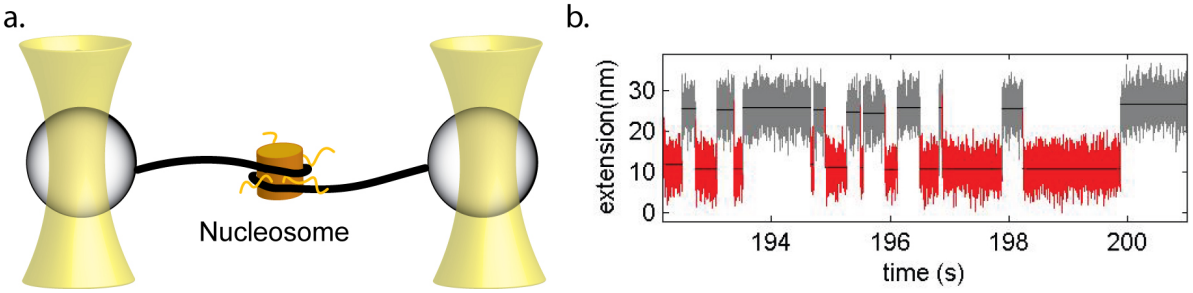

Figure 3.7: **Nucleosome dynamics in the absence of Pol II. (a)** Experimental setup for single-molecule nucleosome dynamics experiments. **(b)** Example of hopping events between the wrapped state (red) and the unwrapped one (gray).

In order to answer these questions, we probed the stability of nucleosomes by observing them unwrap and rewrap under force in the absence of the polymerase, using the setup shown in Figure 3.7a. In these type of experiments, as the force is increased, the DNA unwraps from the nucleosome in two steps (see Figure 2.3). The first step, which happens around 3-4 pN, is called the outer wrap, and is identified as the release of DNA from the H2A/H2B dimers. The second step, which takes place at a higher force (above 5 pN) and is called the inner wrap, corresponds to the central DNA wrapped around the H3/H4 tetramer. In other words, the outer wrap represents the entry/exit regions of the nucleosome and inner wrap represents the central region [47]. By holding the positions of the optical traps fixed such that the force applied to the nucleosome is in the correct range, we can observe unwrapping and rewrapping transitions of the nucleosome (Figure 3.7b) either of the entry-exit regions or of the central region. The results of these experiments are summarized in Tables 3.2 and 3.3.

We observe the same changes (shown in red) in the wrapped equilibrium constant of modified nucleosomes compared to unmodified ones $K_w^{Relative}$, as we saw in our transcription

Table 3.2: **Nucleosome hopping rates in the entry region**

	k_u (s^{-1})	F (pN)	N	k_w (s^{-1})	F (pN)	N	K_w	$K_w^{Relative}$
unmodified	0.18±0.18	4.0±0.6	131	1.1±0.2	3.7±0.5	100	6.2±1.7	1
tailless	too fast	-	-	too slow	-	-		0
acetylated	0.15±0.22	3.7±0.6	88	0.4±0.4	3.7±0.4	36	2.4±1.0	0.39
Sin H3	1.20±0.12	3.3±0.2	294	1.0±0.2	3.0±0.2	170	0.9±0.2	0.15
Sin H4	0.15±0.22	3.8±1.1	84	0.9±0.2	4.2±0.9	149	6.1±1.7	0.99

Table 3.3: **Nucleosome hopping rates in the central region**

	k_u (s^{-1})	F (pN)	N	k_w (s^{-1})	F (pN)	N	K_w	K_w^{Rel}
unmodified	0.43±0.15	7.4±0.6	177	4.8±0.2	6.8±0.4	134	11.0±2.6	1
tailless	0.98±0.15	7.4±0.3	172	7.4±0.2	7.3±0.2	142	7.6±1.7	0.69
acetylated	0.45±0.26	7.4±0.2	63	3.2±0.6	6.8±0.2	14	7.2±4.6	0.65
Sin H3	0.33±0.30	7.7±0.5	50	1.6±0.4	7.5±0.4	33	5.0±2.4	0.45
Sin H4	0.39±0.15	7.3±0.8	184	1.2±0.2	7.1±0.6	88	3.1±0.8	0.28

experiments. These results strongly suggest that the polymerase does not directly influence the wrapping/unwrapping equilibrium, but most likely simply exploits local nucleosome fluctuations.

From these experiments we notice that the Sin mutations decreases only the re-wrapping rates of the nucleosome in the central region. This result helps explain why other laboratories have noticed a decrease in the efficiency of upstream histone transfer during transcription for this mutant [97]. Sin H3 also shows an increase in the unwrapping rate in the entry region, which rationalizes the reduced overall pausing we observe in our transcriptional data.

The results with the tailless nucleosomes are perhaps more surprising. First of all, we notice an increase in both the unwrapping and the re-wrapping rates in the central region, but their ratio, which determines the equilibrium constant, is very similar with that of unmodified nucleosomes. This result indicates that the tails affect fluctuations of the nucleosome near the dyad, but don't affect the overall stability of this region, and therefore don't significantly affect transcription. Secondly, in 300 mM KCl we couldn't detect unwrapping under force of the entry/exit regions for the tailless nucleosomes. However, these regions are protected against micrococcal nuclease digestion for the tailless nucleosomes at this salt concentration (data not shown), which means that the entry region of the NPS is associated with histones. We interpret these observations as indicative that this outer region unwraps so fast in 300 mM KCl under the addition of force that it is outside our experimental measurement range.

Since we still wanted to test the trends in stability of the entry region for tailless nucleo-

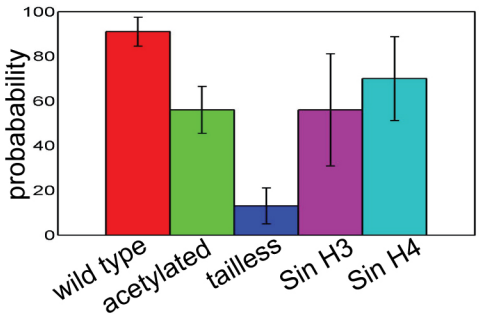

Figure 3.8: **Unwrapping of the nucleosome entry region.** Probability of observing a distinct unwrapping event of the entry region of the nucleosome (measured in 40 mM KCl).

somes, we performed the nucleosome mechanical manipulation experiments at a lower ionic strength, 40 mM KCl. In these conditions, while $91 \pm 6\%$ of unmodified nucleosomes showed an outer wrap, only $56 \pm 10\%$ of the acetylated nucleosomes and as little as $13 \pm 8\%$ of the tailless nucleosomes had an outer wrap (Figure 3.8).

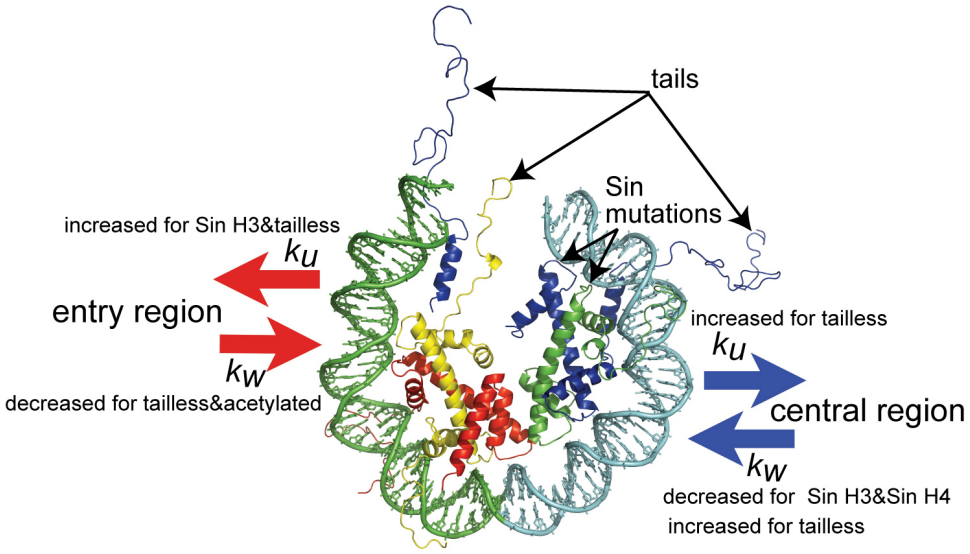

Figure 3.9: **Summary of nucleosome fluctuations changes.** Half of the crystal structure of the nucleosome, with the DNA in the entry region in green, the DNA in te central region in grey, H2A in yellow , H2B in red, H3 in blue, and H4 in green.

These results match our transcription observations that the entry point in the nucleosome is highly destabilized for tailless nucleosomes and moderately for acetylated nucleosomes. As

expected, for the Sin H3 nucleosome we also saw a change in the probability of observing the unwrapping of the outer wrap—only $56 \pm 25\%$ of Sin H3 nucleosomes showed an outer wrap. For the Sin H4 nucleosome, $70 \pm 19\%$ of the molecules showed rip associated with the entry/exit region, which is not significantly different from the unmodified nucleosome, as expected.

Even at this low ionic strength we didn't see any re-wrapping events of the entry region for the tailless nucleosomes, suggesting that while unwrapping happens very fast, re-wrapping is a very slow event in the absence of the tails. This dramatic decrease in the re-wrapping rates in the entry region suggests another function for the tails, that of helping in docking the outer segments of the nucleosomal DNA to wrap around the histone core.

Our findings about the changes we observe in the wrapping-unwrapping fluctuations for the modified nucleosomes are summarized in Figure 3.9.

3.6 The Nucleosome Amplifies Sequence-Dependent Pausing

So far, when analyzing pausing, both on bare DNA and at the nucleosome, we assumed that:

1. The DNA landscape is uniform, with no difference between GC-rich or AT-rich regions.

2. The polymerase can backtrack unimpeded arbitrarily far on the DNA template.

We do not believe that these two assumptions accurately represent reality, but they were convenient, as they helped us coarse grain the problem. Despite these simplifications, we managed to extract important insights from the experimental data presented so far. We discovered that the polymerase does not actively separate DNA from histones, but rather exploits nucleosomal fluctuations to ratchet its way through the nucleosome. We have also shown how these fluctuations are modulated both by the tails and by direct contacts of the histone fold domains with DNA.

However, with the improved precision in determining the position of Pol II on the DNA template gained by aligning the stall site as well as the runoff position (described in detail in Section 2.8), we can now see non-uniformities in pausing even on bare DNA (Figure 3.10). We notice an increase in the number and durations of pauses at two positions on the template—around 3300 and right after 3500—relative to other locations on DNA. The first region is located before the NPS and, since that DNA never interacts with the histones, unsurprisingly, the number and duration of pauses are the same for the experiments performed with or without a nucleosome. The second region of increased pausing occurs right before Pol II reaches the nucleosome dyad. Interestingly, in the latter case, pausing at the nucleosome follows the same trends as pausing on bare DNA, but the magnitude of this sequence specific pausing is amplified by the nucleosome.

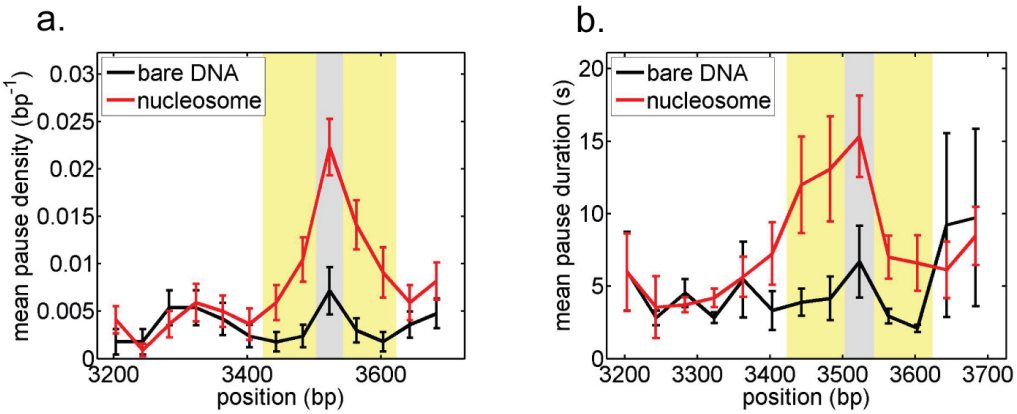

Figure 3.10: **Sequence-dependent pausing.** Pausing characteristics as a function of the position of the active center of the polymerase on the DNA template. **(a)** Pause density with a nucleosome (red line) and on bare DNA (solid black line). **(b)** Mean pause durations with (red) and without (black) a nucleosome present.

We note that the increase in pause number correlates with an increase in pause duration at a given transcriptional position (Figure 3.10). These data imply that at certain positions there is an increased tendency to pause, accompanied by a slow recovery from this paused state.

3.7 A Simple Model of Sequence-Dependent Pausing

What exactly about the DNA or RNA sequence affects the propensity of the polymerase to enter a pause more readily and also prevents recovery from a pause? Since transcriptional pausing is a knob for gene expression control, the subject has been intensely studied and debated over the last 30 years (see [99] for a review). It has been shown that the stability of the RNA-DNA hybrid and the structure of the nascent RNA are important factors in determining sequence-dependent pausing [100–102]. However, despite important progress in kinetic modeling of transcriptional pausing [103, 104], a consensus has not been reached on how to predict pauses given a DNA sequence to be transcribed.

We use our experimental data to add some insight to the pausing problem. As mentioned in Chapter 2, Pol II pauses can be very well described by a backtracking model. In this model, pauses start when the 3' end of the RNA becomes misaligned with the active center of Pol II, followed by a back-and-forth translation of the entire elongation bubble on DNA. A pause ends when Pol II realigns its active center with the 3' end of the DNA.

The initial fraying of the 3' end of the RNA could be promoted by a weak, AU-rich RNA-DNA hybrid. This model would predict an increased pause density on AT-rich sequences.

However, we don't see such a correlation between the AT content in our sequence (Figure 3.11, green dashed) and the measured pause density (Figure 3.10a). In fact, there seems to be a better (though not perfect) correlation between increased pausing and increased GC-content of the DNA sequence. One could argue that because GC-rich sequences are harder to open, they could slow down translocation, and since elongation is in kinetic competition with pause entry, GC-rich regions would thus increase the number of pauses. However, such a model would not explain why pauses also get longer in these regions.

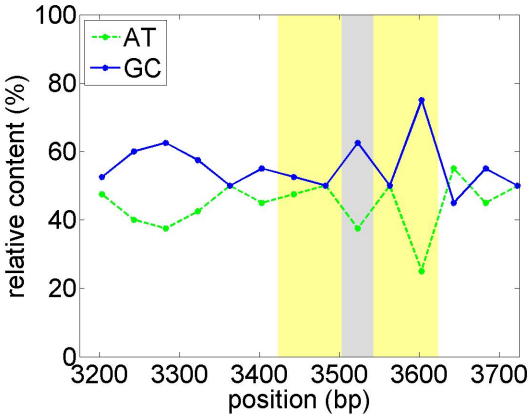

Figure 3.11: **DNA sequence composition.** The percentages of AT and GC base pairs are plotted along the DNA sequence that serves as a template in our experiment. Shading is the same as in Figure 3.10

We now turn to the other candidate for modifying pauses, the structure of the growing RNA chain as it exits behind the polymerase. If an RNA hairpin forms right outside of the RNA exit channel, it prevents the polymerase from backtracking (Figure 3.12). In order to take a step back on DNA, Pol II would have to break at least the first base pair in this structure. The existence of a hairpin would then decrease the number of pauses by preventing the first backtracking step, and at the same time slow the rate of moving back during a backtrack, by placing a barrier to these excursions. So the RNA secondary structure matches our data at least qualitatively. In this model, the increased number of pauses correlates to sequences with no immediate hairpins behind Pol II, while the regions with strong RNA hairpins show decreased pause numbers and durations.

While it is satisfying that we have a hypothesis for sequence-dependent pausing that matches the data qualitatively, we would like to reach a quantitative agreement between hairpin formation and pausing. To accomplish this task properly, we need to be able to predict the folded structure of the RNA behind the polymerase at each step during transcription. This type of computational problem [105] is not trivial for two reasons: because it involves folding of large RNA molecules, and because the RNA behind the polymerase

does not necessarily fold to its minimum energy structure, as depicted in Figure 3.12. At each step during transcription, only a portion of the total transcript exists behind the polymerase, and parts of it will have already folded up to this point in time. As the polymerase adds an additional base to the nascent RNA, this base could base pair with any of the free bases in the intermediate RNA transcript, but the addition of the new base is unlikely to lead to the disruption of already formed RNA structures. The structure of the RNA folded co-transcriptionally in this manner (Figure 3.12b) can look very different from the minimum energy structure obtained were the full RNA allowed to equilibrate starting from the completely unfolded state (Figure 3.12a).

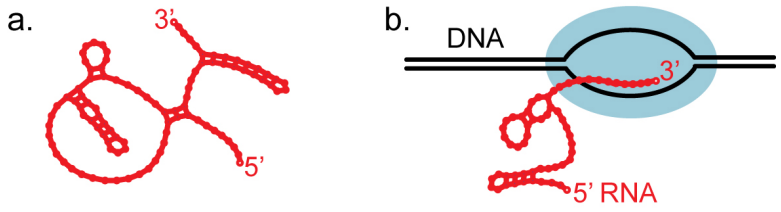

Figure 3.12: **Proposed connection between RNA structure and transcription. (a)** Minimum-energy RNA secondary structure. **(b)** The native RNA chain folds as the polymerase advances on DNA. The secondary structure formed co-transcriptionally differs from the minimum-energy one. The formation of local RNA hairpins behind Pol II prevent its backtracking.

Because of these computational difficulties, for now we turn to coarse graining the problem again (though to a different level this time). Remember that in Chapter 2 we characterized pausing on bare DNA using a single parameter, the diffusion rate of Pol II on the template while backtracked (k_0). We now describe the contribution of the sequence (most likely acting through RNA hairpins) as an average energy barrier to backtracking associated with each region of the template: ΔG_{seq}. In the absence of any external force or bias, and assuming a homogeneous DNA landscape with no hairpins, the forward and backward rates of Pol II moving on DNA during backtracking were taken to be equal:

$$k_f = k_0 \text{ and } k_b = k_0, \tag{3.2}$$

The presence of a hairpin behind Pol II only modifies the backward rate (k_b), in a way that reflects the height of the barrier to breaking the hairpin:

$$k_f = k_0 \text{ and } k_b = k_0 e^{-\Delta G_{seq}/k_B T}, \tag{3.3}$$

In our experiment, we are also applying a forward force on the polymerase, so the rates

of stepping during a backtracked pause become:

$$k_f = k_0 e^{F \cdot d/k_B T} \text{ and } k_b = k_0 e^{-F \cdot d/k_B T} e^{-\Delta G_{seq}/k_B T}, \tag{3.4}$$

Note that instead of introducing a barrier to backtracking through ΔG_{seq}, we could have naively tried to fit the pauses in different regions on the template with a modified k_0. However, this idea doesn't work even at a qualitative level. In the regions with longer pauses, the diffusion on DNA (k_0) has to be small to match the slow pause recovery. But since diffusion into a pause directly competes with elongation, a decrease in k_0 predicts a smaller pause density, in disagreement with the larger number of pauses observed experimentally in these regions.

We first use this sequence-dependent phenomenological model to fit the pause durations in different regions for ΔG_{seq}, as shown in Figure 3.13. We concentrate on the regions around the nucleosome positioning sequence, since our overarching goal is to understand how the nucleosome exploits sequence-dependent pausing to retard the polymerase. For each region (entry, central and exit into the NPS), we fit the pause durations for k_0 and ΔG_{seq}. All the ΔG_{seq} energies are calculated with respect to the sequence of highest pausing (i.e. the central region of the NPS, right before the dyad), which is set to zero by default. We use this approach because we reason that the region with most pausing must have the least amount of structure.

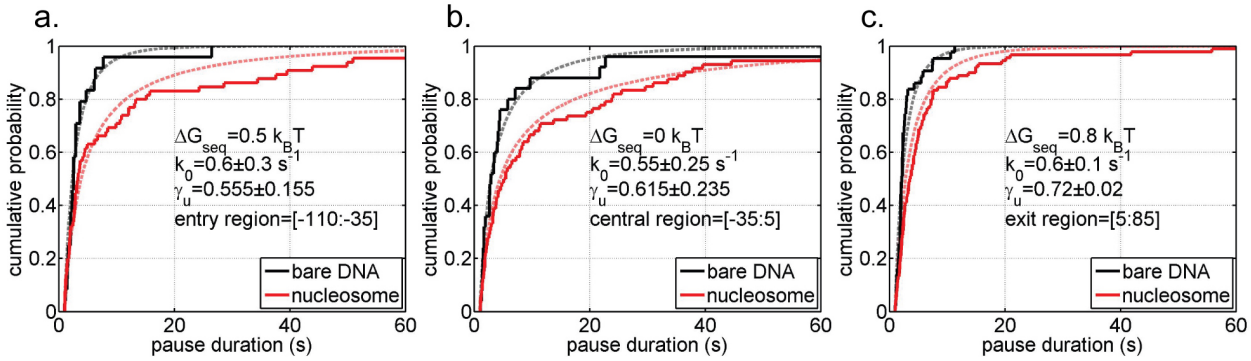

Figure 3.13: **Fits of sequence-dependent pause durations.** Cumulative distributions of pause durations in different regions of the NPS: **(a)** entry, **(b)** central, and **(c)** exit. Fits of the model presented in the text are shown as dashed lines.

It is rewarding that the fitting of the data obtained by this model predicts that, as may be expected, the intrinsic value of Pol II diffusion on DNA (k_0) is the same for all regions. The values of the average barrier that the RNA structure imposes on backtracking are very low—between 0.5 and 0.8 $k_B T$. These sequence-dependent barrier energies correspond to a decrease in the backtracking rate (k_b) by roughly a factor of 2.

For pausing at the nucleosome, we keep the k_0 and ΔG_{seq} determined for each region, and fit for the probability of finding the nucleosome locally unwrapped, γ_u. This value doesn't change much between the entry and central regions, indicating that the histone-DNA interactions are uniform along the NPS. The apparent increase in the unwrapping equilibrium that we see in the exit region might reflect the fact that in some cases, the histones are removed from DNA after Pol II passes the nucleosome dyad (see Chapter 4).

We test our model further by checking the predictions it makes—with the parameters we just determined—for the number of pauses in each region. As shown in Figure 3.14, the pause density predictions capture the correct trend and are within error of the experimental data. As expected, the model performs worst for the nucleosomal data in the exit region. As already mentioned, this discrepancy may be due to the fact that after Pol II passes the dyad the histones have a high probability of detaching from DNA, fact that we have not incorporated into our pausing model.

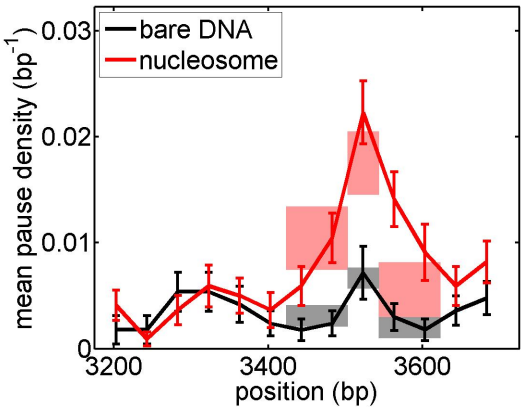

Figure 3.14: **Predictions of sequence model for pause numbers.** Experimental pause density along the DNA template (continuous lines), and predictions of the model described in the text (shaded regions).

While the phenomenological fits we have been performing so far to describe the effect of the sequence on pausing are a useful exercise, what we would really like to have is a direct way of quantitatively connecting the DNA sequence with the experimental data. To achieve this, we plan to perform more experiments that would prove beyond doubt that the effect of the sequence on pausing is mediated through the structure of the nascent RNA chain (see Chapter 5). At the same time, we are trying to develop a more detailed theoretical model to predict sequence-dependent pausing, as described in more detail in Chapter 5.

3.8 Biological Implications and Predictions

We have found that the histone tails mainly gate the access into the nucleosomal region. While removal or acetylation of the tails don't have a dramatic effect on the overall efficiency of transcription, the state of the tails could be an important factor in regulating access to chromatin remodeling factors in this region. For instance, histone chaperones or specific domains of chromatin remodeling factors could bind the tails and sequester them away from the nucleosome core particle, thus opening the gate into the nucleosome for ATP-dependent remodeling factors. Once bound to the nucleosome, these remodeling factors could perturb its structure further, and facilitate the action of RNA and DNA polymerases. Indeed, this might be an important part of the remodeling mechanism of NURF (nucleosome remodeling factor) and SWI/SNF complexes [106].

Our results with the Sin mutants proved that histone-DNA contacts around the dyad are extremely important for the stability of the nucleosome, and thus for the barrier it poses to transcription elongation. Disruption of as little as one contact in the dyad region can greatly weaken the nucleosomal barrier. We speculate that there must exist remodeling/transcription factors that bind to the nucleosome and disrupt one of these contact points in order to make the nucleosomal DNA easily accessible to Pol II and other DNA-translocating motors. The histone chaperone Asf1 (Anti-silencing function 1), which in addition to performing other functions in the cell also mediates chromatin disassembly during transcriptional elongation, is a good candidate for employing such a mechanism [107].

Lastly, our results on sequence-dependent pausing and the amplification of this effect by the nucleosome, while still preliminary, raise the question of about the degree of correlation between nucleosome positioning sequences and pausing sites. We plan to check wether the increased pausing that we see in the central region of the nucleosome is a general characteristic of nucleosome positioning sequences.

3.9 Appendix: Materials and Methods

3.9.1 Purification and Assembly of Modified Nucleosomes

Tailless, mock-acetylated, and Sin yeast histone proteins containing the deletions or substitutions indicated Tables 3.4, 3.5, and 3.6 respectively were cloned and expressed in *E. coli*. They were purified and assembled into octamers in the same manner as unmodified histones [68]. Nucleosomes were assembled onto the high-affinity 601 nucleosome positioning sequence [43] using salt dialysis reconstitution [69].

Table 3.4: **Tailless histones sequences.** Deleted sequences are shown in red.

H2A: MSGGKGGKAGSAAKASQSRSAKAGLTFPVGRVHRLLRRGNYAQRIGSGAP VYLTAVLEYLAAEILELAGNAARDNKKTRIIPRHLQLAIRNDDELNKLLG NVTIAQGGVLPNIHQNLLPKKSAKATKASQEL
H2B: MSAKAEKKPASKAPAEKKPAAKKTSTSTDGKKRSKARKETYSSYIYKVLK QTHPDTGISQKSMSILNSFVNDIFERIATEASKLAAYNKKSTISAREIQT AVRLILPGELAKHAVSEGTRAVTKYSSSTQA
H3: MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGT VALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQ ESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLARRLRGERS
H4: MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGGVKRI SGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYAL KRQGRTLYGFGG

Table 3.5: **Mock-acetylated histones sequences.** L→Q mutations are shown in red.

H2A: MSGGQGGQAGSAAKASQSRSAKAGLTFPVGRVHRLLRRGNYAQRIGSGAP VYLTAVLEYLAAEILELAGNAARDNKKTRIIPRHLQLAIRNDDELNKLLG NVTIAQGGVLPNIHQNLLPKKSAKATKASQEL
H2B: MSAKAEKKPASQAPAEQKPAAKKTSTSTDGKKRSKARKETYSSYIYKVLK QTHPDTGISQKSMSILNSFVNDIFERIATEASKLAAYNKKSTISAREIQT AVRLILPGELAKHAVSEGTRAVTKYSSSTQASSSTQA
H3: MARTKQTARQSTGGQAPRQQLASQAARKSAPSTGGVKKPHRYKPGT VALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQ ESVEAYLVSLFEDTNLAAIHAKRVTIQKKDIKLARRLRGERS
H4: MSGRGQGGQGLGQGGAQRHRQILRDNIQGITKPAIRRLARRGGVKRI SGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYAL KRQGRTLYGFGG

Table 3.6: **Sin mutant histones sequences.** Point mutations are shown in red.

H3 T118H: MARTKQTARKSTGGKAPRKQLASKAARKSAPSTGGVKKPHRYKPGT VALREIRRFQKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAIGALQ ESVEAYLVSLFEDTNLAAIHAKRVHIQKKDIKLARRLRGERS
H4 R45A: MSGRGKGGKGLGKGGAKRHRKILRDNIQGITKPAIRRLARRGGVKAI SGLIYEEVRAVLKSFLESVIRDSVTYTEHAKRKTVTSLDVVYAL KRQGRTLYGFGG

3.9.2 Analysis of Nucleosome Wrapping/Unwrapping Events

In order to determine the wrapping-unwrapping rates of the nucleosomal DNA, we loaded histone octamers on a 2964 bp DNA fragment containing the 601 nucleosome positioning sequence. This DNA fragment was obtained by PCR from a modified pUC19 plasmid [108] using DNA primers modified with biotin and digoxigenin respectively at their 5' ends. These modifications allowed us to tether the DNA containing the mononucleosomes between two beads that were anti-digoxidening and streptavidin-coated respectively. We recorded the nucleosome fluctuations at 2000 Hz using a dual-trap optical tweezers setup (Figure 3.7a), and found the wrapping and unwrapping transitions (Figure 3.7b) by running a t-test analysis on these data, as described previously [109, 110].

Chapter 4

The Fate of the Nucleosome During Transcription

In vivo, transcription elongation can disrupt chromatin [111], with outcomes ranging from partial loss to complete removal and exchange of histones [112–114]. These changes can affect subsequent rounds of transcription as well as other chromatin-related processes. Additionally, the fate of the nucleosome during *in vivo* transcription is complicated by the action of histone chaperones and chromatin remodeling factors.

It is therefore of interest to first understand the fate of the histones during eukaryotic transcription in a simpler, *in vitro* system, with the hope that we can build on this knowledge to understand the *in vivo* complexity of this process. However, because of the difficulties in performing transcription with eukaryotic polymerases (which normally require an entire host of transcription factors to initiate), initial studies inquiring into the fate of the nucleosome used the phage SP6 RNA polymerase. These experiments have shown that upon transcription by the phage polymerase, the histone octamer moves upstream by 40-95 base pairs [115, 116]. It was later proved that transcription by the eukaryotic RNA Polymerase III (Pol III)—the enzyme that transcribes the tRNAs—leads a similar outcome [117]. Surprisingly, more recent experiments suggested that transcription through a nucleosome by Pol II leads to H2A/H2B dimer loss [35, 118], and the formation of a hexamer whose position on DNA is unchanged [35]. Despite extensive work on characterizing the nucleosomal barrier [35, 37, 38, 115–121], there is still little mechanistic understanding of how transcription dynamics affects histone turnover, and little basis for rationalizing the differences in outcome among these polymerases.

Moreover, the histone transfer process for all these polymerases is believed to involve looping of the DNA template, but claims of template looping for Pol II have so far relied on indirect evidence [119].

Portions of this chapter were submitted for publication in 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**. Used with permission.

In this chapter, we follow the fate of the histone proteins during transcription. First, we use a modified version of the single-molecule experiment described in Chapter 2 to check for the presence of the histones behind Pol II after transcription. While informative, this experiment only provides a yes-no type of answer for the histone transfer question. For a more detailed, direct picture of the transcriptional intermediates and different outcomes of transcription, we turn to Atomic Force Microscopy (AFM). We developed methods that make it possible to image concurrently the polymerase and the nucleosome, and to obtain snapshots of individual Pol II-nucleosome complexes before, during and after transcription. This method allows us to answer questions regarding the outcome of transcription as well as the mechanism of histone transfer.

4.1 Histone Transfer is Force-Sensitive

The core histones in a nucleosome can remain associated with the DNA after transcription by being transferred upstream of the polymerase [53, 116, 118, 120]. A transient DNA loop [116, 117] might allow the histones of a partially unwrapped nucleosome to make contact with upstream DNA behind the polymerase. The probability of forming a DNA loop through thermal fluctuations and, therefore, that of histone transfer in this model, should be sensitive to forces as low as 0.2 pN [122]. We designed a construct that stops the polymerase in a mechanically stable conformation after it has passed the nucleosome (Figure 4.1a, see Section 4.8 for details). This strategy allowed us to determine the dependence of histone transfer on applied force.

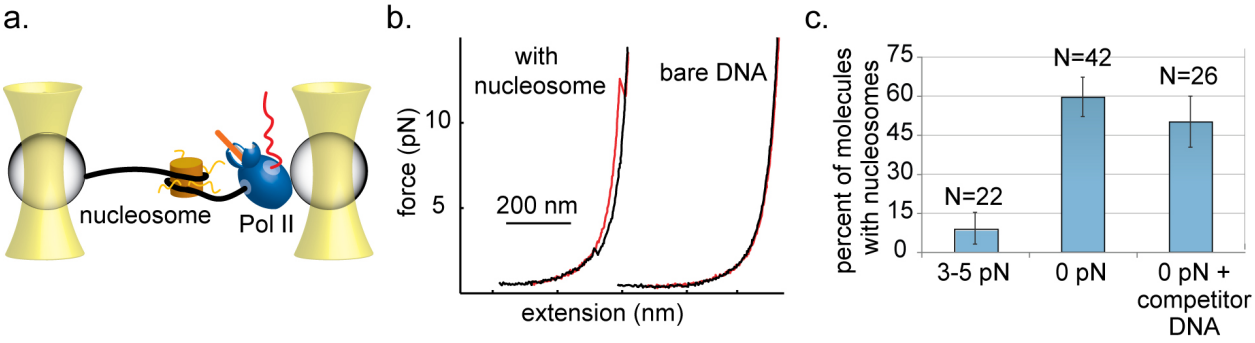

Figure 4.1: **Histone transfer is sensitive to force. (a)** Experimental setup that allows us to check for the presence of the nucleosome after transcription. The polymerase (blue) is stopped at the end of the template by a piece of crosslinked DNA (orange). The nucleosome, if transferred, should be on the upstream DNA between the two beads. **(b)** Force-extension curves of transcribed DNA. Pulling curves are shown in red and relaxation curves are in black. **(c)** Frequency of histone transfer as a function of applied force during transcription.

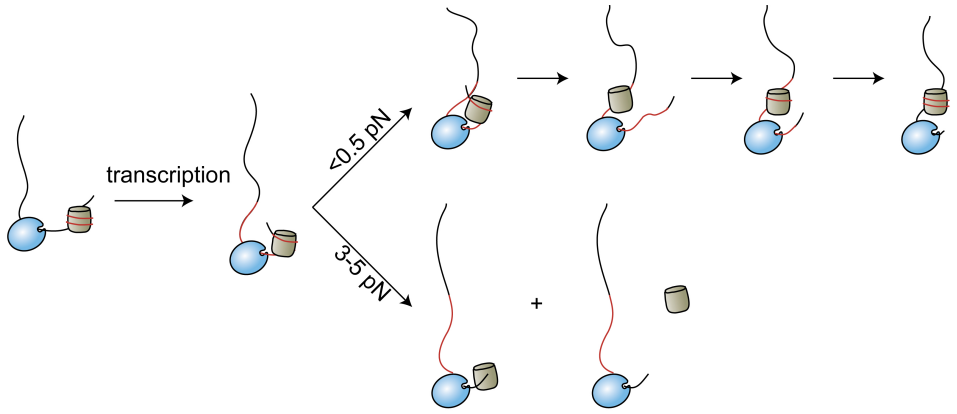

Figure 4.2: **Model for the force dependence of histone transfer.** Nucleosomal transcription begins as Pol II partially unwraps the nucleosome through a ratcheting action that locks local breathing of the nucleosome in the open conformation. At low forces (top pathway), DNA thermal fluctuations lead to a looped state in which the histones contact the DNA upstream of Pol II, allowing for histone transfer behind the polymerase. Histone transfer is followed by rewrapping of the nucleosome behind the polymerase. In contrast, at higher forces (bottom pathway), DNA fluctuations are highly constrained, so a loop is unlikely to form. In this scenario, histone transfer cannot happen; the histones must either slide along the DNA or detach in order for the polymerase to advance. The nucleosomal positioning sequence is shown in red.

We first monitored transcription of the nucleosomal region at forces between 3 and 5 pN, and then pulled on the transcribed DNA when Pol II reached the end of the template. Ramping the force between 1 and 20 pN after halting the polymerase allowed us to obtain force-extension curves of the upstream (transcribed) DNA template (Figure 4.1b). If histone transfer occurred during transcription, then a nucleosome will be positioned upstream of Pol II, and the force-extension curve will show the characteristic unwrapping of the nucleosome's inner wrap.

Very few molecules displayed nucleosome unwrapping transitions (2 out of 20) despite marked transcriptional pausing in the nucleosomal region; instead, most showed reversible monotonic force extension curves, indicating that no nucleosome was present behind the polymerase (Figure 4.1c). In contrast, when transcription proceeded in bulk before tether formation so that the DNA template was not under tension during transcription, we observed a significant increase in the fraction of complexes with nucleosomes upstream of the polymerase (25 out of 42, $p < 0.0025$, Fisher's exact test), indicating that histone transfer had occurred (Figure 4.1c).

Histone transfer was not significantly affected by an eight-fold excess (50 ng/L) of com-

petitor DNA (Figure 4b), a concentration sufficient to capture displaced histones [116]. We conclude that histones are transferred in *cis* to DNA upstream of Pol II upon nucleosomal transcription, as the looping model proposes, and that tension inhibits formation of the looped DNA intermediate necessary for transfer. This interpretation is consistent with the reduced pause density observed as the polymerase advances through the nucleosome (Chapter 2), because either the core histones detach from the DNA, are transferred to DNA upstream of the enzyme, or are pushed to a lower-affinity (i.e. higher γ_u) downstream sequence.

This force dependence supports a model in which histone transfer is mediated through DNA looping between the partially unwrapped nucleosome downstream of the polymerase and the upstream DNA template, as shown in Figure 4.2.

4.2 Direct Imaging of Nucleosomal Transcription

In order to test the looping model of histone transfer more thoroughly, we used direct imaging by AFM. This method allowed us to obtain snapshots of Pol II and the nucleosome during different stages of transcription. Pol II elongation complexes were assembled on a 96 bp DNA template [35, 53] and ligated to 574 bp of DNA containing a single nucleosome loaded on the 601 nucleosome positioning sequence (NPS) [43]. Complexes prepared in this manner were incubated either in the absence ("stalled sample," Figure 4.3a) or presence ("chased sample," Figure 4.3b) of nucleotide triphosphates (NTPs), then fixed with formaldehyde, deposited on mica, and imaged using AFM, as described in detail in Section 4.8.

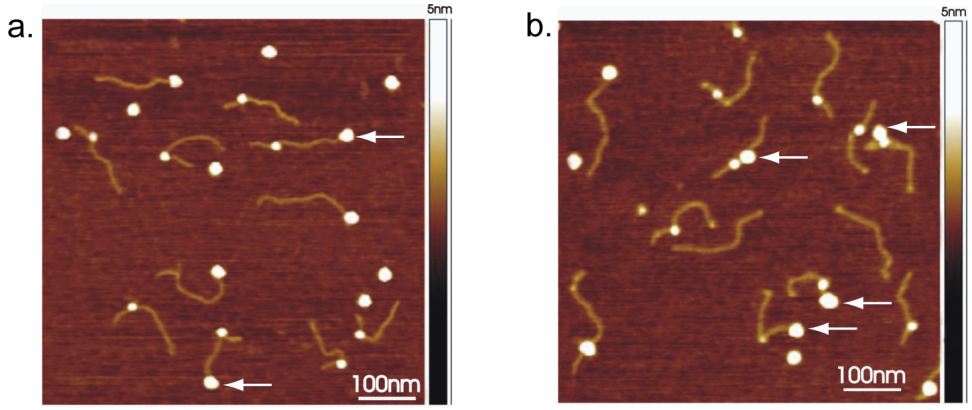

Figure 4.3: **Snapshots of transcription.** **(a)** stalled (no NTPs added) and **(b)** chased samples (all four NTPs added).

We measured the lengths of the different segments of free DNA (i.e. the DNA not covered by protein), as well as the heights of the proteins for complexes that have both the nucleosome

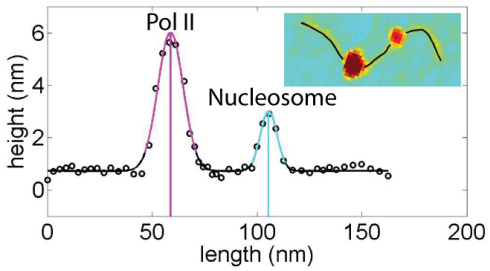

Figure 4.4: **Lengths and heights measurements.** The height profile of an example complex (inset) is plotted along the DNA path as black circles. The Pol II and nucleosome heights are fit to Gaussians shown in purple and blue respectively. The free DNA segment lengths (black part of the fit) are defined as the lengths of the paths that start two standard deviations away from the centers of the proteins.

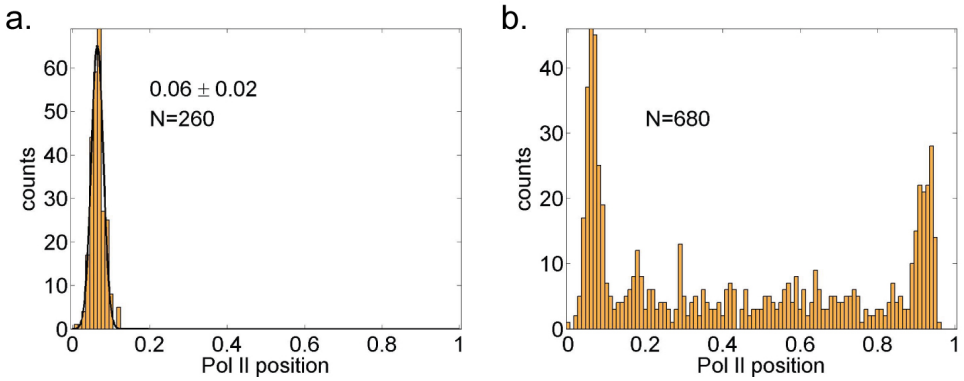

Figure 4.5: **Pol II activity observed by AFM.** The relative position of the polymerase on DNA is plotted for Pol II-nucleosome complexes in **(a)**the stalled samples or **(b)** the chased ones. Here, relative position is calculated as the length of the upstream arm (measured from the center of the polymerase to the end of DNA) divided by the total length of the DNA.

and the polymerase, as shown in Figure 4.4 (see Section 4.8 for details).

Because Pol II has a significantly larger molecular weight (∼550 kDa) than the nucleosome (∼190 kDa), it is possible to unambiguously distinguish the two complexes by their different sizes in the images (Figure 4.3 and Figure 4.4).

The position of the polymerase in stalled samples is centered at the start site of transcription (Figure 4.5a). In contrast, after addition of all four nucleotides, Pol II is distributed along the entire length of the DNA template (Figure 4.5b), indicating that transcription has ensued.

4.3 Changes in Position of the Histones

To determine which complexes in the chased sample have completed transcription through the nucleosome, we made use of the fact that the DNA upstream of the nucleosome, containing the start site of Pol II, is about three times longer than the DNA downstream of the nucleosome. When Pol II is on the long arm and does not contact the nucleosome, we reason that transcription has not yet proceeded into the nucleosomal region, and label these nucleosomes as "untranscribed". Conversely, when Pol II is on the short arm, we infer that transcription through the nucleosome was completed, and we deem these nucleosomes "transcribed" (Figure 4.6).

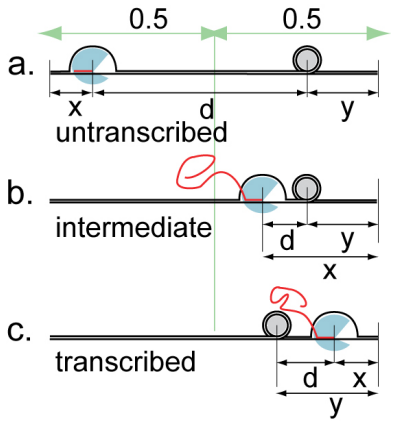

Figure 4.6: **Separating transcribed and untranscribed nucleosomes.** All distances (x, d, y) are measured as fractions of the total DNA length. x and y are the positions of Pol II and the nucleosome to the closest DNA end. d is the distance between the polymerase and the nucleosome. **(a)** If x+d+y=1, the nucleosome is identified as untranscribed. **(b)** If x=d+y, the complex is tagged as intermediary and is used for the free length calculations in Figure 2c if Pol II and the nucleosome are within 5 nm of each other. **(c)** If y=d+x, the nucleosome is identified as transcribed.

Note that in order to correctly identify transcribed nucleosomes, we assume that even if their positions change from the original NPS, they remain on the same half of DNA after transcription. This assumption is justified because for complexes identified as untranscribed in the chased sample, the position of the nucleosome is unchanged compared to untranscribed nucleosomes imaged in the absence of Pol II (Figure 4.7a, $p = 0.3$, t-test).

In order to get an accurate measurement of the changes in position of the nucleosomes (see Section 4.8 for details), we compared the length of the free DNA segment upstream of transcribed nucleosomes (Figure 4.7b, red) to that of untranscribed nucleosomes from a sample without Pol II (Figure 4.7b, blue). The distribution for transcribed nucleosomes is

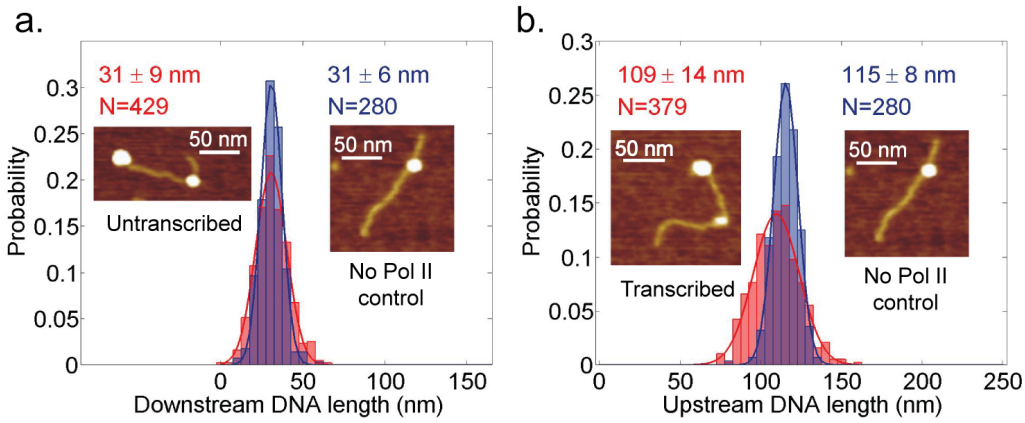

Figure 4.7: **Nucleosome position. (a)** The length of the downstream free DNA segments for untranscribed nucleosomes (red) and complexes without Pol II (blue). **(b)** The length of the upstream free DNA segment for transcribed nucleosomes (red) and complexes without Pol II (blue).

broader and there is a modest but statistically significant shift to shorter lengths (6 nm, $p < 0.001$, t-test). The broadening of the transcribed distribution and the partial overlap with the corresponding distribution from untranscribed nucleosomes suggests that while some of the nucleosomes are placed at the same location after transcription, others (approximately 20%) move immediately upstream of their original position (Figure 4.7b). The average shift of the population lying to the left of the original distribution is 24 nm (72 bp).

The shift in the positions of some of the transcribed nucleosomes that we observe might seem contradictory with previous results [35] that report no change in position. However, in those experiments, the position of the nucleosome was measured using a gel shift assay. With that method, the type of broadening in the distribution of nucleosome positions that we observe would appear as an increased thickness of a band in a gel; the thickness of the nucleosomal band is very sensitive to contrast in the image and is thus not usually reported in these types of assays. In addition, the previous experiments use the 5S nucleosome positioning sequence, which has two possible positions for the nucleosome; this makes any results about nucleosome position changes after transcription harder to disambiguate. In contrast, we use the 601 sequence, which uniquely positions the nucleosome. Overall, we report that the majority of the nucleosomes don't change position, in agreement with previous results. The small changes that we see in the position distribution represent a refinement of the previous result, and point to the advantage of using single-molecule assays.

This upstream relocation of the histones is consistent with previously proposed looping models of histone transfer [53, 116, 120].

4.4 Looping of the Template During Histone Transfer

In models of DNA looping during nucleosomal transcription, the histones from a partially unwrapped nucleosome situated downstream of the transcribing polymerase are assumed to simultaneously contact a DNA segment upstream of the polymerase forming a loop. According to such models, this process eventually leads to the transfer of histones behind the polymerase and permits transcription to resume. In agreement with this idea, we find many intermediate complexes where Pol II is in the process of transcribing the nucleosome that show the histones contacting the DNA segments both upstream and downstream of Pol II (Figure 4.8a).

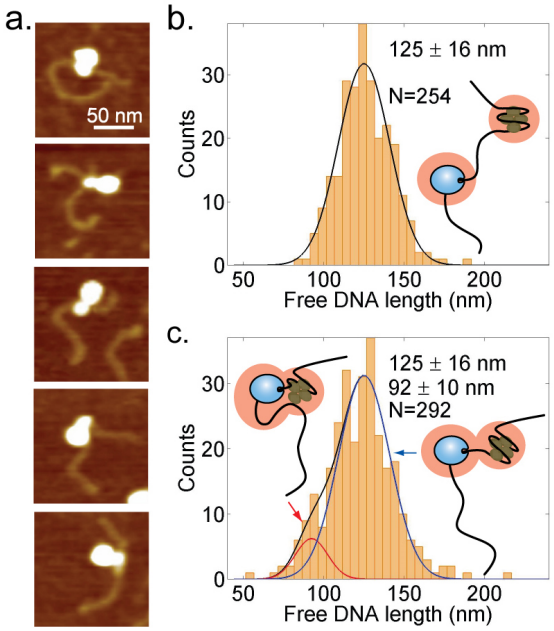

Figure 4.8: **DNA looping during histone transfer.** **(a)** AFM images of complexes where the histones contact both the upstream and downstream DNA. **(b)** Free DNA length in the presence of NTPs (all concentrations) for complexes where Pol II has not yet reached the nucleosome and **(c)** for complexes where Pol II is in the process of transcribing the nucleosome. Numbers indicate the means and standard deviations. Insets show the presumed structures of each population, with Pol II in blue, DNA in black, and histones in brown; the pink shading reflects broadening of the molecules due to the AFM tip.

The distribution of total free DNA lengths for intermediate complexes where Pol II is in the process of transcribing the nucleosome is different from that of complexes where Pol II has started transcribing but has not yet reached the nucleosome ($p = 0.009$, Kolmogorov-Smirnov test). Because the free-DNA lengths distribution for these intermediate complexes is

not well described by a single Gaussian ($p = 0.03$, Lilliefors test), we fit this distribution with two Gaussians (Figure 4.8c). The main peak is identical with the corresponding distribution for complexes where Pol II has started transcribing but has not yet reached the nucleosome (Figure 4.8b). The second peak corresponds to an additional population of intermediates where the DNA outside the polymerase-nucleosome complex is shorter by ∼30 nm. We interpret this shortening as evidence that the template in the proximity of the nucleosome participates in a loop that facilitates histone transfer behind the polymerase, and that cannot be resolved because of the broadening effect of the AFM tip (inset, Figure 4.8c). The estimated size of these DNA loops (∼90 bp) is smaller than the persistence length of DNA (∼150 bp) and they may be facilitated by the putative 90° bend that Pol II introduces in its DNA template [20, 21, 118].

4.5 Dimer Dissociation

As Pol II advances onto the nucleosomal template, the DNA is being detached from the core histones, exposing them to the surrounding conditions. Since the octamer consists of a collection of positively charged histone proteins, it is unstable at salt concentrations under 1 M [123, 124]. Thus, unless the core histones contact another piece of DNA that could neutralize their charges and stabilize their association, the octamer may dissociate with partial loss of its components. Indeed, loss of an H2A/H2B dimer and the formation of a hexasome upon transcription by Pol II has been reported previously [35, 118]. Consistent with these results, we observe a reduction in the apparent physical size of transcribed nucleosomes (Figure 4.9).

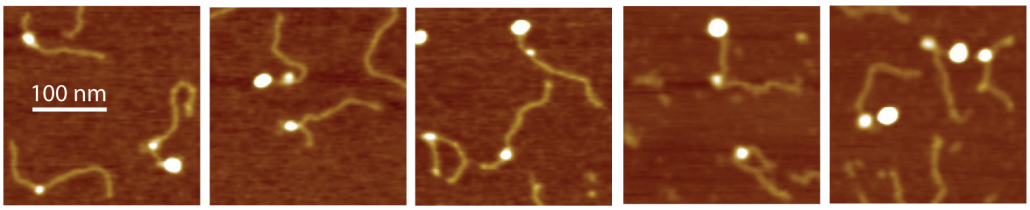

Figure 4.9: **Particles with reduced heights.** Images illustrating changes in the size of the nucleosomes upon transcription. The nucleosomes with reduced height are shown next to regular size nucleosomes, for comparison.

The height of the transcribed nucleosomes (Figure 4.10b) consists of two populations: one similar to untranscribed nucleosomes (3±0.4 nm, Figure 4.10a) and the other one corresponding to subnucleosomal particles with lower height (2.1±0.3 nm). In order to check the identity of the transcribed particles with decreased height, we reconstituted and imaged

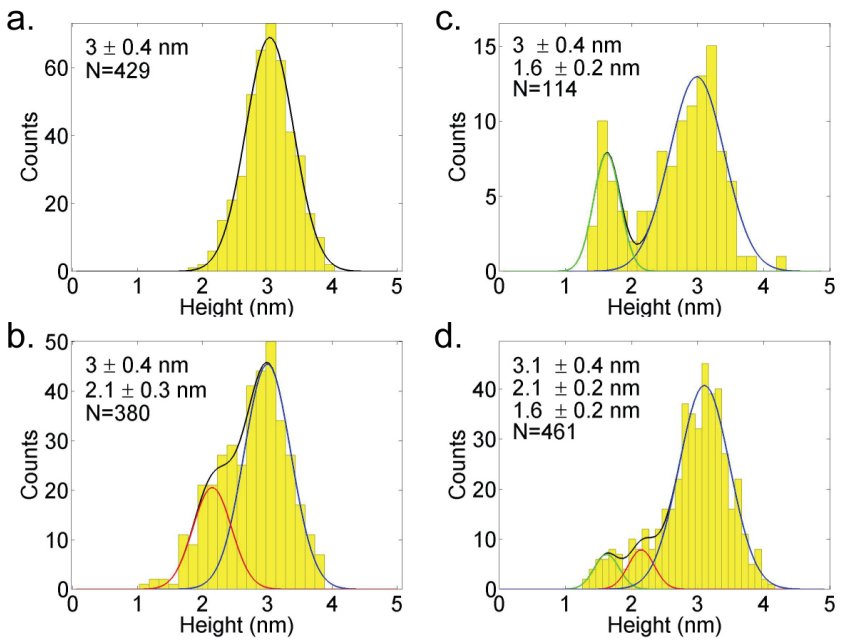

Figure 4.10: **Transcription leads to hexamer formation.** Histograms of nucleosome heights for **(a)** untranscribed nucleosomes and **(b)** transcribed nucleosomes at all NTPs concentrations. **(c)** Heights of tetramers compared with those of octamers. **(d)** Height of nucleosomes destabilized by incubation in high salt (1 M KCl). We identify the three peaks as tetramers, hexamers and octamers.

tetramers, using the same methods as for octamers, except we omitted H2A and H2B addition. As shown in Figure 4.10c, the height of tetramers, 1.6±0.2 nm, is significantly lower than that of the transcribed particle, 2.1±0.3 nm. Moreover, when we destabilize the nucleosomes by incubating them in 1 M KCl, we obtain 3 nucleosomal species, consistent with octamers, hexamers and tetramers (Figure 4.10d). The height of the middle peak, which we identify as a hexamer, matches the height of the subnucleosomal particles resulting from transcription.

Most significantly, we find that the fraction of smaller subnucleosomal particles observed after transcription depends on the rate of elongation. When transcription was performed at low NTP concentrations (100 μM), only 10±3% of the transcribed nucleosomes were converted to hexasomes (Figure 4.11d). Increasing the NTPs concentration to 200 μM NTP augments the percentage of hexasomes to 17±3% (Figure 4.11e). At saturating NTPs (1000 μM), 25±5% of the transcribed nucleosomes were converted to hexasomes (Figure 4.11f). No changes were observed in the sizes of untranscribed nucleosomes in these samples (Figure 4.11a-c).

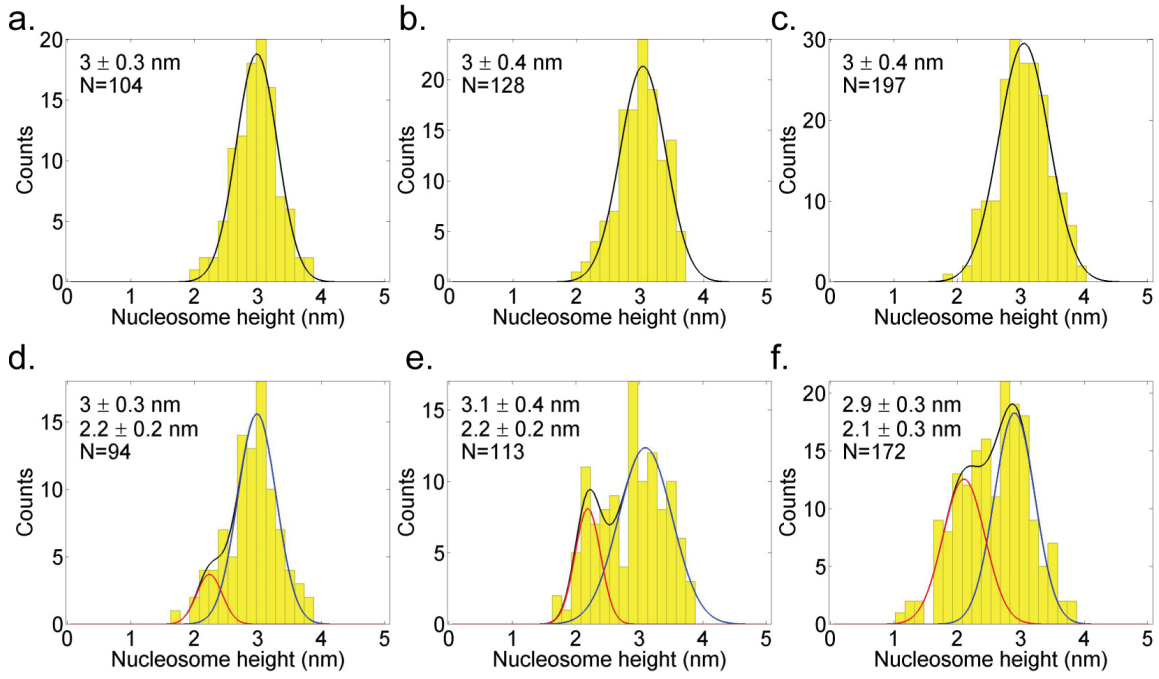

Figure 4.11: **Histone transfer outcome depends on the speed of transcription.** Nucleosome height for molecules where Pol II has not passed the nucleosome in **(a)** 100 μM NTPs, **(b)** 200 μM NTPs, and **(c)** 1000 μM NTPs. Height of transcribed nucleosomes at **(d)** 100 μM NTPs, **(e)** 200 μM NTPs, and **(f)** 1000 μM NTPs.

4.6 Kinetic Model of Histone Transfer

These findings indicate that a kinetic competition exists between the net rate of nucleosome unwrapping during elongation (k_{ue}), octamer transfer (k_t), and dimer dissociation (k_d) during transcription through the nucleosome (Figure 4a). Initially, as the nucleosome partially unwraps during Pol II advancement, enough of the histone core is exposed to allow contact with the upstream DNA through a temporary DNA loop, but not so much as to cause H2A/H2B dissociation. During slow transcription (100 μM NTPs), this partially exposed histone intermediate lasts long enough to allow transfer of the intact octamer onto the upstream DNA. However, if the rate of transcription is increased slightly, more of the nucleosome will unwrap and, as enough of the histone core becomes exposed, dimer dissociation starts competing with octamer transfer to the upstream DNA. Under these conditions, representative for transcription at 200 μM and 1000 μM NTPs, both octamers and hexamers can be found as a result of transcription. Finally, when the rates of transcription are even higher, enough DNA is unwrapped from the surface of the histone core that the complete histone detachment from DNA greatly outcompetes the rates of histone transfer

and histone-histone dissociation, thus leading to bare DNA formation.

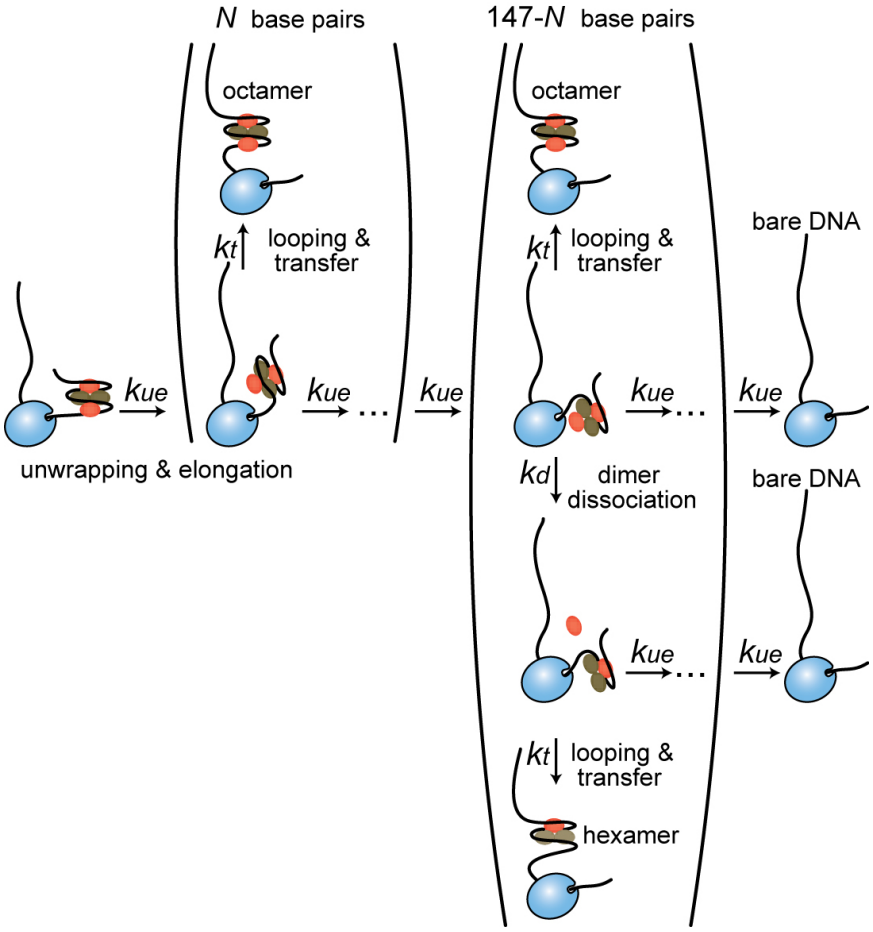

Figure 4.12: **Histone transfer model.** Kinetic scheme of transcription and histone transfer. Pol II in blue, H2A/H2B dimers in red, H3/H4 dimers in brown, DNA in black.

From the kinetic scheme shown in Figure 4.12, we can compute the probabilities of octamer, hexamer, and bare DNA formation upon transcription as a function of three competing rates: the net rate of nucleosome unwrapping during elongation (k_{ue}), octamer transfer (k_t), and dimer dissociation (k_d). The competition between these rates occurs at every step during transcription, so we also need to include in the model the number of base pairs that need to be unwrapped to allow octamer transfer, but no dimer dissociation (N). During the unwrapping of these first N base pairs of the nucleosome, the histone core is stable and H2A/H2B dissociation doesn't take place until further unwrapping occurs. Since the nucleosome wraps 147 base pairs [16], the model assumes that for the remaining $M = 147 - N$ base pairs, the histone core can do one of three things: transfer as an octamer, undergo

dimer dissociation or further peel away from the DNA in an elongation-dependent manner.

We can compute the probabilities of each outcome – octamer, hexamer, or bare DNA– by calculating the probability of each possible path in the kinetic scheme, and then summing over these paths. For instance, in order to obtain a hexamer, we have to first unwrap and elongate through the first N base pairs of the nucleosome. The probability of choosing further unwrapping and elongation versus octamer transfer in this first part of the kinetic scheme is $[k_{ue}/(k_{ue}+k_t)]^N$. Once the dimer can dissociate, we move in the second part of the kinetic scheme, where both transfer and dissociation compete with continued DNA unwrapping from the surface of the histones followed by elongation. In this part of the scheme, Pol II can transcribe a variable number of base pairs m, with probability $[k_{ue}/(k_t+k_{ue}+k_d)]^m$, before the dimer dissociates with probability $k_d/(k_t+k_{ue}+k_d)$. Afterwards Pol II can transcribe another variable number of base pairs i, with probability $[k_{ue}/(k_{ue}+k_t)]^i$, before histone transfer leads to hexamer formation behind the polymerase with probability $k_t/(k_{ue}+k_t)$. Constructed in this manner, the probabilities of obtaining hexamer (P_{hex}), octamer (P_{oct}), or bare DNA (P_{bare}) are:

$$P_{hex} = \sum_{m=0}^{M-1} \sum_{i=0}^{M-1-m} \left(\frac{k_{ue}}{k_{ue}+k_t}\right)^N \left(\frac{k_{ue}}{k_t+k_{ue}+k_d}\right)^m \left(\frac{k_d}{k_t+k_{ue}+k_d}\right) \left(\frac{k_{ue}}{k_t+k_{ue}}\right)^i \left(\frac{k_t}{k_{ue}+k_t}\right)$$

$$P_{oct} = \sum_{n=0}^{N-1} \left(\frac{k_{ue}}{k_{ue}+k_t}\right)^n \left(\frac{k_t}{k_t+k_{ue}}\right) + \sum_{m=0}^{M-1} \left(\frac{k_{ue}}{k_{ue}+k_t}\right)^N \left(\frac{k_{ue}}{k_t+k_{ue}+k_d}\right)^m \left(\frac{k_t}{k_t+k_{ue}+k_d}\right)$$

$$P_{bare} = 1 - P_{oct} - P_{hex}$$

Note that the first sum in the expression for octamer formation (P_{oct}) comes from transcribing the first N base pairs of the nucleosome (the first part of the kinetic scheme), while the second sum represents transcription through the rest of the nucleosome (second part of the kinetic scheme).

If we close the sums, we can rewrite:

$$P_{hex} = \left(\frac{k_{ue}}{k_t+k_{ue}}\right)^N \left(\frac{k_d}{k_t+k_d} + \frac{k_t}{k_t+k_d} \cdot \left(\frac{k_{ue}}{k_t+k_{ue}+k_d}\right)^M - \left(\frac{k_{ue}}{k_t+k_{ue}}\right)^M\right) \tag{4.1}$$

$$P_{oct} = 1 - \left(\frac{k_{ue}}{k_t+k_{ue}}\right)^N \left(\frac{k_d}{k_t+k_d} + \frac{k_t}{k_t+k_d} \cdot \left(\frac{k_{ue}}{k_t+k_{ue}+k_d}\right)^M\right) \tag{4.2}$$

$$P_{bare} = \left(\frac{k_{ue}}{k_{ue} + k_t}\right)^{N+M} \tag{4.3}$$

In our case, since the nucleosome contacts 147 base pairs of DNA, $M = 147 - N$, the probabilities of observing a hexamer (P_{hex}), an octamer (P_{oct}), or bare DNA (P_{bare}) after transcription can be re-written as:

$$P_{hex} = \left(\frac{k_{ue}}{k_t + k_{ue}}\right)^{N} \left(\frac{k_d}{k_t + k_d} + \frac{k_t}{k_t + k_d} \cdot \left(\frac{k_{ue}}{k_t + k_{ue} + k_d}\right)^{147-N}\right) - \left(\frac{k_{ue}}{k_t + k_{ue}}\right)^{147} \tag{4.4}$$

$$P_{oct} = 1 - \left(\frac{k_{ue}}{k_t + k_{ue}}\right)^{N} \left(\frac{k_d}{k_t + k_d} + \frac{k_t}{k_t + k_d} \cdot \left(\frac{k_{ue}}{k_t + k_{ue} + k_d}\right)^{147-N}\right) \tag{4.5}$$

$$P_{bare} = \left(\frac{k_{ue}}{k_{ue} + k_t}\right)^{147} \tag{4.6}$$

These expressions show a direct competition between the average rate of elongation at each base pair and histone transfer on one hand, and between dimer dissociation and histone transfer on the other. As elongation becomes faster, and thus more disruptive, the probability of complete histone removal and the resulting production of bare DNA (P_{bare}) should increase monotonically, while the production of transferred octamers (P_{oct}) should instead decrease monotonically. Significantly, this model predicts that the probability of hexamer formation should increase from low and moderate Pol II elongation rates because as the rate of elongation-dependent octamer unwrapping increases, the probability of histone dissociation effectively competes with that of octamer transfer, enhancing the production of hexamers as observed in this study (first term in P_{hex} dominates). However, as the rate of elongation and nucleosome unwrapping increases further, the rate of histone dissociation is outcompeted by the rate of complete histone removal and the production of hexamers should attain a maximum and eventually decrease (the last term in P_{hex} dominates).

Note that for small k_{ue} (as is the case for Pol II), we can ignore terms of order $N + M$, so the probability of transcription resulting in bare DNA is negligible, and the probabilities of octamer and hexamer transfer simplify to:

$$P_{hex} \cong \left(\frac{k_{ue}}{k_t + k_{ue}}\right)^{N} \frac{k_d}{k_t + k_d} \text{ and } P_{oct} \cong 1 - \left(\frac{k_{ue}}{k_t + k_{ue}}\right)^{N} \frac{k_d}{k_t + k_d}$$

4.7 Predictions and Implications of the Kinetic Model

To test this model, we sought to determine the rates involved in this process. The net rate of nucleosome unwrapping during elongation is equal to the average overall velocity of transcription through the nucleosome, which at saturating NTPs (1000 μM) is $k_{ue} = 1$ bp/s [53]. Since the average Km of Pol II for NTP hydrolysis is 100 μM (Section 4.9), we can use Michaelis-Menten kinetics to estimate the net rates of transcription through the nucleosome at 200 μM and 100 μM NTPs; this analysis yields about, 0.7 bp/s and 0.5 bp/s for these two rates, respectively. Finally, to determine the rate of H2A/H2B dimer loss for pre-assembled octamers directly exposed to the salt concentration used in these studies (300 mM KCl), we performed an ensemble FRET-based assay with fluorescently labeled H2B and H4 [124]. These experiments gave $k_d = 0.027 \pm 0.001$ s^{-1} (Section 4.9).

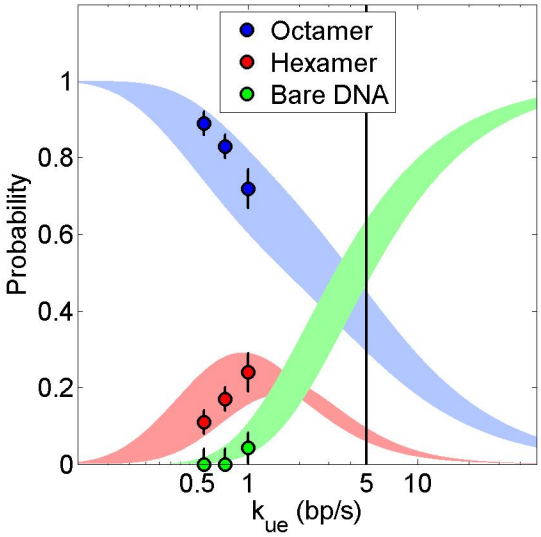

Figure 4.13: **Predictions of the histone transfer model.** Hexamer (red) and octamer (blue) transfer probabilities, as well as bare DNA formation (green) are plotted as a function of the net nucleosome unwrapping rate during elongation. Experimental data are shown as circles; the shaded areas represent the model predictions, with the width reflecting uncertainties in k_t and N, as indicated in the text. The black line marks the elongation rate for the faster polymerase Pol III.

Fit of the experimental data (Figure 4.13) shows that this simple competition model captures correctly the details of hexamer and octamer transfer probabilities, as well as that of complete histone removal, if the initial DNA unwrapped region allowing only octamer transfer is $N = 40 \pm 5$ bpand the rate of histone transfer is $k_t = 0.02 \pm 0.005$ s^{-1}. Notice

that the value of N obtained here is consistent with the amount of DNA contacted by the H2A/H2B dimer (∼30 bp), and that the rate of transfer is quite comparable to that of dimer dissociation. For histone transfer, the rate-limiting process is most likely the actual hand-off of the histones from the downstream to the upstream DNA, as looping resulting from DNA bending fluctuations is known to be much faster [125].

This competition model also explains why faster polymerases produce a mix of octamers and bare DNA but yield little or no hexamers upon transcription. For instance, the majority of Pol III complexes complete transcription through a nucleosome in approximately 30 seconds ($k_{ue} = 5$ bp/s [117], vertical black line, Figure 4.13), so we predict that octamer transfer is likely – approximately 40%. To obtain hexasomes, on the other hand, two slow processes have to occur before Pol III can finish transcription: dimer dissociation and histone transfer, making the probability of hexamer transfer very unlikely, approximately 10%, under these fast transcription conditions. The rest of the molecules (about 50%) are predicted to lead to complete removal of the histones, in agreement with the published results [117]. Moreover, transcription by the even faster SP6 polymerase also leads to the formation of octamers and bare DNA, without hexamer formation, with the percentage of bare DNA increasing as the speed of elongation is increased [116].

Gene regulation *in vivo* may result from the modification of any one of the competing rates involved in elongation on a nucleosomal template. In particular, elongation factors that increase the net transcription rate of Pol II through the nucleosome would result in an increased probability of complete histone removal from DNA. Alternatively, dimer dissociation from the partially unwrapped octamer could be faster for certain histone variants of H2A [126], or in the presence of histone chaperones that bind the dimer, increasing the probability of hexasome formation. Such transcription-induced alterations in chromatin structure may affect gene expression by reducing/eliminating nucleosomal barriers for future transcription elongation events [118, 119] or by altering the accessibility of transcription factor binding sites [127]. The findings communicated here might also be relevant to other processes that involve advancement of molecular motors through DNA wrapped in nucleosomes, such as processive DNA replication and chromatin remodeling [128].

4.8 Appendix: Materials and Methods

4.8.1 Mechanically Stable Arrest of Pol II for Pulling Curves

We developed a way to halt the polymerase immediately after observing it transcribe nucleosomal DNA by ligating an inter-strand linked DNA at the end of the DNA template. Inter-strand linked DNA was obtained by incubating 10 nM 200-bp HindIII-digested DNA with 0.1 mg/mL trioxsalen at 25°C under UV exposure at 247 nm for 40 minutes. This DNA was purified with a HiSpeed Plasmid Mini Kit (Qiagen), concentrated with SpeedVac (Fisher/Thermo Scientific) and ligated to the DNA template 98 bp downstream of the NPS

(at position 3709 bp) in 10-fold excess before preparation of elongation complexes.

For comparison of nucleosome unwrapping counts after transcription, we applied Fisher's exact test (the two-tailed version). This nonparametric statistical test also requires no knowledge of the underlying probability distributions, and is suitable for comparisons between categorical data sets of small sample sizes.

4.8.2 Sample Preparation for AFM

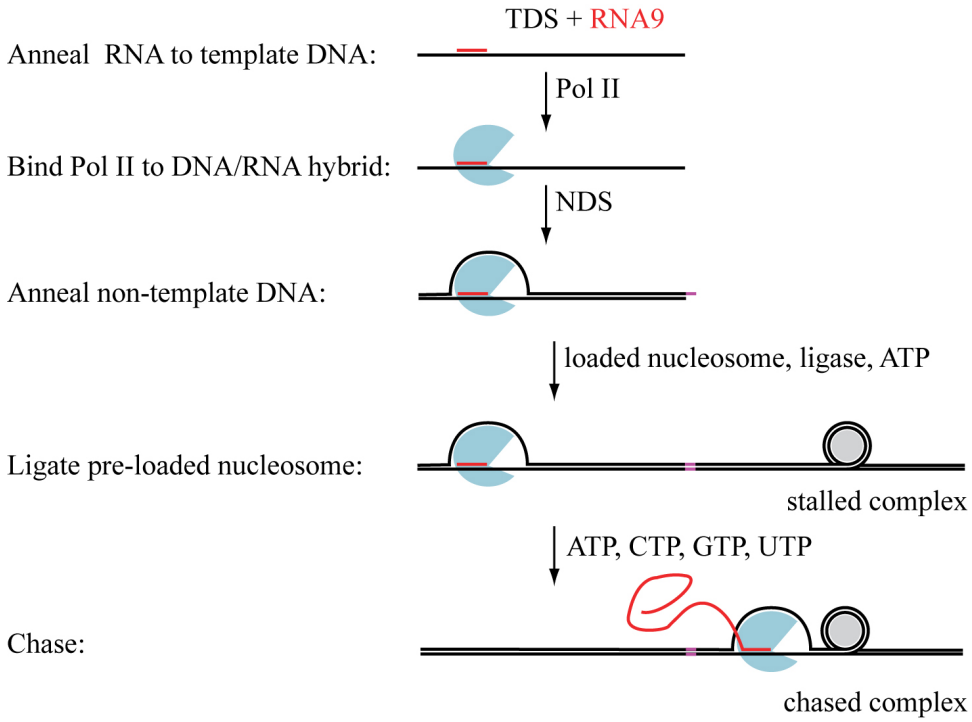

Figure 4.14: **Assembly of Pol II elongation complexes with nucleosome.**

His-tagged Pol II (*S. cerevisiae*) [67] was assembled on a 96 bp DNA fragment as previously described [53], and the resulting elongation complexes were ligated to downstream DNA containing a nucleosome as shown in Figure 4.14. The 574 bp DNA template was prepared by PCR from a modified pUC19 plasmid containing the 601 nucleosome positioning sequence [108]. Octamers were reconstituted from recombinant yeast histones [68] and loaded onto the template DNA using salt dialysis [69]. The assembly and ligation were performed in TB40 (20 mM HEPES pH 7.8, 40 mM KCl, 10 mM MgCl2, 10 μM ZnCl2, 1 mM β-mercaptoethanol). Chasing was performed in TB300 (similar to TB40, except with 300 mM KCl instead of 40 mM KCl), supplemented with 1000 μM NTPs (unless otherwise specified) and 1 μM pyrophosphate, at 25°C for 30 minutes. Following transcription, samples

were fixed with 1% (w/v) formaldehyde for 2 hours at room temperature, and then dialyzed into TB40.

In the sample referred to as untranscribed nucleosomes, the 96 bp double stranded DNA was ligated in excess to the 574 bp nucleosomal DNA in the absence of Pol II.

4.8.3 Image Acquisition and Processing

For deposition, the samples were diluted in TB40 to 2 nM DNA concentrations, placed on freshly cleaved ultra-clean mica (Grade V, Ted Pella, Inc.), and incubated at room temperature for ∼2 minutes. The mica discs were then rinsed with purified 18.2 MΩ deionized water and dried using a gentle N_2 gas flow, perpendicular to the mica surface. AFM measurements were performed with a Multimode AFM Nanoscope V (Veeco Instruments Inc., Santa Barbara, CA, USA) equipped with a type E-scanner (10-micron x 10-micron x 2.5-micron vertical range). The samples were imaged in tapping mode using a commercial silicon cantilevers (Nanosensors), with a high-resonance frequency in the range of 260-410 kHz and a spring constant of 46 N/m. Images (512 x 512 pixels) were captured in the trace direction, at a scan size of 1.5 μM x 1.5 μM, with a scan rate of 1.5 Hz. The imaging amplitude (amplitude set-point) of the cantilever was maintained by the feedback circuitry to 80-85% of the free oscillation amplitude and the scan angle was maintained at 0. All samples were measured at room temperature in air, at a relative humidity of 30%.

Image processing and data analysis were performed using Matlab (The MathWorks, Natick, MA), with software based on ALEX [129, 130]. We imported the images into Matlab, automatically masked all the points higher than 9 nm, and flattened the images by subtracting from each line a polynomial of degree 2 fit to that line. We then identified all the objects higher than 0.2 nm in this flattened image, used these points as a mask, and performed a new line by line flattening on the original with a polynomial of degree 4.

For each complex, the DNA path (passing through the proteins) was digitized and fit by a polynomial of degree 3 [129]. The polymerase and nucleosome center positions were recorded as the centers of the highest and second highest Gaussians respectively along the identified DNA path (Figure 4.3c). The nucleosome heights were measured as the maximum height in a 4-nm box centered at the position of the nucleosome, to correct for cases when the DNA path does not pass through the center of the nucleosome. To account for small height variation among different depositions, we corrected the height of transcribed nucleosomes using the height of untranscribed nucleosomes as a standard. In the chased samples, we first identified the molecules that have Pol II but where Pol II has not yet crossed the nucleosome, as described in Figure 4.6. We fit the heights of these untranscribed nucleosomes with a Gaussian function for each sample (different tip and deposition), and shifted all these distributions such that the Gaussian peak of each one was at 3 nm (which is the height we get when imaging nucleosomes alone with high frequency tips). Finally, we shift all the other nucleosome heights in that sample by exactly the same amount as the untranscribed nucleosomes. In general, this correction shift is between 0.2-0.8 nm for each sample, with

the higher shifts for samples imaged with low frequency tips.

We use the length of free DNA (DNA not covered by protein) to estimate the position of the nucleosome on the template. Before Pol II passes the nucleosome, it covers part the upstream arm of the nucleosome; in contrast, after transcription, the upstream arm of the nucleosome is completely unobscured. Therefore, in order to obtain an accurate measurement of the length of the upstream arm of the nucleosome before transcription, we imaged a sample that lacks the polymerase, but has a full length template containing the nucleosome.

Percentages of hexamers were calculated as the fraction of transcribed particles with heights under 2.4 nm. Amounts of bare DNA are estimated as the percentage of molecules without nucleosomes in the chased samples minus the corresponding percentage in the stalled samples.

4.9 Appendix: Measurements of the Kinetic Rates

4.9.1 Elongation Rate Dependence on the NTPs Concentration

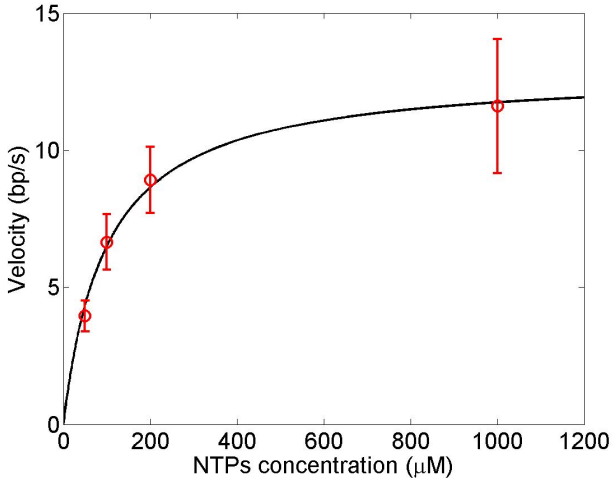

Figure 4.15: **Elongation rate as a function of NTPs concentration.** Overall velocity of Pol II on bare DNA is measured as a function of NTP concentration (red). The theoretical fit to the Michaelis-Menten equation is shown in black.

We measure the dependence of transcription velocities on the NTPs concentration from single-molecule trajectories obtained using optical tweezers as described in Chapter 2. These experiments are performed in TB300, supplemented with $1\mu M$ PPi and 50, 100, 200 or 1000 μM NTPs. We fit the experimental data (Figure 4.15) with the Michaelis-Menten equation:

$$v = v_{\max}\frac{[NTP]}{[NTP]+K_m}, \text{ and determine } K_m = 99 \pm 50\ \mu\text{M}.$$

Even though the dependence of transcription velocity on the NTPs concentration is measured on bare DNA, we expect the K_m for NTPs hydrolysis to remain the same for transcription thorough the nucleosome. The average overall velocity of transcription through the nucleosome is equal to the net rate of nucleosome unwrapping during elongation (k_{ue}). This rate depends on the time spent in active elongation ($t_{\text{elongation}}$) and the time spent pausing (t_{pausing}) over a stretch of DNA that contains L base pairs:

$$k_{ue} = \frac{L}{t_{\text{elongation}} + t_{\text{pausing}}}.$$

When we increase the NTPs concentration, thus increasing the nucleotide addition rate (and implicitly the active elongation rate, $k_e = \frac{L}{t_{\text{elongation}}}$) by a factor of f, the elongation time and the pausing time both decrease by the same factor f, as follows:

$$k_e^{new} = f \cdot k_e \Rightarrow t_{\text{elongation}}^{new} = t_{\text{elongation}}/f$$

While this change is obvious for the elongation time, which depends directly on the nucleotide addition rate, the proof requires slightly more work for the pausing time. The average pausing time is equal to the number of pauses (N_{pause}) multiplied by the average pause duration (t_{pause}):

$$t_{\text{pausing}} = N_{\text{pause}} \cdot t_{\text{pause}}$$

The average pause duration doesn't depend on the NTP concentration, since pause recovery is independent of nucleotide addition:

$$t_{\text{pause}}^{new} = t_{\text{pause}}$$

However, the number of pauses changes with NTPs concentration, since pause entry (k_p) competes directly with active elongation (k_e):

$$N_{\text{pause}} \propto \frac{k_p}{k_e + k_p} \cong \frac{k_p}{k_e}, \text{ since the pausing rate is much smaller than the elongation rate.}$$

This means that when we increase the nucleotide addition rate by a factor f, we decrease the number of pauses by f:

$$k_e^{new} = f \cdot k_e \Rightarrow N_{\text{pause}}^{new} = N_{\text{pause}}/f$$

Therefore, the total time spent pausing changes by the same factor as the nucleotide addition rate when we modify the NTPs concentration:

$$t^{new}_{\text{pausing}} = N^{new}_{\text{pause}} \cdot t^{new}_{\text{pause}} = (N_{\text{pause}}/f) \cdot t_{\text{pause}} = t_{\text{pausing}}/f$$

We can now compute the new unwrapping and elongation rate in terms of the old one:

$$k^{new}_{ue} = \frac{L}{t^{new}_{\text{elongation}} + t^{new}_{\text{pausing}}} = \frac{L}{t_{\text{elongation}}/f + t_{\text{pausing}}/f} = f \cdot k_{ue}.$$

In summary, if we change the rate of elongation on bare DNA by a factor f by modifying the NTP concentration, the overall rate of transcription through the nucleosome changes by the same factor:

$$k^{new}_{e} = f \cdot k_e \Rightarrow k^{new}_{ue} = f \cdot k_{ue}.$$

This result allows us to use the K_m determined on bare DNA to estimate the elongation rates through the nucleosome at different NTPs concentrations based on the rate of elongation through the nucleosome at saturating NTPs.

4.9.2 Measurement of H2A/H2B Dimer Dissociation Rate

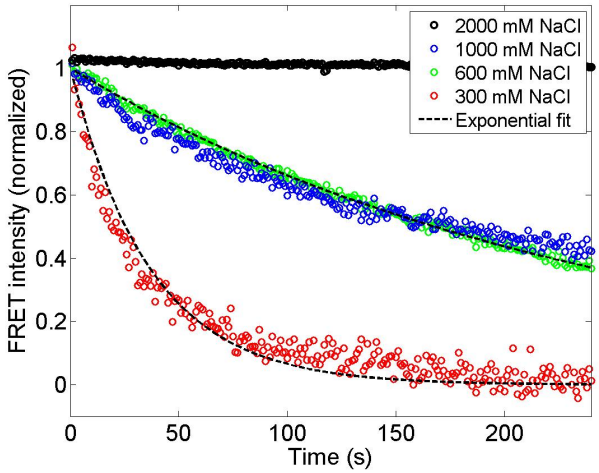

Figure 4.16: **Kinetics of dimer dissociation.** FRET-based assay for observing the kinetics of octamer dissociation in the absence of a complete DNA wrap.

In order to determine the rate of histone octamer dissociation in the absence of complete DNA wrapping, we used previously described techniques to FRET-label recombinant

yeast histones [124]. Briefly, we cloned, expressed, and purified two recombinant yeast mutant histones: H4 T71C and H2B T112C, then separately labeled H4 T71C histones using fluorescein-5-maleimide (F5M), and H2B T112C histones with 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM). By convention, residue numbering is given relative to the homologous residues in the *Xenopus* histones. FRET-labeled histone octamers were prepared from both F5M- and CPM-labeled mutant histones using salt dialysis reconstitution.

FRET-labeled histone octamers were incubated in a buffer containing 20 mM tris-HCl, pH 7.4, 5 mM 2-mercaptoethanol, and variable salt concentrations (2 M NaCl, 1 M NaCl, 600 mM NaCl, or 300 mM NaCl). Measurement of the dissociation rate was made by observing the increase in normalized donor fluorescence, with 385-nm excitation (2-nm slit width), and donor emission monitored at 460 nm (2-nm slit width). The concentration of histone octamers was 0.15 μM, and we estimate our mixing time to be under 5 s.

Increase in donor fluorescence (i.e. decay of FRET) was not observed at 2 M NaCl, indicating that the histone octamers were completely stable at 2 M NaCl without a complete DNA wrap. At 1 M NaCl and 600 mM NaCl, we observed an exponential decay consistent with a rate of 0.0041 ± 0.0001 s^{-1}. In contrast, a much faster decay of 0.027 ± 0.001 s^{-1} was observed at 300 mM NaCl, indicating that 2-3% of the histone octamers dissociate every second in the absence of DNA.

Chapter 5

Conclusions and Future Directions

5.1 Sequence-Dependent Pausing

The work on the sequence dependence of pausing presented in Chapter 3 is still a project in progress. We are currently performing single-molecule transcription experiments on AT-rich and GC-rich templates in order to tease out the effects of the sequence composition on transcription.

We also intend to confirm experimentally the importance of the RNA structure on the dynamics of transcription (especially through the nucleosome) by adding ribonucleases (RNases) to our single-molecule transcription assay. It has already been proven that digestion of the nascent RNA chain prevents backtracking [37]. If our hypothesis that most of the sequence-dependent pausing we observe is mediated by the structure of the RNA is correct, we should see a disappearance of the peaks in pausing on bare DNA. Pausing should become uniform along the sequence in this case. We also predict that pausing at the nucleosome, while increased compared to bare DNA, should be much more uniform, lacking the strong peak right before the dyad.

At the same time we are developing a theoretical model that builds on previous attempts to predict the regions of increased transcriptional pausing given a DNA sequence [103–105].

5.2 Transcription Elongation Factors

In vivo, transient pauses and arrests at nucleosomal DNA are suppressed by elongation factors, which increase the net rate of transcription. Several eukaryotic elongation factors have been identified, among them TFIIS, TFIIF, ELL, Elongin, and FACT [29, 30, 131, 132]. Some, like TFIIS, interact directly with the polymerase to rescue pauses [39, 45]. Others, like FACT, are chromatin-specific elongation factors that act upon nucleosomes to facilitate transcription [133, 134].

It would be interesting to investigate the effect of these factors on the dynamics of nucle-

osomal transcription at the single-molecule level. These experiments would allow for measurement of individual pause times, transcription velocity, pause density and backtracking extent. Such data could lead to the development of mechanochemical models that describe the mechanisms by which these factors impact the dynamics of nucleosomal transcription.

5.3 Chromatin Remodeling Factors

The strong barrier that nucleosomes pose for Pol II allows for rich regulation of transcription through chromatin *in vivo*, where histone chaperones, chromatin assembly factors, and ATP-dependent chromatin remodeling factors promote elongation [30]. Some chromatin related factors, such as FACT, use their histone chaperone activity to facilitate transcription through chromatin by removing one H2A/H2B dimer [134, 135].

On the other hand, the mechanism of ATP-dependent chromatin remodeling factors during transcription is more complex and less understood. Chromatin remodelers, such as SWI/SNF and RSC, are involved in transcriptional regulation *in vivo* [136], and mutations of SWI/SNF are synthetically lethal with TFIIS mutations, suggesting a complementary role for chromatin remodelers and elongation factors [137]. It has also recently been shown that SWI/SNF and RSC greatly increase the rate of transcription through mononucleosomes *in vitro* [138]. These chromatin remodelers translocate along DNA and move nucleosomes by a combination of intranucleosomal loop formation and nucleosome sliding [108, 139], making nucleosomal DNA more accessible. In addition, they can also remove one H2A/H2B dimer [128, 140], lowering the barrier to transcription.

However, regulation of transcription through chromatin might involve more than interactions of remodelers with nucleosomes. For instance, RSC increases the probability of nucleosomal passage to a greater extent than SWI/SNF even though the two are homologous proteins [138]. This difference may arise because RSC interacts directly with the Rpb5 subunit of Pol II [141].

The single-molecule methods presented in the preceding chapters could serve as tools for studying the effect of various remodeling factors on the dynamics of transcription through the nucleosome.

5.4 Concluding Thoughts

As a physicists approaching biological problems, I work by the principle that Lord Kelvin stated so eloquently more than 100 years ago [142]: "When you can measure what you are speaking about and express it in numbers you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the stage of science, whatever the matter might be."

As scientists in general strive to attain a better quantitative description of complex biological phenomena, it is very satisfying to discover apparently intricate and perplexing processes that can be rationalized and even made predictable by simple underlying physical principles. Each of the main chapters of this work provides one such story.

We find that nucleosome dynamics have a significant effect on transcription by RNA polymerase, and that, in turn the elongation rate of the polymerase affects the fate of the nucleosome. Our observations and analyses add a new stone to the foundation of gene expression understanding.

References

[1] C. R. Darwin. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. John Murray, 1859. ISBN 1592242863.

[2] J.G. Mendel. Versuche über pflanzenhybriden verhandlungen des naturforschenden vereines in brünn. *Abhandlungens*, 47:3, 1865.

[3] C.T Mendel, J.G. (Druery and William Bateson). Experiments in plant hybridization. *Journal of the Royal Horticultural Society*, 26:1, 1901.

[4] E. Mayr. *The growth of biological thought : diversity, evolution, and inheritance*. Belknap Press, Cambridge, Mass., 1982.

[5] G.E. Allen. *Life science in the twentieth century*. Cambridge University Press New York, 1978. ISBN 0521218640.

[6] J.D. Watson and F.H. Crick. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737738, 1953.

[7] F.H. Crick. On protein synthesis. In *Symposia of the Society for Experimental Biology*, volume 12, page 138, 1958.

[8] F. Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

[9] R. Dawkins. *The selfish gene*. Oxford University Press, Oxford; New York, 30th anniversary edition, 2006.

[10] S.J. Lolle, J.L. Victor, J.M. Young, and R.E. Pruitt. Genome-wide non-mendelian inheritance of extra-genomic information in arabidopsis. *Nature*, 434(7032):505–509, 2005. ISSN 0028-0836.

[11] M. Rassoulzadegan, V. Grandjean, P. Gounon, S. Vincent, I. Gillot, and F. Cuzin. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature*, 441(7092):469–474, 2006. ISSN 0028-0836.

[12] J.A. Arai, S. Li, D.M. Hartley, and L.A. Feig. Transgenerational rescue of a genetic defect in long-term potentiation and memory formation by juvenile enrichment. *Journal of Neuroscience*, 29(5):1496, 2009.

[13] E. Jablonka and M.J. Lamb. *Epigenetic Inheritance and Evolution: the Lamarckian Dimension*. John Wiley & Sons, Inc., Oxford University Press., 1995.

[14] K. Robertson. Epigenetic mechanisms of gene regulation. *DNA Methylation and Cancer Therapy*, pages 13–30, 2005.

[15] R.D. Kornberg. Chromatin structure: a repeating unit of histones and DNA. *Science (New York, NY)*, 184(139):868, 1974.

[16] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. Crystal structure of the nucleosome core particle at 2.8 $\mathring{A}$ resolution. *Nature*, 389:251–260, 1997.

[17] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell*. Garland Science, New York, 4th edition, 2002.

[18] B.D. Strahl and C.D. Allis. The language of covalent histone modifications. *Nature*, 403(6765):41–45, 2000. ISSN 0028-0836.

[19] J. Hurwitz. The discovery of RNA polymerase. *Journal of Biological Chemistry*, 280 (52):42477, 2005. ISSN 0021-9258.

[20] P. Cramer, D. A. Bushnell, and R. D. Kornberg. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, 292(5523):1863–1876, 2001.

[21] 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 $\mathring{A}$ resolution. *Science*, 292(5523):1876–1882, 2001.

[22] M.C. Thomas and C.M. Chiang. The general transcription machinery and general cofactors. *Critical reviews in biochemistry and molecular biology*, 41(3):105–178, 2006. ISSN 1040-9238.

[23] E. Nudler. RNA polymerase active center: the molecular engine of transcription. *Annual review of biochemistry*, 78:335, 2009.

[24] R.J. Sims, R. Belotserkovskaya, and D. Reinberg. Elongation by RNA polymerase II: the short and long of it. *Genes & development*, 18(20):2437, 2004. ISSN 0890-9369.

[25] R. D. Kornberg and Y. Lorch. Irresistible force meets immovable object: transcription and the nucleosome. *Cell*, 67(5):833–6, 1991.

[26] R.D. Kornberg. Eukaryotic transcriptional control. *Trends in Genetics*, 15(12):M46–M49, 1999. ISSN 0168-9525.

[27] R.J. Sims et al. Recent highlights of RNA-polymerase-II-mediated transcription. *Current opinion in cell biology*, 16(3):263–271, 2004. ISSN 0955-0674.

[28] R. D. Kornberg. The molecular basis of eukaryotic transcription. *Proc Natl Acad Sci USA*, 104(32):12955–12961, 2007.

[29] B. Li, M. Carey, and J.L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–719, 2007. ISSN 0092-8674.

[30] O.I. Kulaeva, D.A. Gaykalova, and V.M. Studitsky. Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 618(1-2):116–129, 2007. ISSN 0027-5107.

[31] V. G. Allfrey and A. E. Mirsky. Evidence for the complete DNA-dependence of RNA synthesis in isolated thymus nuclei. *Proc. Natl. Acad. Sci.*, 48(9):1590–1596, 1962.

[32] G. C. Yuan, Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando. Genome-scale identification of nucleosome positions in S. cerevisiae. *Science*, 309(5734):626–630, 2005.

[33] T. N. Mavrich, C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters, S. J. Zanton, L. P. Tomsho, J. Qi, R. L. Glaser, S. C. Schuster, D.S. Gilmour, I. Albert, and B. F. Pugh. Nucleosome organization in the Drosophila genome. *Nature*, 2008. 0028-0836 10.1038/nature06929 10.1038/nature06929.

[34] M. G. Izban and D. S. Luse. Transcription on nucleosomal templates by RNA polymerase II in vitro: inhibition of elongation with enhancement of sequence-specific pausing. *Genes & Development*, 5(4):683, 1991.

[35] 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. *Molecular Cell*, 9(3):541–552, 2002.

[36] W. Walter, M.L. Kireeva, V.M. Studitsky, and M. Kashlev. Bacterial polymerase and yeast polymerase II use similar mechanisms for transcription through nucleosomes. *Journal of Biological Chemistry*, 278(38):36148, 2003. ISSN 0021-9258.

[37] M. L. Kireeva, B. Hancock, G. H. Cremena, W. Walter, V. M. Studitsky, and M. Kashlev. Nature of the nucleosomal barrier to RNA polymerase II. *Molecular Cell*, 18: 97–108, 2005.

[38] V. A. Bondarenko, L. M. Steele, A. Újvári, 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. *Molecular Cell*, 24(3):469–479, 2006.

[39] R. N. Fish and C. M. Kane. Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta: Gene Struct. Exp.*, 1577(2):287–307, 2002.

[40] G. Li, M. Levitus, C. Bustamante, and J. Widom. Rapid spontaneous accessibility of nucleosomal DNA. *Nature structural & molecular biology*, 12(1):46–53, 2004. ISSN 1545-9993.

[41] I. Sidorenkov, N. Komissarova, and M. Kashlev. Crucial role of the RNA: DNA hybrid in the processivity of transcription. *Molecular Cell*, 2(1):55–64, 1998. ISSN 1097-2765.

[42] M.L. Kireeva, N. Komissarova, D.S. Waugh, and M. Kashlev. The 8-nucleotide-long RNA: DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *Journal of Biological Chemistry*, 275(9):6530, 2000. ISSN 0021-9258.

[43] P. T. Lowary and J. Widom. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *Journal of molecular biology*, 276(1):19–42, 1998.

[44] J.W. Shaevitz, E.A. Abbondanzieri, R. Landick, and S.M. Block. Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature*, 426 (6967):684–687, 2003. ISSN 0028-0836.

[45] 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(7137):820–823, 2007.

[46] C. Bustamante, J. F. Marko, E. D. Siggia, and S. Smith. Entropic elasticity of lambda-phage DNA. *Science*, 265(5178):1599–1600, 1994.

[47] S. Mihardja, A. J. Spakowitz, Y. Zhang, and C. Bustamante. Effect of force on mononucleosomal dynamics. *Proceedings of the National Academy of Sciences*, 103(43):15871, 2006.

[48] I. Samkurashvili and D.S. Luse. Translocation and transcriptional arrest during transcript elongation by RNA polymerase II. *Journal of Biological Chemistry*, 271(38): 23495, 1996. ISSN 0021-9258.

[49] D. Angelov, J. M. Vitolo, V. Mutskov, S. Dimitrov, and J. J. Hayes. Preferential interaction of the core histone tail domains with linker DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 98(12):6599, 2001.

[50] N. Komissarova and M. Kashlev. Transcriptional arrest: Escherichia coli RNA polymerase translocates backward, leaving the 3' end of the RNA intact andextruded. *Proceedings of the National Academy of Sciences*, 94(5):1755–1760, 1997.

[51] 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(24):15329–15338, 1997.

[52] M. Depken, E.A. Galburt, and S.W. Grill. The origin of short transcriptional pauses. *Biophysical journal*, 96(6):2189–2193, 2009.

[53] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science*, 325 (5940):626, 2009.

[54] P.H. von Hippel and E. Delagoutte. A general model for nucleic acid helicases and their coupling within macromolecular machines. *Cell*, 104(2):177, 2001. ISSN 0092-8674.

[55] 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(2):183–193, 2005. ISSN 0092-8674.

[56] G. Nogues, S. Kadener, P. Cramer, D. Bentley, and A. R. Kornblihtt. Transcriptional activators differ in their abilities to control alternative splicing. *Journal of Biological Chemistry*, 277(45):43110–43114, 2002.

[57] A.R. Kornblihtt, M. DE LA, J.P. Fededa, M.J. Munoz, and G. Nogues. Multiple links between transcription and splicing. *Rna*, 10(10):1489, 2004. ISSN 1355-8382.

[58] M. Yonaha and N.J. Proudfoot. Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Molecular cell*, 3(5):593–600, 1999. ISSN 1097-2765.

[59] N.J. Park, D.C. Tsao, and H.G. Martinson. The two steps of poly (A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Molecular and cellular biology*, 24(10):4092, 2004. ISSN 0270-7306.

[60] Ann E. Rougvie and John T. Lis. The RNA polymerase II molecule at the 5' end of the uninduced hsp70 gene of D. melanogaster is transcriptionally engaged. *Cell*, 54: 795–804, 1988.

[61] A. Krumm, T. Meulia, M. Brunvand, and M. Groudine. The block to transcriptional elongation within the human c-myc gene is determined in the promoter-proximal region. *Genes Dev.*, 6(11):2201–2213, 1992.

[62] L. J. Strobl and D. Eick. Hold back of RNA polymerase II at the transcription start site mediates down-regulation of c-myc in vivo. *EMBO J*, 11(9):3307–3314, 1992.

[63] A. Plet, D. Eick, and J. M. Blanchard. Elongation and premature termination of transcripts initiated from c-fos and c-myc promoters show dissimilar patterns. *Oncogene*, 10(2):319–28, 1995.

[64] D. L. Bentley and M. Groudine. A block to elongation is largely responsible for decreased transcription of c-myc in differentiated HL 60 cells. *Nature*, 321(6071):702–706, 1986.

[65] S. A. Brown, A. N. Imbalzano, and R. E. Kingston. Activator-dependent regulation of transcriptional pausing on nucleosomal templates. *Genes Dev.*, 10(12):1479–1490, 1996.

[66] L.J. Core and J.T. Lis. Transcription regulation through promoter-proximal pausing of RNA polymerase II. *Science*, 319(5871):1791, 2008.

[67] 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. *Methods in enzymology*, 370:138–155, 2003.

[68] J. Wittmeyer, A. Saha, and B. Cairns. DNA translocation and nucleosome remodeling assays by the RSC chromatin remodeling complex. *Methods in enzymology*, 377:322–343, 2004.

[69] A. Thastrom, P. T. Lowary, and J. Widom. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods*, 33(1):33–44, 2004.

[70] J.R. Moffitt, Y.R. Chemla, D. Izhaky, and C. Bustamante. Differential detection of dual traps improves the spatial resolution of optical tweezers. *Proceedings of the National Academy of Sciences*, 103(24):9006, 2006.

[71] P. R. Selvin and Ha T. *Single-Molecule Techniques: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, 2008.

[72] R.V. Dalal, M.H. Larson, K.C. Neuman, J. Gelles, R. Landick, and S.M. Block. Pulling on the nascent RNA during transcription does not alter kinetics of elongation or ubiquitous pausing. *Molecular cell*, 23(2):231–239, 2006. ISSN 1097-2765.

[73] K.C. Neuman, E.A. Abbondanzieri, R. Landick, J. Gelles, and S.M. Block. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell*, 115(4): 437–447, 2003. ISSN 0092-8674.

[74] Wikipedia. Catalan number. `http://en.wikipedia.org/wiki/Catalan_number`, retrieved October 12, 2010.

[75] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, Inc., New York, 2 edition, 1971.

[76] J. L. Workman and R. E. Kingston. Alteration of nucleosome structure as a mechanism of transcriptional regulation. *Ann. Rev. Biochem.*, 67(1):545–579, 1998.

[77] V. G. Allfrey, R. Faulkner, and A. E. Mirsky. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci.*, 51 (5):786–794, 1964.

[78] T. R. Hebbes, A. W. Thorne, and C. Crane-Robinson. A direct link between core histone acetylation and transcriptionally active chromatin. *The EMBO Journal*, 7(5): 1395, 1988.

[79] Y. Zhang and D. Reinberg. Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.*, 15(18): 2343–2360, 2001.

[80] S. I. S. Grewal and D. Moazed. Heterochromatin and epigenetic control of gene expression. *Science*, 301(5634):798–802, 2003.

[81] J. C. Rice and C. D. Allis. Histone methylation versus histone acetylation: new insights into epigenetic regulation. *Curr. Opin. Cell Biol.*, 13(3):263–273, 2001.

[82] J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, and S. I. S. Grewal. Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science*, 292(5514):110–113, 2001.

[83] V. M. Weake and J. L. Workman. Histone ubiquitination: Triggering gene activity. *Mol. Cell*, 29(6):653–663, 2008.

[84] P. Cheung, C. D. Allis, and P. Sassone-Corsi. Signaling to chromatin through histone modifications. *Cell*, 103(2):263–271, 2000.

[85] K. Robzyk, J. Recht, and M. A. Osley. Rad6-dependent ubiquitination of histone H2B in yeast. *Science*, 287(5452):501–504, 2000.

[86] S. Rea, F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun, M. Schmid, S. Opravil, K. Mechtler, C. P. Ponting, and C. D. Allis. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature*, 406:593–599, 2000.

[87] T. Jenuwein and C. D. Allis. Translating the histone code. *Science*, 293(5532):1074–1080, 2001.

[88] X. de la Cruz, S. Lois, S. Sanchez-Molina, and M. A. Martinez-Balbas. Do protein motifs read the histone code? *BioEssays*, 27(2):164–175, 2005.

[89] M. J. Bottomley. Structures of protein domains that create or recognize histone modifications. *EMBO Reports*, 5(5):464–9, 2004.

[90] 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. Biol.*, 20(23): 8866–8878, 2000.

[91] 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. *Journal of Biological Chemistry*, 283(47):32236, 2008.

[92] 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. *The EMBO Journal*, 13(8):1823, 1994.

[93] R. T. Simpson. Structure of chromatin containing extensively acetylated H3 and H4. *Cell*, 13(4):691–699, 1978.

[94] B. Brower-Toland, D. A. Wacker, 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(1):135–146, 2005.

[95] 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 and development*, 9(22):2770, 1995.

[96] 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 proteinDNA interactions. *The EMBO journal*, 23(2):260–271, 2004.

[97] F. K. Hsieh, M. Fisher, A. Újvári, V. M. Studitsky, and D. S. Luse. Histone Sin mutations promote nucleosome traversal and histone displacement by RNA polymerase II. *EMBO reports*, 2010.

[98] 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 Research*, 2010. ISSN 0305-1048.

[99] R. Landick. The regulatory roles and mechanism of transcriptional pausing. *Biochemical Society Transactions*, 34:1062–1066, 2006.

[100] R.G. Keene, A. Mueller, R. Landick, and L. London. Transcriptional pause, arrest and termination sites for RNA polymerase II in mammalian N-and c-myc genes. *Nucleic acids research*, 27(15):3173, 1999. ISSN 0305-1048.

[101] M. Palangat and R. Landick. Roles of RNA: DNA hybrid stability, RNA structure, and active site conformation in pausing by human RNA polymerase II1. *Journal of Molecular Biology*, 311(2):265–282, 2001. ISSN 0022-2836.

[102] P.J. Hawryluk, A. Újvári, and D.S. Luse. Characterization of a novel RNA polymerase II arrest site which lacks a weak 3' RNA–DNA hybrid. *Nucleic acids research*, 32(6): 1904, 2004. ISSN 0305-1048.

[103] V.R. Tadigotla, D.Ó. Maoiléidigh, A.M. Sengupta, V. Epshtein, R.H. Ebright, E. Nudler, and A.E. Ruckenstein. Thermodynamic and kinetic modeling of transcriptional pausing. *Proceedings of the National Academy of Sciences of the United States of America*, 103(12):4439, 2006.

[104] L. Bai, A. Shundrovsky, and M.D. Wang. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *Journal of molecular biology*, 344(2):335–349, 2004. ISSN 0022-2836.

[105] M. Geis, C. Flamm, M.T. Wolfinger, A. Tanzer, I.L. Hofacker, M. Middendorf, C. Mandl, P.F. Stadler, and C. Thurner. Folding kinetics of large RNAs. *Journal of molecular biology*, 379(1):160–173, 2008. ISSN 0022-2836.

[106] M. Vignali, A.H. Hassan, K.E. Neely, and J.L. Workman. ATP-dependent chromatin-remodeling complexes. *Molecular and cellular biology*, 20(6):1899, 2000. ISSN 0270-7306.

[107] 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. ISSN 0092-8674.

[108] 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. *Molecular cell*, 24(4):559–568, 2006.

[109] NJ Carter and RA Cross. Mechanics of the kinesin step. *Nature*, 435(7040):308–312, 2005. ISSN 0028-0836.

[110] 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. ISSN 0028-0836.

[111] C. Thiriet and J. J. Hayes. Replication-independent core histone dynamics at transcriptionally active loci in vivo. *Genes and Development*, 19(6):677, 2005.

[112] H. Kimura and P. R. Cook. Kinetics of core histones in living human cells little exchange of H3 and H4 and some rapid exchange of H2B. *Journal of Cell Biology*, 153 (7):1341–1354, 2001.

[113] M. A. Schwabish and K. Struhl. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Molecular cell*, 22(3):415–422, 2006.

[114] A. Jamai, R. M. Imoberdorf, and M. Strubin. Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Molecular cell*, 25(3):345–355, 2007.

[115] D. J. Clark and G. Felsenfeld. A nucleosome core is transferred out of the path of a transcribing polymerase. *Cell*, 71(1):11, 1992. 0092-8674 doi: DOI: 10.1016/0092-8674(92)90262-B.

[116] V. M. Studitsky, D. J. Clark, and G. Felsenfeld. A histone octamer can step around a transcribing polymerase without leaving the template. *Cell*, 76:371–382, 1994.

[117] V. M. Studitsky, G. A. Kassavetis, E. P. Geiduschek, and G. Felsenfeld. Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science*, 278 (5345):1960, 1997.

[118] O. I. Kulaeva, F. K. Hsieh, and V. M. Studitsky. RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. *Proceedings of the National Academy of Sciences*, 2010.

[119] O. I. Kulaeva, D. A. Gaykalova, N. A. Pestov, V. V. Golovastov, D. G. Vassylyev, I. Artsimovitch, and V. M. Studitsky. Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. *Nature Structural and Molecular Biology*, 2009.

[120] J. Bednar, V. M. Studitsky, S. A. Grigoryev, G. Felsenfeld, and C. L. Woodcock. The nature of the nucleosomal barrier to transcription direct observation of paused intermediates by electron cryomicroscopy. *Molecular Cell*, 4(3):377–386, 1999.

[121] W. Walter and V. M. Studitsky. Facilitated transcription through the nucleosome at high ionic strength occurs via a histone octamer transfer mechanism. *Journal of Biological Chemistry*, 276(31):29104–29110, 2001.

[122] J. Yan, R. Kawamura, and J. F. Marko. Statistics of loop formation along double helix DNAs. *Physical Review E*, 71(6):61905, 2005.

[123] T. H. Eickbush and E. N. Moudrianakis. The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry*, 17(23):4955–4964, 1978.

[124] Y. J. Park, P. N. Dyer, D. J. Tremethick, and K. Luger. A new fluorescence resonance energy transfer approach demonstrates that the histone variant H2AZ stabilizes the histone octamer within the nucleosome. *Journal of Biological Chemistry*, 279(23): 24274, 2004.

[125] J. Huang, T. Schlick, and A. Vologodskii. Dynamics of site juxtaposition in supercoiled DNA. *Proceedings of the National Academy of Sciences*, 98(3):968, 2001.

[126] Y. Bao, K. Konesky, Y. J. Park, S. Rosu, P. N. Dyer, D. Rangasamy, D. J. Tremethick, P. J. Laybourn, and K. Luger. Nucleosomes containing the histone variant H2A.Bbd organize only 118 base pairs of DNA. *The EMBO journal*, 23(16):3314, 2004.

[127] J. J. Hayes and A. P. Wolffe. Histones H2A/H2B inhibit the interaction of transcription factor IIIA with the Xenopus borealis somatic 5S RNA gene in a nucleosome. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (4):1229, 1992.

[128] M. Bruno, A. Flaus, C. Stockdale, C. Rencurel, H. Ferreira, and T. Owen-Hughes. Histone H2A/H2B dimer exchange by ATP-dependent chromatin remodeling activities. *Molecular cell*, 12(6):1599–1606, 2003.

[129] C. Rivetti, M. Guthold, and C. Bustamante. Scanning force microscopy of DNA deposited onto mica: Equilibration versus kinetic trapping studied by statistical polymer chain analysis. *Journal of molecular biology*, 264(5):919–932, 1996.

[130] C. Rivetti and S. Codeluppi. Accurate length determination of DNA molecules visualized by atomic force microscopy: evidence for a partial B to A-form transition on mica. *Ultramicroscopy*, 87(1):55–66, 2001.

[131] G. Orphanides and D. Reinberg. RNA polymerase II elongation through chromatin. *Nature*, 407(6803):471–476, 2000. ISSN 0028-0836.

[132] A. Shilatifard. Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim. Biophys. Acta: Gene Struct. Exp.*, 1677(1-3):79–86, 2004.

[133] 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, 1998.

[134] R. Belotserkovskaya, S. Oh, V. A. Bondarenko, G. Orphanides, V. M. Studitsky, and D. Reinberg. FACT facilitates transcription-dependent nucleosome alteration, 2003.

[135] G. Orphanides, W.-H. Wu, W.S. Lane, M. Hampsey, and D. Reinberg. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature*, 400(6741):284–288, 1999.

[136] M. A. Schwabish and K. Struhl. The Swi/Snf complex is important for histone eviction during transcriptional activation and RNA polymerase II elongation in vivo. *Mol. Cell. Biol.*, 27(20):6987–6995, 2007.

[137] J. K. Davie and C. M. Kane. Genetic interactions between TFIIS and the Swi-Snf chromatin-remodeling complex. *Mol. Cell. Biol.*, 20(16):5960–5973, 2000.

[138] M. Carey, B. Li, and J. L. Workman. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular Cell*, 24(3):481–487, 2006.

[139] B. R. Cairns, Y. Lorch, Y. Li, M. Zhang, L. Lacomis, H. Erdjument-Bromage, P. Tempst, J. Du, B. Laurent, and R. D. Kornberg. RSC, an essential, abundant chromatin-remodeling complex. *Cell*, 87:1249–1260, 1996.

[140] X. Yang, R. Zaurin, M. Beato, and C. L. Peterson. Swi3p controls SWI/SNF assembly and ATP-dependent H2A-H2B displacement. *Nature Structural and Molecular Biology*, 14(6):540, 2007.

[141] J. Soutourina, V. Bordas-Le Floch, G. Gendrel, A. Flores, C. Ducrot, H. Dumay-Odelot, P. Soularue, F. Navarro, B. R. Cairns, and O. Lefebvre. Rsc4 connects the chromatin remodeler RSC to RNA polymerases. *Mol. Cell. Biol.*, 26(13):4920–4933, 2006.

[142] B.W.T. Kelvin and Institution of Civil Engineers (Great Britain). *Electrical units of measurement*. Published by the Institution, 1883.