2021

Single-Molecule Investigation of Chromatin-Associated Factors in Genome Organization and Epigenetic Maintenance

Rachel Leicher

SINGLE-MOLECULE INVESTIGATION OF CHROMATIN-ASSOCIATED FACTORS IN GENOME ORGANIZATION AND EPIGENETIC MAINTENANCE

A Thesis Presented to the Faculty of
The Rockefeller University
in Partial Fulfillment of the Requirements for
the degree of Doctor of Philosophy

by
Rachel Leicher
June 2021

SINGLE-MOLECULE INVESTIGATION OF CHROMATIN-ASSOCIATED FACTORS IN GENOME ORGANIZATION AND EPIGENETIC MAINTENANCE

Rachel Leicher, Ph.D.
The Rockefeller University, 2021

The central dogma of biology has laid the foundation for understanding gene expression through the mechanisms of transcription and translation. However, another layer of eukaryotic gene regulation lies in the complex structure of chromatin. This scaffold of structural proteins and enzymatic regulators determines what genes are expressed at what times, leading to cell differentiation, cell fate, and often disease.

Currently, the field of chromatin biology has relied on basic biochemistry and cellular assays to identify key epigenetic regulators and their role in genomic maintenance. For this thesis work, I have developed a biophysical platform to study chromatin-associated factors at the single-molecule level (Chapter 2). This methodology allows us to extract key mechanistic details often obscured by standard bulk methodologies. Using this platform, we posed the question of how epigenetic factor, Polycomb repressive complex 2 (PRC2) engages with chromatin (Chapter 3). PRC2 is a major epigenetic machinery that maintains transcriptionally silent heterochromatin in the nucleus and plays critical roles in embryonic development and oncogenesis. It is generally thought that PRC2 propagates repressive histone marks by modifying neighboring nucleosomes in a strictly linear progression. However, the behavior of PRC2 on native-like chromatin substrates remains incompletely characterized, making the precise mechanism of PRC2-mediated heterochromatin maintenance elusive. Our understanding of this process was limited by the resolution of structural techniques that fail to identify PRC2-binding modes on long chromatin substrates. In short, we found direct evidence that PRC2 can simultaneously engage nonadjacent nucleosome pairs. The demonstration of PRC2's ability to bridge noncontiguous chromosomal segments furthers our understanding of how Polycomb complexes spread epigenetic modifications and compact chromatin.

In addition to this single-molecule chromatin binding technology, I also created a single-molecule platform harnessing correlative force and fluorescence microscopy to assay the material properties of phase separated condensates (Chapter 2). This assay combined methodology to visualize condensate formation at the single-molecule level, in addition to optical trapping of individual droplets to investigate their material properties. Utilizing this technology, we interrogated the role of linker histone H1 (Chapter 4). The linker histones are the most abundant group of chromatin-binding proteins that bind and organize eukaryotic chromatin. However, roles for the diverse and largely unstructured H1 proteins beyond chromatin compaction remain unclear. We used correlative single-molecule force and fluorescence microscopy to directly visualize the behavior of H1 on DNA under different tensions. Unexpectedly, our results show that H1 preferentially coalesces around nascent, relaxed single-stranded DNA. *In vitro* bulk assays confirmed that H1 has a higher propensity to form phase-separated condensates with single-stranded DNA than with double-stranded DNA. Furthermore, we dissected the material properties of different H1:DNA condensates by controlled droplet fusion with optical tweezers, and found that increased DNA length and GC content result in more viscous, gel-like H1 condensates. Overall, our findings suggest a potential role for linker histones to sense and coacervate single-stranded nucleic acids in the nucleus, forming reaction

hubs for genome maintenance. This work also provides a new perspective to understand how various H1 subtypes and disease-associated mutations affect chromatin structure and function.

In summary, we have gained a greater understanding of the biophysical basis for chromatin regulation by both PRC2 and histone H1. Both of the biophysical platforms created for these studies can be applied to various new targets in chromatin biology. They will enable the investigation of a multiplicity of binding interactions, regulatory mechanisms, and material properties of protein-nucleic acid complexes (Chapters 5 & 6). I believe single-molecule techniques will become a major toolset to study chromatin biology, identifying the intricacies and interactions between epigenetic factors and our genome.

Dedicated to women in science. Remember that you belong and to be bold.

ACKNOWLEDGMENTS

This doctoral thesis was a highly collaborative effort. I have many individuals to thank for their contributions during the last few years, perhaps too many to list here, and will try to be as comprehensive as possible. First and foremost, I am thankful to Shixin Liu, my thesis advisor, who took a chance on me as his first graduate student. Shixin has taught me invaluable skills in data analysis, science communication, and experimental design. I must also thank the Liu lab members for their insightful analyses including Mike Wasserman, Ling Wang, Sai Li, Xiangwu Ju, Tuan Nguyen, Gabby Chua, Jeremy Chang, and Rochelle Shih. The students have especially given me much needed energy over the last year. Thank you to Gabby in particular for her continuation of the H1 project and thoughtful advice.

I would like to thank my thesis committee, Dave Allis, Seth Darst, and Scott Blanchard, who have been very supportive during my time at Rockefeller and provided me with much needed insight concerning my scientific endeavors and career goals. I also thank David Rueda for serving as my external referee.

In many ways, single-molecule experiments require collaboration from a wide array of biologists, biophysicists, and computational chemists to create a meaningful story. I have had the honor to collaborate with many scientists from Rockefeller and beyond for all of my projects. First, I would like to thank Tom Muir and Eva Ge for their collaborative work on the PRC2 story and immense assistance with protein purification. Thank you to many members of the Allis lab, including Agata Lemiesz, Leah Gates, Ben Nacev, Liling Wan, and Lijuan Feng for their collaborative efforts on PRC2, oncohistones, ENL and insights into all things epigenetics. Thank you to Tom Walz and his lab members for overseeing my PRC2-chromatin electron microscopy experiments. Next, I would like to acknowledge Bin Zhang and his lab members Xingcheng Lin, Wejun Xie, and Andrew Latham for their computational contributions to the PRC2 and H1 stories. I would like to also thank Yael David and her lab members, Wola Osunsade and Qingfei Zheng, for extensive collaboration on the H1 story and histone glycation project. Also, to members of the Alushin Lab, including Greg Alushin, Lin Mei, Ayala Carl, Santiago Espinosa, in addition to Mark Ebrahim and Johanna Sotiris from the EM core, for structural collaborations on the PRC2 story and actin-force sensing paper. I am indebted to two talented rotation students, Matthew Reynolds and John Watters, who wrote excellent data analysis software and were ever so patient with my minimal computational background. Finally, I would like to thank the scientists at LUMICKS for their scientific expertise, especially Jordi Cabanas-Danes who provided instrumental training and knowledge about the C-Trap and single-molecule force acquisition.

The Tri-Institutional Program in Chemical Biology (TPCB) has provided me with an amazing opportunity for my graduate studies under its administration, Derek Tan, Kathleen Pickering, and Margie Hinonangan-Mendoza. Their leadership has been ever so responsive to the needs of our student body, not only providing a supportive community, but also welcoming social justice and change within the program. TPCB has also given me countless friends and colleagues that I will treasure for years to come.

Lastly, I could not have done this Ph.D. without tremendous personal support from friends and family. My parents, Bruce and Ellen Leicher, have been extraordinarily supportive, helping me every step of the way. I am thankful for the lifelong friends I have made in TPCB, including Emma Garst, Lina Miller, and Darren Johnson, in addition to my ever so patient

roommates Malik Chaker-Margot and Jordan Mattheisen. To all my other friends from TPCB, Harvard, Wesleyan, and beyond, thank you for your support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. Introduction: chromatin biology and single-molecule methods

1.1 Scope: chromatin and epigenetics, an introduction

DNA is packaged into the nucleus of cells in eukaryotes via the chromatin fiber which condenses to form individual chromosomes (**Figure 1.1**). The chromatin fiber is made up of the repeating unit, termed the nucleosome, where approximately 147 base pairs of DNA are wrapped 1.65 times around a histone octamer containing core histones H2A, H2B, H3 and H4. The core nucleosome is often bound to linker histones (i.e. linker histone H1), forming a unit called the chromatosome. Histones are decorated with a variety of post-translational chemical modifications, called post-translational modifications, that interact with genomic machinery to regulate gene expression. These mechanisms contribute to epigenetics, or the heritable phenotypes that take place without alterations to the DNA sequence.

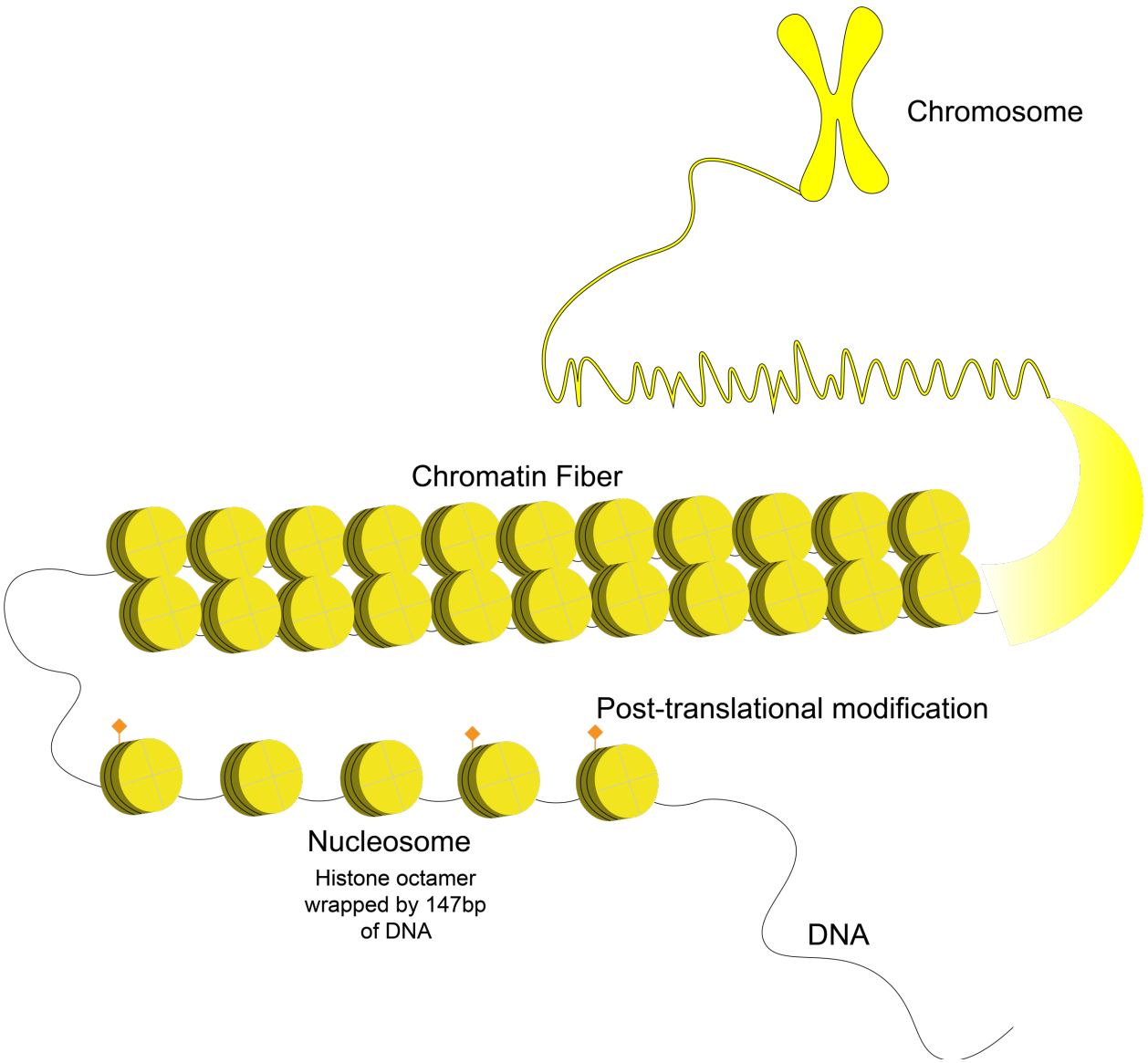


Figure 1.1. Chromatin organization cartoon. DNA is packaged into chromosomes in the nucleus via compaction of the chromatin fiber. Chromatin is made up of the repeating using of the nucleosome, which contains a histone octamer wrapped roughly twice by 147 base pairs of DNA. Histone tails in the nucleosome are modified by post-translational modifications that infer gene expression patterns.

Epigenetic maintenance has been long studied since the distinction between euchromatin (transcriptionally active chromatin) and heterochromatin (inactive chromatin) was discovered by Heitz in 1928, leading to studies on non-Mendalian inheritance in *Drosophila melanogaster* [1] as well as X-inactivation [2] and imprinting. Early staining studies using DNA intercalating dyes revealed spatially separated regions of euchromatin (light staining) and heterochromatin (dark staining) [3, 4]. A further distinction in condensation states and transcriptional activity was discovered between facultative heterochromatin (fHC) and constitutive heterochromatin (cHC).

cHC remains silent and is found mostly at centromeres, where fHC has the ability to be transcriptionally activated and regulated by a variety of epigenetic factors [5].

Further discoveries in DNA methylation, nucleosome organization, and histone modifications [6] in flies, yeast and plants led to the identification of key factors that regulate switches between active and silent chromatin. Publication of the X-ray crystal structure of the nucleosome [7] helped visualize the proposed nucleosomal organization of chromatin [8]. Through the identification of histone modifications [9] and subsequent enzymes [10] that worked in tandem to catalyze the addition and removal of these marks, hypotheses for gene regulation and the formation and maintenance of heterochromatin arose. For example, HP1 occupancy and H3K9 methylation were identified as signatures of constitutive heterochromatin [11, 12], whereas a whole host of different histone modifications (i.e. H3K27me3) have been found to mark facultative heterochromatin [5] or euchromatin (H3K4me3, H3K27ac).

In what is often called "the modern era of epigenetic research," the histone code hypothesis [13] was born. It was proposed that combinations of different histone post-translational modifications would lead to alternative outcomes in gene expression, regulated by effector proteins termed writers, readers, and erasers, working in *cis* and in *trans* to perform specific chemical reactions to regulate these marks. One open question in the field is how these marks are maintained throughout development and cell division upon the generation of nascent histones [14, 15]. How are these marks continually spread by different writers, readers and erasers?

Epigenetic marks play a key role in regulating gene expression. These post-translational modifications contribute to heterochromatin and euchromatin states, which control gene expression profiles that can determine cellular differentiation and development. Failure to maintain the proper epigenetic landscape during cell differentiation and division can result in disease states such as cancer. The players that facilitate the maintenance of these reversible chromatin marks are often essential proteins in the cell. Uncovering the mechanisms of these molecular machines and their dynamic interplay is imperative to understand how epigenetics plays a role in genomic maintenance.

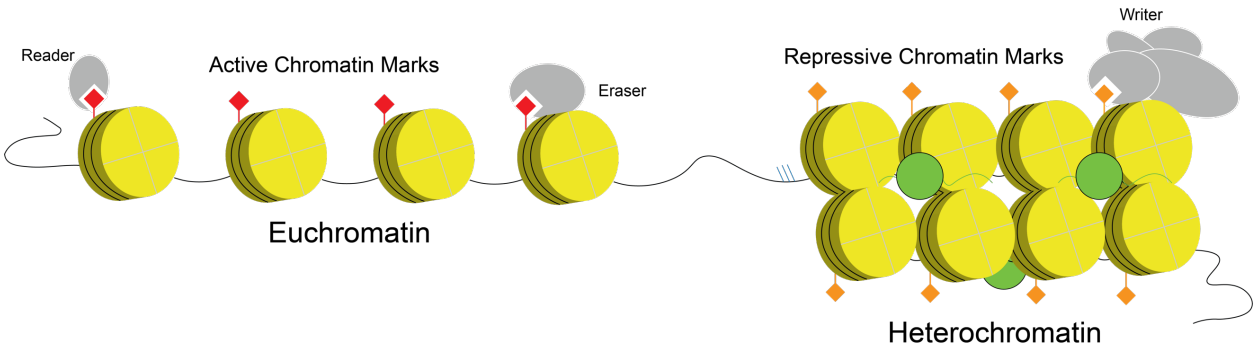


Figure 1.2. Chromatin organization by post-translational modifications. Broadly, chromatin is categorized into two types; euchromatin, containing active marks, which is actively transcribed, and heterochromatin, which contains repressive marks, and is the silenced form of chromatin. Writers, readers, and erasers maintain and bind these epigenetic marks.

There are many key writers, readers, and erasers that have been identified as major regulators of gene expression. Early work identified how histone acetylation and deacetylation were coupled to on and off states of gene regulation [16] through catalytic activities of histone acetyltransferases (HATs) and deacetylases (HDACs) [17]. Key repressors were identified, including many that modify histone lysine residues and contain an evolutionarily conserved SET domain [18]. Examples of these include Suv39H1, E(z) complexes, and Trithorax proteins, from which the SET name is derived. Suv39H1 methylates H3K9 [19], which is a key mark of constitutive heterochromatin [20]. H3K9me3 subsequently recruits HP1, a chromatin reader, which promotes chromatin compaction and silencing [21, 22]. Other key histone lysine methyltransferases include Trithorax and MLL (mixed-lineage leukemia) proteins [23-25] which add the activating H3K4me3 mark, and Polycomb Repressive Complex 2 (PRC2) [26], which adds the repressive H3K27me3 mark, a hallmark of facultative heterochromatin [27]. A comprehensive discussion of PRC2 and its role in gene repression can be found in Chapter 3, as PRC2-chromatin engagement is a major topic of this thesis work.

In addition to the epigenetic factors of readers, writers, and erasers, structural proteins may regulate key processes in the nucleus, such as DNA replication, damage, and repair. Later in this thesis (Chapter 4), we will discuss the role of histone H1, the linker histone that binds to the core nucleosome. While previously only known for chromatin compaction [28], this structural protein may influence other cellular processes, as it is the most abundant chromatin protein in the nucleus.

1.2 Techniques to study chromatin organization and regulatory mechanisms

Many different techniques have been used to study epigenetic maintenance and histone modifications. From early studies utilizing yeast genetics, *in vitro* biochemistry, and *Drosophila* models, to improvements in next-generation ChIP-seq and *in vivo* chromatin mapping via RICC-seq [29], great strides have been made to understand the processes underlying gene regulation. Recent advances in chemical biology [30] have also been utilized to understand the intricacies of epigenetic effector proteins.

Structural techniques have been extensively employed to study chromatin function. Many hypotheses about the chromatin structure and compaction of DNA into chromosomes have been published in the last century. Electron microscopy images of chromatin fibers [31, 32] revealed that chromatin forms a distinct 30nm fiber [33] as an intermediary step to fully condensing. Advances in cryo-electron microscopy have discovered the chromatin fiber repeating unit as the tetranucleosome which alternates in a zig-zag fashion, and is modulated by the presence of linker histone H1 [34]. Recent studies, utilizing alternative techniques, have suggested that the chromatin fiber is a more disordered chain, with alternative structures between 5-24nm [35]. Given the disagreement in the field, there is much more to be learned about chromatin structure, especially in the presence of epigenetic factors and modifications.

Groundbreaking techniques, such as chromatin conformation capture, have identified spatially separated chromatin domains in cells, termed topologically associated domains (TADs) [36]. TADs have specific characteristics in size, structure, and protein composition across varied organism types, and perpetuate through cell division and lineage differentiation [37]. Super-resolution imaging techniques, including STORM, have allowed the visualization of nanoscale organization of the antagonistic epigenetic marks, H3K4me3 and H3K27me3, into discrete compartments [38].

Despite a breadth of knowledge, specifics about the mechanism of chromatin-associated proteins and their molecular interactions with the chromatin scaffold have remained elusive. Throughout this thesis, I propose that using single-molecule methods will provide answers to these specific questions through precise manipulation of chromatin substrates and direct visualization of chromatin factors at the single-molecule level.

1.3 Single-molecule techniques: optical trapping

As discussed above, chromatin biology has been studied through techniques ranging from bulk biochemistry to chromatin immunoprecipitation, chromatin capture and *in vivo* studies. While these findings are imperative to understanding the chromatin landscape, they cannot access shorter time scales (milliseconds to seconds) or the dynamics of molecules in the nanometer scale, including structural fluctuations, effector interactions, and chromatin state spreading facilitated by epigenetic regulators and nucleosome substrates [39]. Single-molecule techniques overcome limitations in biochemistry due to ensemble averaging which masks dynamic information and mechanistic details.

Optical trapping was discovered by Arthur Ashkin and others, identifying the principle that a beam of light could trap a dielectric particle [40], earning the Nobel Prize in 2018. Since that discovery, force microscopy – including optical trapping, magnetic trapping, and atomic force microscopy – has been a powerful tool to investigate the movement and forces of different biological molecules, motors, and enzymes [41]. In particular, optical trapping techniques provide the ability to study a wide array of biological substrates, including early studies of trapping bacteria [42], kinesin directly acting on microtubules [43], enzymes translating on tethered DNA substrates [44], viral packaging [45, 46], and protein unfolding measurements [47]. Early work in the field studied the properties of DNA extension [48, 49], where DNA manipulation by optical traps lead to the discovery of a worm-like chain model. DNA is a polymer that stretches and deforms under applied force, and at very high forces, can undergo "overstretching" or DNA melting [50], exposing regions of single-stranded DNA. In light of these discoveries, the field has broadened the knowledge of DNA-based motor proteins and processes such as DNA packaging, supercoiling, and modification by epigenetic marks [51-54].

1.3.1 Using force microscopy to study chromatin biology

Recent advances in structural biology have yielded a growing number of high-resolution structures of chromatin-associated complexes [53]. However, determination of the binding configuration of multi-subunit holo complexes on polynucleosome arrays remains challenging, not least due to the considerable conformational heterogeneity of these large assemblies.
In particular, single-molecule force manipulation methods are powerful tools to decipher the mechanical properties of nucleosomes, chromatin fibers, and whole chromosomes [55-66]. By studying the unfolding dynamics of nucleosomes and chromatin arrays, we are able to decipher the mechanical fingerprints of epigenetic modifiers, such as PRC2. The basic principle of chromatin unfolding is as follows; when pulling on nucleosome arrays with optical traps, the outer wrap of the nucleosome comes off first simultaneously at lower forces, while the inner wrap of the nucleosomes is removed stochastically at higher forces. Any additional transitions we observe in the presence of epigenic chromatin-bound factors could be attributed to their

distinct binding modes. We have applied this model to decipher a diverse repertoire of PRC2 binding modes on chromatin.

1.3.2 Combining force and fluorescence microscopy

Recent advances in single-molecule techniques have combined fluorescence microscopy, optical tweezers, and microfluidics [67] to allow for the manipulation and observation of individual molecules with high spatial, temporal, and force-scale resolution. Such correlative single-molecule fluorescence-force microscopy has allowed the discovery of replication fork activation mechanisms [68], how cohesion specifically tethers DNA [69], and off-target Cas9 activity [70]. This powerful technique not only allows the manipulation of DNA and chromatin substrates, but also the simultaneous tracking of bound fluorescent proteins. In this thesis, we discuss how this technology can be used to dissect interactions of chromatin modifiers and binders, including PRC2, histone H1, eleven-nineteen-leukemia protein (ENL) and others.

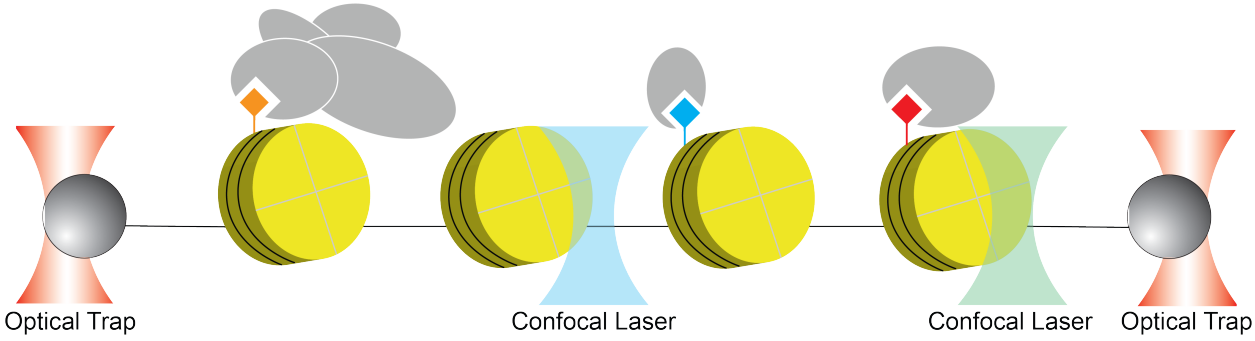


Figure 1.3. Correlative force-fluorescence microscopy of chromatin biology. A cartoon of a single chromatin array tethered between beads positioned in optical traps (nucleosomes not drawn to scale). Additional confocal scanning lasers illuminate bound epigenetic factors (grey).

1.4 Phase Separation

Recently, the formation of membrane-less cytoplasmic and nuclear compartments via phase separation has been widely studied [71]. Phase separation is a well-known phenomenon in polymer chemistry [72]. In the most basic sense, it is a physical process that occurs when a supersaturated solution of components spontaneously separates into two phases, a dense phase, and a dilute phase, that stably coexist. In a biological context, protein phase transitions can span across liquid, to gel, and even solid-like states. Some membrane-less organelles, such as the nucleolus, have been studied previously, however the application of phase separation to such a large array of biomacromolecules has been a more recent development [73]. The most commonly studied process is liquid-liquid phase separation, or LLPS. Many cellular proteins have the propensity to phase separate into liquid droplets, often regulated by intrinsically disordered regions, ionic interactions, and molecular crowding [71]. Phase separated compartments can aid in the buffering of components. For example, stress granules formed by RNA and RNA-binding proteins provide a specific environment for processing or suppressing certain reactions in response to stimuli [74], such as high temperatures or oxidative stress. LLPS can also regulate

activation mechanisms [75], protein localization [76], formation of mechanical filters [77], and the generation of mechanical forces [78]. Consequently, dysregulation of LLPS can lead to disease, such as in ALS [79].

Specifically in chromatin biology, LLPS has been proposed as the basis for heterochromatin formation, given the high concentration of chromatin components in the nucleus and their charged nature [80, 81]. Examples of phase separation for both repressive [82, 83] and active [84] epigenetic factors have emerged, including instances of modulation by post-translational modifications [85]. To that end, the phase separation properties of chromatin components, such as HP1 [83, 85, 86], nucleosome arrays, BRD4, and linker histone H1 have been reported [87]. Further interactions between post-translational modifications and their effectors, for example H3K9me3 and the Suv39H1-HP1 reader-writer complex [88], can nucleate LLPS droplets *in vivo*. The biophysical properties of chromatin compartments have also been theorized through various computational models such as LLPS, polymer-polymer phase separation [89], and nonspecific bridging [90]. Despite many examples of chromatin associated LLPS, a recent study suggested chromatin is a gel-like solid structure in the nucleus [91]. However, chromatin-mixing and liquidity may be hard to differentiate at the mesoscale. A full understanding of the mechanisms of the nature of chromatin phase separation is still lacking.

CHAPTER 2. Single-molecule methods development for studying chromatin biology and phase separation

Key innovations in methods development enabled the innovative single-molecule studies in this thesis. In this chapter, I will expand upon two key methods, (1) single-molecule method development for studying chromatin biology, and (2) single-molecule adaptation of correlative force-fluorescence spectroscopy for studying condensates. Both invoke single-molecule force spectroscopy as a powerful tool to analyze the architecture and interaction of large macromolecular assemblies that are refractory to high-resolution structural interrogations. In the first section, I describe an optical-tweezers-based platform for extracting the mechanical fingerprints of individual nucleosome arrays bound with chromatin-associated complexes, such as PRC2. This platform comprehensively characterizes the possible binding modes of PRC2 on chromatin, measures their mechanical strengths, and is broadly applicable to the studies of other epigenetic machineries. In the second section, I describe an optical-tweezers and confocal microscopy-based platform to assay phase separated condensates at the single-molecule level. Please note that all other experimental details can be found in the methods and materials chapter (Chapter 7).

2.1 Single-molecule method development for studying chromatin biology

Adapted from *Leicher et al*.: Probing the interaction between chromatin and chromatin-associated complexes with optical tweezers.

Eukaryotic chromatin is made up of the repeating unit called the nucleosome, which contains a histone octamer wrapped by ~147 base pairs of DNA [7, 92]. The structure and dynamics of nucleosomes and chromatin fibers provide an important layer of regulation for genome organization and gene expression [54]. Moreover, a plethora of nuclear factors bind, remodel, and modify chromatin, driving cell cycle progression and cell fate transition. Recent advances in structural biology have yielded a growing number of high-resolution structures of chromatin-associated complexes [53]. However, determination of the binding configuration of multi-subunit holo complexes on polynucleosome arrays remains challenging, not least due to the considerable conformational heterogeneity of these large assemblies.

Because of their unique ability to detect transient intermediates and heterogeneous populations, single-molecule techniques have been successfully utilized to study chromatin biology [66]. In particular, single-molecule force manipulation methods, such as optical tweezers, magnetic tweezers, and atomic force microscopy, are powerful tools to decipher the mechanical properties of nucleosomes, chromatin fibers, and whole chromosomes [55-58, 61-65]. In this chapter, we describe a workflow aimed at extending the usage of single-molecule force spectroscopy to the studies of diverse nuclear factors that engage with chromatin, including pioneer transcription factors [93], chromatin remodelers [94], and epigenetic modifiers [6]. By probing the binding configurations of these factors and the mechanical fingerprints of chromatin-associated assemblies, we are able to gain fresh insights into the mechanisms by which the genetic information stored in chromatin is accessed, decoded, and modified. We also introduce data analysis tools developed in parallel to extract information from the force-extension curves yielded by the single-molecule experiments.

We adopted this workflow to investigate the behavior of PRC2 on customized nucleosome arrays using a dual-trap optical tweezers instrument [67]. PRC2 is a major epigenetic machinery that maintains transcriptionally silent heterochromatin in the nucleus and plays critical roles in embryonic development and oncogenesis [95]. It is generally thought that PRC2 propagates the repressive H3K27me3 mark by modifying neighboring nucleosomes. Surprisingly, our results revealed a diverse repertoire of PRC2 binding modes on chromatin, including a previously underappreciated mode in which PRC2 engages with a pair of non-adjacent nucleosomes. The prevalence and stability of these binding modes are regulated by PRC2 auxiliary subunit composition and histone posttranslational modification state [96]. These findings have implications for the mechanism of epigenetic spreading and chromatin compaction. In the future, by combining single-molecule fluorescence and force spectroscopy [68, 97], this experimental platform can be utilized to directly visualize the dynamics of chromatin-associated complexes while monitoring their response to mechanical perturbations.

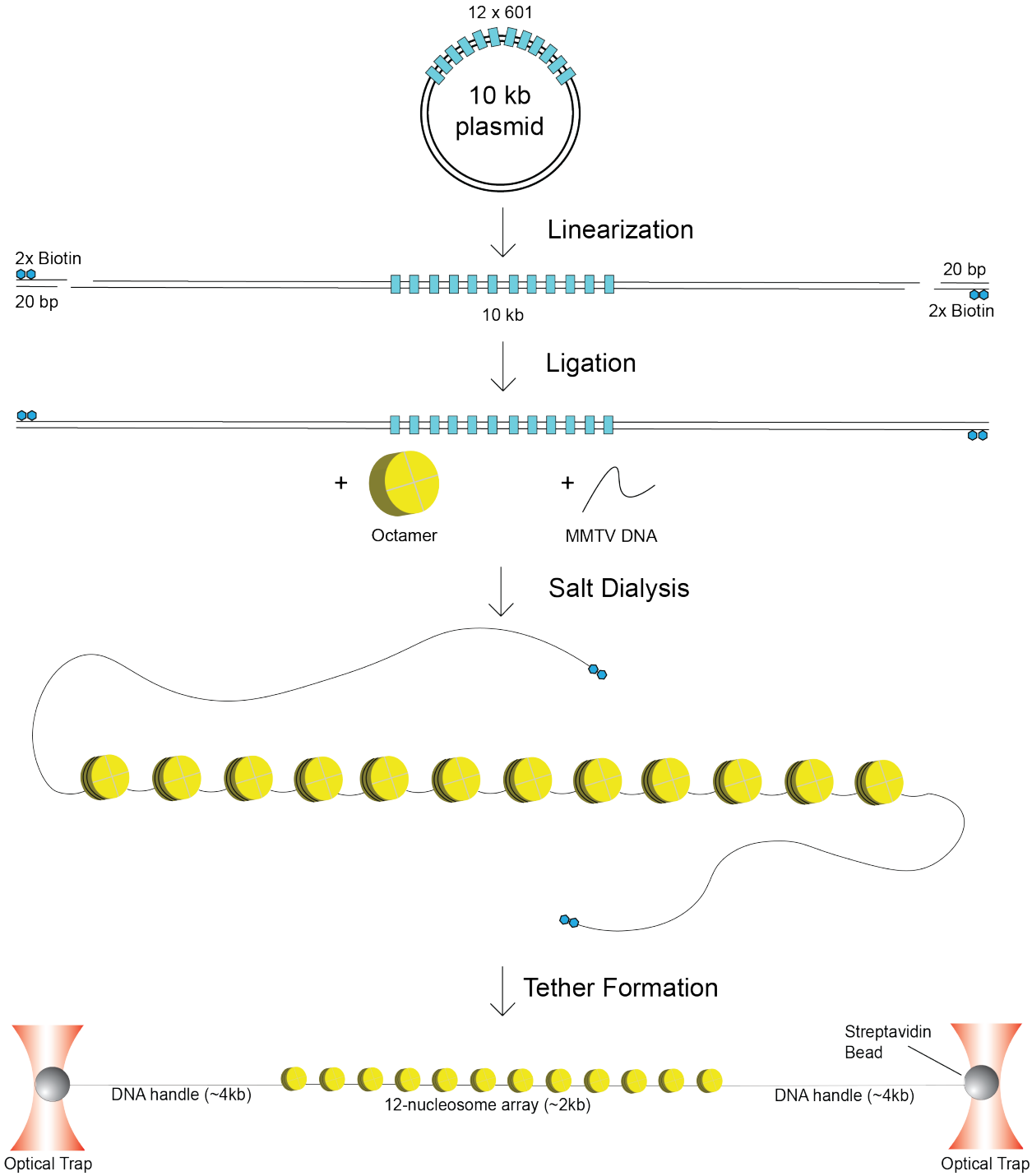


Figure 2.1. Construction of nucleosome array substrates for single-molecule manipulation. 12-mer nucleosome arrays are formed by linearization of a 10-kb plasmid containing twelve '601' nucleosome positioning sequences, ligation to 20-bp biotinylated adapters, and chromatinization with reconstituted histone octamers in the presence of competitor DNA via salt dialysis. Individual arrays are tethered between a pair of optically trapped streptavidin-coated beads and subjected to increased pulling force.

2.1.1 Materials

A. Construction of nucleosome array substrates

A1: Linearization of nucleosome array templates

1. *E. coli* DH5α competent cells (New England Biolabs)
2. Lysogeny broth (LB) (Fisher Scientific)
3. ZymoPURE Plasmid Gigaprep kit (Zymo Research)
4. BglI (New England Biolabs)
5. 10× NEBuffer 3.1 (New England Biolabs)
6. Sodium acetate (NaOAc) solution (3 M, pH 5.2)
7. Ethanol
8. Nanodrop 2000 UV-Vis spectrophotometer (Thermo Scientific)

A2: Ligation of nucleosome array template to biotinylated adapters

1. Linearized nucleosome array template (see section **2.1.1**)
2. Oligonucleotides for the adapters (**Table 1**)
3. 10× Annealing buffer (100 mM Tris-HCl pH 7.5, 500 mM NaCl, 100 mM EDTA)
4. T4 DNA Ligase (New England Biolabs)
5. 10× Ligase buffer (500 mM Tris-HCl pH 7.5, 100 mM $MgCl_2$, 10 mM ATP, 100 mM DTT)
6. DNA Clean & Concentrator kit (Zymo Research)
7. Thermocycler (Eppendorf)
8. Nanodrop 2000 UV-Vis spectrophotometer (Thermo Scientific)

Table 2.1. Sequences of adapter oligonucleotides

RLO33	5'P-ATGATACCCGGGCGGCGACCT
RLO34	5'Biotin(2×)-AGGTCGCCGCCCGGGTATCATTGC
RLO35	5'Biotin(2×)-ATGATACCCGGGCGGCGACCTGCA
RLO36	5'P-AGGTCGCCGCCCGGGTATCAT

A3: Purification of mouse mammary tumor virus (MMTV) DNA

1. Plasmid containing repeats of MMTV DNA sequence [98]:
   ACTTGCAACAGTCCTAACATTCACCTCTTGTGTGTTTGTGTCTGTTCGCCA
   TCCCGTCTCCGCTCGTCACTTATCCTTCACTTTCCAGAGGGTCCCCCCGCA
   GACCCCGGCGACCCTGGTCGGCCGACTGCGGC ACAGTTTTTTG
2. ZymoPURE Plasmid Gigaprep kit (Zymo Research)
3. EcoRV (New England Biolabs)
4. 10× CutSmart buffer (New England Biolabs)
5. 40% Polyethylene glycol (PEG) 6000 (Sigma-Aldrich)
6. Chloroform:isoamyl alcohol (CIA) 24:1 (Sigma-Aldrich)
7. 3 M NaOAc solution, pH 5.2
8. Ethanol

A4: Reconstitution of nucleosome arrays

1. Initial buffer (10 mM Tris-HCl, 1.4 M KCl, 0.1 mM EDTA, 1 mM DTT)

2. Dilution buffer (10 mM Tris-HCl, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT)
3. Assembly buffer (10 mM Tris-HCl, 2 M or 4 M KCl, 0.1 mM EDTA, 1 mM DTT)
4. Histone octamers (at least 5 μM) (purified in house or purchased from Epicypher)
5. Slide-A-Lyzer MINI dialysis units (7,000 MWCO) (Thermo Scientific)
6. Peristaltic pump (LongerPump)
7. Stir plate

B. Evaluation of nucleosome arrays

B1: Magnesium-induced precipitation

1. Composite APAGE gel (2% acrylamide, 1% agarose)
2. Magnesium chloride ($MgCl_2$) (Sigma-Aldrich)
3. TEK buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA, 10 mM KCl, 1 mM DTT)
4. 50× Tris-Acetate-EDTA (TAE) buffer (2 M Tris base, 100 mM EDTA, 1 M acetic acid, pH 8.5)
5. Sucrose (Sigma-Aldrich)
6. SYBR Gold nucleic acid gel stain (Thermo Fisher)
7. Composite APAGE gel (0.4 g agarose, 0.2 mL 50× TAE buffer, 2 mL acrylamide, 37.7 mL water, 200 μL 10% ammonium persulfate (APS), 16.5 μL tetramethylethylenediamine).

B2: Micrococcal nuclease (MNase) digestion

1. Micrococcal nuclease (MNase) (New England Biolabs)
2. 10× MNase digest buffer (500 mM Tris-HCl pH 7.9, 50 mM $CaCl_2$)
3. Quench buffer (0.4 M NaCl, 0.2% (w/v) SDS, 20 mM EDTA)
4. 5% TBE gel
5. DNA Clean & Concentrator kit (Zymo Research)

B3: ScaI digestion

1. ScaI (New England Biolabs)
2. 5× ScaI digest buffer (50 mM Tris-HCl pH 7.5, 2.5 mM $MgCl_2$, 0.5 M KCl)

C. Optical tweezers assay

1. Imaging buffer (10 mM Tris-HCl pH 8.0, 200 mM KCl, 0.5 mM $MgCl_2$, 0.1 mM EDTA, 0.1% Tween)
2. 1 mM dithiothreitol (DTT)
3. 3.23-μm streptavidin-coated polystyrene beads (Spherotech)
4. 1× phosphate buffered saline (PBS), pH 7.4
5. Nucleosome array substrates (see section **2.1**)
6. Chromatin-associated complexes (e.g. PRC2)
7. Dual-trap optical tweezers (LUMICKS C-Trap)

D. Single-molecule data analysis

1. Custom python script: Force-Extension Analyzer
2. Commercial software: Origin (OriginLab), Prism (GraphPad)

2.1.2 Methods

Our method harnesses the strong nucleosome positioning sequence, the Widom '601' sequence [99], to create an array of nucleosomes with a distinct nucleosome repeat and linker length. By adding flanking biotinylated DNA handles through plasmid extraction and subsequent short-adapter ligation, we can dialyze histone octamers directly onto the DNA template to easily adapt the nucleosome arrays for single-molecule manipulation. The general workflow is depicted in **Figure 2.1**.

A. Construction of nucleosome array substrates

A1: Linearization of nucleosome array templates

The DNA template for nucleosome arrays is ~10 kbp in length, containing 2, 4, or 12 tandem repeats of the 147-bp-long '601' sequence flanked by ~4-kbp of random DNA sequences. Four template plasmids were created (**Table 2.1**).

1. Isolate plasmid DNA from 2 L of *E. coli* DH5α cells grown in the LB medium using the ZymoPURE Plasmid Gigaprep kit.
2. Digest the DNA with 10 μL of BglI and 25 μL of 10× NEBuffer 3.1 in a total volume of 250 μL at 37˚C overnight (see **Note 1**).
3. Precipitate the DNA by adding 25 μL of 3 M NaOAc solution and 688 μL of cold ethanol. Incubate for at least 30 min at -20˚C. Spin at 14,000 rpm and 4˚C for 10 min, remove supernatant, add 700 μL of 70% ethanol, repeat spin, and remove supernatant.
4. Resuspend the DNA in 30 μL of water. Take A_{260} reading with Nanodrop. Store at 4˚C.

Table 2.2. Plasmids containing '601' DNA sequences for making nucleosome arrays

pRL9	12-nucleosome array (30-bp linker)
pRL10	4-nucleosome array (30-bp linker)
pRL11	2-nucleosome array (30-bp linker)
pRL14	12-nucleosome array (50-bp linker)

A2: Ligation of nucleosome array template to biotinylated adapters

1. Anneal complementary adapter oligonucleotides (RLO33 + RLO34; RLO35 + RLO36) using a thermocycler. In each tube, add 4 μL of each oligonucleotide (100 μM), 2 μL of 10× Annealing buffer and 1 μL of water. Starting at 94˚C, decrease temperature by 2˚C every minute.
2. Ligate annealed adapters to the nucleosome array template at a 100:1 ratio overnight at room temperature. For one reaction, mix 1 μL of T4 DNA ligase, 1 μL of 10× Ligase buffer, 6 μL of each annealed adapter (80 μM), 2 μL of digested DNA template (6 μg/μL), and add water up to 20 μL. Perform 20 reactions.
3. Clean up the ligation product with a DNA Clean & Concentrator kit. Use one column per 5 ligation reactions. Elute each column with 25 μL of water. Take A_{260} reading with Nanodrop.

A3: Purification of MMTV DNA

1. Isolate MMTV plasmid DNA from 2 L of *E. coli* DH5α cells grown in the LB medium using the ZymoPURE Plasmid Gigaprep kit.
2. Digest MMTV DNA with 10 μL of EcoRV and 25 μL of 10× CutSmart buffer in a total volume of 250 μL at 37˚C overnight (see **Note 1**).
3. Perform PEG extraction. Add 104 μL of 5 M NaCl, 146.25 μL of 40% PEG6000, digested MMTV, and water to 650 μL. Vortex briefly and incubate at 37˚C for 5 min. Spin for 5 min at 14,000 rpm and room temperature. Save supernatant.
4. Perform CIA extraction. Extract supernatant with an equal volume of CIA. Mix vigorously and spin for 3 min at 14,000 rpm and 4˚C. Remove the aqueous fraction and repeat extraction.
5. Perform ethanol precipitation by adding 0.1× volume of 3 M NaOAc solution and 2.75× volume of cold ethanol. Incubate for at least 30 min at -20˚C. Spin at 14,000 rpm and 4˚C for 10 min, remove supernatant, add 700 μL of 70% ethanol, repeat spin, and remove supernatant.
6. Resuspend in 30 μL of water and take A_{260} reading.

A4: Reconstitution of nucleosome arrays

We use the salt dialysis strategy to form nucleosome arrays [100]. The purpose of adding MMTV DNA is to sequester excess histone octamers. To form arrays with site-specifically mutated or posttranslationally modified histones, see **Note 2**.

1. Prepare 200 mL of Initial buffer at 4˚C.
2. Prepare 1 L of Dilution buffer at 4˚C.
3. Prepare two 1-mL aliquots of Assembly buffer, one with 2 M KCl and the other with 4 M KCl.
4. Mix octamers, DNA template, MMTV DNA, and assembly buffers (2 M and 4 M KCl) in a final volume of 50 μL (see **Note 3**).
5. Equilibrate Slide-A-Lyzer MINI dialysis units (7,000 MWCO) in 200 mL of Initial buffer at 4˚C for 10 min.
6. Add the octamer:DNA mixture to the dialysis units. Equilibrate in Initial buffer at 4˚C for 1 hour.
7. Pump Dilution buffer into Initial buffer at a flow rate of 1 mL/min for 6 hours at 4˚C.
8. Stop the pump and dialyze for 1 hour at 4˚C.
9. Replace buffer with 300 mL of fresh Dilution buffer and dialyze overnight at 4˚C.
10. Replace buffer with 300 mL of fresh Dilution buffer and dialyze for 1 hour at 4˚C.
11. Collect the nucleosome array product from the dialysis units and store at 4˚C (see **Note 4**). The array product should be at a volume between 50 – 100 μL.

B. Evaluation of nucleosome arrays

B1: Magnesium-induced precipitation

If nucleosome arrays are successfully formed, a sizable shift should be visible between DNA and array samples on an APAGE gel.

1. Add $MgCl_2$ to a final concentration of 4 mM to the nucleosome arrays. Incubate for 10 min on ice.
2. Centrifuge for 10 min at 17,000 × g and 4˚C.
3. Remove supernatant and add 30 μL of TEK buffer (or buffer of choice). Let sit for 10 min before resuspending.
4. Pour an APAGE gel. Heat agarose with TAE and water prior to adding acrylamide, APS, and TEMED. Heat plates to 70˚C prior to pouring.
5. Pre-run APAGE gel in 0.25× TAE buffer for 3 hours at 120 V.
6. Load ~100 ng of sample in 10 μL volumes. Add 2 μL 50% sucrose, array sample, and water to a volume of 10 μL. Run for 75 min at 120 V.
7. Stain gel with SYBR Gold.

B2: MNase digestion
MNase digests DNA that is not protected by nucleosomes, thus estimating the saturation level of the arrays.

1. Mix nucleosome arrays (~2 picomole of '601' sites), 1 μL of 10× MNase buffer, 1 μL of MNase, and water to a final volume of 10 μL.
2. Digest on ice for 30 sec.
3. Quench with 10 μL of Quench buffer.
4. Purify products using a DNA Clean & Concentrator kit.
5. Analyze products on a 5% TBE gel.

B3: ScaI digestion
ScaI restriction sites are present between adjacent '601' sequences in our nucleosome array template. ScaI digestion should produce mononucleosomes if positioning is correct.

1. Mix nucleosome arrays (~1 picomole of '601' sites), 0.75 μL of 5× ScaI buffer, 0.5 μL of ScaI, and water to a final volume of 3.75 μL.
2. Incubate overnight at room temperature.
3. Analyze products on an APAGE gel and stain with SYBR Gold.

C. Optical tweezers assay
Single-molecule experiments are performed in a 5-channel flow cell (**Figure 2.2**) on a dual-trap optical tweezers instrument.

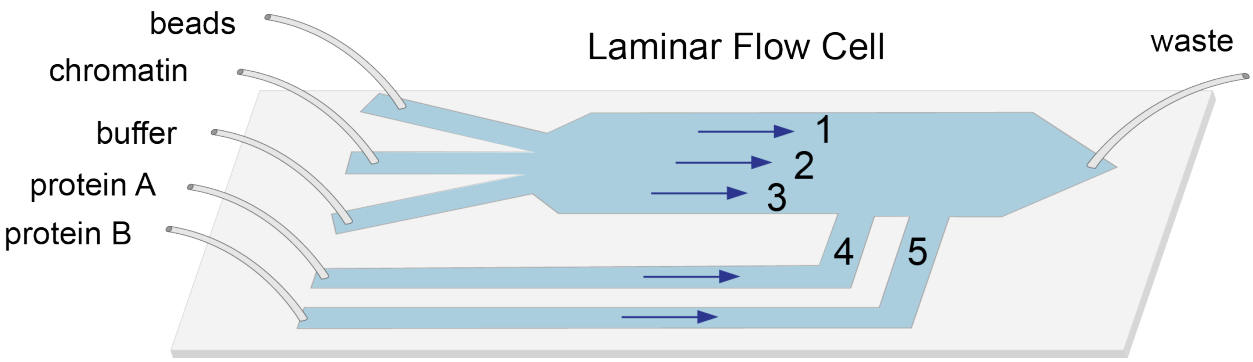


Figure 2.2. Schematic of the microfluidic chamber for the optical tweezers assay. Channels 1-3 are separated by laminar flow. Adapted with permission from [68].

C1: Sample and instrument preparation

1. Channel 1 solution. Dilute 4 μL of 3.23-μm streptavidin-coated beads in 1 mL of PBS buffer. Vortex.
2. Channel 2 solution. Add half of a nucleosome array aliquot to 1 mL of Imaging buffer. Add fresh DTT to a final concentration of 1 mM.
3. Channel 3 solution: 1 mL of Imaging buffer with fresh DTT (1 mM).
4. Channel 4/5 solution: Add protein of choice (e.g. 500 nM PRC2) to 200 μL of Imaging buffer with fresh DTT (1 mM).
5. Load the solutions into syringes connected to the corresponding channels. Flow 100 μL of solution through each channel.
6. Turn on the trapping laser. Catch one bead in each trap in channel 1.
7. Move traps to channel 3 and turn off flow.
8. Calibrate the traps according to the manufacturer's or in-house protocol.

C2: Recording force-extension trajectories of single chromatin tethers

1. Capture one bead in each optical trap in channel 1.
2. Move traps to channel 2. Bring beads close together (< 0.5 μm apart) to form a tether (see **Note 5**).
3. To stretch bare nucleosome arrays, move the tether to channel 3.
4. To stretch nucleosome arrays bound with other proteins, move the tether to channel 4/5 that contains the protein of interest (e.g. PRC2). Incubate for 5-10 sec prior to stretching to allow for binding.
5. Record the force-extension trajectory of the tether while moving one trap away from the other at a constant velocity (e.g. 0.1 μm/sec).

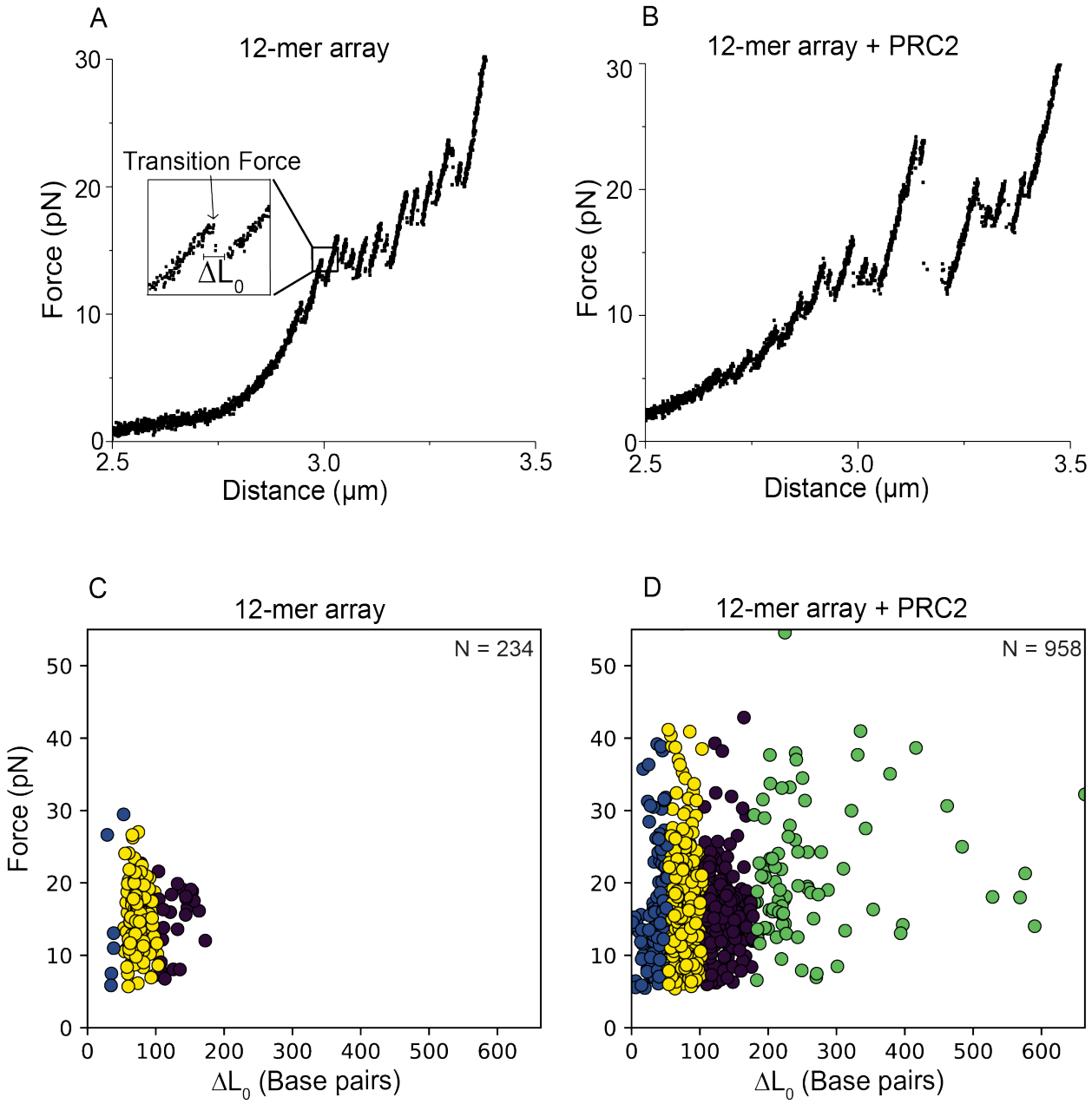


Figure 2.3. Example single-molecule force spectroscopy data. (**A**) A representative force-extension curve for a 12-mer nucleosome array. (*Inset*) Zoomed-in view of a force-induced transition. ΔL_0 represents the contour length change associated with the transition. (**B**) A representative force-extensive curve for a 12-mer array incubated with 500 nM PRC2. (**C**) Distribution of ΔL_0 (x-axis) and transition force (y-axis) for transitions found in the force-extension curves of 12-mer arrays. Each data point corresponds to one transition. (**D**) Distribution of ΔL_0 and transition force for PRC2-bound 12-mer arrays. Transitions are categorized into four distinct clusters based on their ΔL_0 values (blue: $\Delta L_0 < 51$ bp; yellow: 51 bp $< \Delta L_0 < 104$ bp; purple: 104 bp $< \Delta L_0 < 179$ bp; green: $\Delta L_0 > 179$ bp). Adapted from Leicher et al. [96].

D. Single-molecule data analysis

Force-extension trajectories are processed using a custom 'Force-Extension Analyzer' software. Two parameters are extracted for each force-induced transition: the contour length change (ΔL_0) and the force at which the transition occurred.

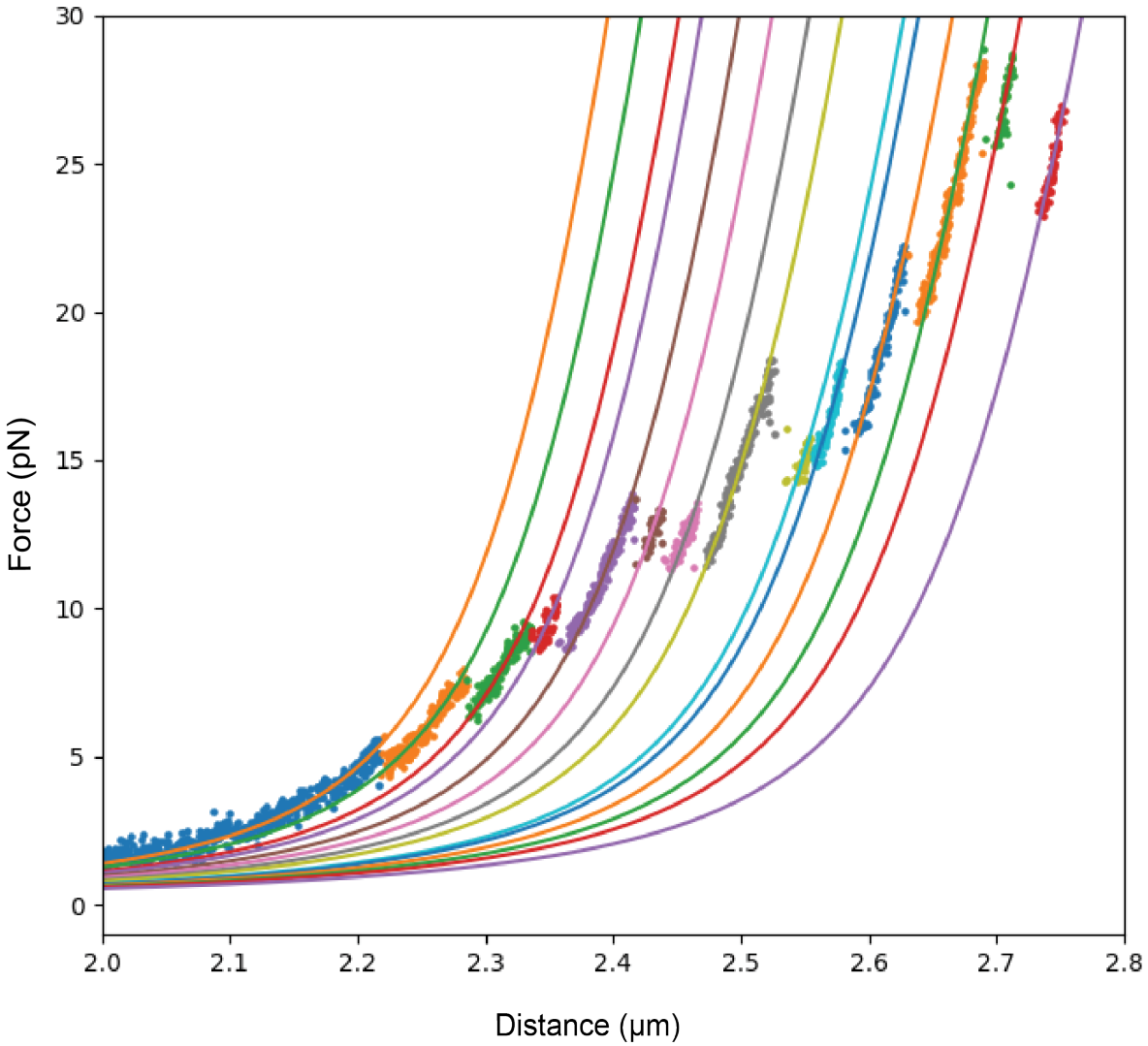


Figure 2.4. GUI of the Force-Extension Analyzer. Example trace from Force-Extension Analyzer for segmenting and fitting individual transitions from a single 12-nucleosome array tether.

D1: Force-extension analysis

1. Screen for single tethers based on signatures in the force-extension curve (see **Note 6**).
2. Load a force-extension trace (.asc or .csv file) into the Force-Extension Analyzer GUI.
3. Divide each trace into segments separated by force-induced transitions (**Figure 2.4**) (see **Notes 7 and 8**).

4. Fit an extensible worm-like-chain model [101] ($F = \frac{k_B T}{L_p}\left[\frac{1}{4(1-x/L_0+F/K_0)^2} - \frac{1}{4} + \frac{x}{L_0} - \frac{F}{K_0}\right]$) to each segment of the unfiltered force-extension data and obtain a characteristic contour length value (L_0) (see **Notes 9 and 10**). Convert the contour length change (ΔL_0) in distance to the number of DNA base pairs released upon transition using known DNA elasticity [102].
5. Export relevant parameters (number of segments, transition force, ΔL_0) for each trace. Plot traces in Origin or Prism.

D2: Cluster assignment

Employ a Gaussian mixture model (GMM) to group all observed transitions into distinct clusters based on their ΔL_0 values.

1. Generate putative *M*-component GMMs with an increasing number of components (e.g. $M = 1—10$) to fit the experimental data.
2. Assess each model with its Bayesian information criterion (BIC) score.
3. Perform bootstrapping to evaluate the uncertainties of the BIC scores, i.e. samples are randomly drawn with replacement from the original dataset to form a new dataset. This process is repeated many times (e.g. 100 times) to compute 95% confidence intervals for the BIC scores.
4. Choose the most parsimonious model for subsequent cluster assignment (see **Note 11**). Assign each transition to its cluster of maximum likelihood based on the GMM.

2.1.3 Notes

1. DNA products of all digestion and extraction steps can be evaluated on a 0.5% agarose gel.
2. Octamers can be prepared with mutated or posttranslationally modified histones. Mutant histones can be cloned using site-directed mutagenesis. Posttranslationally modified histones can be obtained using solid-phase peptide synthesis and native chemical ligation [103] or from commercial sources. Heterogeneous nucleosome arrays can be created using a mixture of canonical, modified, and mutant octamers.
3. The octamer:DNA ratio for making 12-mer arrays was 3:1. Octamer should be at a final concentration of ~1 μM for 12-mer arrays for a 50-μL sample volume. The ratio for making 4-mer and 2-mer arrays was 1.3:1 and 1.15:1, respectively. Because of fluctuations in the octamer and DNA purity, it is useful to try a series of ratios to avoid over- or under-loading of nucleosomes.
4. Avoid centrifugation and magnesium precipitation of the chromatin substrate before single-molecule experiments, as those procedures cause aggregation and clumps of the nucleosome arrays. It is best to use the microfluidic chamber to separate individual nucleosome array tethers from excess components by dragging the tether to a buffer channel.
5. Nucleosome array tethers were formed between two beads under flow to reduce the changes of sticking. To form a tether, rub the beads in the y-direction, then slowly separate the beads in the x-direction. Do not apply more than 10 pN of force during this process, or the inner-wrap DNA of the nucleosomes will be disrupted, causing an irreversible and ejection of histones.

6. A single tether features an initial force increase at the correct extension in the force-extension curve (e.g. 5 pN at ~2.7 μm for our 10-kb tether), as well as a single-step tether breakage. Exclude unqualified tethers from subsequent analysis.
7. Transitions are identified by applying a Butterworth low-pass filter to the trace and assigning transitions when the force reduction exceeds a certain threshold (e.g. 0.2 pN). This threshold is specified depending on the noise level of the force data. A rule of thumb is that a reasonable threshold can capture virtually all transitions discerned by visual inspection without yielding false transitions from the negative control (e.g. pulling on bare DNA).
8. Sometimes fewer transitions were observed when the tether was pulled again after relaxation, indicating nucleosome dissociation during pulling, which could change the spacing of the array and the interaction pattern of the chromatin-associated complex. For this reason, it is recommended that only the force-extension data from the first pull of each tether be analyzed. It took longer to build sufficient statistics with this protocol, but helps eliminate ambiguities in data interpretation.
9. While it is in principle possible to perform global fitting over the entire force-extension curve to obtain the persistence length (L_p), elastic modulus (K_0), and the contour length (L_0) of each segment, such a procedure usually yields non-convergent fits due to too many free parameters. Therefore, for each trace, L_p is determined by fitting the Marko-Siggia worm-like-chain model to the first segment in the low-force regime [104]. Keeping this L_p value fixed, K_0 is then determined by fitting the last segment, in the high-force regime, to the extensible worm-like-chain model. Using these L_p and K_0 values, each segment is then fit to the extensible worm-like-chain model to determine its L_0. Alternatively, one could apply a pre-chosen L_p and K_0 values and perform global fitting to obtain the L_0
10. If the segmentation is not good, the 'smooth factor' and 'prominence' can be adjusted. If the first segment does not fit correctly, and the determined L_p and K_0 poorly fit the other curves, the first segment can be temporarily excluded using the 'Exclude left' field in GUI. Once properly measured values of L_p and K_0 are obtained, these values may be manually entered into the GUI. The code and tutorial of the Force-Extension Analyzer software is available at https://github.com/alushinlab/ForceExtensionAnalyzer.
11. In general the model with the best BIC score should be chosen. If the BIC score difference between models is statistically insignificant, the model with the fewest component should be chosen as it constitutes the most parsimonious framework to describe the data. Independent knowledge or analysis specific to the system of interest could also aid in assessing the robustness of the clustering. For example, a four-component model was selected to perform the cluster assignment of transitions observed in the force-extension curves of PRC2-bound 12-mer nucleosome arrays (**Figure 2.3**). This selection was validated by force-extension data for other conditions such as 12-mer without PRC2, PRC2-bound bare DNA, and PRC2-bound dinucleosomes data [96].

2.2 Single-molecule adaptation of correlative force-fluorescence spectroscopy for studying condensates

Phase separation of biological substrates has been studied by a variety of techniques [78, 105] and bioinformatic tools have been developed to identify the molecular signatures of proteins with propensity for LLPS. Experimentally, the presence of droplets can be detected visually by light microscopy (brightfield, DIC, etc.), ensuring properly formed droplets, and not

aggregated protein molecules. Droplet formation can also be measured using turbidity readings, as mesoscale assemblies from 10s to 100s of nanometers in diameter (when present in solution) can scatter visible light at specific wavelengths (i.e. 350nm). This technique can determine concentration dependence of droplet formation. Centrifugation can also be used to separate the light and dense phases of a phase-separated solution. For these *in vitro* experiments, it is important to maintain physiological salt concentrations and avoid aggregation of protein and nucleotide molecules. Often, molecular crowding agents, such as poly-ethylene glycol (PEG), are added to mimic the crowded environment of the cell.

The material properties of phase separated droplets can vary substantially from liquid to solid [106], determined by their molecular components. Fluorescence recovery after photobleaching (FRAP) and droplet fusion assays are widely used to determine the liquidity of droplets [106]. FRAP can measure the mobility of individual components of the droplets while fusion can determine if the droplet has any gel-like properties. Both of these techniques are commonly used *in vitro*, but also *in vivo* to determine phase separation capabilities in cells by fluorescently labeling components.

While these techniques are a good foundation for studying LLPS, they fail to fully characterize the biophysical properties of the droplets and the formation of condensates at the nanoscale. Recently, the development of techniques to further characterize condensate formation and the material properties of phase separated droplets has been driven by biophysicists. Real time observation of protein and DNA condensation can be visualized using fluorescent microscopy [107]. Optical trapping technology can also be utilized to capture phase separated droplets due to their dielectric nature and small size (0-5μm), which allows precise manipulation and measurement of individual condensates [108-111]. Based on these developments, we engineered two techniques that harness optical trapping technologies; (1) the first, to observe condensate formation on nucleic acids at the single-molecule level, and (2) the second, to observe material properties of liquid-liquid separated droplets. Both of these techniques can serve as platforms to investigate the phase separation properties of a variety of proteins and nucleic acids in the cell.

2.2.1 Materials

A. Preparation of biotinylated lambda DNA

1. Lambda DNA (New England Biolabs)
2. Klenow 3'-5' exo-polymerase (New England Biolabs)
3. 10X NEB2 Buffer (New England Biolabs)
4. 1mM dGTP (Thermo Fisher Scientific)
5. 1mM biotin-11-dUTP (Thermo Fisher Scientific)
6. 0.4mM biotin-14-dCTP (Thermo Fisher Scientific)
7. 1mM dATP (Thermo Fisher Scientific)
8. Water
9. Thermocycler (Eppendorf)
10. Ethanol
11. Sodium acetate (NaOAc) solution (3 M, pH 5.2)
12. TE buffer (10 mM Tris-HCl pH 8, 1 mM EDTA)

B. Preparation of labeled oligonucleotides

1. 0.5 M sodium chloride (Fisher scientific)
2. Cy5 Mono-reactive dye pack (GE)
3. MicroSpin G-25 Columns (GE)
4. 0.1 M sodium bicarbonate pH 8.5 (Fisher Scientific)
5. Ethanol
6. Sodium acetate (NaOAc) solution (3 M, pH 5.2)
7. DMSO (Thermo Fisher Scientific)
8. Oligonucleotide (Integrated DNA Technologies) with 5' amino C6 tag
9. Speed vacuum
10. Nanodrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific)

Table 2.2. Single-stranded DNA oligonucleotide for fluorescent labeling

$ssDNA_{30}$	/5AmMC6/GCATGTATTGAAGTACTTCTAGAGGCCGCC

C. Single-molecule condensate formation assay

1. Imaging buffer (10 mM Tris-HCl pH 8.0, 200 mM KCl, 0.5 mM $MgCl_2$, 0.1 mM EDTA, 0.1% Tween)
2. 3.23-μm streptavidin-coated polystyrene beads (Spherotech)
3. 1× phosphate buffered saline (PBS), pH 7.4
4. Biotinylated lambda DNA
5. Labeled protein of choice (e.g. Cy3-H1.4)
6. Dual-trap optical tweezers (LUMICKS C-Trap)

D. Single-molecule droplet manipulation assays

D1: Prepare homemade slide containing droplet solution

1. VistaVision Microscope Slides (VWR)
2. Micro Cover Glasses, Rectangular, L x W = 24x30 mm (VWR)
3. Clear Nail Polish
4. Double Sided Tape
5. Protein of interest (e.g. 25 μM H1.4)
6. Nucleotide (e.g. 100 μM $ssDNA_{30}$, $ssDNA_{75}$)
7. Fluorescent nucleotide (e.g. 73 μM Cy5-$ssDNA_{30}$)
8. 10X H1 buffer (10 mM Tris-HCl pH 7.5, 200 mM NaCl)
9. 40% poly-ethylene glycol MN 8000 (PEG8000) (Sigma-aldrich)

Table 2.3. Single-stranded DNA oligonucleotides

$ssDNA_{30}$	GCATGTATTGAAGTACTTCTAGAGGCCGCC
$ssDNA_{75}$	CCAAAACTAATCTTTATATAAATGGGAGACTCACTATCATGGCTACTACGA CTACTAAACGAGGTTAAAATGAAT

D2: Fluorescent droplet manipulation and imaging

1. Dual-trap optical tweezers with confocal scanning (LUMICKS C-Trap)

D3: Droplet fusion assay

1. Dual-trap optical tweezers with confocal scanning (LUMICKS C-Trap)

D4: Droplet fluorescence recovery after photobleaching (FRAP) assay

1. Dual-trap optical tweezers with confocal scanning (LUMICKS C-Trap)
2. FRAP Droplet Imaging script. (Directions available on the Lumicks Harbor platform https://harbor.lumicks.com/single-script/3a796fac-dbb3-4fe1-8ce7-8b0cf8c25ad9).

D5: Droplet image analysis

1. C-Trap .h5 File Visualization GUI software. (Detailed instructions available on Harbor (https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f). This software was developed by rotation student John Watters.

D6: Droplet fusion force analysis. Python script fusion_force_log_fitting.py These scripts were developed by rotation student John Watters.

D7: Droplet FRAP analysis

1. FIJI (Fiji also known as ImageJ) with FRAP profiler plugin installed. (Hardin lab http://worms.zoology.wisc.edu/research/4d/4d.html).

2.2.2 Methods

A. Preparation of biotinylated lambda DNA

1. For a 50 μL reaction, mix 31.5 μL 500 ng/μL lambda DNA, 5 μL 10X NEB2 buffer, 1.65 μL 1 mM dATP, 1.65 μL 1 mM dGTP, 1.65 μL 1 mM biotin-11-dUTP, 4.13 μL 0.4 mM biotin-14-dCTP, 1 μL 5units/ μL Klenow 3'5' exo-polymerase, and 3.42 μL water.
2. Run reaction in thermocycler, 45 minutes at 37˚C, 20 minutes at 75˚C, and then cool to 4˚C.
3. Ethanol precipitate by adding the 50 μL reaction to 5 μL 3M NaOAc pH 5.2 and 125 μL cold ethanol. Incubate overnight at -20˚C.
4. To recover DNA, spin 30 minutes at 15,000 rpm, remover supernatant, wash with 1 mL cold 70% ethanol, spin 15K rpm for 15 minutes, remove supernatant, and resuspend pellet in 100 μL TE buffer.

B. Preparation of labeled oligonucleotides

1. Dissolve oligonucleotide (IDT) in 0.5M sodium chloride to a final concentration of 100 μM.
2. Vortex G-25 spin column and then spin at 700 x g for 1 minute to remove resin liquid. Place column in new Eppendorf tube and add 100 μL (10 nmole) of your oligo. Elute for 2 minutes at 700 x g.
3. Speed vacuum the samples to remove any liquid (usually ~ 5 hours).

4. Dissolve dried oligonucleotide in 40 μL 0.1 sodium bicarbonate pH 8.5.
5. Dissolve mono-reactive dye pack in 10 μL DMSO.
6. Add 5 μL dye to dissolved oligo and incubate in the dark for 2-4 hours. Mix tube every 30 minutes.
7. Run reaction through G-25 column again to remove free dye as described in step 2.
8. Ethanol precipitate with 4.5 μL 3M NaOAc pH 5.2 and 120 μL cold ethanol. Incubate overnight at -20°C.
9. To recover DNA, spin 10 minutes at 15,000 rpm, remove supernatant, wash with 200 μL cold 75% ethanol, spin 15K rpm for 10 minutes, remove supernatant, and resuspend pellet in 30 μL water.
9. Take optical density (OD) using a Nanodrop 2000 UV-Vis spectrophotometer at 260 nm and 650 nm (for Cy5 dye, or 532 nm for Cy3 dye).
10. Use the following equations to calculate labeling percentage and oligonucleotide concentration:
    $[DNA_{Cy5}] = [A_{260}-(0.05*A_{650})] / \varepsilon_{260}$
    $[Cy5] = A_{650}/250{,}000$
    $[DNA_{Cy3}] = [A_{260}-(0.08*A_{532})]/\varepsilon_{260}$
    $[Cy3] = A_{532}/150{,}000$

C. Single-molecule condensate formation assay

We engineered a platform to observe condensate formation on single-DNA tethers. To do this, we captured a single lambda DNA molecule between two beads that are positioned in optical traps. We then applied tension to the tether by moving the right trap at a constant velocity in the X-direction while simultaneously scanning with a confocal laser. In our experiments, we were interested in observing H1 condensate formation on single-stranded DNA. When stretching DNA to forces above 60 piconewtons, we observe force-induced DNA melting [48], which leads to one single strand of DNA under tension, and one relaxed strand. In our assay, we observe condensate formation of Cy3-H1.4 on the relaxed single strand of DNA. See **Figure 2.2** for single-molecule assay setup.

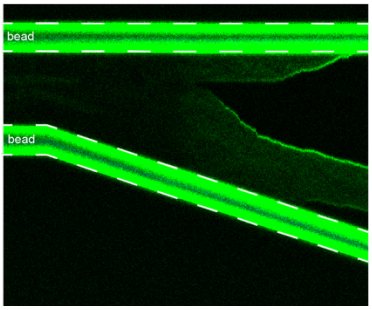


Figure 2.5. Example kymograph of condensate formation. Kymograph of Cy3-H1.4 condensate formation on single-stranded DNA.

C1: Sample and instrument preparation

1. Channel 1 solution. Dilute 4 μL of 3.23-μm streptavidin-coated beads in 1 mL of PBS buffer. Vortex.
2. Channel 2 solution. Add biotinylated-lambda to 1 mL PBS buffer.
3. Channel 3 solution: 1 mL of PBS buffer.

4. Channel 4 solution: Add protein of choice (e.g. 15nM Cy3-H1.4) to 200 μL of Imaging buffer.
5. Load the solutions into syringes connected to the corresponding channels. Flow 100 μL of solution through each channel.
6. Turn on the trapping laser. Catch one bead in each trap in channel 1.
7. Move traps to channel 3 and turn off flow.
8. Calibrate the traps according to the manufacturer's or in-house protocol.
9. Turn on the confocal scanning laser (e.g. 532nm green laser) at 10% power.

C2: Recording kymographs of fluorescent DNA binding and condensate formation.

1. Capture one bead in each optical trap in channel 1.
2. Move traps to channel 2. Bring beads apart and together to catch a DNA tether using laminar flow.
3. Move to channel 3 to check that one tether is caught. Relax DNA.
4. Start recording a kymograph and move tether to protein channel 4. Stretch DNA using a force-distance curve at a constant velocity of 0.1 μm/s. Observe any protein binding and condensate formation. Continue stretching tether to at least 60 pN to observe single-stranded DNA binding.

D. Single-molecule droplet manipulation assays

Combined force-fluorescence microscopy can be utilized to trap liquid droplets for manipulation and imaging. A variety of different recipes for droplet imaging can be used, from different buffers, molecular crowders, nucleic acids, and proteins. The protocol for determining phase separation conditions of H1 and nucleic acids can be found in materials and methods (Chapter 7). Here, we will detail one of those specific conditions with a fluorescently labeled 30 base oligonucleotide ($ssDNA_{30}$). Any fluorescently labeled component (either protein or nucleic acid) was diluted to 10% of its final concentration in combination with the unlabeled component to avoid aggregation caused by dye hydrophobicity. Our droplet trapping assay has multiple advantages to typical droplet imaging. First, droplets captured in optical traps can be manipulated with fine precision. This allows us to step the traps at regular intervals and image droplets that are fixed in space. Second, in the droplet fusion assay, we can simultaneously detect force exerted by the droplets. This force can be tracked temporally (**Figure 2.7**) and allows us to calculate average droplet fusion time (**Figure 2.8**).

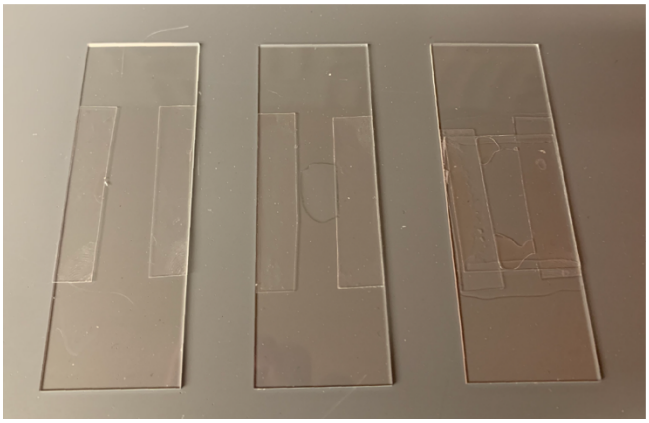


Figure 2.6. Home slide preparation. The slides were prepared in the following steps: double-sided tape application, sample application, coverslip addition, followed by sealing with nail polish to prevent evaporation.

D1: Prepare homemade slide containing droplet solution

1. Prepare a 20 μL droplet sample. (This will vary depending on your components). For our studies we combined 2 μL 25μM H1.4, 1.8 μL 100 μM ssDNA$_{30}$, 5 μL 40% PEG8000, 1.6 μL 10X H1 buffer, and 9.3 μL water. Pipet thoroughly to mix.
2. Place two small strips of double-sided tape in the center of a glass slide on each edge, leaving room in the center for a sample.
3. Pipet the 20 μL sample in the center of the tape strips. Place cover slip over the top, seal edges with nail polish, and allow to dry.

D2: Fluorescent droplet manipulation and imaging

1. Follow standard protocols for aligning trapping lasers on the Lumicks C-Trap.
2. Adjust condenser by eye using the LED brightfield laser, slowly lowering it until droplets are visible.
3. Turn trapping laser power down to 5% or lower and capture droplets (high laser power can damage them). Record brightfield images or confocal scanning images. Droplets can be captured and moved by manipulating the traps around the homemade flow cell. The z-position of the lasers or stage may need to be adjusted for optimal imaging conditions.

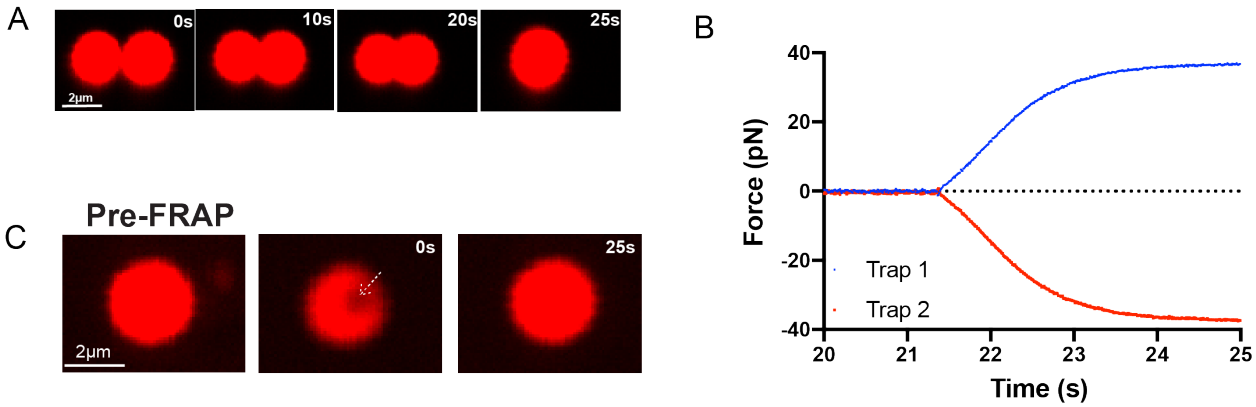


Figure 2.7. Droplet fusion and FRAP. (**A**) Example images of droplet stepping and fusion over time. (**B**) Example trace of force acquisition during fusion assay. Force from each trap is recorded and increases in equal and opposite directions upon fusion until reaching a steady state. (**C**) Example droplet before and after bleaching during a FRAP experiment. White arrow indicates bleached region.

D3: Droplet fusion assay

1. Position one droplet in each of the dual-optical traps.
2. Zero the force.
3. Record continuous confocal images at regular time intervals (e.g. every 5 seconds).
4. After each image manually step one trap towards the other trap at regularly spaced intervals (e.g. 200 μm). Proceed until droplets fuse.

D4: Droplet fluorescence recovery after photobleaching (FRAP) assay

1. Capture one larger droplet in an optical trap. Make sure there are no other droplets nearby that may fuse during the FRAP experiment.
2. Run the following script: FRAP Droplet Imaging. The script first takes one confocal scan, photobleaches a small spot with a point scan, and then takes confocal scans over regular intervals for 25 seconds. More details about the script and how to change specific parameters (laser power, laser wavelength, bleaching time) are described on Harbor.

D5: Droplet image analysis

1. Export droplet images as .h5 files from Lumicks Bluelake acquisition software.
2. Load .h5 files in the C-Trap .h5 File Visualization GUI. Directions to run the script are available at harbor. Export files for Image J. Export fusion force traces.

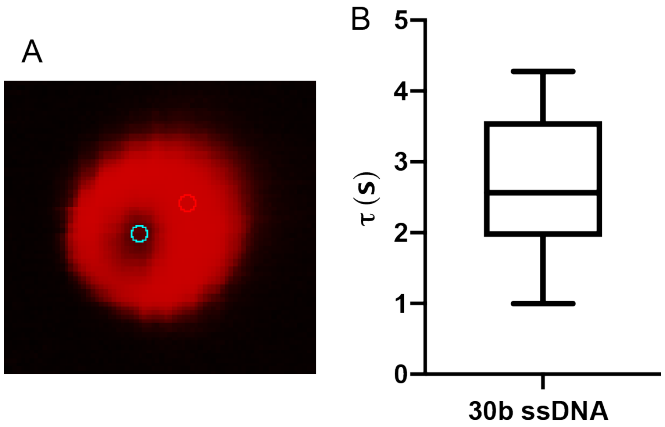


Figure 2.8. Droplet fusion and FRAP analysis. (**A**) Example selection of regions of interest (ROIs) before using the FRAP profiler plugin on ImageJ/FIJI. (**B**) Average fusion time for droplets formed with ssDNA and H1.

D6: Droplet fusion force analysis

1. Run fusion_force_log_fitting.py. (See Chapter 7 for more details).

D7: Droplet FRAP analysis

1. Place all exported image files from one droplet FRAP acquisition from the C-Trap .h5 File Visualization GUI in one folder.
2. Drag the folder into FIJI and import as a stack.
3. Open the ROI manager and on the second image (the first image taken after photobleaching) create 2 circular ROIs. Draw one at the center of the bleached spot and one at another unbleached portion of the droplet. Use the specify function to create a uniform circular shape and size of 9.2 (equal to the 800 nm point scan of the laser on the C-Trap).
4. Run the FRAP profiler plugin.

CHAPTER 3. Single-molecule and in silico dissection of the interaction between Polycomb repressive complex 2 and chromatin

3.1 PRC2: An introduction

Polycomb repressive complex 2 (PRC2) is a major epigenetic machinery that maintains transcriptionally silent heterochromatin in the nucleus and plays critical roles in embryonic development and oncogenesis. It is generally thought that PRC2 propagates repressive histone marks by modifying neighboring nucleosomes in strictly linear progression. However, the behavior of PRC2 on native-like chromatin substrates remains incompletely characterized, making the precise mechanism of PRC2-mediated heterochromatin maintenance elusive. Our understanding of this process was limited by the resolution of structural techniques that could identify PRC2-binding modes on long chromatin substrates. Therefore, we designed a single-molecule platform to interrogate PRC2 engagement of nucleosome arrays.

3.1.1 PRC2 components and mechanism

Polycomb (Pc) group proteins [112] were originally discovered in *Drosophila* by mutants that displayed improper body segmentation where they acted as a negative regulator of homeotic genes required for proper segmentation [113]. Activities of these PcG proteins were regulated by antagonistic properties of the trithorax proteins, leading to proper maintenance of homeotic gene expression throughout *Drosophila* development [114]. Several families of PcG proteins are found across different organisms including the PRC1, PRC2, PhoRC, and PR-DUB complexes, each comprising different core components and having different mechanisms of action [115, 116]. Evidence of Polycomb group protein targeting to specific Polycomb response elements (PREs) has been identified in *Drosophila [117]*, but presence of these sites in mammalian cells has yet to be identified, leading to open questions about PcG recruitment.

In particular, PRC2 is a "reader-writer" complex that installs the hallmark heterochromatin mark H3K27me3. PRC2 is comprised of four major core proteins; EZH1/2, EED, SUZ12, and RbAp46/48 [118]. EZH1/2 is the methyltransferase writer harboring a SET domain, which catalyzes the addition of methyl groups onto lysine 27 of the H3 tail [27]. EED, is the recognition reader, comprised of WD-40 repeats folded into a seven-bladed β-propeller domain, which recognizes pre-existing H3K27me3 marks [119]. SUZ12 and RbAp46/48 complete the PRC2 complex aiding in conformational change and DNA binding, respectively. PRC2 propagates the hallmark repressive chromatin mark H3K27me3 through a "read-and-write" allosteric mechanism by binding pre-existing H3K27me3 marks with EED and methylating nascent H3K27 substrates with EZH2 [120].

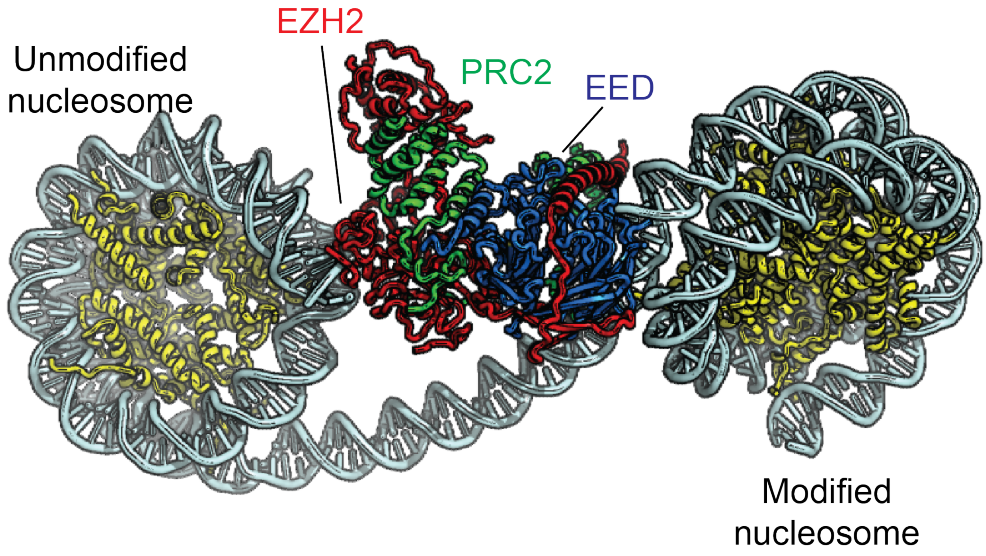


Figure 3.1. Structure of PRC2 bridging a H3K27me3-modified and unmodified dinucleosome. EED binds the nucleosome with the H3K27me3-modified H3 tail while EZH2 is positioned for catalytic activity on the unmodified nucleosome. Adapted from Poepsel et al. [121]

Structural studies of the PRC2 complex [122-124] have indicated ordering of the SRM helix to facilitate allosteric activity. Recent structural models of PRC2 bound to a dinucleosome show the ability for EED to engage with a methylated nucleosome, while EZH2 is directly positioned for methylation on the neighboring unmodified nucleosome [121] (**Figure 3.1**). PRC2 also engages the H3 tail directly through the ordering of the EZH2 bridge helix [125]. Biochemically, PRC2 shows a high binding affinity for linker DNA and longer chromatin substrates, compared to a relatively low affinity for mononucleosomes [126]. More nuanced regulation of PRC2 binding was not detected by bulk biochemistry assays.

Alternative compositions of PRC2 with EZH1 or EHZ2 regulate the specific activity of the complex *in vitro* and *in vivo* [127]. Specifically, EZH2 is the dominant methyltransferase and shows altered activity in the presence of different nucleosome repeat lengths, enhancing the activity presumably when PRC2 is positioned between a stimulatory dinucleosome pair. Alternatively, EZH1 showed a higher sensitivity to the presence of naked DNA but a lower response rate to EED-binding or SAM-based inhibitors. EZH1 is also found to dimerize PRC2 molecules, changing nucleosome binding conformations and promoting chromatin compaction. A key difference between EZH1 and EZH2 is the SANT1L loop of EZH1, which enhances nucleosome binding and catalytic activity [128]. Given these important findings, substrate and complex composition of PRC2 must have a significant role in regulating its activity in the cell.

Multiple accessory proteins have been found to bind the PRC2 complex and modulate its function including AEBP2, JARID2, and the PCLs. AEBP2 is a zinc finger protein often found in complex with PRC2 [129] that co-localizes with it at target sites [130]. AEBP2 stimulates the activity of PRC2 through a mechanism independent of and additive to allosteric regulation, where the KR motif is responsible for enhanced nucleosome binding and PRC2 stimulation. AEBP2 regulates DNA binding of PRC2 with a tighter binding affinity overall and specificity to

mCpG methylated DNA [126]. JARID2 is a regulator of PRC2 activity and has a large gene target overlap with the complex [131]. JARID2 regulates the binding of PRC2 in embryonic stem cells [132] and stimulates H3K27me3 deposition during cell differentiation through methylation of K116 on JARID2 by PRC2 itself [133]. In-depth biochemical work shows how JARID2 increases the rate of PRC2 H3K27 methylation, and its phosphorylation can partially regulate its ability to stimulate PRC2 activity [98]. JARID2 and AEBP2 bound PRC2 can also bind H2AK119ub1-modified nucleosomes (installed by PRC1), showing a modest increase in methyltransferase activity, and suggesting an avenue for PRC2 recruitment [125]. PCL1, PCL2, and PCL3 are the mammalian homologs of the *Drosophila* Polycomb-like proteins, and share the following protein motifs: a tudor domain, two plant homeodomains, and a PCL extended domain [134]. PCL interacts with PRC2 through the EZH2 and SUZ12 subunits, enhancing the activity of the complex for H3K27 tri-methylation activity in *Drosophila* [135].

Extensive work has been done to characterize the allosteric mechanism of PRC2 H3K27me3 deposition as well as binding of chromatin substrates. Work by the Cech group found PRC2 binds longer chromatin substrates with higher affinity [126], a potential mechanism for why dense chromatin activates PRC2 methylation *in vivo* [136]. They also demonstrated that RNA disrupted PRC2-nucleosome binding, hence affecting PRC2 catalytic activity. They identified that histone mutations have only small effects on PRC2 affinity and that histone-free linker DNA dictates PRC2 binding, due to the high affinity of PRC2 for nucleic acids. These properties were not altered by addition of JARID2 or replacement with EZH1. In addition to DNA binding, PRC2 can bend and loop DNA in solution at the single-molecule level, which may lead to heterochromatin formation [137]. Despite the advances in PRC2 enzymology, it remains unclear how PRC2 propagates H3K27me3 on long chromatin substrates.

3.1.2 PRC2: methylation, localization, and recruitment.

H3K27me3 is present in a variety of multicellular organisms maintaining repression of important genes, including the Hox genes during somatic development, cell cycle genes during germline development in *Drosophila* [27, 138], and somatic and germ line cells in mammals [139]. H3K27-methylated histones transmit the memory of repression through several rounds of cell division across generations, which is then perpetuated by PRC2 methylation [140].

Methylation of H3K27 is processive from mono, to di, to tri-methylation [141, 142], yielding the stable mark of H3K27me3. Embryonic stem cells have a distribution of methylation states with roughly 50% of the H3 histone being dimethylated, 15% trimethylated and 15% monomethylated [142]. Histone marks contribute to gene expression by regulating the chromatin compaction state or recruiting other factors. The literature has suggested H3K27me3 operates by recruiting these other factors, such as PRC1 for chromatin compaction. PRC1 has also been thought to recruit PRC2, which regulates the activity of the PRC2 complex to aid in continual spreading of the mark. Overall, the enrichment of H3K27me3 correlates with gene silencing [143].

PRC2 has been found to specifically localize to targets of H3K27me3 deposition, including the *Hox* genes and sequences encoding for other developmental regulators [144, 145]. Specifics regarding PRC2 recruitment still remain unclear, given the lack of PREs in mammalian cells. Leaders in the field suggest a combination of pre-existing H3K27me3 marks, DNA binding interactions, CpG islands, and long non-coding RNAs may contribute to PRC2 occupancy [95].

Overall, PRC2 recruitment and mark spreading during development remain large unknowns in the chromatin field.

3.1.3 H3K27me3 in the context of epigenetics

In addition to H3K27me3, another hallmark heterochromatin mark, H3K9me3 has been well studied. H3K9me3 is usually found at constitutive heterochromatic regions, and is regulated by the writer, Suv39H1, and the reader HP1 [11, 12, 19, 146]. These proteins work in tandem to condense and silence chromatin. In particular, Suv39H1 follows a similar "read and write" mechanism to PRC2, where it is activated by preexisting marks, suggesting a paradigm for other epigenetic modifiers. In tandem, HP1 has been shown to compact chromatin through dynamic, multivalent interactions [147, 148] and nucleosome bridging [149] in studies using single-molecule techniques and bulk biochemistry.

In contrast, marks such as H3K27ac, H3K36me3, and H3K4me2/3 are characteristics of euchromatin. H3K27ac works antagonistically to H3K27me3, as a mark for active enhancers [150]. H3K36me3, a mark correlated with transcription elongation, has distinct localization from the repressive H3K27me3 marks and inhibits the catalytic activity of PRC2 [151, 152]. Opposingly, H3K4 methylation has been found to coexist within larger regions of H3K27me3 forming "bivalent chromatin," where genes silenced during development can be poised for activation in embryonic stem cells (ESCs) [139]. Further studies found that nucleosomes can be asymmetrically modified, with modifications for H3K4me3 and H3K36me3 present on H3K27me3 modified nucleosomes, albeit on different tails *in vivo* [153]. PRC2 activity is directly inhibited by symmetric active H3K4me3 and H3K26me2/3 marks [154] through an allosteric mechanism. The study suggested that while modified mononucleosomes did not contribute to changes in PRC2 binding affinity, these interactions might affect PRC2 on a larger chromatin scale. However, recent studies have found that the presence of JARID2 and AEPB2 reduce the inhibitory effects of H3K4me3 and H3K36me3 on PRC2 methyltransferase activity when bound to H2AK119ub1-modified nucleosomes (and stimulated by K116me3) [125]. Using our single-molecule chromatin set-up, we have been able to detect how more nuanced changes in methylation state effect chromatin binding by PRC2. This technique would be a powerful tool to further probe H3K4me3/H3K36me2/3 interaction with PRC2 on longer chromatin substrates.

3.1.4 Chromatin compaction and Polycomb group proteins

Another Polycomb group complex, PRC1, has been implicated in chromatin condensation. PRC1 mediates H2A ubiquitination (K119) and is recruited to maintain X-inactivation [155]. Positively charged regions of PRC1 have shown the ability to compact chromatin arrays *in vitro* using bulk assays and electron microscopy [156]. Specifically, the Ring1B component can compact chromatin and repress genes *in vivo* in ESCs [157], while altering the PRC1 compaction region can disrupt Polycomb-mediated axial patterning in mice [158]. Regulatory mechanisms between PRC1 and PRC2 still remain unclear, but it is proposed that H3K27me3 could act as a docking site to recruit PRC1 [159, 160]. Transcriptional repression via H3K27 *in vivo* leads to the accumulation of PRC1 at H3K27me sites [161]. However, due to the variety of compositions of the PRC1 complex, its role is not fully understood. Overall, PRC1 most likely contributes to silencing through localization at specific genes and through the H2A ubiquitylation pathway [162].

A direct role for PRC2 involvement in chromatin compaction has not been identified as clearly as PRC1. Knockdown of PRC2 components in stem cells did not show a clear loss of chromatin compaction [163]. In the chromatin field, it is often debated whether epigenetic marks are a physical determinant of chromatin structure, or whether recruitment of subsequent factors leads to (de)condensation, such as with HP1. Throughout this thesis, we will address the local physical effects of these marks on chromatin structure in the absence and presence of different epigenetic modifiers.

3.1.5 PRC2: cancer and drug discovery

Various oncogenic mutations have been found in EZH2 leading to prostate cancer [164], breast cancer [165], melanoma and lymphoma [166, 167], in addition to the overall deregulation of PRC2 and H3K27me3 deposition in cancer cells. One groundbreaking study also identified the H3K27M mutation in PRC2's substrate causing pediatric glioblastoma [168]. The authors found a global reduction of H3K27me3 when only a small percentage of H3K27 harbored this methionine mutation, suggesting PRC2 is a global regulator that can be inhibited by the H3K27M mutation. Further biochemical work has characterized additional inhibitory behavior by bivalent H3K27M-H3K27me3 nucleosomes [169], but more work is necessary to characterize this effect fully.

Many first wave EZH2 inhibitors have been developed with limiting success leading to drug resistance. However, second wave EZH2 inhibitors from Constellation, Epizyme, GSK [170] and others have shown promise in clinical trials for inhibiting EZH2 overactivity through mechanisms such as prolonged residence time [171]. An alternative approach with allosteric inhibitors of EED showed promising success from AbbVie [172], Novartis [173], and others [174]. Currently, the focus in the field of PRC2 drug discovery has shifted to inhibiting protein-protein interactions between PRC2 and its substrate, H3K27me3 [175]. A future in drug discovery for targeting modifier-oncohistone specific interactions may show promise for more specialized on-target medicines.

3.1.6 PRC2 Research Outlook

Despite the extensive knowledge about the role of PRC2 in gene repression, the behavior of PRC2 on native-like chromatin substrates remains incompletely characterized, making the precise mechanism of PRC2-mediated heterochromatin maintenance elusive. More specifically, the interaction between PRC2 and polynucleosome arrays has thus far been refractory to structural interrogation due to the prohibitive conformational heterogeneity associated with the assembled complexes. As a result, the occurrence of various possible binding modes of PRC2 on long chromatin substrates and their relative frequency remain unclear. In particular, we question the theory that PRC2 modifies nucleosomes in a sequential manor and hypothesize an alternative mechanism where PRC2 may skip non-adjacent nucleosomes (**Figure 3.2**).

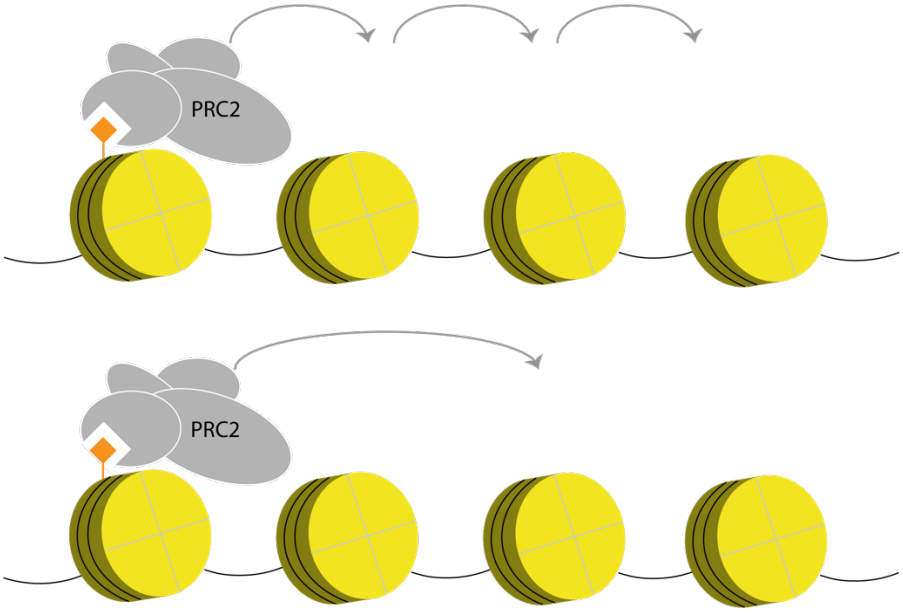


Figure 3.2. PRC2 chromatin spreading model. It is generally thought that PRC2 propagates repressive histone marks by modifying neighboring nucleosomes in strictly linear progression (**Top schematic**). However, the behavior of PRC2 on native-like chromatin substrates remains incompletely characterized, making the precise mechanism of PRC2-mediated heterochromatin maintenance elusive. We hypothesize an alternative mechanism where PRC2 may skip non-adjacent nucleosomes (**Bottom schematic**).

In this chapter, we will thoroughly discuss PRC2 engagement with polynucleosome substrates yielding the mechanical fingerprints of individual PRC2-bound chromatin complexes. The experimental and computational platforms developed here provide a framework for understanding the molecular basis of epigenetic maintenance mediated by Polycomb group proteins and can be extended to understand a whole host of epigenetic machinery.

3.2 Single-molecule and in silico findings of PRC2-chromatin engagement

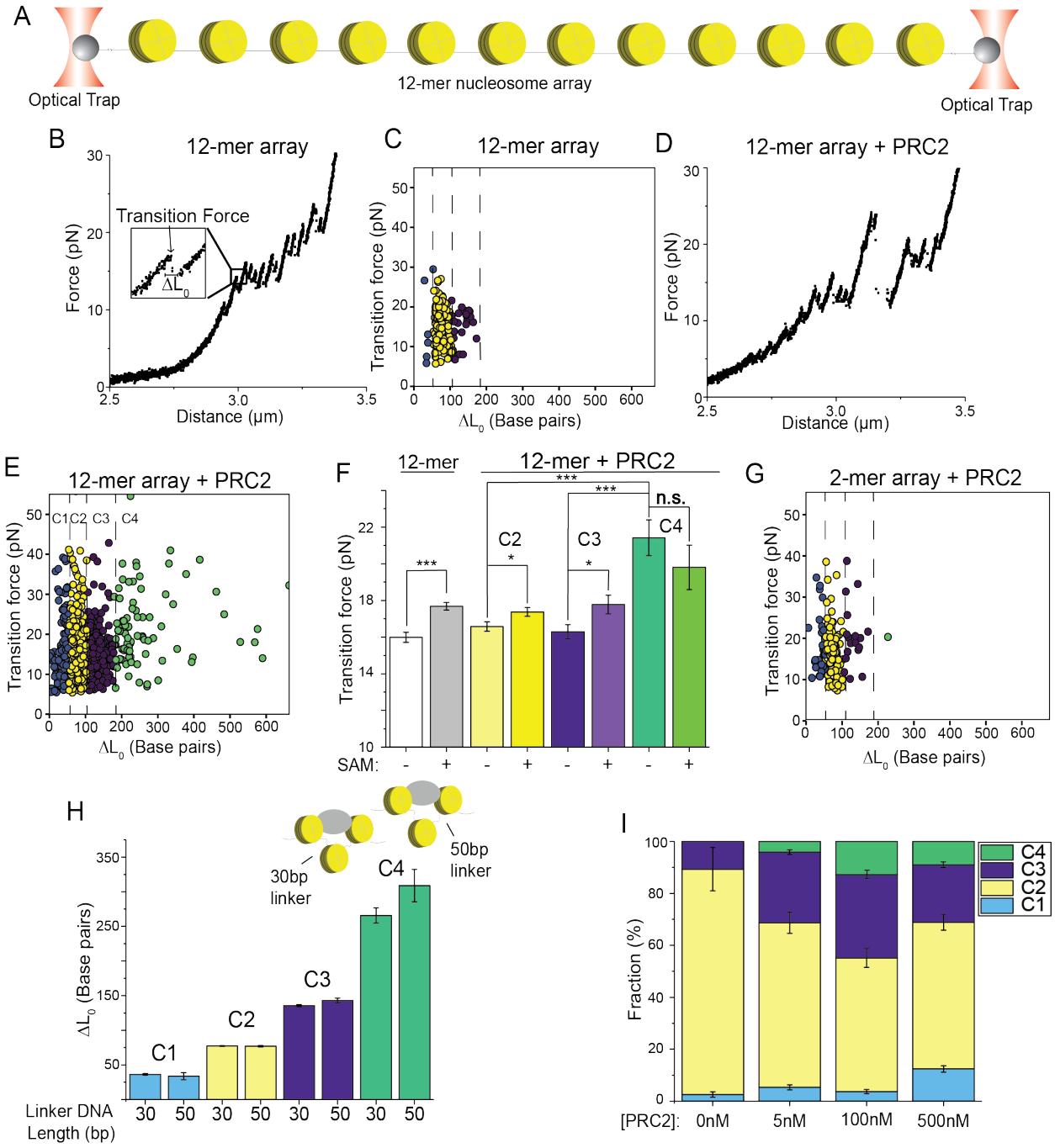
The results from our findings were reported in Leicher et al. [96] where we used single-molecule force spectroscopy and computational modeling to dissect the interactions between PRC2 and polynucleosome arrays. Our results provide direct evidence that PRC2 can simultaneously engage non-adjacent nucleosome pairs. The demonstration of PRC2's ability to bridge non-contiguous chromosomal segments furthers our understanding of how Polycomb complexes spread epigenetic modifications and compact chromatin.

Computational modeling was performed by Xingcheng Lin and Wenjun Xie from Bin Zhang's lab. PRC2 purification was performed by Eva J. Ge from Tom Muir's lab. Force extension, clustering algorithms and electron microscopy analysis was developed by Matthew J. Reynolds. Electron microscopy acquisition was supervised by Thomas Walz.

3.2.1 A single-molecule platform to dissect PRC2-chromatin interactions

We constructed a DNA template harboring twelve repeats of the "601" nucleosome positioning sequence, each separated by 30 base pairs (bp) of linker DNA. After reconstitution with histone octamers, individual nucleosome arrays were tethered and mechanically stretched on a dual-trap optical tweezers instrument (**Figure 3.2A**). Single-molecule force-extension curves displayed signature sawtooth patterns (**Figure 3.3B**) consistent with previous results [176], with each abrupt transition signifying the stochastic unwrapping of a single nucleosome. The average force at which these transitions occurred is 16.0 ± 0.3 pN (mean ± SEM; see **Table 8.1**) and the average number of transitions observed on each array is 12 ± 1, suggesting that octamers were sufficiently loaded onto the DNA template. The contour length change (ΔL_0) per transition, which reports the amount of DNA released upon force-induced disengagement, is 76 ± 1 bp (mean ± SEM), in good agreement with the unraveling of the inner DNA wrap around the histone octamer (**Figure 3.3C** and **Figure 8.1**). The standard deviation of ΔL_0 is 11 bp, which likely reflects the structural pliability of nucleosomes [177]. The unpeeling of the outer DNA wrap occurred in a gradual fashion at low forces (< 5 pN) under our experimental condition, consistent with previous observations [176, 178].

Figure 3.3. Single-molecule force spectroscopy dissects PRC2 binding modes on nucleosome arrays. (**A**) Schematic of the dual-trap optical tweezers setup (not drawn to scale). A 12-mer nucleosome array is tethered to a streptavidin-coated bead via a 4-kbp-long biotinylated DNA handle on each side. (**B**) A representative force-extension curve for a 12-mer array. (Inset) Zoomed-in view of a force-induced transition. ΔL_0 represents the contour length change associated with the transition. (**C**) Cluster analysis of individual transitions observed in the force extension curves of 12-mer arrays. N denotes the number of transitions. (**D**) A representative force-extension curve for a 12-mer array incubated with 500 nM PRC2. (**E**) Cluster analysis of transitions for PRC2-bound 12-mer arrays. (**F**) Effect of the PRC2 cofactor SAM on the average transition force for 12-mer arrays and for different transition clusters of PRC2-bound 12-mer arrays. (**G**) Cluster analysis of transitions found in PRC2-bound dinucleosome substrates. (**H**) Average ΔL0 of each transition cluster for PRC2-bound 12-mer arrays with a 30-bp linker DNA versus those with a 50-bp linker DNA. (**I**) Cluster distribution for 12-mer arrays incubated with PRC2 at different concentrations. Data are presented as mean ± SEM. P-values are determined by two-sample t-tests (n.s., not significant, $P \geq 0.05$; *$P < 0.05$; ***$P < 0.001$).


A
Optical Trap
12-mer nucleosome array
Optical Trap
B
12-mer array
Transition Force
ΔL0
Force (pN)
Distance (μm)
C
12-mer array
Transition force (pN)
ΔL0 (Base pairs)
D
12-mer array + PRC2
Force (pN)
Distance (μm)
E
12-mer array + PRC2
C1 C2 C3 C4
Transition force (pN)
ΔL0 (Base pairs)
F
12-mer
12-mer + PRC2
***
***
n.s.
C2
C3
C4
*
*
Transition force (pN)
SAM: - + - + - + - +
G
2-mer array + PRC2
Transition force (pN)
ΔL0 (Base pairs)
H
30bp linker
50bp linker
C4
C3
C2
C1
ΔL0 (Base pairs)
Linker DNA Length (bp)
30 50 30 50 30 50 30 50
I
C4
C3
C2
C1
Fraction (%)
[PRC2]: 0nM 5nM 100nM 500nM

Next, we incubated the 12-mer arrays in a relaxed state (i.e., under zero force) with PRC2 core complexes at a concentration of 500 nM (**Figure 8.2**). At this concentration, the majority of nucleosomes in the array is expected to be PRC2 bound [179]. In the meantime, some PRC2 complexes are expected to bind the free DNA as well [180]. The tethered enzyme–substrate assemblies were then subjected to single-molecule pulling. Force-extension trajectories for these assemblies exhibited transitions of diverse sizes (**Figure 3.3D**). Using a clustering algorithm based on the Bayesian information criterion (**Figure 8.3**), we categorized all transitions into four distinct clusters (**Figure 3.3E** and **Table 8.1**). The first cluster (C1) includes transitions with ΔL_0 smaller than 51 bp. We posited that PRC2 sequesters, perhaps bends, a stretch of free DNA located in the flanking or linker DNA regions of the nucleosome array, and that the C1 transitions are due to force-induced dissociation of this stretch of DNA from PRC2. This interpretation is supported by the force-extension curves obtained with bare DNA substrates incubated with PRC2, which exhibited exclusively these small transitions (**Figure 3.4**). A recent atomic force microscopy (AFM) study also reported DNA looping by PRC2 [180], which likely represents the same phenomenon as shown here. Our data provide additional insight that this interaction is stable, withstanding an average force of 17 pN.

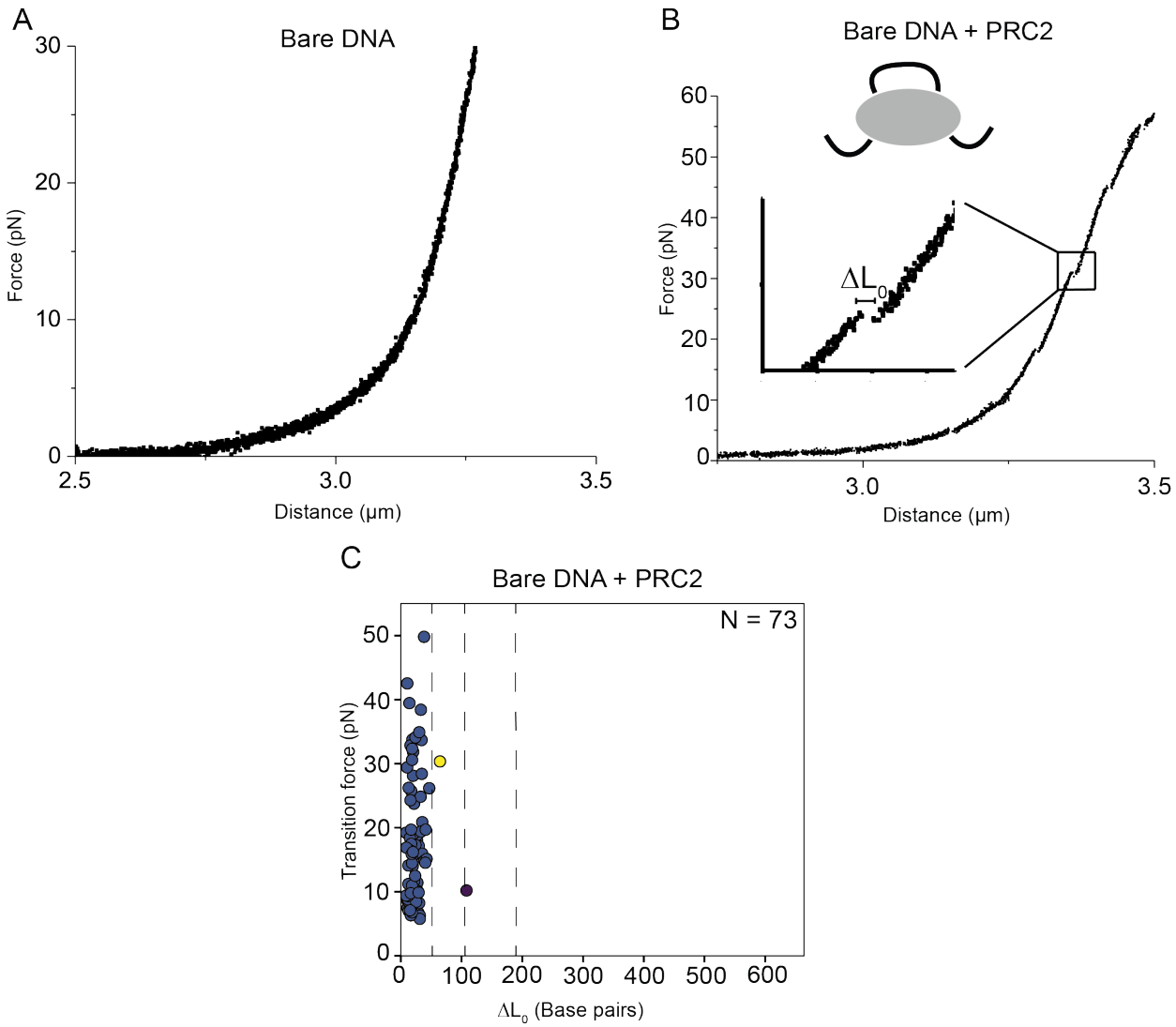


Figure 3.4. Interaction of PRC2 with bare DNA. (**A**) A representative force-extension curve for a piece of 10-kbp-long bare DNA tethered between two optically trapped beads. (**B**) A representative force-extension curve for the same type of DNA bound with PRC2 core complexes. (Inset) Zoomed-in view of one example transition signifying PRC2 disengagement. (**C**) Cluster analysis of all transitions observed in the force-extension curves of PRC2-bound bare DNA. The average size of transitions for bare DNA + PRC2 is 25 ± 2 bp, somewhat lower than that of C1 transitions for 12-mer + PRC2 (36 ± 2 bp). This difference is likely due to the crosstalk between C1 and C2 clusters for the 12-mer + PRC2 condition (see **Figure 8.3E**).

The second cluster (C2 in **Figure 3.3E**) contains transitions between 52 and 104 bp in size with an average ΔL_0 of 77 ± 1 bp, mimicking transitions observed in the absence of PRC2 that represent the unraveling of mononucleosomes (**Figure 3.3C**). The average transition force for C2 was not significantly altered by the presence of PRC2 (16.2 ± 0.3 pN without PRC2 vs. 16.6 ± 0.3 pN with PRC2). Nonetheless, given the high PRC2 concentration used in this experiment, the C2 transitions likely contained disassembly events of both PRC2-bound mononucleosomes and unbound nucleosomes.

The third cluster (C3 in **Figure 3.3E**) encompasses transitions between 105 and 179 bp in size. Besides the fraction that can be attributed to two mononucleosomes coincidentally unraveling at the same time (also seen in the absence of PRC2 but occurring with a lower frequency: 22% with PRC2 vs. 11% without PRC2), this cluster also corresponds to the binding mode in which PRC2 simultaneously engages two neighboring nucleosomes as seen in the PRC2:dinucleosome cryo-EM structure [181]. The rationale for this interpretation is as follows: disruption of the PRC2:dinucleosome complex would release approximately 180 bp of DNA—two outer DNA wraps (approximately 150 bp) plus 30 bp of linker DNA; alternatively, it would release between 105 and 180 bp of DNA if all or part of the outer wrap from one of the PRC2-interacting nucleosomes is not subjected to PRC2 sequestration, hence already undone at a low force (**Figure 8.4A**).

Importantly, we also observed a significant population of transitions (86 out of 958) with ΔL_0 greater than 180 bp (C4 in **Figure 3.3E**). This is in stark contrast to the transition distribution for 12-mer without PRC2, where no C4 transitions were identified (0 out of 234). The average rupture force for C4 transitions is significantly higher than those for C2 and C3 (16.6 ± 0.3 pN for C2, 16.3 ± 0.4 pN for C3, 21.4 ± 1.0 pN for C4; **Figure 3.3F** and **Table 8.1**), indicating that they represent a distinct PRC2 binding mode on chromatin. The large sizes of these transitions entail concurrent engagement of PRC2 with two distal sites. Specifically, if PRC2 bridges two nucleosomes that are separated by one spacer nucleosome ("$Nuc_{1\text{-}3}$" mode), disruption of this linkage would release between 210 and 285 bp of DNA (**Figure 8.4B**). If PRC2 bridges a pair of nucleosomes that are separated by two or more spacer nucleosomes ($Nuc_{1\text{-}4}$, $Nuc_{1\text{-}5}$, etc.), transitions larger than 300 bp would be expected (**Figure 8.4C**). Because the majority of C4 transitions (78%) were between 180 and 300 bp in size, we speculated that $Nuc_{1\text{-}3}$ is the preferred mode of PRC2–nucleosome interaction in this cluster.

Next, we examined the effect of S-adenosylmethionine (SAM)—the methyl donor for PRC2-catalyzed H3K27 methylation—on the transition forces. We found that SAM had a noticeable effect on the stability of nucleosome itself independent of PRC2, raising the average transition force from 16.0 ± 0.3 to 17.7 ± 0.2 pN (**Figure 3.3F**). This difference is markedly larger than the uncertainty of our force measurements associated with the instrumental noise and analysis algorithm (< 0.3 pN), hence representing a real change in the transition force. Therefore, the modest stabilization effect of the SAM cofactor on C2 and C3 transition forces in the presence of PRC2 most likely came directly from SAM (**Figure 3.3F**). In contrast, the C4 transition force was not significantly affected by SAM, reinforcing the notion that C4 represents a distinct class of PRC2–nucleosome interactions.

3.2.2 Validation of the ability of PRC2 to bridge non-adjacent nucleosome pairs

Besides the model in which PRC2 makes specific contacts with a pair of distal nucleosomes, there is an alternative model that could also explain the observed large transitions in C4. In this model, PRC2 engages with two distal segments of the chromatin and slides on them until running into a pair of steric blocks (**Figure 8.5A**). To differentiate between these models, we placed two well-separated dCas9 roadblocks in the DNA template through sgRNA-mediated targeting (**Figure 8.5B**). The non-specific roadblock model predicts large force-induced transitions when the dCas9-bound DNA is incubated with PRC2, while the nucleosome-specific model does not. We did not observe such large transitions, even though there was clear evidence that dCas9 was stably bound to DNA (**Figure 8.5C-E**). This result disfavors the non-specific

roadblock model and suggests that the C4 transitions are largely attributed to specific PRC2-nucleosome contacts rather than steric hindrance.

The interpretation that C4 transitions originated from PRC2 bridging two non-neighboring nucleosomes makes a few testable predictions. First, since this mode of PRC2 interaction requires a substrate harboring at least three nucleosomes, it should not be observed on arrays containing only two nucleosomes. We therefore constructed a dinucleosome substrate (**Figure 8.1**) and, satisfyingly, found that C4 was essentially absent from this substrate when incubated with PRC2 (**Figure 3.3G** and **Table 8.1**). The difference between the transition cluster distributions for PRC2-bound 12-mer and 2-mer arrays is very significant (**Figure 3.3E** vs. **Figure 3.3G** ; $P = 5.9 \times 10^{-27}$, Kolmogorov-Smirnov test). This result rules out the possibility that the C4 cluster originates from PRC2 engaging purely free DNA, or engaging one nucleosome and one free DNA segment. It also argues against an altered pathway of DNA unwrapping from the nucleosome caused by PRC2 being the source of these large transitions.

Second, our interpretation of C4 transitions predicts that the amount of DNA released per transition should be dependent on the length of linker DNA between nucleosomes. To test this, we constructed a 12-mer nucleosome array with longer linkers (50 bp) and conducted single-molecule pulling experiments with this substrate (**Table 8.1**). Our measurements showed that the linker DNA length selectively modulates the size of C4 transitions and that the average ΔL_0 in C4 for the 50-bp-linker array is larger by approximately 40 bp than that for the 30-bp-linker array (**Figure 3.3H**). This magnitude of size change again suggests that the $\text{Nuc}_{1\text{-}3}$ mode, which releases two linker DNA lengths upon disruption, is the predominant mode of PRC2 engagement in C4. We note that the difference in average ΔL_0 for C3 transitions between 30-bp-linker and 50-bp-linker arrays is approximately 7 bp, smaller than the predicted 20 bp (**Figure 3.3H**). As mentioned above, some of the C3 transitions represent two mononucleosomes coincidentally unraveling together, independent of PRC2 engagement, which explains the relative insensitivity of the average ΔL_0 in this cluster to the linker length.

Third, since the occurrence of C4 transitions strictly relied on the presence of PRC2 (**Figure 3.3C** vs. **Figure 3.3E**), we reasoned that the frequency of their occurrence should depend on the amount of chromatin-bound PRC2. We thus conducted pulling experiments with varying concentrations of PRC2 and indeed observed such dependence (**Figure 3.3I**): the relative population of C4 transitions increased as the PRC2 concentration was raised from 5 nM to 100 nM, and appeared to become saturated afterwards. On the other hand, the relative population of C1 transitions, which represent free DNA sequestration by PRC2, continued to rise at higher PRC2 concentrations (**Figure 3.3I**). This is in accordance with the higher K_D value (approximately 900 nM) for PRC2-induced DNA looping obtained from the AFM study [180]. Taking these pieces of experimental evidence together, we concluded that the most plausible interpretation for the C4 transitions is a non-adjacent nucleosome bridging ability possessed by PRC2.

3.2.3 In silico characterization of PRC2-chromatin interactions

To further corroborate the experimental results, we developed a computational workflow that combined molecular dynamics simulations to sample polynucleosome configurations with rigid docking to explore the binding poses of individual PRC2 complexes (**Figure 8.6A**). In the following, we used this in silico platform to determine the set of low-energy poses that PRC2 adopts when in contact with tetranucleosome substrates.

We first performed 25 independent molecular dynamics simulations to collect a total of 1,000 representative structures for the tetranucleosome with 30-bp linker DNA. To ensure that both extended and collapsed configurations of the tetranucleosome were included in this structural ensemble, we used harmonic restraints to bias each simulation towards different spatial distances between the 1-3 and 2-4 nucleosome pairs. These simulations were initialized with a tetranucleosome configuration obtained by sequentially extending the dinucleosome cryo-EM structure [181] with 30-bp linker DNA segments and mononucleosomes. A coarse-grained force field that models protein [182, 183] and DNA [184] molecules at a single-residue and single-base-pair resolution was used to provide an accurate description of protein–protein and protein–DNA interactions. Prior studies showed that this level of resolution is sufficient to accurately model the energetics of nucleosomal DNA unwrapping [185, 186] and inter-nucleosome interaction [187]. To further improve the computational efficiency, we modeled the core region of each nucleosome as rigid bodies while maintaining the flexibility of the outer layer DNA. As shown in **Figure 3.5A**, these simulations covered a wide range of tetranucleosome configurations and allowed us to determine the free-energy surface for chromatin folding.

Next, for each one of the tetranucleosome configurations, we applied a rigid docking procedure to determine the set of lowest energy PRC2 binding poses. A structural model of the PRC2 core complex (SUZ12, EZH2, EED, and RBBP4) was constructed from several published crystal and cryo-EM structures [188-190] using homology modeling [191]. The combination of several partially resolved structures ensured that full-length core subunits are represented in the model, including the N-terminal part of SUZ12 that is important for PRC2 recruitment [188, 192]. Close examination of the docking results revealed that PRC2 can interact with adjacent nucleosomes using EED and EZH2 subunits ($Nuc_{1\text{-}2}$ mode), consistent with the published cryo-EM structure [181]. Repeating the docking procedure using the same PRC2:dinucleosome system as in the cryo-EM study led to similar structural configurations (**Figure 8.6B**). The computational model succeeded in reproducing the experimental structure as the top-ranking configuration. The agreement between simulated and experimental PRC2 binding poses lends support to the use of the computational protocol implemented here to study PRC2–chromatin interactions.

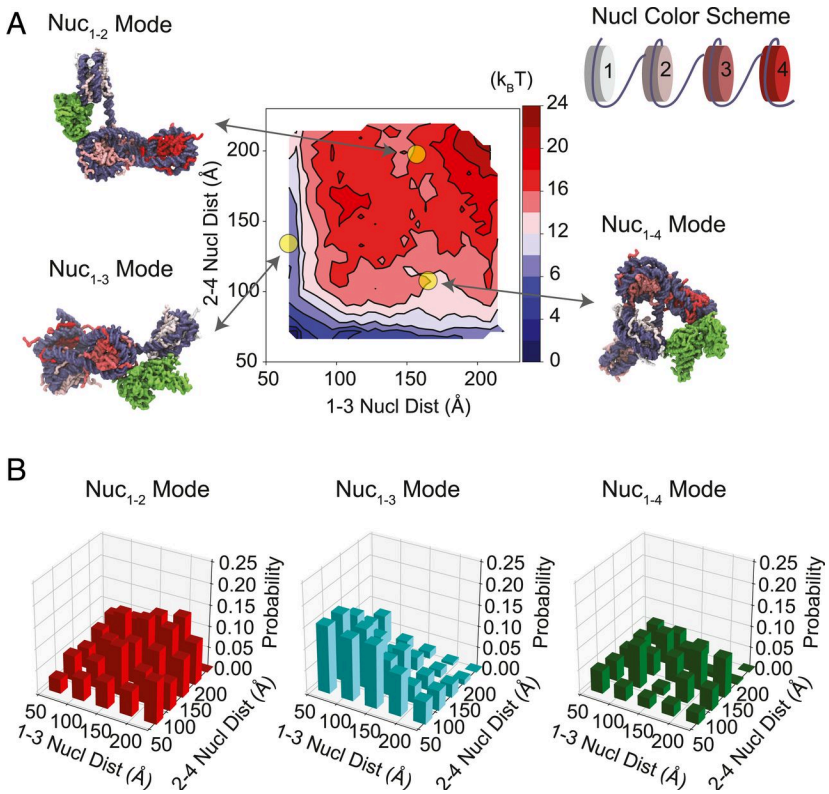


Figure 3.5. In silico analysis reveals the preferred mode of interaction between PRC2 and nucleosome pairs. (**A**) Thermodynamic stability (free energy) for the tetranucleosome with a 30-bp-long linker DNA as a function of the spatial distances between the 1-3 and 2-4 nucleosome pairs. The impact of PRC2 was not included when computing the free energy. Example chromatin configurations are shown on the side, with DNA in purple and the color of histone proteins varying from white to red as the nucleosome index increases. The PRC2 complex is colored in green and shown in its lowest energy pose in the three examples. (**B**) Fraction of PRC2 engaging in different nucleosome-bridging modes (Nuc1-2, Nuc1-3, and Nuc1-4) at given tetranucleosome configurations grouped by its inter-nucleosome distances.

At many tetranucleosome configurations, we found that PRC2 can bridge nucleosome pairs with one (Nuc_{1-3} mode) or two (Nuc_{1-4} mode) spacers using EED and EZH2 subunits of the same complex. For small inter-nucleosome distances, the Nuc_{1-3} mode is favored over Nuc_{1-2} and Nuc_{1-4} modes (**Figure 3.5B**). We note that the results at different inter-nucleosome distances cannot be compared directly against each other since they were obtained by simple counting of the docking results. With proper thermodynamic reweighting, however, the results can indeed be combined to provide a global estimation of the various PRC2 binding modes across the entire phase space. We reweighted each PRC2-tetranucleosome assembly to account for both the binding energy between the two complexes and the free energy cost for the tetranucleosome to adopt the conformation in the assembly. The relative population of the Nuc_{1-3} mode was estimated to be 18%, in comparison with 4% for Nuc_{1-2} and 7% for Nuc_{1-4}, respectively. The rest of the population displays a mononucleosome binding or free DNA binding pose of PRC2. These population estimations are for chromatin under zero tension. A better agreement with the experimental results could be obtained if only extended chromatin configurations, which are

presumably more prevalent in optical trapping experiments, were included in the population estimation.

In addition, we applied the same modeling procedure to a tetranucleosome with 50-bp linker DNA. We again found ample examples in which PRC2 forms stable assemblies with non-adjacent nucleosome pairs. Extending the linker length to 50 bp generally increases the distances between nucleosomes and reduces the probability of PRC2 engaging with multiple nucleosomes. This reduction is more significant for the Nuc_{1-2} and Nuc_{1-4} modes than for Nuc_{1-3} (**Figure 8.7**), a trend that was observed in experimental results as well (C4 populates 9% of total transitions for the 30-bp-linker array vs. 15% for the 50-bp-linker array; **Table 8.1**). Taken together, the computational modeling combined with the thermodynamic estimation corroborates the single-molecule data and suggests that bridging non-adjacent nucleosomes is an inherent property of the PRC2 core complex.

3.2.4 Differential regulation of PRC2-chromatin interaction by AEPB2 and JARID2

Next, we used the single-molecule experimental platform to assess the effects of two accessory subunits, AEBP2 and JARID2, on the binding of PRC2 to chromatin (**Figure 3.6A-B** and **Figure 8.2**). Interestingly, the addition of a stoichiometric amount of AEBP2 largely abolished C1 transitions, which correspond to free-DNA sequestration by PRC2 (Fig. 3*C* and **Table 8.1**). There is a statistically significant difference between the transition cluster distributions for 12-mer arrays incubated with PRC2 core complexes and with PRC2-AEBP2 complexes (Fig. 1*E* vs. 3*A*; $P = 0.00071$, Kolmogorov-Smirnov test). Thus, AEBP2 either reduces DNA binding by PRC2 or alters its binding geometry such that it no longer sequesters DNA. The latter scenario is supported by the recent AFM study, which showed that PRC2-AEBP2 mostly bends DNA without substantial wrapping [180]. PRC2-AEBP2 can still bind nucleosomes and bridge nucleosome pairs (**Figure 3.6C**). However, AEBP2 significantly lowered the average transition force (14.5 ± 0.3 pN for 12-mer + PRC2 + AEBP2 vs. 17.0 ± 0.3 pN for 12-mer + PRC2 core; **Figure 3.6D**), indicating that AEBP2 destabilizes the engagement of PRC2 with chromatin.

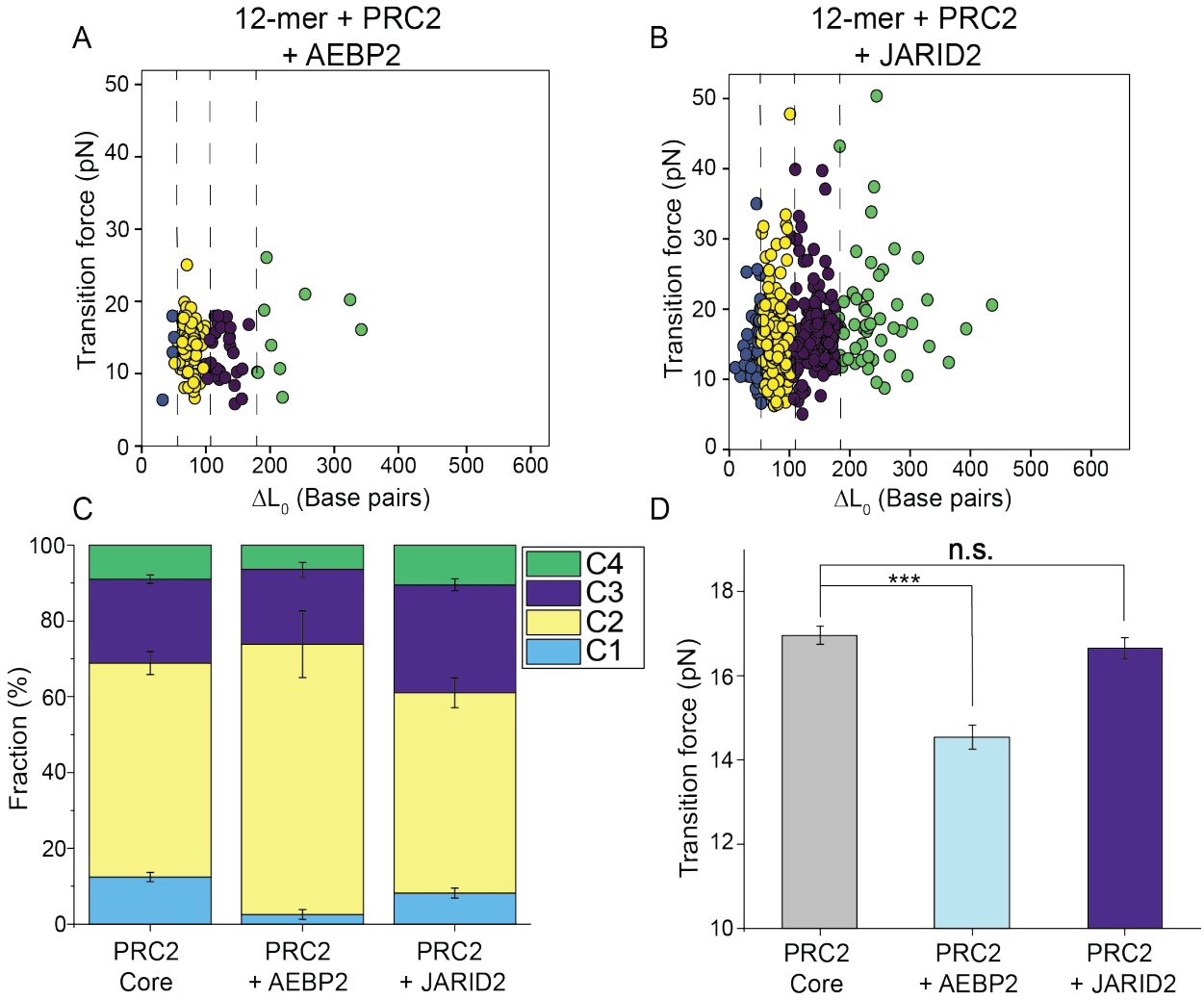


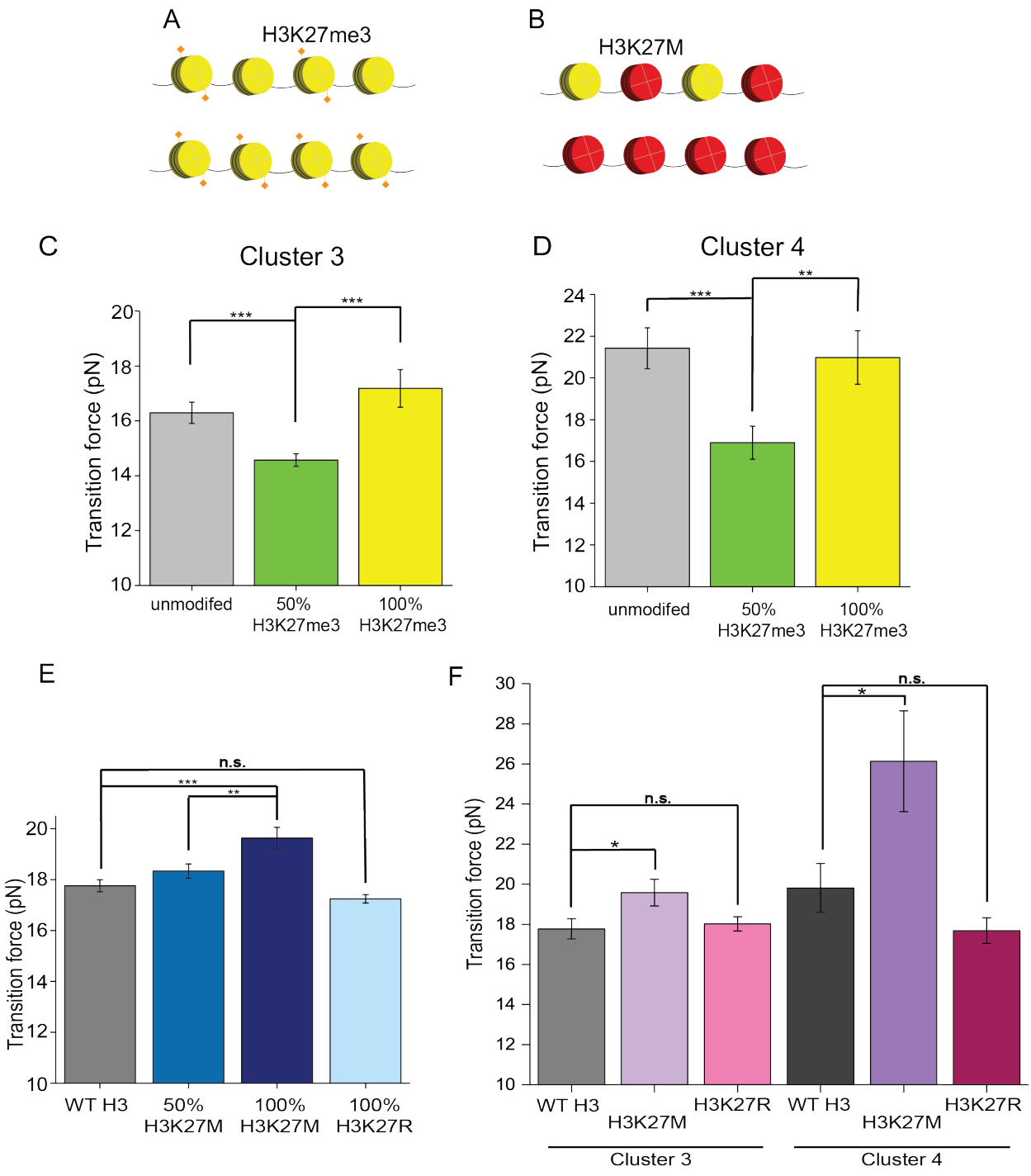
Figure 3.6. Chromatin interaction of PRC2 core complex is differentially modulated by accessory subunits AEBP2 and JARID2. (**A**) Cluster analysis of transitions observed in the force-extension curves for 12-mer arrays incubated with PRC2-AEBP2 complexes. (**B**) Cluster analysis of transitions for 12-mer arrays incubated with PRC2-JARID2 complexes. N denotes the number of transitions. (**C**) Cluster distribution for 12-mer arrays incubated with PRC2 core, PRC2-AEBP2, and PRC2-JARID2 complexes. (**D**) Average transition force for 12-mer arrays bound to PRC2 core, PRC2-AEBP2, and PRC2-JARID2 complexes. Data are presented as mean ± SEM. n.s., $P \geq 0.05$; ***$P < 0.001$.

On the contrary, the addition of JARID2 did not cause significant changes in the transition cluster distribution, nor did it significantly affect the average transition force (**Figure 3.6B-D** and **Table 8.1**). These results are consistent with prior biochemical data suggesting that the effect of JARID2 on PRC2 function is mainly manifested in the allosteric regulation of the catalytic activity of EZH2 [179, 189, 193, 194]. We do not, however, rule out that JARID2 may exert distinct effects under different chromatin contexts, synergizing with or antagonizing other regulatory signals [195, 196].

3.2.5 H3K27 methylation and mutation levels modulate PRC2 interaction with chromatin

To examine how preexisting H3K27me3 marks affect the nucleosome-binding properties of PRC2, we constructed 12-mer arrays with purely H3K27me3-containing nucleosomes, as well as arrays with an equimolar mixture of unmodified and H3K27me3-containing nucleosomes that were randomly incorporated into the array (**Figure 3.7A**). When pulling on the 50% H3K27me3 arrays incubated with PRC2, we found a significant reduction in the average rupture force compared to unmodified arrays for both C3 and C4 transitions (C3: 14.6 ± 0.2 pN for 50% H3K27me3 vs. 16.3 ± 0.4 pN for unmodified; C4: 16.9 ± 0.8 pN for 50% H3K27me3 vs. 21.4 ± 1.0 pN for unmodified; **Figure 3.8B-C** and **Table 8.1**). Interestingly, 100% H3K27me3 arrays brought the transition force back to the level of unmodified arrays (**Figure 3.8B-C** and **Table 8.1**).

Figure 3.7. Methylation and mutation levels of H3K27 affect the behavior of PRC2 on chromatin. (**A**) Cartoon of a nucleosome array containing 50% (Top) or 100% (Bottom) H3K27me3-modified nucleosomes. (**B**) Average C3 transition force for unmodified, 50% H3K27me3-modified, and 100% H3K27me3-modified 12-mer arrays in the presence of PRC2. (**C**) Average C4 transition force for unmodified, 50% H3K27me3-modified, and 100% H3K27me3-modified 12-mer arrays in the presence of PRC2. (**D**) Cartoon of a nucleosome array containing 50% (Top) or 100% (Bottom) H3K27M mutant nucleosomes. (**E**) Average overall transition force for wild-type, 50% H3K27M, 100% H3K27M, and 100% H3K27R 12-mer arrays incubated with PRC2 and SAM. (**F**) Average C3 and C4 transition forces for wild-type, 100% H3K27M, and 100% H3K27R 12-mer arrays in the presence of PRC2 and SAM. The methylated and mutant nucleosomes were randomly incorporated into the arrays at indicated molar ratios. Data are presented as mean ± SEM. n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.


A
H3K27me3
B
H3K27M
C
Cluster 3
Transition force (pN)
20
18
16
14
12
10
***
***
unmodifed
50%
H3K27me3
100%
H3K27me3
D
Cluster 4
Transition force (pN)
24
22
20
18
16
14
12
10
***
**
unmodifed
50%
H3K27me3
100%
H3K27me3
E
Transition force (pN)
20
18
16
14
12
10
n.s.
***
**
WT H3
50%
H3K27M
100%
H3K27M
100%
H3K27R
F
Transition force (pN)
30
28
26
24
22
20
18
16
14
12
10
n.s.
*
n.s.
*
WT H3
H3K27M
H3K27R
WT H3
H3K27M
H3K27R
Cluster 3
Cluster 4

The lysine-to-methionine substitution of H3K27 (H3K27M) is prevalently found in pediatric brain cancers and associated with a global decrease of H3K27me3 via inhibition of PRC2 [197, 198]. To evaluate how this oncogenic mutation impacts PRC2–nucleosome interaction, we compared the force-extension trajectories for PRC2-bound 12-mer arrays that contain 0, 50%, and 100% H3K27M mutant copies of H3 (**Figure 3.8D**). We also included SAM in this experiment, as the inhibitory effect of H3K27M has been shown to rely on the presence of SAM [198, 199]. We found that the mutation increased the overall average transition force in a dose-dependent manner (17.8 ± 0.3 pN for wild-type, 18.3 ± 0.3 pN for 50% H3K27M, 19.6 ± 0.5 pN for 100% H3K27M; **Figure 3.8E** and **Table 8.1**). An increased transition force by the K27M mutation was also observed when C3 and C4 transitions were selectively analyzed (C3: 17.8 ± 0.5 pN for wild-type vs. 19.6 ± 0.7 pN for 100% H3K27M; C4: 19.8 ± 1.2 pN for wild-type vs. 26.1 ± 2.6 pN for 100% H3K27M; **Figure 3.8F**). These results suggest that H3K27M stabilizes the engagement of PRC2 with nucleosomes, lending support to the model in which H3K27M-containing nucleosomes trap PRC2 locally via strong contacts, thereby inhibiting the propagation of H3K27me3 marks [189, 197, 200]. We also performed experiments with 12-mer arrays harboring 100% H3K27R nucleosomes as a non-substrate, non-inhibitor control [197, 199]. As expected, the lysine-to-arginine mutation did not elicit an appreciable effect on the transition forces, in contrast to the pathogenic K27M mutation (**Figure 3.8E-F** and **Table 8.1**).

3.2.6 PRC2-mediated compaction of nucleosome arrays

The ability of PRC2 to bridge pairs of non-adjacent nucleosomes entails a chromatin-compacting activity of the complex. To test this activity, we examined the architecture of wild-type 12-mer arrays using negative-stain EM (**Figure 3.8A-B** and **Figure 8.8**). We found that the arrays generally adopted more compact configurations when incubated with PRC2 (**Figure 3.8C**). We quantified the level of compaction by measuring the average distance between nucleosomes within the same array. This analysis revealed that PRC2 condensed 12-mer arrays in a concentration-dependent manner (**Figure 3.8D**). We then repeated the EM assay with tetranucleosome substrates and observed the same trend of PRC2-dependent compaction (**Figure 3.8E-H**). These results demonstrate that PRC2 is able to mediate chromatin compaction—a property conferred by bridging distal chromosomal segments. We note that, according to previous work, the chromatin-compacting capability of PRC2 can be further enhanced by replacing the EZH2 subunit with its homolog EZH1 [201].

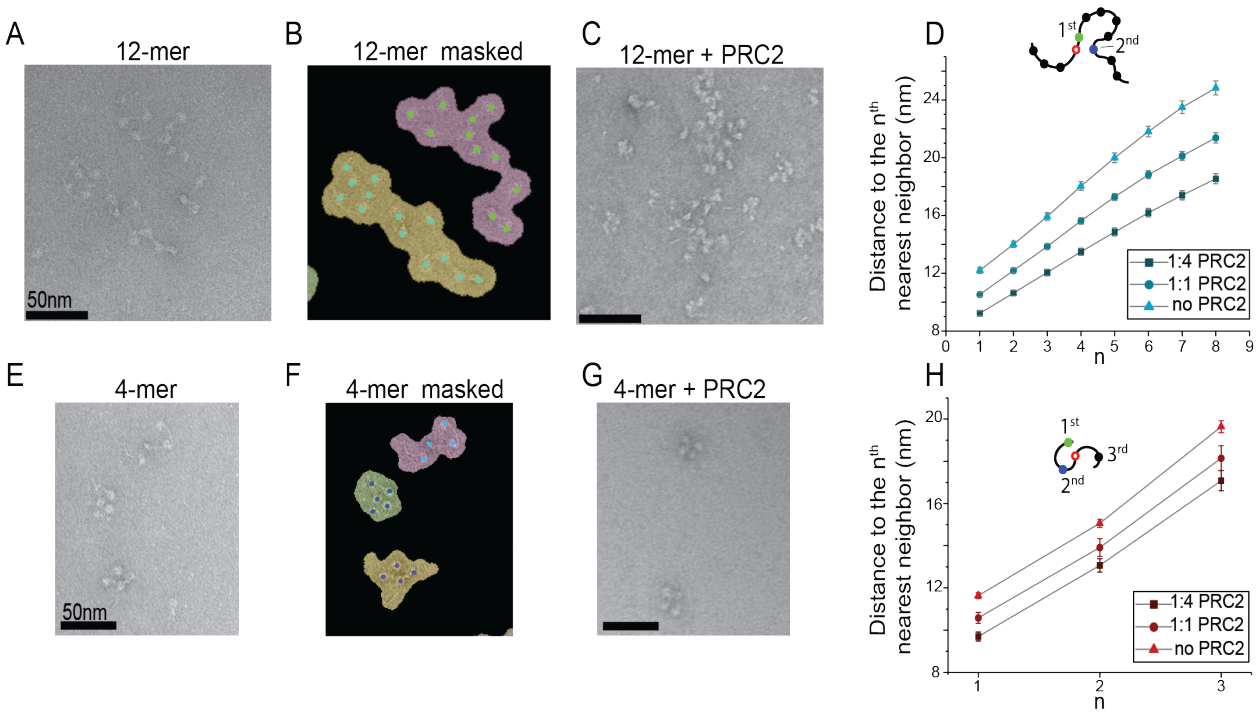


Figure 3.8. PRC2 mediates the compaction of polynucleosomes. (**A**) A representative negative-stained electron micrograph of 12-mer nucleosome arrays. (**B**) The same micrograph as in (A) with binary masks depicting individualized nucleosome arrays. Nucleosomes inside their corresponding masks are shown as colored dots. (**C**) A representative micrograph of 12-mer arrays incubated with PRC2 at 1:4 molar ratio (150 nM array and 600 nM PRC2). (**D**) Average distance between a reference nucleosome (open circle) and its nth nearest neighbor within the same 12-mer array measured at different array:PRC2 mixing ratios. A total of 95 masks from 14 micrographs were analyzed for the 1:4 array:PRC2 ratio, 121 masks from 10 micrographs for the 1:1 array:PRC2 ratio, and 98 masks from 14 micrographs for the no-PRC2 condition. A higher PRC2 level correlates to a shorter distance, indicating a higher degree of array compaction. (**E**) A representative electron micrograph of tetranucleosome arrays. (**F**) The same micrograph as in (E) with binary masks depicting individualized arrays. (**G**) A representative micrograph of tetranucleosome arrays mixed with PRC2 at 1:4 molar ratio. (**H**) Average distance from a reference nucleosome (open circle) to its first, second, and third nearest neighbor within the same tetranucleosome array measured at different array:PRC2 mixing ratios. A total of 103 masks from 10 micrographs were analyzed for the 1:4 array:PRC2 ratio, 99 masks from 12 micrographs for the 1:1 array:PRC2 ratio, and 371 masks from 19 micrographs for the no-PRC2 condition. Data are presented as mean ± SEM.

3.3 PRC2 conclusions

PRC2 maintains transcriptional silencing of a large number of genes involved in development and cancer [202]. Its function and regulation have been subjected to extensive investigations. However, conflicting results still exist regarding the mode of action of PRC2 on chromatin and the roles of diverse regulatory factors. In this work, we employ single-molecule force spectroscopy to extract the structural fingerprints of individual PRC2–polynucleosome assemblies. We observed a distinct cluster of large transitions (C4) that also feature higher

rupture forces than transitions in other clusters for wild-type 12-mer arrays incubated with PRC2 core complexes, suggesting that these C4 transitions are associated with a low-energy PRC2–nucleosome interaction mode. To our knowledge, this finding provides the first direct evidence for a stable PRC2–chromatin complex in which PRC2 mediates inter-nucleosome contacts beyond nearest neighbors. Our data further suggest that this interaction geometry can accommodate different linker DNA lengths, even though its propensity is influenced by the linker DNA. Our experimental and computational analyses established the trinucleosome unit as a favored substrate for PRC2 engagement. The zigzag chromatin topology, which positions every other nucleosome into close proximity [203], may underlie this preferred mode of PRC2 binding.

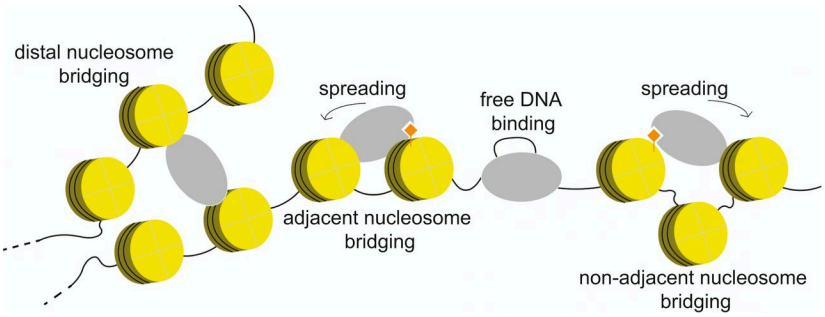


Figure 3.9. Working model for PRC2-mediated heterochromatin formation. PRC2 interacts with chromatin in multiple modes, carrying out enzymatic modification and non-enzymatic compaction of the chromatin, both of which contribute to transcriptional repression. In particular, the ability of PRC2 to simultaneously engage non-adjacent nucleosomes within the same domain, or perhaps distal nucleosomes from different domains, facilitates efficient spreading of repressive histone marks and mediates local or global chromatin condensation.

These long-range contacts have significant implications for the mechanism of H3K27me3 spreading (**Figure 3.9**). It argues against models in which PRC2 propagates H3K27me3 along nucleosome arrays in strictly linear progression. Instead, PRC2 can bypass spacer nucleosomes and perhaps other roadblocks when carrying out the read-and-write function. As such, repressive marks can be spread to nucleosomes far away in sequence from the nucleation site. This feature could serve as the mechanistic basis for several epigenetic phenomena, such as memory and bistability [204, 205], and is necessary for establishing the H3K27me3 pattern in vivo [206]. Another consequence of distal nucleosome bridging by PRC2, which is separable from its catalytic activity, is the physical compaction of nucleosome arrays (**Figure 3.8**). This activity, combined with that of DNA compaction via bending and looping [180], likely constitutes another mechanism besides chemical modification that contributes to PRC2-dependent transcriptional repression. It is plausible that engaging non-contiguous chromosomal segments is a common characteristic of Polycomb group proteins [207].

By subjecting PRC2–chromatin assemblies to mechanical perturbations, we obtained fresh insights into the strengths of different binding modalities for this important epigenetic machinery. Recent studies reported that PRC2–chromatin interactions are transient, allowing the enzyme to readily translocate on substrates for efficient spreading of repressive marks [198, 208]. Our data here show that PRC2–nucleosome linkages can also withstand high tensions, exhibiting mechanical stability comparable to or higher than that of the nucleosome itself. It can be envisioned that PRC2 is performing a balancing act: both stable and transient interactions are at work in order to accomplish de novo installation and self-propagation of H3K27me3. The

strong interaction may be required for PRC2 recruitment to nucleation sites and for the deposition of initial H3K27me3 foci [192, 208], presumably with slow kinetics; whereas the weak contacts induced by, for example, the coexistence of activating (modified) and substrate (unmodified) nucleosomes, facilitate efficient spreading by making PRC2 more labile on chromatin. The latter scenario is directly supported by our single-molecule force measurements, which revealed that PRC2 binds less stably to partially H3K27me3-modified nucleosome arrays than to naïve and fully modified arrays, suggesting that PRC2 residing at the boundary between unmodified and modified nucleosome domains adopts a binding configuration that destabilizes its nucleosome engagement and allows its efficient spreading. In the future, it would be valuable to visualize the behaviors of PRC2 on chromatin substrates at different modification states using structural and single-molecule approaches.

In conclusion, our results unveil a diverse repertoire of PRC2 binding modes on chromatin. The in vivo concentration of PRC2 is estimated to range between 50 and 1000 nM [198], similar to what we used in the experiments. Thus, these interaction modes are expected to contribute significantly to PRC2 function inside the nucleus, which is further regulated by the epigenetic and genetic states of the chromatin, as well as by the availability of various accessory factors. Future experiments are warranted to answer whether these interaction modalities are mediated by PRC2 monomers, dimers, or higher-order multimers [180, 209, 210]. The experimental and computational platforms established in this work will facilitate continued exploration of the elaborate mechanisms that PRC2 and other epigenetic machineries utilize to achieve chromatin targeting and maintenance.

CHAPTER 4. Single-stranded nucleic acid sensing and coacervation by linker histone H1

4.1 The forgotten histone: linker histone H1

Another key player in chromatin biology is the linker histone H1, which binds linker DNA on the nucleosome. One of the earlier studies of linker histone fractionation showed various H1 subtypes [211]. Today it is known that many linker histone variants are present in mammalian cells [212], however their individual functions remain uncharacterized. There are three main categories of H1 subtypes: somatic replication-dependent subtypes (H1.1-H1.5) expressed in S-phase, somatic replication-independent subtypes (H1.0, H1X) expressed during the complete cell cycle, and germline specific subtypes expressed in either the testes or oocytes (H1.6 TS, H1.7 TS, H1.8 OO, H1.9 TS). All variants are highly basic and lysine rich, with a short unstructured N-terminal tail, a structured globular domain, and a long unstructured c-terminal tail, which harbors many of the differences between variants [213].

Linker histone H1 binds the entry and exit site DNA of the nucleosome forming the chromatosome [214], and plays a key role in regulating chromatin condensation [215]. H1 can condense tri-nucleosomes *in vitro*, binding nucleosomes tightly by recognizing specific DNA geometries [216]. *In silico* experiments demonstrated H1 aids in the formation rigid chromatin fibers, compared to a more flexible chromatin structure in depleted H1 models, which have a higher propensity for forming enhanced long-range nucleosome-nucleosome interactions [217]. *In vivo*, H1 has been identified in physically separated regions of the nucleus, compacting constitutive (H3K9me3) and facultative (H3K27me3) heterochromatin [28]. Structural data has elucidated the binding of the globular domain to the entrance and exit site DNA [34, 218, 219], however, resolution of the disordered tails remains elusive. These tails contain many basic residues and are hypothesized to impart much of the functionality for linker histone chromatin compaction as well as differences in function between variants.

Linker histones bind tightly to DNA and chromatin substrates [220], however *in vivo* FRAP studies suggest the rapid exchange of linker histones where binding only occurs for one minute [221, 222]. This exchange is most likely due in part to the presence of linker histone chaperones. One chaperone, ProTα, was found to immune-isolate with linker histone H1 [223]. Transfection of ProTα in mammalian cells made them more susceptible to micrococcal nuclease (MNase) digestion, a hallmark of chromatin de-condensation [224, 225]. ProTα incubation with metaphase chromosomes also led to unraveling [226]. Specific to H1, ProTα has been shown to facilitate displacement of H1 from chromatin substrates as well as deposition on chromatin *in vivo* and increased linker histone exchange [227]. Linker histones may also be chaperoned by core histone chaperones Nap1, NASP, and NPM1 [228]. There are links between linker histones and disease, including cancers [229] such as ovarian [230] and prostate [231] as well as Alzheimer's disease [232]. Recently, mutations in variants H1.2 and H1.4 have shown to be major drivers of lymphoma, partially through impaired chromatin compaction abilities [217].

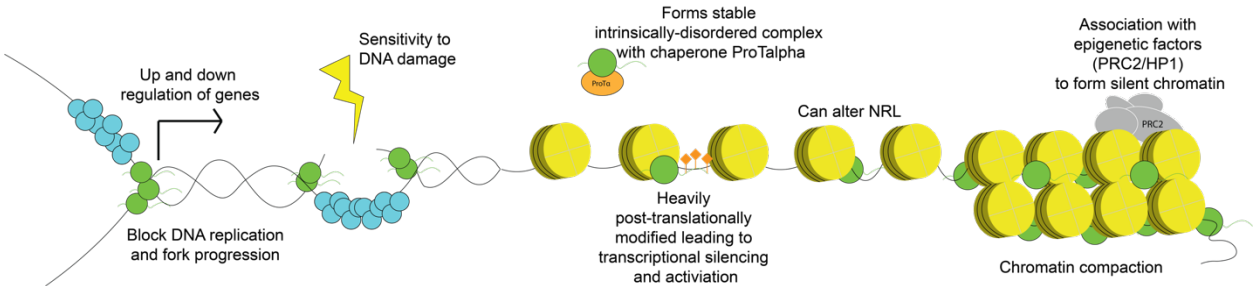


Figure 4.1. Illustration of reported roles of H1. Outside of chromatin compaction, there are multiple proposed roles of the linker histone H1including gene regulation, DNA replication, DNA damage, and interactions with epigenetic factors.

4.1.1 Role of H1 in gene expression and chromatin structure

Numerous classical studies have signified the importance of linker histone H1 in overall chromatin organization due to its role as a 30 nm fiber compactor and a hallmark of mitosis in its hyperphosphorylated state [233]. Early studies in *Tetrahymena* displayed H1 as nonessential gene but an important chromatin compactor [234], showing H1 dephosphorylation correlated with the formation of condensed chromatin and transcriptional silencing during development of macronuclei, which are typically the transcriptionally active somatic nuclei [235]. Later, imaging phosphorylated H1 across entire cells revealed the presence of localized H1 on mitotic chromosomes. The replication-dependent phosphorylation of H1 during S-phase [236, 237], compared to a more diffuse staining pattering during interphase, suggested the likelihood for alternative roles of histone H1 beyond compaction such as gene activation and DNA replication [238]. Later work in *Xenopus* detailed the enrichment of H1 on duplicated chromosomes, demonstrating its necessity in proper chromosome alignment and segregation [239]. It is interesting to think about how phosphorylated H1 may play a role in mitosis given the innate electrostatic shield repulsion of phosphorylation marks and DNA. These conflicting data suggest a role for H1 beyond solely chromatin compaction, in higher order chromatin organization and nucleic acid segregation.

Collectively, linker histones have emerging roles in gene expression, heterochromatin regulation and TAD organization. Different histone variants have been shown to regulate transcription in a locus-specific manor [240, 241]. Multiple studies have aimed to map out H1 localization within cells, finding a characteristic binding pattern of low H1 occupancy at transcription start sites at evolutionarily conserved active genes [212]. Linker histones can also extensively remodel chromatin during sex differentiation in germ cell development, suggesting different variants can mediate epigenetic reprogramming leading to specific cellular states [242]. Overall, linker histones have been implicated in the upregulation and downregulation of a variety of genes, suggesting the differentiated roles of each of the variants.

Linker histones are extensively post-translationally modified [243] with marks including phosphorylation, acetylation, methylation, ubiquitination, and ADP ribosylation [244]. These marks may contribute to the differences in gene regulation [245, 246] by H1 variants. The most studied PTM has been C-terminal phosphorylation, mainly linked to chromatin condensation. H1.4 phosphorylation is also associated with transcriptional activation [247]. However, other studies have found it to disrupt the association between H1 and the repressive chromatin reader HP1 [248]. Further work has suggested the placement of these phosphorylation marks on the C-terminal tail will affect its secondary structure [249], dictating an array of complex functions.

Early work on H1 phosphorylation, as discussed earlier, suggests a potential role of linker histones in chromatin organization.

The roles of other H1 modifications have been initially characterized. One modification that has been well studied is H1.4K34 acetylation, which plays a role in H1 mobility as well as transcription activation [246]. Another broad hypothesis of interactions between neighboring methylation, phosphorylation, acetylation marks [250], has led to the discovery of their dynamic interplay on linker histone H1 [251]. Additionally, H1 ADP-ribosylation has been implicated in the relaxation of chromatin structure [252] and active transcription through PARP-1 interactions [253]. There is also evidence for different modifiers of different H1 subtypes which adds additional functionality of the individual variants [254]. Overall, the interplay between chemically modified H1 variants and epigenetic regulators still remains an open question.

H1 has been shown to regulate heterochromatin formation in *Drosophila [255]*, for example, H1 depletion led to failed development of adult flies. H1 was also shown to be a strong suppressor of PEV (position-effect variegation) and H3K9me2/3 patterning associated with HP1-foci and chromatin structure loss. It has further been shown that H1 directly recruits Su(var)3-9, which adds the repressive H3K9me3 mark to establish heterochromatic regions [256]. In mammals, H1 associates with HP1 [257] and condenses chromatin [258], strengthening the role of H1 in heterochromatin formation. Contrastingly, variant H1.2 has been shown to be a mediator of transcriptional regulation by selectively binding the Cul4A ubiquitin ligase, which sets off a cascade to methylate H3K4 and H3K79, both active marks associated with active transcription [259]. However, alternative roles have been suggested for H1's interference with Set7/9 "active" mark deposition on H3K4 [260] and interaction with DNA methyltransferases DNMT1 and DNMT3B for recruitment, suggesting the role of H1 in heterochromatin formation and gene expression in general is quite complex. Different variants may play different roles in heterochromatin formation, active gene expression, recruitment, and signaling processes.

4.1.2 Linker histone H1 in DNA damage and repair

Different histone variants play various roles in DNA damage response (DDR), especially H2A, H3, and H1 [261]. Conflicting evidence on the role of linker histones in DNA damage suggests depletion of different variants may either increase [262] or decrease [263] the cell's sensitivity to DNA damage, again suggesting different variants may have varied roles *in vivo*. Certain variants, including H1R, have shown increased sensitivity to DNA damage [262], likely due to a more open chromatin structure [263]. Furthermore, deletion of H1.2, H1.3, and H1.4 led to a hyperactivation of the DNA damage checkpoint and increased resistance to DNA-damaging agents [263]. The presence of different variants and post-translational modifications on these proteins can modulate the overall chromatin structure, which will have many downstream effects on DNA damage and response pathways in terms of accessibility and recruitment of remodeling and repair factors. Some work has been done to suggest that displaced damaged DNA or replication occurs on the periphery of heterochromatic regions [264]. It will be imperative to study all of these intricate interactions in greater depth given that other heterochromatic factors such as HP1 [265] and Polycomb group proteins [266] have also been reported to localize to DNA lesions. Physical interactions between linker histones and DNA repair machinery have also been identified [267, 268]. Overall, the role of H1 in chromatin compaction suggests a barrier for DNA lesions and overall damage.

Linker histone Hho1p in *S. cerevisiae* has been shown to repress DNA repair via homologous recombination [269]. Additional work has shown the role of H1 in the double strand break (DSB) repair pathway as a key target of RNF8-UBC13, allowing the ubiquitination of chromatin targets leading to chromatin remodeling and loosening to facilitate repair [268]. It has also been proposed that the post-translational modification state of linker histones may contribute to the DNA damage response [270], where low levels of phosphorylation could eject H1 from the chromatin, causing de-condensation [249] and granting access to repair machinery. Overall, linker histones are involved in various processes involving DNA damage and repair.

4.1.3 Linker histone H1 in DNA replication

Five of the histone variants are replication dependent (H1.1, H1.2, H1.3, H1.4, H1.5). This leaves the open question why these variants are only synthesized at this particular point in the cell cycle, suggesting they may have a specific role in regulating DNA replication. H1 has shown varying abilities to block DNA replication, inhibiting pre-replication complex formation in cell extracts [271] and retardation of replication fork progression through recruitment of SUUR (suppressor of under replication) *in vivo [272]*.

Later in this thesis, we will discuss the relationship between linker histone H1 and Replication Protein A (RPA), a single-stranded DNA-binding protein involved in DNA damage response and replication. RPA consists of a heterotrimer of three tightly associated subunits and is required for processes surrounding DNA metabolism, including replication, recombination, damage checkpoints and repair [273]. Specifically, it is involved in nucleotide excision repair (NER), base excision repair (BER), DNA mismatch repair (MMR) and double strand break repair (SDB) [274]. In contrast to cell-cycle dependent phosphorylation, RPA undergoes extensive phosphorylation upon DNA damage, playing a role in DNA damage pathways [275]. This hyperphosphorylation has been hypothesized to shift the RPA pool towards DNA damage and repair pathways and away from DNA replication during times of stress [276]. Biochemical evidence has shown that RPA binds to single-stranded DNA in a sequential manor with 5' to 3' polarity [277] and nanomolar affinity [278]. It has also been shown to interact with repair helicases (ex: BLM), activating their velocity and bidirectional activity [279]. Open questions remain to how phosphorylation of RPA regulates DNA binding and downstream replication, damage and repair processes.

4.1.4 Linker histone research outlook

Recently, great strides have been made in linker histone purification [280], allowing the biochemical characterization of linker histone chromatin compaction for multiple human variants. Human histone variants have different binding properties *in vitro* and *in vivo* [281] and can alter the nucleosome repeat length [282]. Evidence *in vivo* suggests linker histones may participate in a variety of additional protein-protein interactions, associated with many additional mechanisms not already discussed including mRNA metabolism, stress response, apoptosis, and ribosome function [283]. Much is to be learned about linker histone association with other players in the cell. Recent works suggests their intrinsically disordered nature may actually contribute to the formation of protein-protein complexes and interactions, demonstrated by a unique complex formed with its chaperone ProTα through facilitated diffusion mechanisms [284, 285]. Additionally, linker histones have demonstrated phase separation properties [87, 88,

286, 287], most likely driven by their intrinsically disordered nature, basic charge, and nucleic acid binding properties. Given their proposed role in DNA replication, repair and nucleic acid binding properties, we hypothesized that linker histone H1 might serve as a hub for nucleic acid organization for various cellular processes.

4.2 H1 nucleic acid sensing results using single-molecule, in silico, and cellular techniques

The results from our findings were submitted in Leicher et al. where we utilize correlative single-molecule force and fluorescence microscopy to simultaneously observe and manipulate the behavior of H1 when associated with DNA under different tensions. Unexpectedly, our results reveal that H1 preferentially coalesces around nascent, relaxed single-stranded (ss) DNA. *In vitro* bulk assays and molecular dynamics (MD) simulations confirmed that H1 has a higher propensity to form phase-separated condensates with ssDNA than double-stranded (ds) DNA. Furthermore, we dissected the material properties of different H1:DNA condensates by controlled droplet fusion with optical tweezers and found that increased DNA length and GC contect result in more viscous, gel-like H1 condensates. Finally, we observed H1 puncta in live cells and determined the H1 C-terminal tail as critical for droplet formation in a native context. Co-transfection studies with Replication Protein A (RPA) suggest it may compete with H1 for binding sites on ssDNA. Overall, our findings indicate a potential role for linker histones in sensing and coacervating single-stranded nucleic acids in the nucleus, thereby organizing the interchromatin space and helping to regulate reaction hubs for genome maintenance.

Computational modeling was performed by Andrew Latham from Bin Zhang's lab. H1 purification and cellular assays were performed by Wola Osunsade from Yael David's lab. Droplet analysis software was developed by John Watters. Additional support with figures was provided by Gabby Chua.

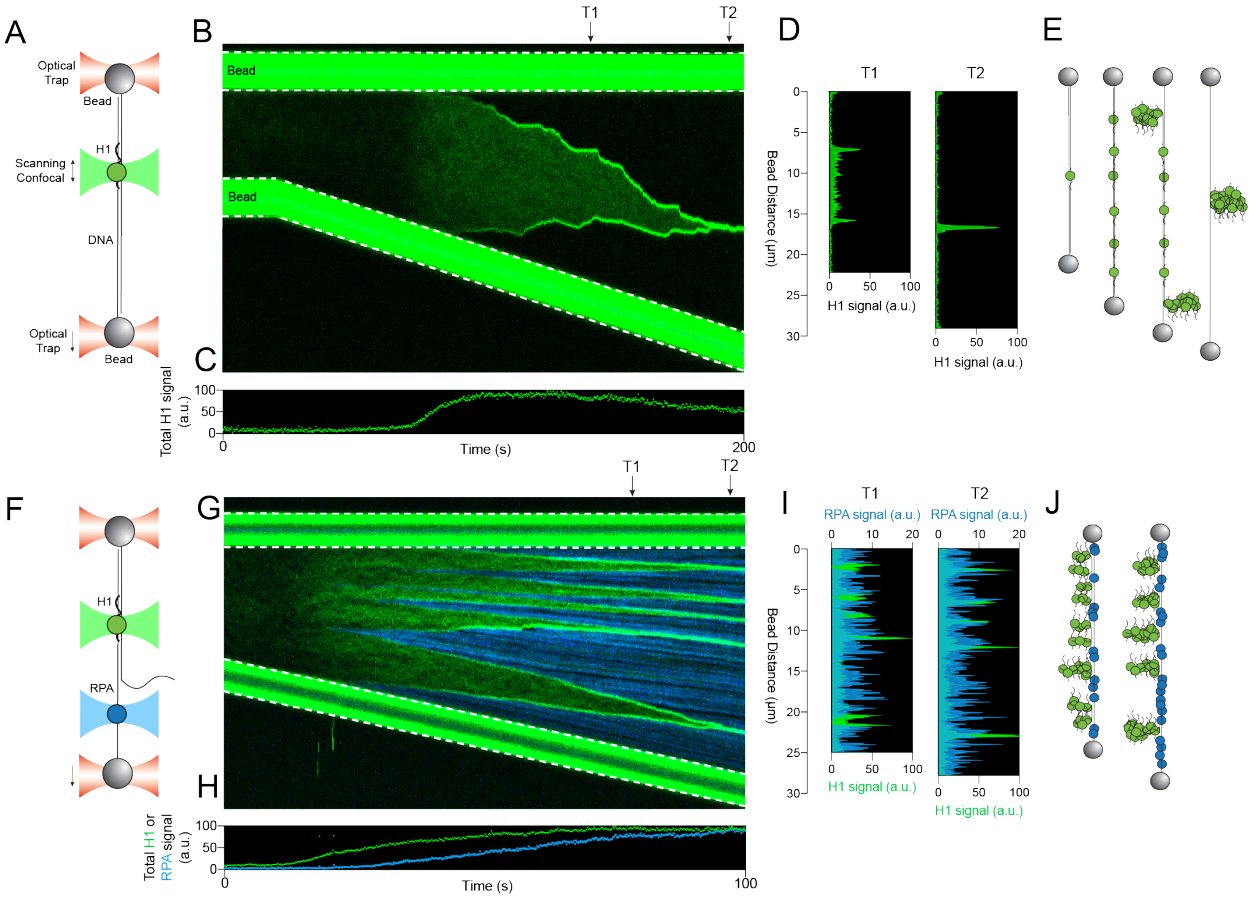


Figure 4.2. Single-molecule force spectroscopy reveals H1 condensate formation on ssDNA (**A**) Schematic of the dual-trap optical tweezers setup. A biotinylated 48.5 kb piece of lambda DNA is tethered to a streptavidin-coated bead on each side. Cy3-H1 is in solution and can bind to DNA. (**B**) A representative kymograph of Cy3-H1 binding to DNA over time as tension is increased. (**C**) Normalized plot of Cy3 intensity over time for kymograph in **B**. (**D**) Normalized plot of Cy3 intensity versus bead distance, occurring at two particular time points, T1 and T2, as indicated by the arrows in **B**. (**E**) Schematic of Cy3-H1 binding to ssDNA. H1 binds to double-stranded DNA, increasing occupancy at higher tensions, and forms condensates on relaxed single strands. (**F**) Schematic of similar setup as in **A**, but including Alexa488-RPA that can bind ssDNA under tension. (**G**) A representative kymograph of Cy3-H1 and Alexa488-RPA binding to DNA over time as tension is increased. (**H**) Normalized plot of Cy3 and Alexa488 intensity over time for kymograph in **G**. (**I**) Normalized plot of Cy3 and Alexa488 intensity versus bead distance, occurring at a particular time point, T1, as indicated by the arrow in **G**. (**J**) Schematic of Cy3-H1 and Alexa488-RPA binding to ssDNA. H1 forms condensates on relaxed single strands, whereas RPA binds to ssDNA under tension.

4.2.1 H1 coalesces around nascent DNA

To directly visualize the behavior of H1 on DNA *in vitro*, we purified human H1.4 (hereafter referred to as H1), one of the major H1 subtypes in human cells, and labeled it with a Cy3 fluorophore. We then added 15 nM of Cy3-H1 to a biotinylated phage lambda genomic DNA (48.5 kbp) tethered between two laser-trapped beads and monitored H1 binding along the DNA using scanning confocal microscopy (**Figure 4.2A**). At low force (< 1 pN), H1 binds minimally to the dsDNA tether (**Figure 4.2B**). When we applied increasing tension to the DNA by moving the beads apart from each other, H1 binding markedly intensified as evidenced by a higher total Cy3 signal across the tether (**Figure 4.2C**). The diffuse nature of the signal indicates largely nonspecific binding of H1. Strikingly, when the dsDNA was pulled into the overstretching regime (> 60 pN), H1 accumulated at both ends of the tethered DNA as observed in the emergence of two fluorescent foci (**Figure 4.2B,D**; "T1"). As the inter-bead distance continued to increase, the foci became brighter, migrating towards each other, and eventually merging into one singular spot (**Figure 4.2B,D**; "T2"). Because the DNA was attached to the bead via only one strand on each end, we posited that H1 preferentially binds to the untethered ssDNA created by force-induced unpeeling [288] (**Figure 4.2E**). Notably, H1 did not bind the other single DNA strand that was attached to the bead [i.e. we detected no fluorescence signal between the H1 focus and its proximal bead (**Figure 4.2 B,E**)], and therefore H1 does not appear to bind ssDNA under tension. In other examples, we also observed H1 foci forming internally, presumably around relaxed ssDNA originating from the internal nicks occurring within the DNA substrate (**Figure 8.10**). In addition, we found that this behavior is unique to the linker histone, as core histones such as H2B exhibit consistent binding from low to high forces (0-70 pN) (**Figure 8.11**). Indeed, H1 proteins are structurally more disordered and have a higher net positive charge compared to the core histones [289] which may account for this effect.

To confirm this interpretation, we used AlexaFluor488-labeled Replication Protein A (RPA), a well-studied eukaryotic ssDNA-binding protein that is capable of interacting with ssDNA under tension [290] (**Figure 4.2F**). Indeed, both H1 and RPA signals on DNA increased with tension (**Figure 4.2G,H**), but with drastically distinct localization patterns. H1 foci were consistently observed on the flanks of RPA-bound regions of the tether (**Figure 4.2G**), and the H1 and RPA fluorescence signals appeared to be anticorrelated (**Figure 4.2I**). This finding suggests that H1 does not bind to RPA:ssDNA regions that are under tension, whereas RPA is excluded from H1 foci around the relaxed ssDNA (**Figure 4.2J**).

Next, we examined whether the formation of H1 foci is reversible by sequentially stretching and relaxing the tether. Interestingly, we found that the majority of H1 foci dissolved upon tether relaxation (**Figure 8.12A**), presumably induced by the energetically favorable re-annealing of the two complementary ssDNA. In contrast, in the presence of RPA, most of the H1 foci persisted at low forces (**Figure 8.12B**), suggesting that RPA poses a barrier against the re-annealing of ssDNA and the dissolution of H1 foci. Overall, our single-molecule observations reveal that H1 preferentially binds to and coalesces around relaxed ssDNA, forming condensate-like complexes (**Figure 4.2F**).

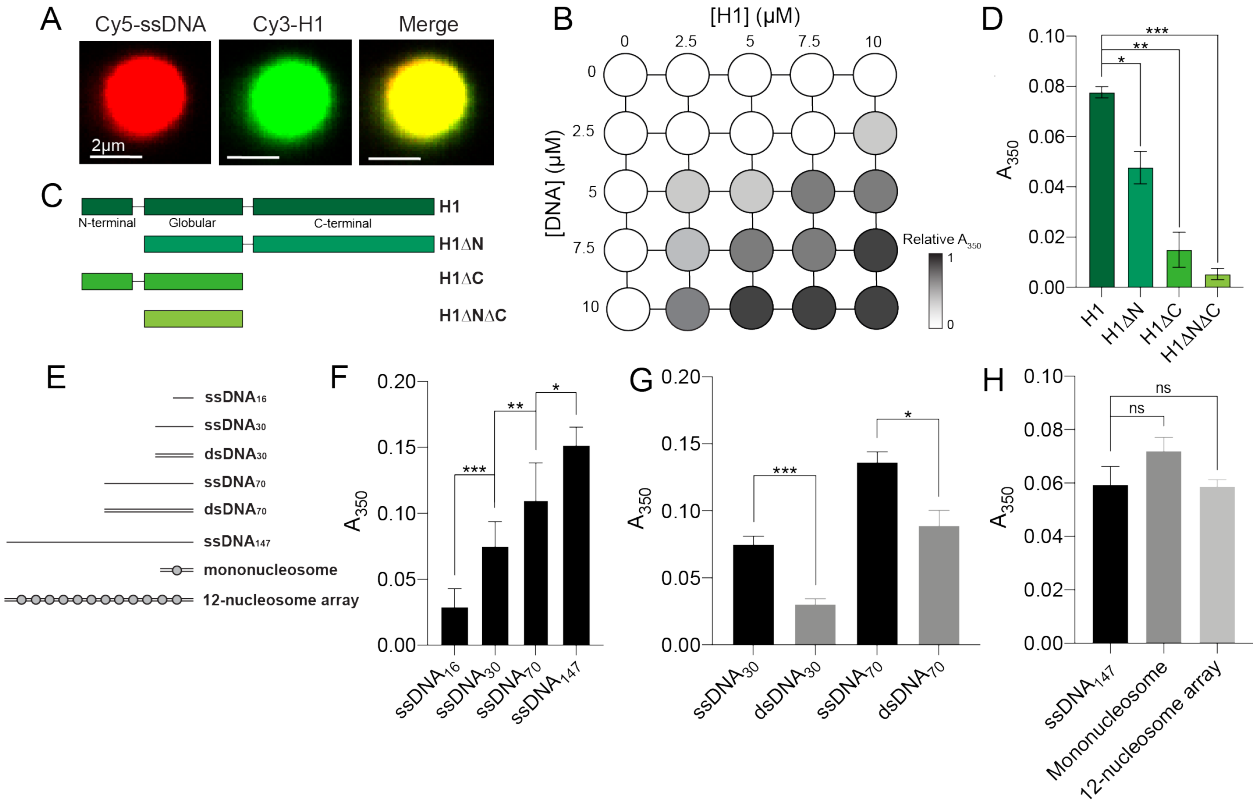


Figure 4.3. Phase separation of H1 and DNA is driven by single-stranded character, GC content, and intrinsically disordered regions. (**A**) Representative confocal images of representative droplet with 10 μM Cy5-ssDNA$_{75}$ (10% labeled) and 2.5 μM Cy3-H1. All droplet experiments performed with 2.5 μM H1 unless otherwise noted. (**B**) Matrix phase diagram of viscosity values for increasing H1 and ssDNA concentrations. (**C**) Schematic of H1 domain structure and truncation mutants. (**D**) Comparison between the average A_{350} turbidity values for full length H1:ssDNA (10 μM ssDNA$_{75}$) droplets and truncation mutants. (**E**) Schematic of ssDNA, dsDNA, and chromatin substrates used in H1 phase separation experiments in **F-H**. (**F**) Comparison between the average turbidity values for H1:ssDNA droplets of different DNA lengths (43.8 μM ssDNA$_{16}$, 23.3 μM ssDNA$_{30}$, 10 μM ssDNA$_{70}$, and 4.8 μM ssDNA$_{147}$). (**G**) Comparison between the average turbidity values for H1 DNA droplets with 10 μM ssDNA$_{30}$, ssDNA$_{70}$, dsDNA$_{30}$, dsDNA$_{70}$. (**H**) Comparison between the average turbidity values for H1 DNA droplets with ssDNA$_{147}$ (5 μM), mononucleosomes (5 μM), and 12-nucleosome arrays (600 nM). Data are mean ± SEM of at least three independent measurements.

4.2.2 H1 differentially phase separates with ssDNA and dsDNA

Based on the above observations, we hypothesized that H1 forms phase-separated condensates preferentially with ssDNA. It is worth noting that, although LLPS of linker histones with dsDNA and nucleosomes has been previously reported [291], H1 interaction with single-stranded nucleic acids has not been investigated. To this end, we mixed full-length H1 with 75-nt-long ssDNA under physiological buffer conditions, which indeed yielded distinct phase-separated droplets (**Figure 8.13A**). We further validated the enrichment of H1 and ssDNA within the same droplets via fluorescence imaging using Cy3-labeled H1 and Cy5-labeled ssDNA

(**Figure 4.3A**). The observed LLPS propensity increased as a function of H1 and ssDNA concentrations, as shown by solution turbidity measurements (absorbance at 350 nm, A_{350}) (**Figure 4.3B**). Droplet formation required the presence of DNA, but not the other buffer components such as polyethylene glycol (PEG), nor was it affected by the H1 chaperone prothymosin α (ProT-α) which was shown to form a picomolar affinity complex with H1 while remaining intrinsically disordered [284] (**Figure 8.13A-C**).

To determine the effect of the disordered tails of H1 on its LLPS behavior, we truncated either the NTD or CTD or both (H1ΔN, H1ΔC, H1ΔNΔC, respectively) and assessed the capacity of these truncated H1 proteins to phase separate with ssDNA (**Figure 4.3C**). We found that H1:ssDNA LLPS is significantly impaired by deletion of the longer CTD and to a lesser extent, deletion of the NTD (**Figure 4.3D**). When both tails were removed, the globular domain alone showed minimal phase separation with ssDNA (**Figure 4.3D**). These results are consistent with the expectation that the LLPS ability of H1 is mediated by its intrinsically disordered tails.

Next, we examined how properties of the DNA may impact its tendency to phase separate with H1. Using ssDNA of different lengths ranging from 16 to 147 nt (**Figure 4.3E**), we found that longer ssDNA promotes H1:ssDNA droplet formation, even after normalizing for the total amount of nucleotides (**Figure 4.3F**). Then we compared pairs of ssDNA and dsDNA of 30 or 70 nt/bp in length ($ssDNA_{30}$ vs. $dsDNA_{30}$; $ssDNA_{70}$ vs. $dsDNA_{70}$). A_{350} measurements showed that ssDNA possesses a significantly higher capacity to phase separate with H1 compared with dsDNA of the same length and sequence (**Figure 4.3G**), in accordance with the single-molecule results shown in **Figure 4.2**. This difference is even more pronounced when we normalized the DNA concentrations to the total number of nucleotides between ssDNA and dsDNA (**Figure 8.13D**). In addition, we examined chromatin substrates, since they were shown to phase separate with H1 [291, 292] and found that ssDNA without any core histones possesses a similar ability to form droplets with H1 compared to mononucleosomes and nucleosome arrays (**Figure 4.3E,H**). All oligonucleotide sequences can be found in **Table 8.2**.

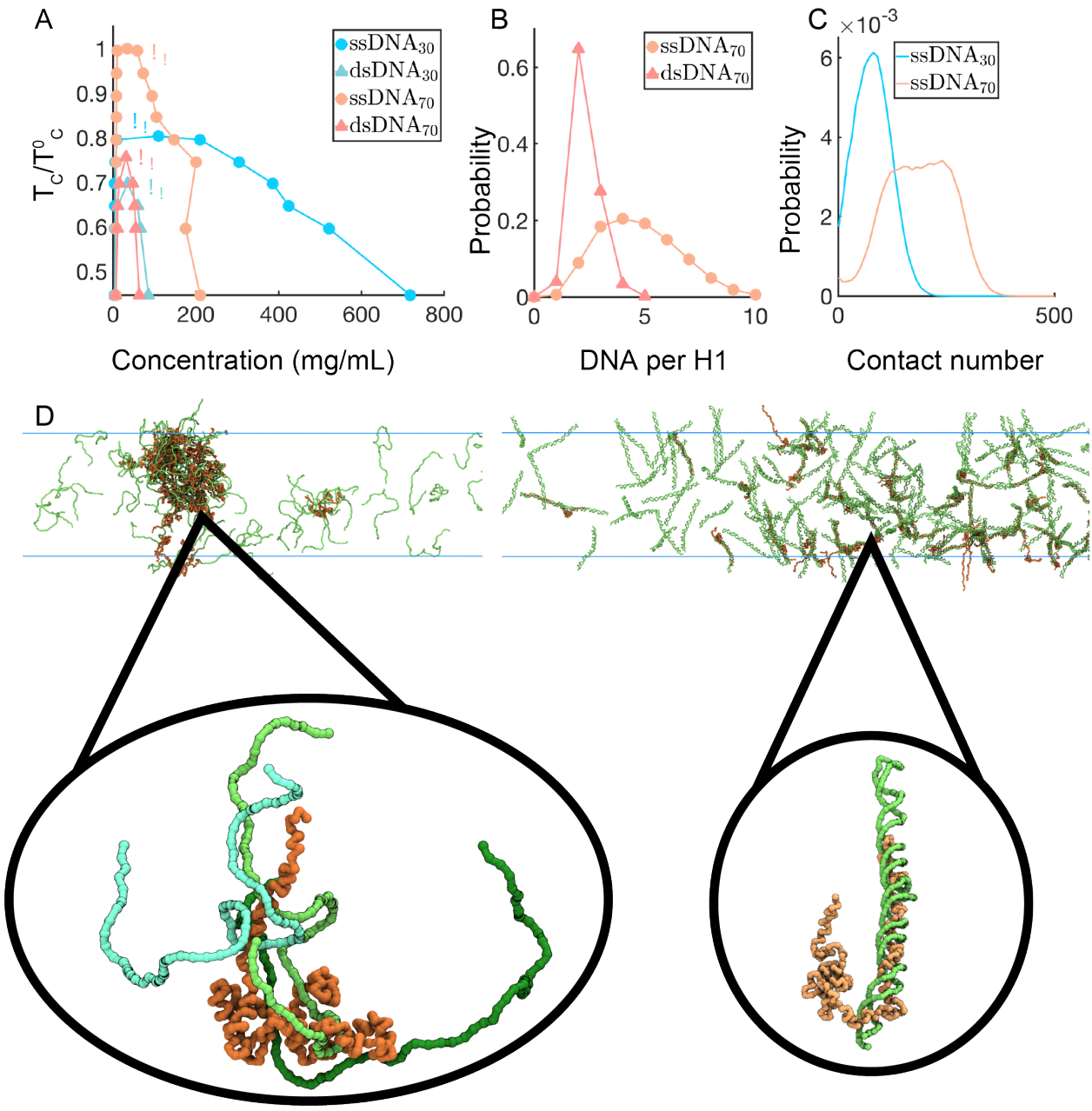


Figure 4.4. Multivalency and interaction strength influence phase separation. (**A**) Computational phase diagram as a function of simulation temperature. For each system, we calculate the upper critical temperature for phase separation (T_C). The displayed values are normalized by the TC with $ssDNA_{70}$, which is 336 K. (**B**) Probability distribution of the number of DNA bound to each H1. (**C**) Probability distribution of the number of contacts to H1 formed by a single DNA molecule. (**D**) Example configurations of the condensates formed between 70b ssDNA (left) and 70bp dsDNA (right) with H1. The insets highlight the different protein-DNA binding modes.

4.2.3 MD simulations reveal H1:ssDNA multivalent interactions contribute to enhanced coacervation

We also carried out coarse-grained MD simulations to provide a detailed structural characterization of the H1:DNA condensates[293, 294]. The force fields used in these simulations were fine tuned to capture sequence-specific protein-protein interactions and the persistence length of ssDNA and dsDNA. The simulated critical temperatures (T_c) of phase separation in different systems match the A_{350} trends recorded experimentally (**Figure 4.4A**), supporting the usefulness of the *in silico* model for mechanistic exploration. Close examination of the simulated condensed phase revealed drastic differences in their internal organization. H1 molecules coacervate with ssDNA into a much higher density than with dsDNA (**Figure 4.4B**), explaining the difference in their viscosity and kinetic rate for droplet merging. dsDNA molecules have a higher persistence length and are more rigid, rendering them more resistant for dense packing. Furthermore, H1 molecules engage in distinct binding modes with the two DNA types. As shown in the insets of **Figure 4.4D**, while they coil around dsDNA to form tight binding, the interactions between H1 molecules and ssDNA are more transient [295]. However, fewer contacts with each individual ssDNA allow each H1 to contact multiple ssDNA molecules simultaneously (**Figure 4.4C**). Such multivalent interactions are essential for their stronger phase separation potential seen experimentally in these systems than those formed with dsDNA, consistent with previous studies [296]. Moreover, in agreement with the experimental observations that the disordered tails of H1 control its phase separation, we found that the DNA interacts with H1 primarily through the CTD, although the NTD also contributes significantly to the multivalent interactions (**Figure 4.4D**).

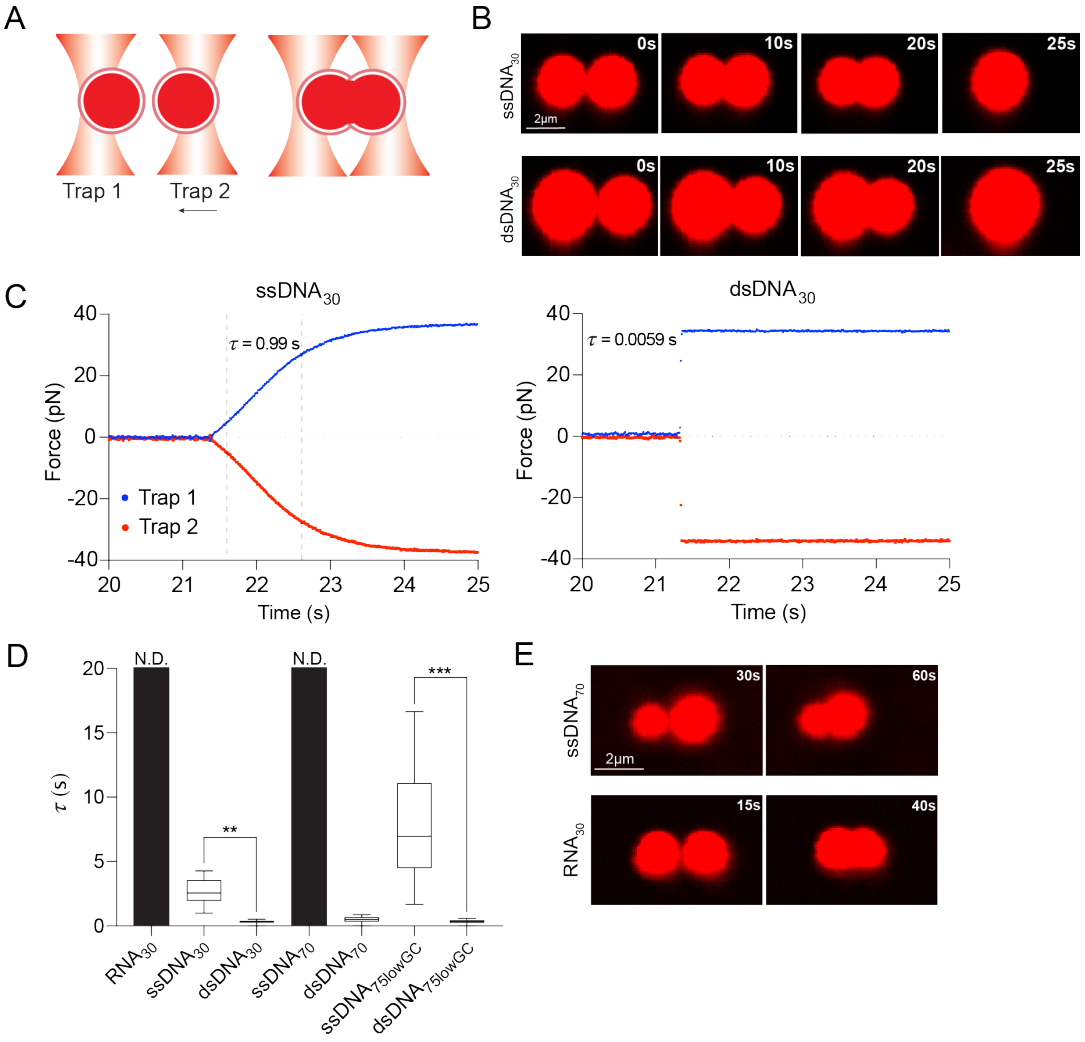


Figure 4.5. H1:DNA droplets show different material properties. (**A**) Schematic of droplets caught in optical traps pre- and post-fusion assay. (**B**) Representative images of H1:Cy5-ssDNA$_{30}$ (10 μM, 10% labeled) droplet fusion (top) and H1:Cy5-dsDNA$_{30}$ (10 μM, 10% labeled) droplet fusion (bottom). All droplet fusion experiments performed with 2.5 μM H1 unless otherwise noted. (**C**) Representative force trajectory of two droplets in each trap during fusion for H1:Cy5-ssDNA droplets (left) and H1:Cy5-dsDNA droplets (right). Same conditions as in **B**. (**D**) Average fusion time for H1 droplets with Cy5-ssDNA, Cy5-dsDNA, and Cy5-RNA (all DNA and RNA are 10% labeled). At least 12 fusion events were quantified for each condition. (**E**) 50% GC Cy5-ssDNA$_{70}$ and H1 fusion images. (**F**) Cy5-RNA and H1 fusion images.

4.2.4 H1:ssDNA and H1:dsDNA droplets exhibit distinct material properties

To further dissect the difference in physicochemical properties between H1:ssDNA and H1:dsDNA droplets, we leveraged the ability of optical tweezers to trap and manipulate protein droplets due to their different refractive index relative to the surrounding medium [297, 298]. We first generated H1:ssDNA droplets containing fluorescently labeled ssDNA for real-time visualization. Two micron-sized droplets were captured and brought into proximity by moving

one trap towards the other (**Figure 4.5A**). Consistent with the droplet's liquid nature, we observed that the two droplets eventually fused into a singular one after contact (**Figure 4.5B**). Meanwhile, the force experienced by the droplets can be recorded during the fusion process, displaying a gradual increase as the droplets were being pulled away from their respective trap center (**Figure 4.5C** and **Figure 8.14**). When fusion was completed, the merged droplet was caught between two traps, yielding a steady-state force readout. This force measurement allowed us to quantify the timescale of the fusion process, which on average lasted 2.6 s for the H1:ssDNA$_{30}$ droplets (**Figure 4.5D**).

We then performed the droplet fusion experiment with H1:dsDNA$_{30}$ droplets (**Figure 4.5B**). We noticed that dsDNA droplets fused much faster than ssDNA ones (**Figure 4.5C,D** and **Figure 8.14**). The contrast in the fusion kinetics suggests that H1:ssDNA droplets are more viscous and gel-like, whereas H1:dsDNA droplets are more liquid-like, which is supported by our MD simulation results showing different internal organizations of H1:ssDNA versus H1:dsDNA condensates (**Figure 4.4D**). To further corroborate this interpretation, we conducted fluorescence recovery after photobleaching (FRAP) experiments and found that the recovery rate for H1:dsDNA droplets was indeed significantly faster than that for H1:ssDNA ones (**Figure 8.15**).

Next, we evaluated how DNA length and base composition affect the material properties of H1 condensates. We found that a longer ssDNA (70 nt) with the same GC content (50%) as ssDNA$_{30}$ renders the droplets unable to fuse over the observation window (**Figure 4.5C**). Conversely, lowering the GC content of the ssDNA liquefies the droplet and accelerates the fusion process (compare ssDNA$_{70}$ and ssDNA$_{75lowGC}$ in **Figure 4.5D**). Deletion of the NTD of H1 also makes the droplets more liquid-like, as evidenced by a shortened fusion time and faster FRAP recovery for H1ΔN-containing droplets (**Figure 8.15**). Deletion of the H1 CTD completely abolished droplet formation, therefore no measurements could be made. Notably, we observed that H1 can also form condensates with single-stranded RNA. However, compared to ssDNA of the same length and sequence, it took much longer for H1:RNA droplets to fuse, if at all (**Figure 4.5D,E**), suggesting that H1 differentially coacervates with DNA and RNA.

Collectively, our single-molecule binding, single-droplet fusion, and solution turbidity experiments, as well as MD simulations, are all in agreement and suggest that H1:ssDNA condensates feature high viscosity and diminished internal dynamics.

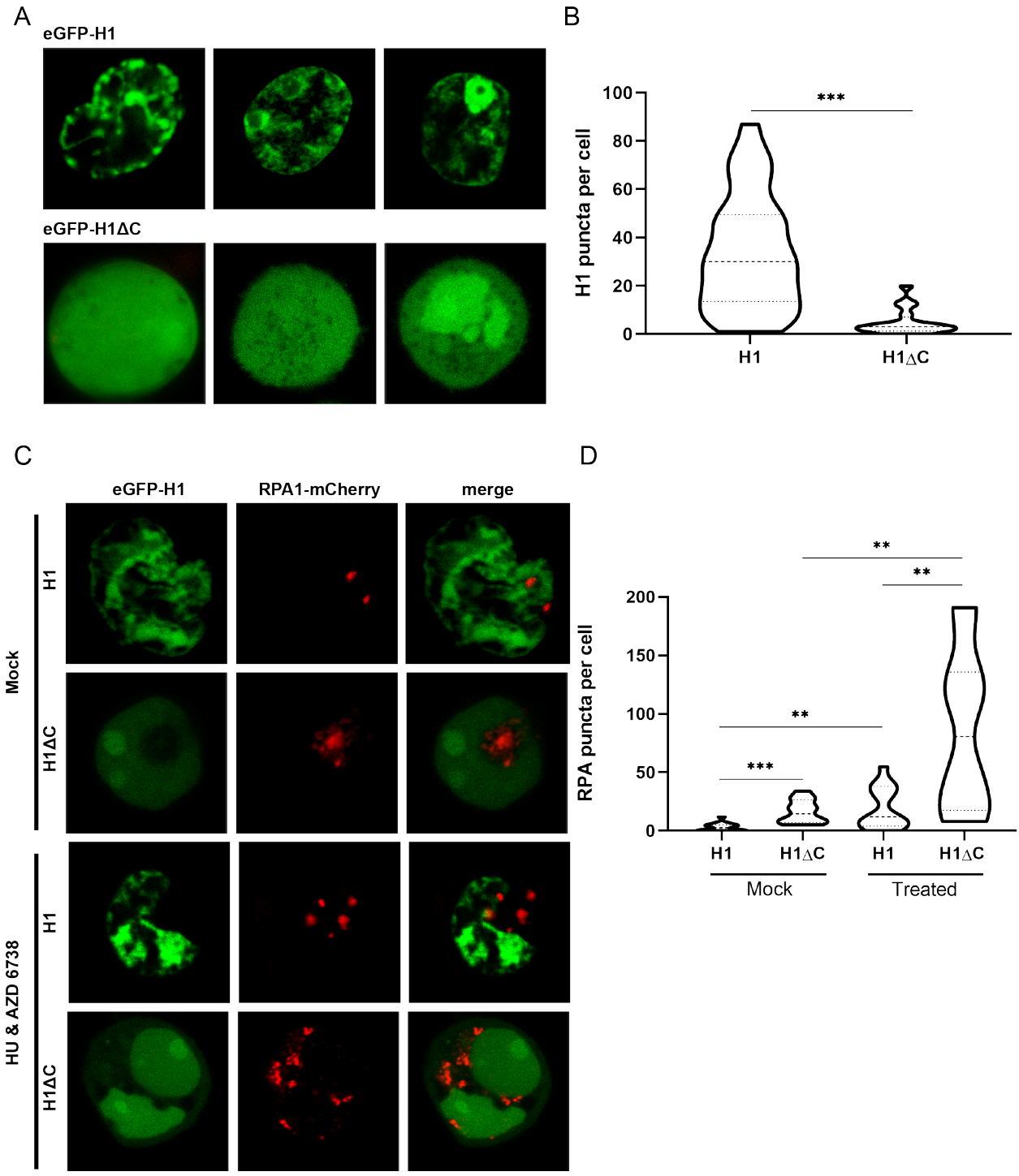


Figure 4.6. Confocal images reveal H1 and RPA droplet formation in live cells. (**A**) Representative images of HEK293T cells transfected with either eGFP-H1 (top) or eGFP-H1ΔC (bottom). (**B**) Violin plot of the distribution of H1 puncta per cell for eGFP-H1 and eGFP-H1ΔC cells. (**C**) Representative images of HEK293T cells transfected with either eGFP-H1 or eGFP-H1ΔC and RPA_{70}-mCherry, and treated with either mock or 2 mM hydrocyurea (HU) and 20 μM ceralasetib (AZD6738) for 12 hours. (**D**) Violin plot of the distribution of RPA-mCherry puncta per cell for eGFP-H1 and eGFP-H1ΔC cells after mock or treatment with HU and AZD.

4.2.5 H1 C-terminal domain mediates droplet formation in nuclear contexts

Finally, to observe the H1 localization pattern *in vivo*, we expressed eGFP-tagged H1.4 in HEK293T cells and performed live-cell confocal imaging. Cells expressing tagged WT H1 (eGFP-H1) displayed multiple puncta (**Figure 4.6A**), consistent with previously observed patterns [291]. We also created the eGFP-tagged CTD-deleted H1 construct and found that cells expressing eGFP-H1ΔC had a marked deficiency of puncta and displayed a diffused signal throughout the nucleus (**Figure 4.6A,B**), consistent with the primary role of CTD in chromatin binding [299]. Of note, we showed that recombinant eGFP-H1 can phase separate with ssDNA *in vitro* and eGFP-H1:ssDNA droplets fuse together with comparable kinetics as untagged H1 droplets (**Figure 8.16**).

To better understand the relationship between ssDNA and H1 puncta formation, we co-transfected the eGFP-H1 cells with RPA1-mCherry. In agreement with our *in vitro* single-molecule results (**Figure 4.2G-I**), H1 and RPA localization appeared to be mutually exclusive (**Figure 4.6C)**. It is possible that H1 occupies the majority of free ssDNA, while RPA only interacts with ssDNA under tension exerted by force-generating motor proteins. In accordance with this view, eGFP-H1ΔC cells contain significantly more RPA puncta than eGFP-H1 cells (**Figure 4.6C,D**), suggesting that more ssDNA sites are available for RPA binding due to the lack of H1 condensates. Furthermore, we treated cells simultaneously with hydroxyurea (HU) and the ATR kinase inhibitor ceralasertib (AZD6738). HU induces replication stress and the accumulation of ssDNA in the cell, while AZD6738 impairs DNA damage response by inhibiting RPA dissociation from ssDNA. Treatment with both 2 mM HU and 20 μM AZD6738 resulted in a significant increase in the number of RPA puncta formed in both eGFP-H1 and eGFP-H1ΔC cells, but the latter group had considerably more RPA puncta in correlation with a lack of H1 puncta **(Figure 4.6C,D)**. Together, these results suggest that the CTD of H1 is critical for H1 puncta formation in cells and that these puncta exclude RPA from binding to ssDNA.

4.2.6 Conclusions

In this work, we provide evidence that the capacity of H1 to form phase-separated condensates is dependent on the physical nature of its binding partner and that ssDNA is preferred over dsDNA for this function, which is modulated by nucleic acid length and sequence. Since the physical properties of chromatin are involved in regulating all DNA-associated processes, the ability of H1 to sense and partition different forms of nucleic acids suggests that it can serve as an organizational hub of various nucleic acids leading to downstream signaling in the nucleus. In this context, H1 may serve to partition nucleic acids of different physical natures into distinct sub-compartments in the cell. Our results may help explain how H1 can simultaneously cause both global changes in chromatin structure and specific changes in gene expression [300]. More generally, different phases that protein-nucleic acids can form (liquids, solids, gels, glasses) [301] are likely applicable to other proteins.

Our work both *in vitro* and in cells suggests that H1 and RPA compete for binding to ssDNA. H1-induced phase separation of ssDNA– to the exclusion of other potential binders such as RPA – may allow for a complex organizational structure in the nucleus leading to ssDNA protection and favorable microenvironments for DNA replication and repair, where significant amounts of ssDNA are generated. Indeed, H1 has been shown to inhibit homologous recombination and its depletion leads to altered sensitivity to DNA damage [302, 303]. These phenotypes have until now been largely attributed to the established role of H1 in chromatin compaction. Our data presented here provide a new perspective via H1:ssDNA condensate formation, which may be

prevalent given the high nuclear concentration of H1. It did not escape our attention that H1:RNA interactions are also unique from H1:DNA interactions and that, given the abundance of ssRNA, the observed gel- or solid-like H1:RNA condensates could have important biological implications.

We further demonstrate that this property of H1 is mediated, at least in part, by its intrinsically disordered C- and N-terminal domains. Human H1 subtypes diverge in homology in their CTD composition and while these variants have been shown to have distinct abilities to bind and compact chromatin [280], the extent of this functional variability beyond these roles remains unclear. We show here that the CTD of H1 is a key mediator of LLPS. It is therefore possible that the divergent CTDs in H1 variants may attenuate their ability to phase separate with different nucleic acids and result in distinct roles for each subtype in the compartmentalization of the nucleus. The experimental and computational platforms established in this study pave the way for a systemic interrogation of the effects of H1 subtypes, disease-associated mutations, and posttranslational modifications on H1:DNA interaction and cellular behavior. The possibility emerges for a highly regulated network of nuanced organizational activity in the nucleus that can modulate the overall chromatin landscape. Our work contributes to the understanding of the multifaceted nuclear functions of linker histones.

CHAPTER 5. Other chromatin biology interests

In addition to the major players of PRC2 and histone H1, the platforms developed in this thesis are amenable to studying various epigenetic systems. Below, I will discuss other projects for studying chromatin biology at the single molecule level. While some of these projects are incomplete, they are great starting points for additional exploration and demonstrate the wide array of targets we can study using these single-molecule platforms.

5.1 Histone glycation

Adapted from Zheng et al. [304]

Cellular proteins continuously undergo non-enzymatic covalent modifications (NECMs) that accumulate under normal physiological conditions and are subjected to changes in cellular environments. Non-enzymatic chromatin modification, such as histone glycation, has been identified in cells, causing DNA damage in diabetes and cancer patients [305]. Histone proteins have exceptionally long half-lives, making them highly susceptible to these modifications *in vivo*. One major form of glycation is caused by the small molecule methyglyoxal (MGO), a byproduct of anaerobic glycolysis that non-enzymatically reacts with thiol and amino groups in proteins and DNA [306, 307]. In this section, I will discuss how we utilized force microscopy to assess the role of glycation on chromatin unfolding and condensation. We propose that glycation enhances the stability of mononucleosomes through histone-DNA or histone-histone crosslinking, which is supported by bulk biochemical assays.

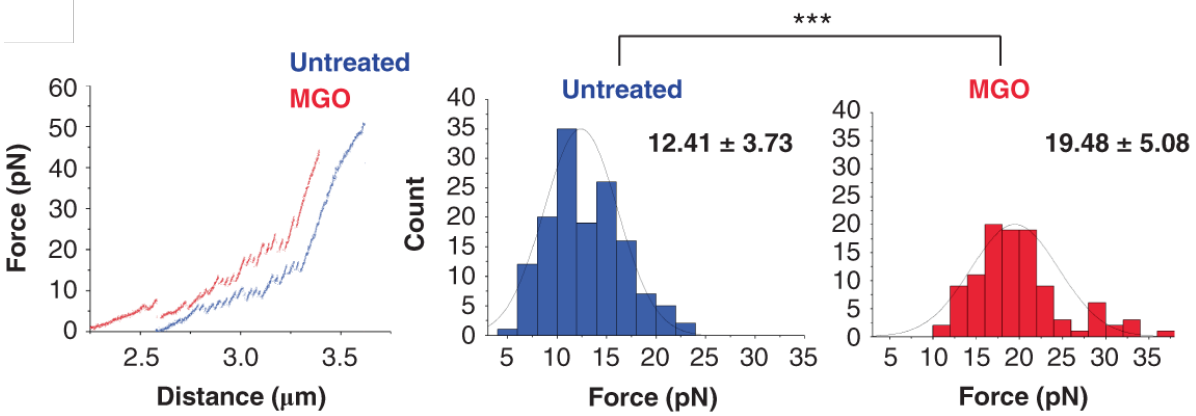


Figure 5.1. Single-molecule MGO-treated nucleosome array unfolding assay. Example traces of nucleosome array unfolding for MGO-treated and untreated arrays. The average transition force for MGO treated arrays was significantly higher.

A detailed analysis of histone glycation *in vitro* and *in vivo* shows global ramifications of histone post-translational modifications, the assembly and stability of nucleosomes, and chromatin architecture [304]. Specifically, the enzyme DJ-1, was identified as a potent deglycase, and is overexpressed in breast cancer tumors. One specific question was how glycation adducts affect nucleosome stability and chromatin architecture. Glycation was found to induce DNA-histone cross-linking on individual nucleosomes. Therefore, we hypothesized that this could affect the architecture of chromatin on both a local and global scale. To examine this

possibility, we utilized reconstituted nucleosomal arrays composed of 12 repeats of 601 DNA. To gain a quantitative measurement of NCP stability in an array context, we used a single-molecule high-resolution optical tweezers assay. This assay measures the precise force required to unravel each nucleosome in an array by tethering a single array between two beads – each held by a laser trap – and applying force to pull them apart. Performing this assay on untreated and MGO-treated arrays indicated that more force was required to unravel nucleosomes in treated arrays, strongly suggesting that on a single nucleosome level MGO-treated arrays are more stable, presumably due to additional intra-nucleosomal bonding.

5.2 Oncohistones

Mutations in epigenetic pathways are found in various different cancers. Classic oncohistone mutations have been previously identified including H3K36M and H3K27M. However, until recently, the high prevalence of somatic histone mutations in all cancer types (4%), on the level of BRCA2 or NOTCH1, had not been characterized. Multiple studies [308, 309], including a study from the Allis group uncovered the vast array of histone oncogenic mutations found in patient tumor samples. Many of these mutations are found not only on the N-terminal H3 tail, but also in the globular domains of the core histones, especially in the acidic patch of H2A or H2B, disrupting the H2B-H4 interface, and affecting the function of SWI/SNF remodeling complex [310]. Given that many of these globular domain mutations might affect nucleosome structure, we hypothesized that their presence might alter nucleosome unfolding dynamics. Therefore, single-molecule force microscopy would be a key tool to identify how oncohistones contribute to structural alternations in chromatin arrays.

5.2.1 Oncohistones in the globular histone domains

Hypotheses generated in conjunction with Allis lab members Ben Nacev, Agata Lemiesz, and Lijuan Feng.

The first region of the nucleosome they identified included mutations in the acidic patch, a region between H2A and H2B, which may have implications in chromatin condensation and folding. The second area they identified were mutations in the H3 and H4 globular domains (Sin-mutations) which have previously abrogated the need for SWI/SNF chromatin remodelers in yeast, hypothesizing that these mutations may play a role in chromatin folding and stability. The third group consists of mutations that might alter nucleosome structure within the globular core. While these mutations are distal in primary sequence, they are closely spaced in the folded nucleosome structure. For example, oncogenic mutant H2B (E76K/Q) has been shown to destabilize nucleosomes and perturb the local structural arrangement of H4 R92 [311]. Given these hypotheses, I proposed testing the stability of these oncogenic mutations in the context of chromatin. I hypothesized that the presence of these mutations would change how chromatin unfolds under force.

'Sin' mutants (H3.1 E105K, E105Q) were originally identified in yeast by virtue of their ability to abrogate the need for SWI/SNF. The mutations fall in two relatively small regions in the globular domains of H3 and H4. Within the tumor mutation dataset, H3 E105 is the single most mutated residue and corresponds to the position of one of the original yeast mutants. Therefore, the Allis lab hypothesized that this class of mutations may make nucleosome positioning less stable (i.e. more susceptible to sliding), which could be determined by single-

molecule force experiments of nucleosome-array unwinding. Furthermore, 'Sin' mutants of H4 residues (H4R45 and H4G42) originally discovered in yeast are also identified in cancer, including HG42R/V in metastatic prostate cancers. H4G42R/V might enhance the histone eviction by changing the acetylation of a neighboring H3 lysine. Given that these mutants might disrupt the histone-DNA interactions, and therefore are less stable and bypass the need of remodeler (SWI/SNF), they may have lower stability in nucleosome arrays.

Interestingly, the most highly mutated residues in H2B (H2BE71, E76) and H4 (H4R92, D68) occur at the region of the H2B-H4 tetramer-dimer interface. Based on structural predictions, H2BE71 forms a salt bridge with H4K91, and H2BE76 can form a hydrogen bond with H4R92 and H4D68. The Allis lab hypothesized that mutations within such an important part of the nucleosome will disrupt nucleosome integrity leading to global disruption in the chromatin structure [310]. FRAP experiments of cells overexpressing eGFP tagged H2BE71K and H2BE76K showed the increased mobility of these oncohistones. Additionally, nucleosomes containing mutated H2BE71K/Q; H2BE76K/Q and H4K91E/R showed a decrease in temperature of dimer dissociation in thermal shift assays. Given these results, H2BE71K/Q, H2BE76K/Q, and H4K91E/R may influence the stability of the nucleosome arrays in single-molecule force experiments.

5.2.2 Oncohistones on the histone H3 tail

Another class of oncohistone mutants that were identified include arginine mutants on the H3 tail (R2, R8, R26). The Allis and Muir labs have shown that these arginine mutants cause changes *in cis* for H3K27 methylation. I am interested in testing how these mutants affect PRC2 binding and nucleosome engagement. Despite the difficulty in drugging protein-protein interactions, the prevalence of oncohistones and their interactions with key epigenetic machineries provides a unique and selective avenue for cancer therapeutics. The study of these modifier-oncohistone interactions could uncover novel mechanisms and lead to very specific targets in oncology.

Based on analysis of tumor samples, histone H3 N-terminal arginine residue mutations occur at frequencies similar to mutations at residues that harbor 'classical' oncohistone mutations (i.e H3K27, H3G34, and H3K36). Given the proximity of these arginine residues to sites of key regulatory PTMs (i.e. H3K4 and H3K27), the Allis lab hypothesized that mutating the arginine might disrupt the writing of histone marks at these residues ([310] and unpublished studies). They expressed transgenic H3.1 that harbors mutations at candidate residues (H3.1 R2C, R8C, R17C, or R26C). Cysteine substitutions were selected due to the prevalence of this particular substitution in tumors and an R17 mutation was included as a 'negative' control since that residue is not commonly mutated in tumors. They observed that expression of H3.1 R26C completely abrogates H3K27me3 on the transgenic protein as do H3.1 R2C and H3.1 R8C (but not R17C). The effects of H3.1 R2C and H3.1 R8C were particularly striking, given the relatively large distance between the altered residue and H3K27. Given these effects, we hypothesized that these mutations may alter the biophysical properties of PRC2-chromatin engagement in an optical tweezers set-up.

5.2.3 Oncohistone preliminary results

Given these hypotheses, we set out to test the biophysical stability of mutated histones in nucleosome arrays using optical tweezers. Cloning and protein expression were done in collaboration with a visiting summer student to identify biophysical differences in nucleosome array structure. The following plasmids were donated from the Muir lab, amplified, and frozen into glycerol stocks: H2B E71K/Q, H2B E76K/Q, H3 E97K, H3 E105K/Q, H4 G42R/V, H4 K91E/R. H3.1 R17C and R26C were also cloned.

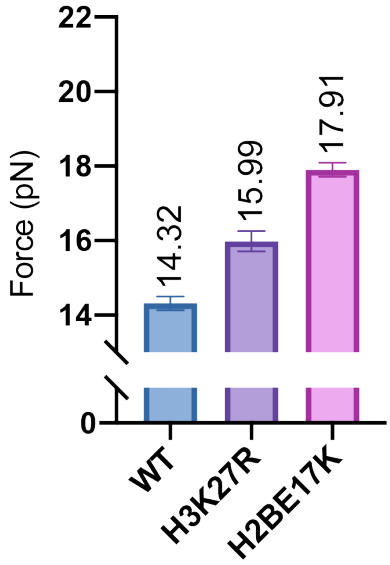


Figure 5.2. Transition forces for oncohistone 12-mer nucleosome arrays.

H2B E71K was successfully purified and formed into nucleosome arrays (see methods in Chapter 2). Force-distance curves were acquired on the optical tweezers by pulling at a constant velocity (0.1 μm/s), and contour length change and force of inner wrap transitions were calculated (using the force-extension analyzer). We hypothesized that this mutant would destabilize nucleosome arrays, lowering the force of inner wrap nucleosome transitions. The force data acquired for wild type controls by the summer student was not consistent with previous measurements taken on wild type nucleosome arrays. Therefore, the force values for H2BE71K would need to be repeated to provide more conclusive data on whether H2B E71K altered nucleosome array stability. Furthermore, it is possible that these mutations might affect more nuanced changes in the outer wrap unfolding or tetranucleosome dissociation that are not captured by our current analysis software. Future work on this mutation, as well as other oncogenic histone mutations, will be needed to probe whether the presence of these mutants changes the biophysical properties of chromatin structure.

5.3 Chromatin readers (ENL)

Hypotheses generated in collaboration with Leah Gates and Liling Wan.

Eleven-nineteen-leukemia protein (ENL), an acetyl-lysine chromatin reader, has been found to be critical for disease maintenance in acute myeloid leukemia [312]. ENL recruits

transcriptional machinery, which promotes oncogenic gene programs that sustain the leukemia stem cell state. ENL contains the YEATS domain, a chromatin binding domain, that is found in a few acetyl-lysine readers in yeast and humans. The YEATS domain has been structurally characterized as a selective reader for acetylation on histone H3 (ex: H3K9ac, H3K18ac, H3K27ac) [313]. Recently, additional cancer-associated mutations in the ENL YEATS domain were found to confer a gain-of-function in transcriptional activation in pediatric kidney cancer. Specifically, the Allis lab has found that these ENL mutants exhibit enhanced chromatin occupancy at selective genomic loci and that histone acetylation binding by the YEATS domain only partially contributes to this recruitment [312]. To that end, they demonstrated the ability of these ENL mutants to form distinct puncta *in vivo* suggesting that cancer-associated mutations promote self-reinforced aggregation and recruitment at top target sites [314]. Given these findings, our goal is to determine if purified ENL proteins have the same self-association promoting properties and further investigate this self-aggregation mechanism. We aim to use the combination of optical traps and confocal fluorescence microscopy to observe ENL binding to poly-acetylated chromatin.

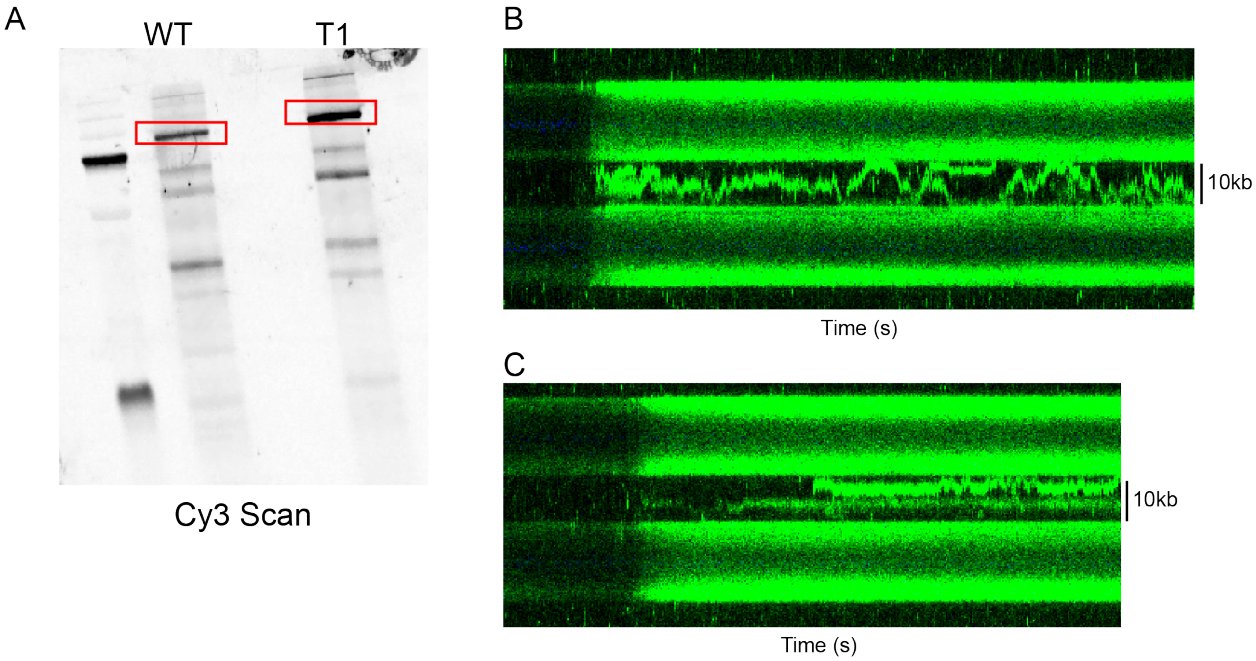


Figure 5.3. Tracking ENL self-association at the single molecule level. (**A**) Non-specific labeling of WT and T2 mutant ENL by Cy3-NHS. (**B**) Example kymograph of Cy3-WT binding to a Poly-Ac H312-nucleosome array. (**C**) Example kymograph of Cy3-T1 binding to a Poly-Ac H3 12-nucleosome array.

We hypothesized that this phenomenon could be tracked *in vitro* using single-molecule methods. We set out to determine if purified ENL proteins have the same self-association promoting properties to further investigate this self-aggregation mechanism. We used a combination of optical traps and confocal fluorescence microscopy to observe ENL binding on poly-acetylated chromatin. To do this, we tethered 12-nucleosome arrays assembled with poly-acetylated H3 nucleosomes. ENL was non-specifically labeled using Cy3-NHS dye to label all free amines. We observed that wild type and mutant (T1) ENL bind to poly-acetylated chromatin

(**Figure 5.3**). The proteins appear to diffuse over large regions of chromatinized and non-chromatinized DNA. Given these preliminary observations, dynamic ENL binding on chromatin and DNA might be an interesting avenue to explore. To do this, we would need to site-specifically label ENL with a single fluorophore to better characterize these interactions.

5.4 Linker histone force data

In collaboration with Wola Osunsade and the David Lab.

Human linker histone H1 variants have been purified *in vitro* and characterized by their compaction abilities [280]. One remaining question is to understand how each of these variants contributes to chromatin array stability. To do this, we first formed 12-nucleosome arrays using a stoichiometric ratio of one histone H1 per 601 localization site. We then characterized the unfolding trajectories of H1-bound chromatin. So far, I have tested somatic histone variants H1.1, H1.2, H1.4, and H1.5. Using our force-extension analyzer software, we have quantified the changes in contour length and force for each of the tested variants, finding no significant differences. However, changes in the lower force regions are inaccessible by our software and may display different stabilities in tetranucleosome dissociation or outer wrap unfolding. Qualitatively, we notice some changes in the low force region for variant H1.2 (**Figure 5.4**). Overall, a more in-depth analysis of the lower force regime is needed to interpret more nuanced differences in chromatin stability between H1 variants.

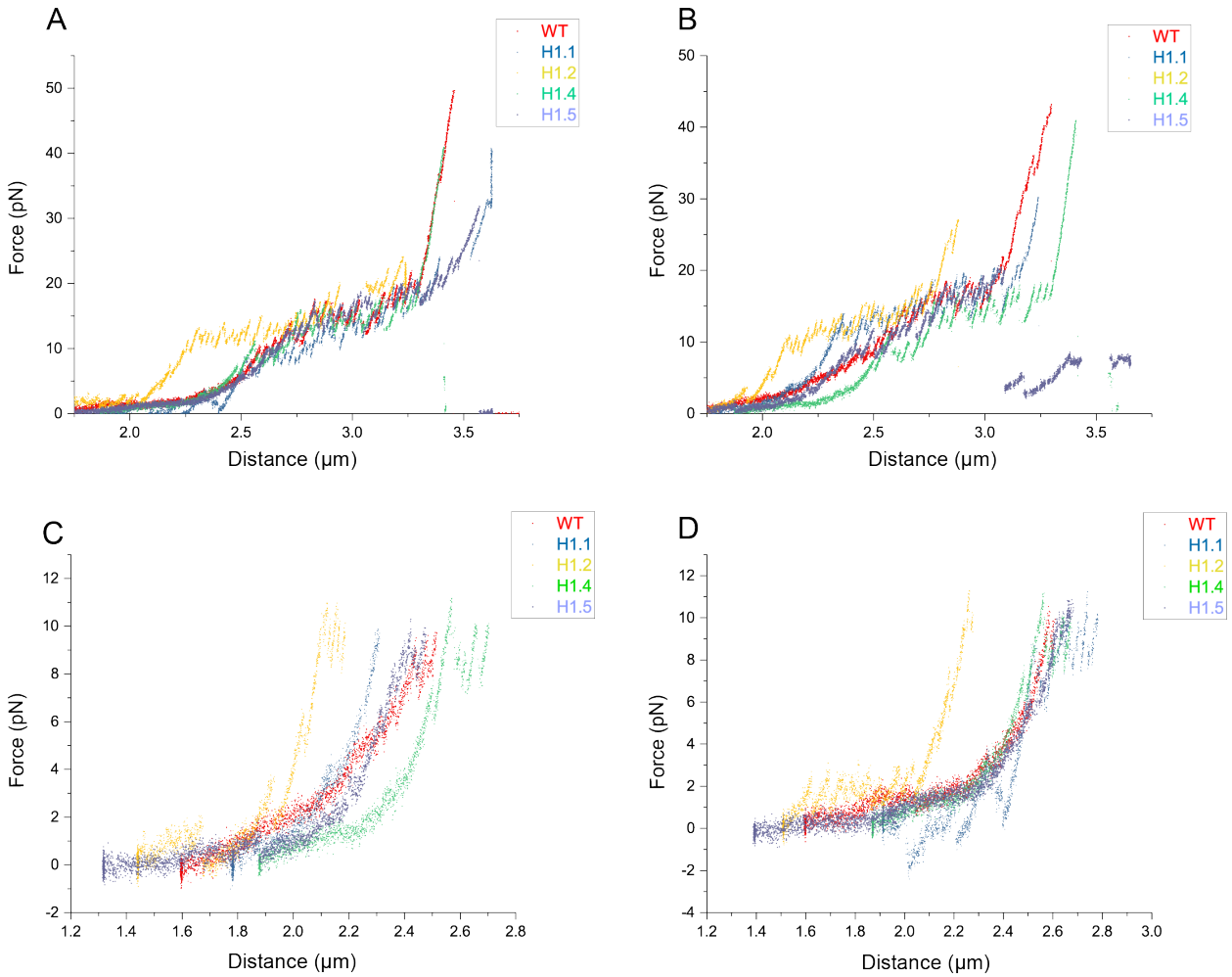


Figure 5.4. Example traces of 12-mer nucleosome array unfolding with H1 variants. (**A-B**) Example traces of 12-nucleosome arrays in the presence of different H1 variants. (**C-D**) Example traces from A zoomed in to the low force region.

5.5 PRC2-nucleosome cryo-electron tomography

In collaboration with Ayala Carl, Santiago Espinosa and the Alushin Lab.

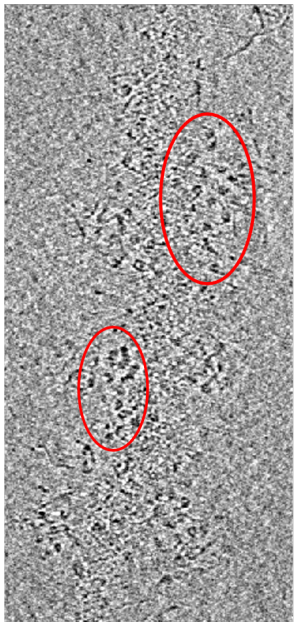


Figure 5.5. Example tomograph of PRC2-bound 12-nucleosome arrays. Red circles highlight the 12-mer nucleosome arrays that could be identified.

Given our single-molecule data on PRC2-nucleosome engagement (Chapter 3), we hypothesized that PRC2 may have the ability to locally compact chromatin. We demonstrated this property using negative stain electron microscopy [96]. Visualizing PRC2 engagement of long chromatin substrates using higher resolution techniques such as x-ray crystallography or cryo-electron microscopy is difficult due to heterogeneity in chromatin conformation and the inability to average structural data. One alternative technique, cryo-electron tomography, allows the observation of 3D structures without averaging by collecting electron density at different grid angles. We aimed to visualize PRC2 binding to 12-nucleosome arrays using this methodology. In a collaboration with the Alushin lab, freezing conditions for PRC2 and 12-nucleosome arrays were optimized to avoid destruction of the arrays upon flash-freezing the grids. Images taken on the Arctica microscope showed some individual nucleosome array molecules but failed to show recognizable PRC2 density. More optimization in the future may yield the first structure of PRC2-chromatin engagement on substrates longer than a dinucleosome [121], and give additional insights on PRC2 binding configurations on chromatin.

5.6 PRC2-DNA binding interactions

Adapted from Lin et al (in review).

Our computational collaborators worked to characterize the different conformational states of the PRC2 complex and its interaction with DNA. They found that PRC2 binds DNA via the EZH2 subunit, which induces DNA bending due to the protein interface's intrinsic curvature. They also found that two copies of PRC2 were shown to bend DNA cooperatively via an allosteric mechanism to induce DNA looping. Furthermore, when PRC2 formed in complex with

AEPB2, it bends DNA less significantly, suggesting a weakening of DNA binding and bending upon AEBP2 association.

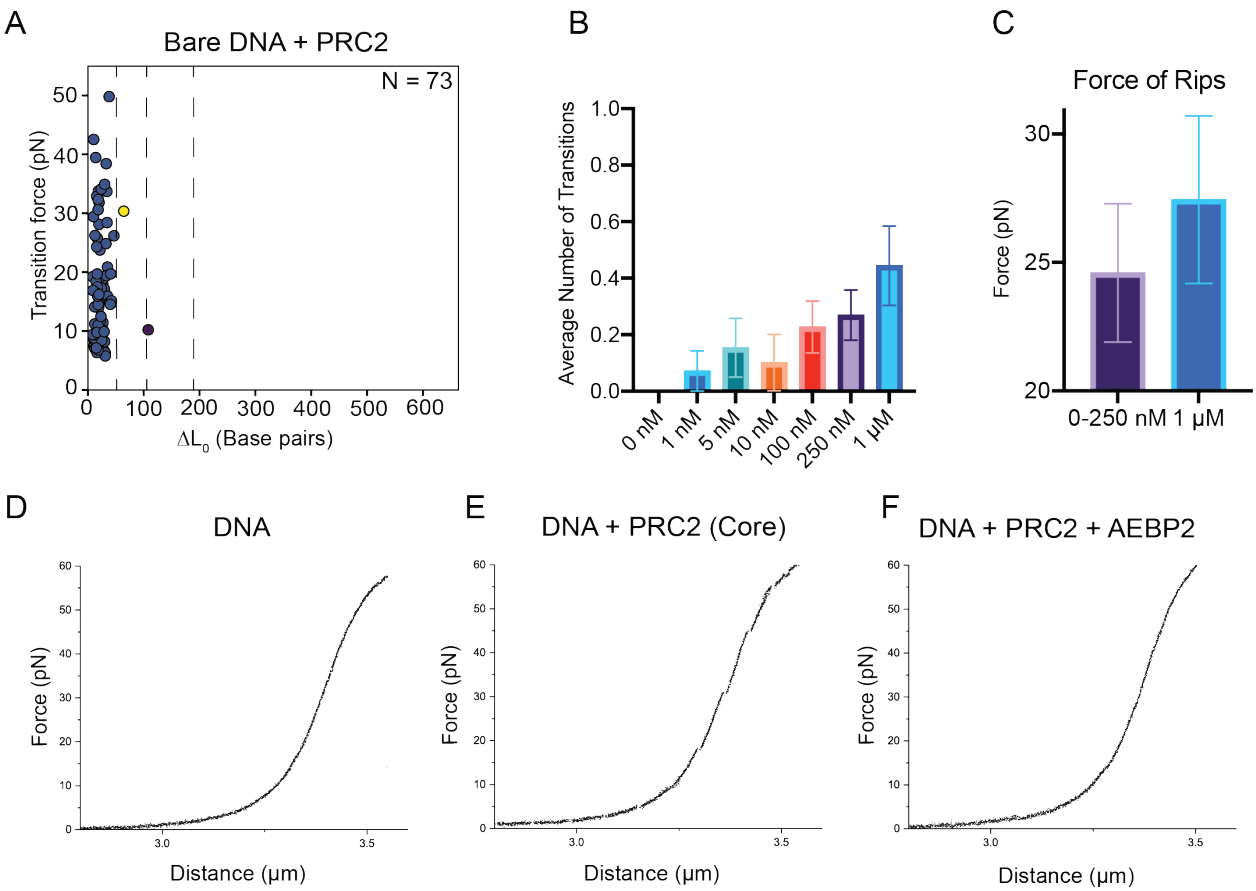


Figure 5.6. Quantification of PRC2-DNA binding transitions. (**A**) Cluster plot of PRC2-DNA rupture forces. (**B**) Histogram of average number of transitions over a series of PRC2 concentrations. (**C**) Comparison of transition forces between two different PRC2 concentrations. (**D-F**) Example traces for Bare DNA, DNA and PRC2, and DNA with PRC2 and AEPB2.

To complement these computer simulations, I performed single-molecule force spectroscopy experiments to characterize PRC2-DNA interactions. Specifically, we incubated a ten kbp-long bare 601-sequence DNA with 500 nM PRC2 core complex under physiological salt conditions. The DNA ends were then pulled by optical tweezers at a constant velocity of 0.1 μm/s. Force extension curves were recorded to measure the pulling force as a function of the distance between the two ends. As shown in **Figure 5.6**, multiple rupture events with a sudden drop in the pulling force can be seen in the force-extension curve. Notably, when the DNA was incubated with PRC2 and AEBP2, rupture events can no longer be detected. Results from the force spectroscopy experiments are consistent with the free energy profiles of PRC2-DNA binding simulations. Additionally, we tested PRC2-DNA binding over a range of concentrations, finding that when concentrations increase, so do binding events. The average force of these binding events is higher; however, the difference is not significant, and more data will need to be collected to solidify the trend. Overall, this data indicates that PRC2 associated proteins may

weaken the DNA binding affinity of the core complex and shift the complex towards specific chromatin regions for targeted histone modification.

CHAPTER 6. Discussion: perspective and outlook

Single-molecule force spectroscopy is a powerful tool to analyze the architecture and interactions of large macromolecular assemblies that are refractory to high-resolution structural interrogations. Using single-molecule methods, combined with alternative techniques, I have explored the specific mechanisms of how effectors engage with chromatin and nucleic acid substrates *in vitro* to broaden our knowledge about their role in the cell. By establishing single-molecule methods to study chromatin factors and their dynamics, I have enabled a large array of future studies to address open questions in the field about specific chromatin-associated proteins and their mechanisms. In this chapter, I will propose multiple avenues for exploration enabled by the technological creations and biological insights of this thesis.

6.1 Further investigation of PRC2 and other chromatin modifiers

Using our single-molecule chromatin platform we identified a multiplicity of PRC2-chromatin binding modes using force spectroscopy. While we have obtained evidence of PRC2-nucleosome engagement through force data, we have not directly visualized PRC2 in our system. Using correlative force-fluorescence microscopy with site-specifically labeled PRC2 would allow direct visualization of protein binding and disengagement. Furthermore, we could calculate additional mechanistic parameters, such as dwell time, on/off rates, as well as track any diffusive movement. Currently, the confocal fluorescence microscope is diffraction limited to around 250 nm. Addition of super resolution technology using a stimulated emission depletion (STED) laser would allow more accurate localization of PRC2 (<35 nm). We could harness this technology to observe PRC2 binding on labeled nucleosomes to view direct engagement by PRC2 or at least eviction of individual fluorescently labeled histone octamers.

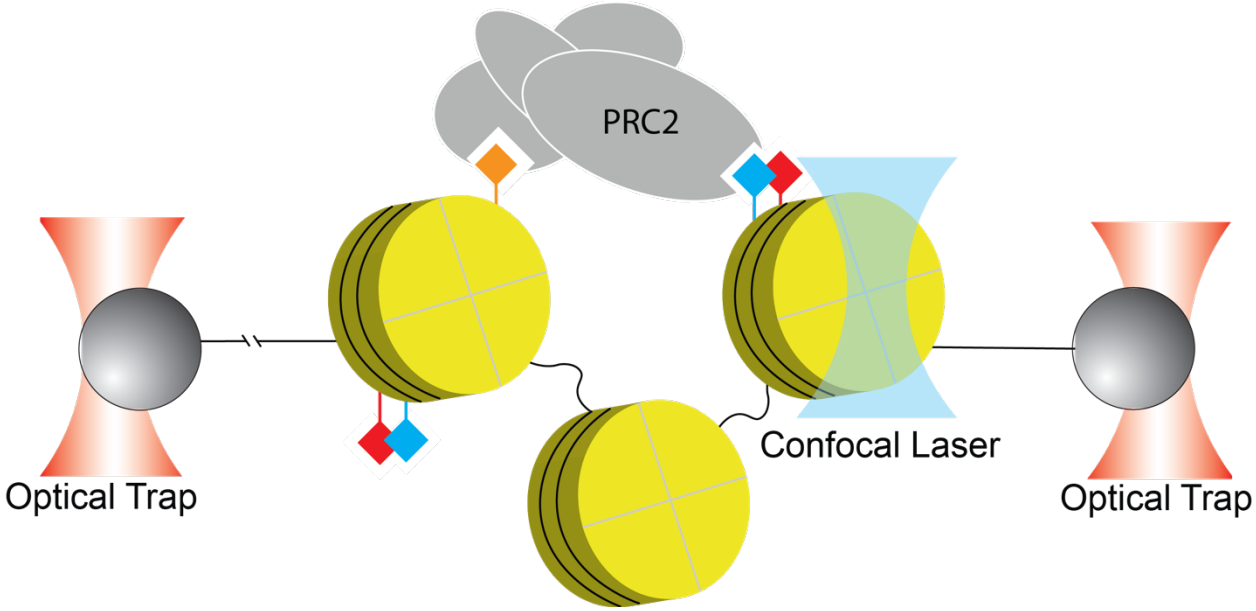


Figure 6.1. Experimental setup for PRC2-binding of active and repressive marks. PRC2 non-adjacent bridging interactions could explain the role of bivalent chromatin and nucleosomes modified with both active and repressive marks on opposing H3 tails.

In addition to just studying PRC2 alone, we could observe how PRC2 engages with other factors, including linker histone H1 (discussed below), or histone post-translational modifications, such as H3K36me3. As previously discussed, H3K36me3 and H3K4me3 are both activating marks found on opposing tails to H3K27me3 that can inhibit PRC2 [154, 315]. Our single-molecule platform may be able to distinguish activating-mark inhibition of PRC2. One mechanism of action for inhibition of PRC2 by these opposing marks could be through non-adjacent bridging of nearby modified-nucleosomes or bivalent domains. A potential experimental setup (depicted in **Figure 6.1**) could outline a role for simultaneous PRC2 binding to active and repressive marks.

Our platform could also be used to decipher any antagonistic effects of H3K27me3 erasers, such as UTX and JMJD3. Overall, there are ample possibilities for studying binding interactions and mark spreading using combined fluorescence-force microscopy for a whole host of chromatin-associated factors.

6.2 Linker histone H1 variants

In Chapter 4, I detailed the investigation of a specific linker histone variant H1.4. However, there are 11 different H1 variants in humans, each with an undefined role, despite differences in sequence, especially in the intrinsically disordered regions. Methodologies developed in this thesis to study phase separation, condensate formation, and material properties of droplets could be extended to test all the H1 variants. Changes in how the variants form condensates might lead to important insights on the differences between them. For example, if one variant formed a more gel-like droplet compared to the liquid-like condensates of H1.4, it could potentially be involved in a more stable phase separated compartment in the nucleus. These different compartments could regulate various processes such as DNA replication, RNA synthesis, or transcription factor binding.

Linker histones are frequently mutated in cancers including lymphoma, leukemia, and colorectal cancer [232]. The expression levels of different histone variants are also altered in cancer cells. Linker histones can also be mediated by various cancer-related proteins, which can govern linker histone affinity to the nucleosome. For example, the C-terminal charge alternation from basic to acidic in mental impairment syndrome occurs through frameshift mutations that might be further studied using these techniques [316]. By continuing to elucidate the role of these variants, we can uncover how they may affect disease.

The role of post-translational modifications in regulating phase separation has been identified for multiple intrinsically disordered proteins [317], such as HP1 [83]. The linker histones are extensively phosphorylated, in addition to containing other PTMs, which may affect how different variants contribute to phase separation properties [244, 248, 249]. It would be interesting to interrogate how PTMs on H1 add an additional layer of regulation.

6.3 Linker histones and PRC2

One interesting interaction to interrogate further is the connection between different chromatin-associated factors. Specific to this thesis, there are numerous links between linker histones and PRC2, facilitating gene expression patterns and heterochromatin maintenance. Collectively, depleted H1 content leads to reduction in histone modification levels such as H4K12ac and H3K27me3 [240], highly specific changes in gene expression, and loss of DNA

methylation in imprinting control regions. Biochemically, the presence of H1 increases the methylation efficiency of PRC2 on chromatin substrates [136] through a direct physical interaction, with the H1.2 variant specifically binding to H3K27me3 nucleosomes *in vitro* and H3K27me3 stimulating H1.2 occupancy in cells, suggesting a positive feedback loop reinforcing heterochromatic silencing [318].

PRC2 has also been shown to catalyze H1.4K26me, which acts as a binding target for HP1, facilitating heterochromatin formation [319]. Studying the interactions between PRC2 and H1 might lead to discoveries about heterochromatin maintenance. On this front, recent studies have shown that H1 stimulates the activity of PRC2-AEBP2 on di-nucleosome substrates while inhibiting the activity of NSD2, a H3K36 dimethyltransferase [28]. This mechanism most likely leads to a loss of H3K27me3 and a gain in H3K36me2 in H1 depleted chromatin *in vivo*, demonstrating the importance of H1 stoichiometry in the distribution of these opposing repressive and activating marks.

Segregation of H3K27me3 and H3K36me3 marks was further demonstrated in cells without variants H1.2 and H1.4. These knockout cells have also shown a phenotype with a reduction of H3K36me2, a mark already known to activate transcription and antagonize PRC2 [320]. H1 depletion causes changes in chromatin structure leading to de-repression of T-cell activation genes [28], while lymphomagenic mutations in variants H1.2 and H1.4 can reverse silencing of PRC2 targets and upregulate genes activated by NSD2 in B-cells and T-cells [217]. These variants also maintain the inactivation of primitive stem cell genes that are silenced during lineage specification and differentiation.

Using our single-molecule chromatin platform, it would be interesting to observe PRC2 binding in the presence of histone H1, and vice versa. Given that variant H1.2 binds H3K27me3, we could test how a methylated array in the presence of both H1.2 and PRC2 affects chromatin engagement by these binders. Methylated chromatin and PRC2 could also be added to our phase separation droplet experiments containing H1 to detect any differences in condensate formation.

6.4 Phase-separated protein-nucleic acid condensates

As discussed in Chapter 2, biophysicists are expanding the array of technologies for studying phase-separated condensates [108, 110, 111]. Harnessing optical tweezers for studying the material properties of phase separated droplets can expand the questions we can ask about the formation of condensates and their properties [110]. Recent studies have suggested phase separation in the nucleus can span beyond liquid-liquid condensates to gel and solid like chromatin scaffolds [91]. Having tools that can quantitatively measure physiochemical properties of droplets can help distinguish the mechanisms and functions of these condensates.

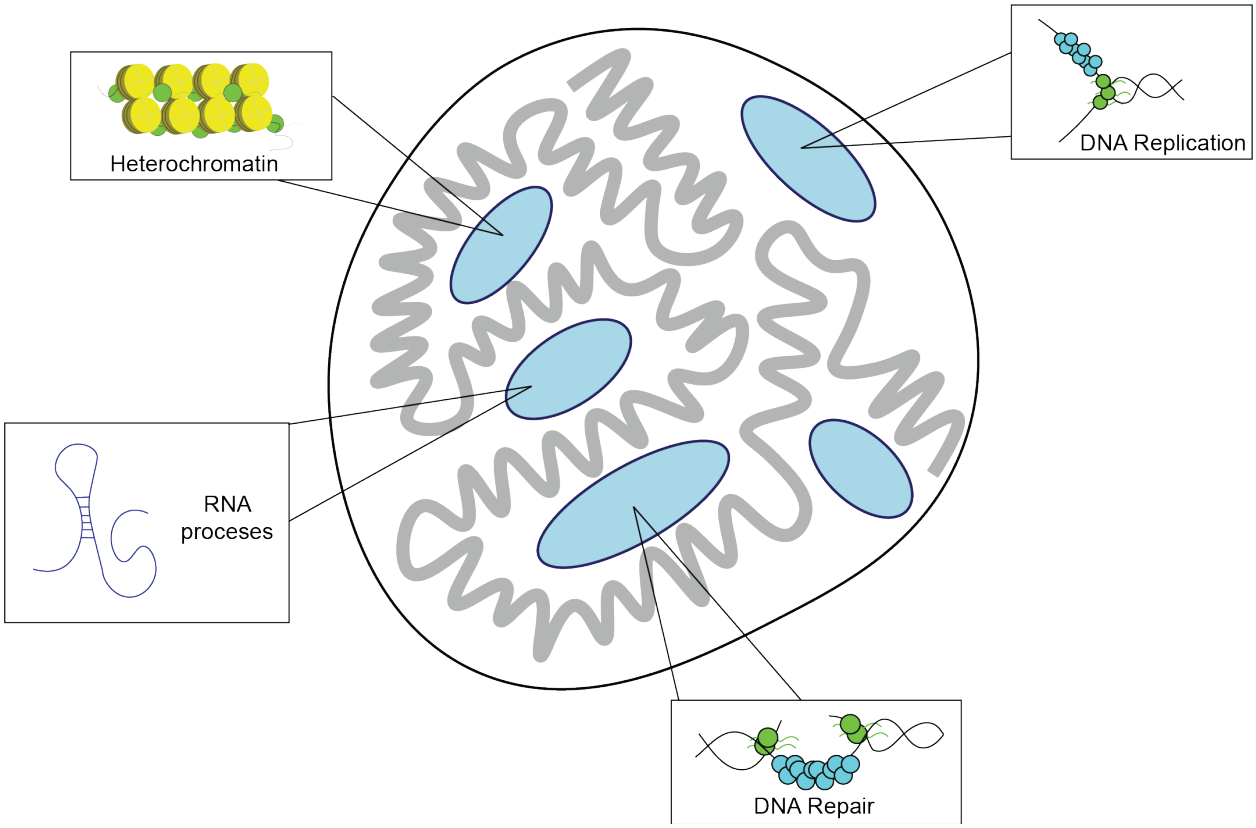


Figure 6.2. Phase separated condensates regulate various nuclear mechanisms. The nucleus may contain many phase separated regions that regulate different nuclear processes such as gene expression, heterochromatin formation, DNA repair, and DNA replication. Scaffold proteins, such as histone H1, may regulate the segregation of nucleic acids and formation of phase-separated compartments.

One major conclusion from our findings about histone H1 phase separation (Chapter 4) is its ability to differentially form droplets with different types of nucleic acids, as their fusion speed and FRAP recovery times are varied with viscosity. This type of nucleic acid segregation (**Figure 6.2**) might also be mediated by other intrinsically disordered proteins. Outside of the realm of H1, segregation of nucleic acids could potentially help the cell distinguish foreign DNA, such as viral DNA, and confine it to phase separated compartments. While some mechanisms for exogenous DNA clearance by the immune system have already been discovered [321], the role that membrane-less organelles might play in this process remains largely unexplored. In addition to DNA, RNA segregation into membrane-less organelles has been long identified, such as in the nucleolus where rRNA is processed during ribosome biogenesis. There is also evidence of additional RNA phase separation mechanisms. Some of these compartments have been identified as paraspeckles, which are protein-rich nuclear organelles built around a specific long non-coding RNA (lncRNA) scaffold. These bodies regulate gene expression through sequestration of specific proteins and RNAs [322]. As the role of lncRNAs is still being elucidated, these large charged molecules may be found to play additional roles in generating compartments throughout the cell [323]. One lncRNA of interest is HOTAIR, an oncogenic RNA known to associate with PRC2 and is necessary for PRC2 occupancy and H3K27me3 deposition on chromatin [324].

Isolation of certain nucleic acids in the cytoplasm may also regulate processes such as RNA degradation. Phase-separated ribonucleoprotein bodies have been found to organize mRNA decay pathways in bacteria [325]. Processing bodies (P-bodies) in eukaryotes have been found as compartments for mRNA storage [326]. Overall, it is interesting to investigate the broader roles for nucleic acid condensate formation and segregation mediated by nuclear and cytosolic factors.

6.5 Role of mechanical force in the nucleus

Much of this thesis focuses on data collected from force microscopy. It is imperative to consider the variety of forces that proteins and nucleic acids experience in the cell when interpreting force measurements. In a broader cellular context, the actin cytoskeleton mediates mechanical coupling between cells and their microenvironments which can be ultimately translated into biochemical signals (mechanotransduction). The architecture and composition of actin networks are modulated by force and regulated by specific proteins, such as α-catenin and vinculin [327]. While not discussed in this thesis, we contributed to the discovery of this force sensing through the application of optical tweezers.

Force-sensing is not only a mechanism in the cytoplasm, but also in the nucleus. For example, RNA polymerase is a powerful motor, exerting forces up to 25 – 30 pN on DNA and ~11 pN x nm of torsional force [328, 329], where DNA polymerase can work against forces up to 50 pN [330, 331]. These examples show that significant forces are present on chromatin in the nucleus. Given these ranges of forces, it is plausible that chromatin unraveling, modifier ejection (e.g. PRC2), or other processes may occur pervasively. As we see in our single-molecule pulling experiments, PRC2 can withstand high tensions on chromatin or DNA substrates. It is interesting to ponder how the generation of force and torque may regulate chromatin structure and gene expression.

The nucleus itself can also act as a mechanosensory organelle, promoting structural rearrangements [332, 333]. Molecular connectivity from the cytoskeleton to the nucleus is regulated by a variety of factors from the LINC (Linker of Nucleoskeleton to the Cytoskeleton) complex to other lamins and crucial connectors [334]. Forces transmitted through these nuclear scaffold networks may be focused on DNA regulatory enzymes and binding factors in addition to regulating transport through the nuclear pore complex. The forces propagated into the nucleus might alter molecular self-assembly [335] due to changes in the conformation or movement of these load-bearing molecules. Stress or strain on these nuclear scaffolds could also alter higher-order chromatin organization, changing the accessibility of genes to transcription factors or other chromatin-associated proteins [336]. Overall, there is preliminary evidence that altering nuclear structure can change gene expression and protein synthesis [337], however, specific mechanisms remain unknown. One hypothesis to how force affects genomic organization might be due to phase separated compartments within the nucleus. In our studies, we observe liquid and gel-like droplets that can withstand high forces (20-40 pN) before fusing. These forces are clearly modulated by the size of the droplets and increase as droplet size widens. Force sensing in the nucleus could be propagated by droplet resistance or fusion, leading to a variety of different compartments for controlling gene regulation.

6.6 Chromatin biology in disease and drug development

Many key targets in cancer biology include proteins involved in epigenetic pathways. PRC2, for example, is indicated in pediatric glioblastoma, colon cancer, and lymphomas, among other diseases [167, 174]. PRC2 is a promising target for cancer therapy [175], with multiple drugs already in the clinic. Most of these candidates bind to the active site of EZH2 or allosterically regulate PRC2 via EED, which binds pre-existing H3K27me3 marks [171, 172]. However, key studies have indicated that some cancers resulting from PRC2-associated processes are caused by oncohistone mutations in its substrate on the H3 tail (H3K27M) [168]. As discussed in Chapter 5, a wide variety of oncogenic histone mutations have been identified [310], however the mechanisms in which they contribute to cancer remain unknown. One of the main reasons why epigenetic inhibitors may be unsuccessful in the clinic, such as active site analogs of SAM, may be due to the fact that only the epigenetic modifier is being targeted. Given the vast array of histone mutations and chromatin-modifier pairs, the interaction between the two in a disease context must be crucial. Despite the difficulty of inhibiting protein-protein interactions, it is key to study oncohistone-modifier pairs as important cancer targets. This starts with uncovering the fundamental role of how oncohistones affect nucleosome stability and chromatin structure, a question well-suited for force spectroscopy. It is then imperative to understand the interaction between oncohistone containing chromatin and epigenetic modifiers. We have shown PRC2-chromatin engagement is strengthened in the presence of H3K27M (Chapter 3) and propose that our methodology (Chapter 2) could be applied to a multitude of chromatin and epigenetic targets.

Another factor to consider is how chromatin, and specifically heterochromatin, might influence small molecule drug delivery as phase-separated condensates in the nucleus [338]. These condensates surrounding certain chromatin regions may regulate the localization and concentration of small-molecule inhibitors. As a preliminary step, our single-molecule platform for assessing the material properties of condensates (Chapter 2) could determine how small-molecule inhibitors change the properties of condensates formed with chromatin and chromatin-associated factors. Understanding how these phase-separated membrane-less compartments serve as a barrier to entry will be imperative for inhibiting epigenetic targets.

6.7 Conclusion

While extensive work in the field of chromatin biology has uncovered the innerworkings of gene expression, key mechanistic details have remained elusive, such as mark spreading, which has been identified as a major regulatory step in chromatin repression. In this vein, we discovered the direct engagement of chromatin by PRC2 to understand its role in binding long chromatin substrates, mark spreading, and local chromatin compaction (Chapter 3). The role of linker histone H1 has also remained elusive outside of its mediation of chromatin compaction. Our findings linking H1 to single-stranded DNA condensation as well as overall nucleic acid organization suggest a more diversified role for this structural protein (Chapter 4). All of the work presented here was enabled by the development of single-molecule methods, allowing the precise manipulation of substrates to identify complex molecular mechanisms (Chapter 2). These platforms will enable the study of a multitude of chromatin-associated factors, identifying their specific mechanisms and propensities to form phase-separated compartments in the nucleus.

CHAPTER 7. Materials and methods

7.1 Molecular cloning, protein expression and purification

Bacterial cell culture

Core histone expression vectors were transformed into *E. coli* BL21 (DE3) cells and grown in lysogeny broth (Fisher Scientific) containing 100 μg/mL ampicillin and 34 μg/mL chloramphenicol at 37 °C, induced mid-log phase with 0.5 mM IPTG, and then harvested after 4 h (6,000 × g, 15 min, 4 °C). Vectors for DNA extraction (12 × 601 and MMTV) were transformed into *E. coli* DH5α cells and grown in lysogeny broth containing 100 μg/mL ampicillin and 50 μg/mL kanamycin at 37 °C and harvested (6,000 × g, 15 min, 4 °C) after 16-18 h following the ZymoPURE protocol.

Insect cell culture

Developed by Eva Ge.

All PRC2 proteins used in this study were prepared in Sf9 cells using a baculovirus system. Flag-tagged EZH2, and His-tagged EED, RBBP4, and SUZ12 were cloned into a pACEBac1 vector using the Multibac system [339]. His-AEBP2 and Flag-JARID2 were prepared in their own individual pACEBac1 vectors. Plasmids were used to generate bacmids according to manufacturer protocol (Multibac, Geneva Biotech). 2.5 μg of bacmid was transfected into 1.0×10^6 attached Sf9 cells in a 6-well plate. Following transfection, cells were overlaid with 2 mL of fresh medium (Sf-900III SFM, Thermo Fisher Scientific) and incubated at 27 °C for 96 h in the dark. The supernatant was collected, filtered through 0.22 μm, and supplemented with 2% v/v fetal bovine serum (FBS) to produce the P1 virus. P2 virus was generated by infection of 10 mL of Sf9 cells (1.5×10^6 cells/mL) with 1 mL of P1 virus solution. Cells were grown at 27 °C in suspension culture until they reached < 50% viability as monitored by trypan blue staining. Culture supernatant was collected, filtered, and supplemented with 2% v/v FBS. To generate the P3 virus, 300 μL of P2 virus solution was added to 50 mL of Sf9 cells (1.5×10^6 cells/mL). Cells were grown at 27 °C in suspension culture until they reached < 50% viability. Culture supernatant was collected, filtered, and supplemented with 2% v/v FBS. For protein expression, a 1:100 dilution of P3 virus was added to Sf9 cells at 2.0×10^6 cells/mL density. After 48 h of incubation at 28 °C in the dark, cells were harvested by centrifugation.

Purification of PRC2 complexes

Developed by Eva Ge.

Sf9 Cells were lysed by a Dounce homogenizer (Wheaton) in HEGN600 buffer (25 mM HEPES pH 7.0, 600 mM NaCl, 1 mM EDTA, 10% v/v glycerol, 0.02% NP-40). Soluble extracts were incubated with anti-Flag M2 affinity gel (100 μL resin per 100 mL cell culture) in HEGN350 (25 mM HEPES pH 7.0, 350 mM NaCl, 1 mM EDTA, 10% glycerol, 0.02% NP-40) for 2 h at 4°C. Bound proteins were eluted with HEGN350 containing 0.25 mg/mL Flag peptide (3 × 20 min at 4 °C). Eluted proteins were pooled, spin concentrated in a Vivaspin centrifugal concentrator (MWCO 30,000, Viva Products), and purified by size-exclusion chromatography on a Superose 6 column (GE Healthcare). Final products were eluted in 25 mM HEPES pH 7.0, 350

mM NaCl, 2.5mM $MgCl_2$, 10% glycerol, 0.02% NP-40, and 1 mM DTT. Fractions containing monomeric PRC2 (as analyzed by SDS-PAGE) were pooled, flash frozen with liquid N_2, and stored at -80 °C. Prior to use, the protein concentration was quantified by A_{280} and a BSA standard curve.

Purification of histone proteins

The human core histones (H2A, H2B, H3, and H4) were expressed in BL21 (DE3) cells. The cells were lysed and inclusion bodies were harvested through rounds of sonication and centrifugation [340]. Histones were extracted from inclusion bodies with DMSO and purified through ion-exchange chromatography. Octamers were reconstituted using salt dialysis and size-exclusion chromatography (Superdex 200 10/300) as described previously [340]. H3K27me3-modified histone octamers were purchased from EpiCypher. H2BT49C, H3K27M, H3K27R, H3R17C, and H3R26C H3 mutant histones were cloned using the NEB Q5 site-directed mutagenesis kit and NEBaseChanger tool.

Purification of recombinant H1.4 A4C

Developed by Wola Osunsade.

Human H1.4 A4C was purified as described previously with minor adjustments[280]. Briefly, Rosetta DE3 cells expressing His-SUMO-H1.4(A4C)-GyrA-His were induced with IPTG and expressed for 12 hours before harvesting. After rod sonication, lysate was incubated with Ni-NTA beads (Bio-Rad) and eluted. The eluent was treated with recombinant Ulp-1 (1:100 w/v) and 500 mM beta-mercaptoethanol, followed by the addition of solid urea to a final concentration of 6 M and the pH was adjusted to 9. This was then loaded onto a HiTrap SP cation exchange column (Cytiva) and subjected to a gradient of 100% H1 purification buffer A (6 M urea, 20 mM Tris pH 9.0, 200 mM NaCl) to 100 % H1 purification buffer B (6 M urea, 20 mM Tris pH 9.0, 1 M NaCl) of an AKTA FPLC system (GE Healthcare). Fractions containing full-length H1 were pooled and purified on a semi-preparative C18 HPLC column on a gradient of 0—70% buffer B and freeze-dried until use.

Reconstitution of nucleosome arrays

DNA templates were digested with BglI (New England Biolabs) from a plasmid containing 12 repeats of 601 nucleosome positioning sequences [341] in a pET28b backbone, and then ligated with T4 DNA Ligase (New England Biolabs) to DNA handles containing two biotins on each end. Nucleosome arrays were formed on the DNA template described above through salt-gradient dialysis from 1.4 M KCl to 10 mM KCl in Slide-a-Lyzer MINI Dialysis units (7,000 MWCO) using a peristaltic pump set at a 1 mL/min flow rate for 6 h. Mouse mammary tumor virus (MMTV) DNA was added during dialysis to prevent octamer overloading. Nucleosome arrays were formed at a target final concentration of 1 μM.

7.2 Protein labeling

Labeling of H1 with Cy3-maleimide

Developed by Wola Osunsade.

H1 labeling with Cy3 was performed based on a previously published protocol[342]. Briefly, H1.4 A4C was dissolved in 4 mL of histone labeling buffer (6 M Guanidine, 20 mM Tris pH 7.5, 0.2 mM TCEP). Three molar equivalents of Cy3-maleimide (ApexBio) in DMF were added and mixed gently at room temperature followed by the additional of 1 mM beta-mercaptoethanol to quench the reaction. The resultant mixture was purified on a semi-preparative reverse-phase C18 HPLC column on a gradient of 0 -70% buffer B, and freeze-dried before resuspension in H1 buffer (20 mM Tris pH 7.5, 200 mM NaCl).

Labeling of H2B T49C with Cy3-Maleimide

H2B was labeled with Cy3 maleimide (GE healthcare) under denaturing conditions, refolded in buffer containing 2 M NaCl, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, and 5 mM beta-mercaptoethanol, and purified on a Superdex 200 10/300 GL column (GE Healthcare) [343].

Non-specific protein labeling of ENL

ENL was labeled non-specifically through amine labeling with NHS (N-hydroxysuccinimide) ester dye. ENL (~low μM range) was dialyzed twice into 1 L labeling buffer (50 mM HEPES pH 7, 500 mM NaCl, 0.25 mM EDTA, and 1 mM DTT). Protein was incubated with 5X dye (Cy3-NHS), previously diluted to 1 mM in DMSO. After 1 hour incubation, reaction was quenched with 25 mM Tris-HCl pH 6.8. Sample was dialyzed twice to remove free dye into 1 L buffer BC-500 (20 mM Tris-HCl pH 8, 500 mM NaCl, 0.2 mM EDTA, 1 mM DTT). Labeling was checked on SDS-PAGE and scanned using a typhoon instrument with a 532 nm laser.

7.3 Oligonucleotide preparation

To create a terminally biotinylated dsDNA template, the 12-base 5' overhang on each end of genomic DNA from bacteriophage λ (48,502 bp; Roche) was filled in with a mixture of natural and biotinylated nucleotides by the exonuclease-deficient DNA polymerase I Klenow fragment (New England BioLabs). Reaction was conducted by incubating 10 nM λ-DNA, 33 μM each of dGTP/dATP/biotin-11-dUTP/biotin-14-dCTP (Thermo Fisher), and 5 U Klenow in 1 × NEB2 buffer at 37 °C for 45 min, followed by heat inactivation for 20 min at 75 °C. DNA was then ethanol precipitated overnight at −20 °C in 2.5 × volume cold ethanol and 300 mM sodium acetate pH 5.2. Precipitated DNA was recovered by centrifugation at 20,000 × g for 15 min at 4 °C. After removing the supernatant, the pellet was air-dried, resuspended in TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and stored at 4 °C.

7.4 Single-molecule methods

A more detailed description can be found in Chapter 2.

Single-molecule chromatin tethering data acquisition for PRC2 experiments

Single-molecule chromatin experiments were performed at room temperature on a LUMICKS C-Trap instrument equipped with dual-trap optical tweezers [344]. A computer-

controlled stage enabled rapid movement of the optical traps within a five-channel flow cell. 3.23-μm streptavidin-coated polystyrene beads (Spherotech) were flown into Channel 1 and captured by optical traps. The traps were then moved to Channel 2 that contained nucleosome array samples diluted in the imaging buffer (10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, 200 mM KCl, 0.5 mM $MgCl_2$, 0.1% Tween, and 1 mM DTT). Nucleosome array tethers were formed between two beads under flow, which reduces the chance of sticking. The tether was then transferred to Channel 3 (containing imaging buffer) or Channel 4 (containing PRC2 samples) for data collection. The PRC2 core complex (or PRC2-AEBP2, PRC2-JARID2) was diluted to 500 nM (unless otherwise noted) in the same imaging buffer. For experiments with PRC2, there was free PRC2 in solution for the entire duration of the experiments. PRC2 was incubated with nucleosome arrays at zero force for 5-10 seconds prior to pulling. Where applicable, SAM was added at a final concentration of 1.4 μM. The tethers were subjected to mechanical pulling by moving one trap relative to the other at a constant velocity (0.1 μm/s), generating force-extension curves. Data were collected at 200 kHz. The force measurement has a resolution (standard deviation of the force over 10-second time windows) of 0.2 pN and a stability (peak-to-peak difference of the force over a 2-minute interval) of < 0.3 pN.

Single-molecule chromatin tethering data acquisition for histone H1 experiments

Single-molecule experiments were performed at room temperature on a LUMICKS C-Trap instrument combining three-color confocal fluorescence microscopy with dual-trap optical tweezers[345]. A computer-controlled stage enabled rapid movement of the optical traps within a four-channel flow cell. Laminar flow separated channels 1—3, which were used to form DNA tethers between 3.23-μm streptavidin-coated polystyrene beads (Spherotech) held in traps with a stiffness of 0.6 pN/nm. Under constant flow, a single bead was caught in each trap in channel 1. The traps were then quickly moved to channel 2 containing the biotinylated DNA of interest. By moving one trap against the direction of flow but toward the other trap, and vice versa, a DNA tether could be formed and detected via a change in the *F-x* curve. The traps were then moved to channel 3 containing only buffer, and the presence of a single DNA was verified by the *F-x* curve. Orthogonal channel 4 served as a protein loading and experimental imaging chamber as described for each assay. Unless otherwise noted, flow was turned off during data acquisition. Force data was collected at 100 kHz. A488 and Cy3 fluorophores were excited by two laser lines at 488 nm and 532 nm. Kymographs were generated via a confocal line scan through the center of the two beads.

For H1 and RPA binding experiments, single DNA tethers were moved to channel 4, where they were incubated with 15 nM Cy3-H1.4 and imaged using line scans to generate a kymograph. DNA was stretched between the beads by moving the right trap at a constant velocity (0.1 μm/s) in the *x*-direction. In experiments with H1 and RPA, 10 nM Alexa488-RPA (prepared as described previously [68]) was added simultaneously to channel 4. In experiments with H2B, single tethers were incubated with 15 nM Cy3-H2B and imaged as described for H1. A more detailed description can be found in Chapter 2.

7.5 Single molecule data analysis

Force-extension analysis

Developed by Matt Reynolds.

Single tethers were screened based on signatures in the force-extension curve (position of the first segment of the curve and a single-step tether breakage). At least 15 single tethers were analyzed for each condition. Only the force-extension data from the first pull of each tether were analyzed in order to eliminate ambiguities related to nucleosome dissociation. Force-extension traces were processed using a custom Force-Extension Analyzer software suite. Each trace was separated into segments of constant contour length (L_0) based on disruption peaks. Disruption peaks were identified by applying a Butterworth low-pass filter to the trace and assigning transitions when the force decreased by a user-specified threshold, typically 0.2 pN. The unfiltered data for each segment were then fit with the extensible worm-like-chain (WLC) model: $F = \left(\frac{k_B T}{L_p}\right)\left[\frac{1}{4(1-x/L_o+F/K_o)^2} - \frac{1}{4} + \frac{x}{L_o} - \frac{F}{K_o}\right]$ [346]. In order to prevent over-fitting and to allow for direct comparison of the array's changes in L_0, the persistence length (L_p) and elastic modulus (K_0) were determined for each nucleosome array based on the first and last segments, respectively. L_p was determined by fitting the Marko-Siggia WLC model to the first segment in the low force regime [347]. Keeping this L_p fixed, the array's K_0 was determined by fitting the last segment, in the high force regime, to the extensible WLC model. Each segment was subsequently fit to the extensible WLC model to determine its L_0, maintaining L_p and K_0 constant for a given trace. The average L_p value extracted from our data is 26 ± 4 nm. The average K_0 value is 1500 ± 200 pN. Relevant parameters (number of segments, transition force, contour length change ΔL_0) for each trace were exported. The code and tutorial of the Force-Extension Analyzer software is available at https://github.com/alushinlab/ForceExtensionAnalyzer.

Cluster assignment

Developed by Matt Reynolds.

To identify the groups composing all observed transitions, the distribution of changes in contour length (ΔL_0) was modeled using a Gaussian mixture model (GMM). Putative models were generated with varying numbers of components, and each model was assessed using the Bayesian information criterion (BIC) score (**Figure 8.3E**). To evaluate the uncertainties of the BIC scores, bootstrapping was performed. Samples were drawn with replacement from the measured contour length changes, followed by clustering with an M-component GMM (M ranges from 1 to 10) and BIC scoring. This process was repeated 100 times to compute 95% confidence intervals for the BIC scores. Models with fewer than four components had significantly worse BIC scores than those with four or more components. Models with five or more components had BIC scores that were worse than or within the uncertainty of the four-component model. Furthermore, the models with four or five components had similar BIC scores and virtually the same cluster boundaries, with the only effective difference being that the cluster of large transitions would be further subdivided in the five-component model. Therefore, a four-component model was used in subsequent analyses. Each transition was then assigned to its cluster of maximum likelihood based on the GMM.

To further validate cluster robustness and to estimate uncertainties in the cluster boundaries, the following bootstrapping procedure was performed: 50% of the transitions were randomly selected 100,000 times with replacement, and the GMM clustering analysis was performed on each subset. We found that 96.3% of the random subsets led to a nearly identical clustering pattern as the aggregated data; for the remaining 3.7% of subsets, Clusters 1 and 2 collapsed into each other. This ambiguity was due to the relatively low frequency of transitions in Cluster 1 in this small percent of randomly subsampled data and, thus, these models can be rejected in context of the bare DNA data shown in Fig. S4. This bootstrapping procedure allowed us to estimate 95% confidence intervals for the cluster boundaries: 50.8 ± 1.3 bp, 104.4 ± 1.4 bp, and 179.0 ± 4.1 bp for the first, second, and third cutoffs, respectively. These results indicate that the four-cluster model is robust and the cluster boundaries are stable. Modeling was implemented in Python using the Scikit-learn library and default settings [348]. Lastly, kernel density estimation was performed to provide a smoothed depiction of the underlying distribution of contour length changes using the Scikit-learn library with a Gaussian kernel. A bandwidth of 6.6 bp was used.

Kymograph analysis
Developed by John Watters.

Bead photon counts were removed from the kymographs using either manual removal or a local search method to maximize the sum over the size of the beads' autofluorescence. The distance and confocal photon count measurements between these removed regions were then used to plot line scans of the kymographs. A photon count threshold value was used to define the foci points in the kymograph traces (110% of the maximum photon count from the dsDNA region of the kymograph). To analyze the reversibility of H1 condensate formation in the optical trap experiments, the lumicks.pylake Python package's greedy line tracking algorithm was applied to define line traces in the regions where the DNA tether was being relaxed[349, 350]. Line traces present in the first twenty percent of the region were analyzed. Each trace was counted as dissolved if the trace ended before the last twenty percent of the region, and traces that were still present in the last twenty percent of the region were counted as retained.

7.6 Negative stain electron microscopy

Electron-microscopy imaging

12-mer and 4-mer nucleosome arrays were formed using salt dialysis as described above. Arrays were purified by $MgCl_2$ precipitation (4 mM final $MgCl_2$ concentration, centrifugation at 15,000 rpm for 10 min at 4 °C). PRC2 and nucleosome arrays were dialyzed into EM buffer (50 mM HEPES pH 7.9, 50 mM KCl, 1 mM TCEP). Samples were adsorbed to glow-discharged carbon-coated copper grids and stained with uranyl formate. Imaging was performed on a CM10 electron microscope at a nominal magnification of 52,000 ×.

Image analysis of electron micrographs
Developed by Matt Reynolds.

The coordinates of nucleosomes in each micrograph were manually picked using the e2boxer tool in EMAN2 [351]. These coordinates were assigned to nucleosome arrays by detecting whether they existed within masks encompassing the arrays. These masks were generated for

each array as follows. Negative-stain micrographs were down-sampled after binning over 4 × 4 pixels to speed up computations. Uneven illumination was corrected by calculating an adaptive, local threshold and subtracting it from the down-sampled micrograph. These images were then entropy-filtered, followed by binarization at a fixed threshold per micrograph. Masks were slightly dilated and returned to the original scale. Picked coordinates within a mask were considered to be part of the same nucleosome array. For 12-mer arrays, only masks associated with five or more picked nucleosomes were selected for further analysis to evaluate the level of compaction.

7.7 Statistical analysis

Errors reported in this study represent the standard error of the mean (SEM). Unless noted otherwise, *P*-values were determined from two-tailed two-sample *t*-tests (n.s., not significant, $P \geq 0.05$; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

7.8 dCas9 roadblock experiments

dCas9 was purified as described previously [352]. Two sgRNA sequences (IDT) were designed to target two sites flanking the 12 repeats of 601 nucleosome positioning sequence within the DNA template that was also used for generating the 12-mer nucleosome arrays. dCas9 and sgRNA were incubated for 10 min at room temperature at a concentration of 1 μM [353]. The dCas9:sgRNA complexes were diluted to a final concentration of 25 nM and flown into Channel 4. DNA tethers were incubated with dCas9:sgRNA for 3 min before pulling. For experiments with PRC2, tethers were incubated with dCas9:sgRNA in Channel 4 and subsequently moved to Channel 5 that contained 500 nM PRC2 in the imaging buffer.

7.9 In silico modeling and simulations

Modeling for PRC2-nucleosome interactions
Developed by Xingcheng Lin.

Structural modeling for PRC2 and tetranucleosomes. We built a structural model for the full-length human PRC2 protein using homology modeling [191]. Only the four core subunits, SUZ12, EZH2, EED, and RBBP4 were included in the structure. Three partially solved EM/X-ray structures of PRC2: PDBID 6C23 [190], 5WAI [188] and 5HYN [189] were used as templates for structural modeling. Missing residues that cannot be found in any of the PDB structures were built as random loops.

An initial configuration for the tetranucleosome was obtained by sequentially extending the dinucleosome cryo-EM structure [181]. A 30-bp-long linker DNA (sequence: TATGACAGTGCATCACGGGGTGAGATCGCT) was used to connect the $2^{nd}/3^{rd}$ and the $3^{rd}/4^{th}$ nucleosomes. The linker DNA segments were constructed in the perfect helical form using 3DNA tool kit [354], and their orientations were dictated by the exiting nucleosomal DNA configuration to minimize bending and twisting. Configurations for the 3^{rd} and 4^{th} nucleosomes were taken from the PDB structure (ID: 3LZ1) and we replaced the protein coordinates with those from 1KX5 to model disordered histone tails. In the final construct, all nucleosomes share the same 601 sequence. We followed a similar procedure to model the tetranucleosome with 50-

bp-long linker DNA (sequence: TATGACAGTGCATCACGGGGTGTGACAGTGCATCACGGGGTGAGATCGCT). We note that our results and conclusions are independent of this initial configuration because of the use of extensive simulations for equilibration.

Coarse-grained simulation of tetranucleosomes. We combined the 3SPN.2C DNA model [184] and the structure-based Cα model [182, 355] to create a coarse-grained force field for accurate and efficient modeling of protein–DNA interactions. We represent each DNA base with three beads and every α-carbon with one bead. The energy function of the system includes contributions from intra-DNA, intra-protein, inter-protein and protein–DNA interactions. Parameters from 3SPN.2C were directly applied to model intra-DNA interactions for the tetranucleosome sequence studied here. When simulating proteins with the structure-based model, we treated each histone octamer as a single unit. Intra-protein interactions therefore refer to all interactions within an octamer, while inter-protein interactions correspond to those between octamers. To ensure the stability of the histone octamer during simulation, we included a list of native contacts for intra-protein interactions. These contacts were generated from the PDB structure (ID: 1KX5) using the Shadow contact map [356]. Two residues were considered in contact if their minimal atomic distance is 6 Å or less, regardless of whether they are from the same protein chain or not. We scaled the energy of the structure-based model by a factor of 2.5 to 0.6 kcal/mol to keep the protein complex from unfolding at a temperature of 300 K. Detailed expressions of the energy function for the protein and DNA models can be found in Ref. [183, 184]. Electrostatic interactions modeled at the Debye-Hückel level were included between charged beads, including DNA phosphates, Lys, Arg, Glu, and Asp residues. A salt concentration of 150 mM was used for the screening effect. In addition, a weak, nonspecific Lennard-Jones potential was applied between all protein–DNA beads. Detailed expressions for these potentials can be found in Ref. [357]. Our model treats water molecules implicitly, and the solvation effect was accounted for when parameterizing the DNA model and protein–DNA interactions. Similar treatments are widely used in coarse-grained modeling of protein–DNA complexes and were shown to accurately model different properties of chromatin, including the energetic cost of nucleosomal DNA unwrapping [185], the twisting of DNA relative to core histones [358], and the interaction strength between a pair of nucleosomes [187]. Since the stability of chromatin conformation and PRC2–chromatin interaction is mostly driven by electrostatic interactions [179], we anticipate the results shown here are robust with respect to our coarse-grained modeling approach. Further studies are needed to investigate DNA-sequence-specific PRC2 binding and PRC2–histone interactions.

To further improve the computational efficiency, we modeled the core region of each nucleosome as rigid bodies. This region includes the folded segments of the histone octamer and the central 107 bp of nucleosomal DNA. The rigid units consist of six degrees of freedom for translation and rotation and all atoms within the same unit move concurrently. Our setup maintains the flexibility of the linker DNA, part of the outer-wrap nucleosomal DNA (20 bp from the entry/exit site of a nucleosome), and disordered histone tails. Since the inner DNA wrap is known to bind tightly to the well-folded histone core in resting nucleosomes under no stress, we anticipate the rigid-body treatment to be a good approximation.

The most stable tetranucleosome configuration is expected to be collapsed under a salt concentration of 150 mM. To explore more expanded configurations in which the tetranucleosome might bind more favorably to PRC2, we carried out 25 independent simulations

of the tetranucleosome in the presence of harmonic biases. These biases were implemented to restrain the distances between the non-adjacent nucleosomes (1-3 and 2-4 nucleosomes) at specified values and adopt the following expression:

$$V_{bias} = \frac{k}{2}[(d_{13} - d_{13}^0)^2 + (d_{24} - d_{24}^0)^2].$$

d_{13} and d_{24} stand for the distance between the 1-3 and 2-4 non-neighboring nucleosomes. d_{13}^0 and d_{24}^0 are the corresponding target values and range from 80 to 200 Å with an increment of 30 Å. The spring constant k = 0.01 kcal/mol/Å^2 was chosen to overcome the free energy barrier at large distances while ensuring sufficient overlap among umbrella windows.

Molecular dynamics simulations were carried out with a time step of 5 fs using the LAMMPS software package to explore tetranucleosome configurations. The Nosé-Hoover thermostat was applied to maintain the simulations at a temperature of 300 K. Periodic boundary condition was enforced with a cubic box of 2,000 Å × 2,000 Å × 2,000 Å in size. The box length is much larger than the size of the tetranucleosome (~300 Å) in our simulation. A total of at least 7.5 million steps were carried out for each simulation and we saved the configurations along the trajectory at every 5,000 steps. Tetranucleosome configurations were well equilibrated in these simulations as evidenced by the convergence of the cumulative averages of the distance between 1-3 and 2-4 nucleosomes. Only data after the first three million steps were used for constructing the free energy profile.

Rigid-body docking for PRC2–tetranucleosome binding. In principle, one can study PRC2 binding with the tetranucleosome via molecular dynamics simulations. However, the slow timescale associated with diffusion and the rugged energy landscape make an exhaustive exploration of different binding modes challenging. To more efficiently study the binding, we applied a rigid docking procedure as detailed below.

First, to account for the conformational flexibility of the tetranucleosome, we selected 1,000 structures from biased simulations introduced in the previous section. The structures were chosen based on a *K*-means clustering over all the simulated configurations to include both collapsed and extended configurations. Each configuration was represented with the six inter-nucleosome distances for clustering.

For each one of the 1,000 structures, we then determined the set of most stable PRC2 binding configurations by evaluating the energy of a large set of structures. Specifically, we selected 577 PRC2 orientations from a uniform sampling of the three Euler angles. For each orientation, we then searched for every possible position on a grid of ~800 Å × 800 Å × 800 Å in size with a spacing of 2 Å using discrete Fourier transform. Since the size of PRC2 is roughly 150 Å, the grid is large enough to enclose all possible PRC2-bound tetranucleosome configurations. Interaction energy between PRC2 and tetranucleosome was evaluated using the same force field introduced in the previous section. To avoid steric clashes, we did not use the disordered regions of PRC2 (SUZ12: residue# 1-78, 150-153, 168-181, 210, 224-227, 254-294, 323-350, 364-422, 549-560, 686-739; EZH2: residue# 1-9, 183-210, 217-219, 250-256, 346-421, 480-513, 741-746; EED: residue# 1-76; RBBP4: residue# 1-2, 94-104, 413-425) for energy evaluation.

Population estimation of PRC2 binding modes. From the docking simulations performed in the previous section, we can estimate the fraction of different PRC2 binding modes at a given

tetranucleosome configuration. These data, however, cannot be combined straightforwardly to estimate an overall probability for various binding modes as the tetranucleosome configurations were collected from biased simulations. Proper thermodynamic reweighting must be carried out before averaging as detailed below [359].

The probability of PRC2 binding in between adjacent nucleosomes ($Nuc_{1\text{-}2}$ mode) at distances d_{13}^o, d_{24}^o for 1-3 and 2-4 nucleosomes can be defined as:

$$p_{12}(d_{13}^o, d_{24}^o) = \frac{1}{Z}\int e^{-\beta H}\delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$
$$= \frac{1}{Z}\int e^{-\beta(E_N+E_P+E_{NP})}\,\delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}.$$

where R_N and R_P correspond to the nucleosome and PRC2 degrees of freedom and $Z = \int e^{-\beta H}\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$ is the partition function. $H = E_\mathrm{N} + E_\mathrm{P} + E_\mathrm{NP}$ is the potential energy, with E_N, E_P, and E_NP being the internal energy of the tetranucleosome, the internal energy of PRC2, and the binding energy between PRC2 and nucleosomes, respectively. $\delta[C_{12}(R_P)]$ represents all the configurations in which PRC2 binds simultaneously with the first and second nucleosomes. $\delta[d_{13}(R_N) - d_{13}^o]$ and $\delta[d_{13}(R_N) - d_{24}^o]$ select out tetranucleosome configurations with 1-3 nucleosome distance at d_{13}^o or 2-4 nucleosome distance at d_{24}^o, respectively.

The above definition, though formally exact, is difficult to calculate in practice, as it requires simulations in the presence of both PRC2 and tetranucleosomes. To make progress, we invoke the mean field approximation and replace E_{NP} with the average value at the given distances d_{13}^o, d_{24}^o for the $Nuc_{1\text{-}2}$ mode, $\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{12}}$. The probability can then be simplified as:

$$p_{12}(\,d_{13}^o, d_{24}^o) \approx \frac{1}{Z}e^{-\beta\langle E_{NP}(d_{13}^o,d_{24}^o)\rangle_{C_{12}}}\int\int e^{-\beta(E_N+E_P)}\,\delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$
$$= \frac{1}{Z}e^{-\beta[\langle E_{NP}(d_{13}^o,d_{24}^o)\rangle_{C_{12}}+F(d_{13}^o,d_{24}^o)+E_P]}\int \delta[C_{12}(R_P)]\delta[d_{13}(R_N) - d_{13}^o]\delta[d_{13}(R_N) - d_{24}^o]\mathrm{d}R_\mathrm{N}\mathrm{d}R_\mathrm{P}$$
$$= \frac{1}{Z}e^{-\beta[\langle E_{NP}(d_{13}^o,d_{24}^o)\rangle_{C_{12}}+F(d_{13}^o,d_{24}^o)+E_P]}\frac{N_{12}(d_{13}^o,\ d_{24}^o)}{N},$$

where $N_{12}(d_{13}^o,\ d_{24}^o)$ is the number of docked PRC2 structures engaging in the $Nuc_{1\text{-}2}$ mode at distances d_{13}^o, d_{24}^o, while N is the total number of docked PRC2 configurations. $F(d_{13}^o,\ d_{24}^o)$ is the free energy of tetranucleosomes with different 1-3, 2-4 nucleosome distances and was calculated with the weighted histogram analysis method (WHAM) from the biased simulations. To obtain the final population estimation independent of inter-nucleosome distances, we integrate over d_{13}^o and d_{24}^o

$$p_{12} = \frac{1}{Z}\int e^{-\beta[\langle E_{NP}(d_{13}^o,d_{24}^o)\rangle_{C_{12}}+F(d_{13}^o,d_{24}^o)+E_P]}\frac{N_{12}(d_{13}^o,\ d_{24}^o)}{N}\mathrm{d}d_{13}^o\mathrm{d}d_{24}^o$$

Population of $Nuc_{1\text{-}3}$ and $Nuc_{1\text{-}4}$ binding modes can be similarly defined as:

$$p_{13} = \frac{1}{Z}\int e^{-\beta[\langle E_{NP}(d_{13}^o,d_{24}^o)\rangle_{C_{13}}+F(d_{13}^o,d_{24}^o)+E_P]}\frac{N_{13}(d_{13}^o,\ d_{24}^o)}{N}\mathrm{d}d_{13}^o\mathrm{d}d_{24}^o$$

$$p_{14} = \frac{1}{Z}\int e^{-\beta\left[\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle_{C_{14}} + F(d_{13}^o, d_{24}^o) + E_P\right]} \frac{N_{14}(d_{13}^o,\ d_{24}^o)}{N} \mathrm{d}d_{13}^o \mathrm{d}d_{24}^o$$

To determine the numerical values of the various populations, the integration was converted into a summation over the 1,000 tetranucleosome structures. For each structure, we used the top 1,000 lowest PRC2 binding configurations determined from docking to estimate $\langle E_{NP}(d_{13}^o, d_{24}^o)\rangle$ and $N_{12}(d_{13}^o,\ d_{24}^o)$. When determining the binding modes, the minimum atomic distance between PRC2 and each of the four nucleosomes was calculated. A cutoff distance of 0.8 nm (approximately one Debye length) was used to determine whether PRC2 and the corresponding nucleosome are in contact.

Modeling for H1:ssDNA phase separation interactions
Developed by Andrew Latham.

Coarse grained molecular dynamics simulations were performed with one bead per protein residue or DNA base using the maximum entropy optimized force field (MOFF) for proteins[294] and the molecular renormalization group coarse-graining model (MRG-CG) for DNA[293]. Structure-based modeling potentials, the strength of which were tuned to reproduce the root-mean-square fluctuations from all-atom simulations[355], were introduced to stabilize tertiary contacts in the ordered domain of H1 molecules. Electrostatic interactions were described with implicit solvent using a distance-dependent dielectric constant and 150 mM salt[294]. To account for this implicit solvent instead of the explicit ions used originally in the DNA model, bonded, angle, and fan interactions were scaled by a factor of 0.9. At 100 mM salt, the simulated persistence lengths for 200 nt/bp of DNA are 3.62 ± 0.28 nm for ssDNA and 47 ± 6 nm for dsDNA, which compare favorably to experimental values of 2.5 nm[360] and 47.8 ± 0.9 nm[361], respectively. Interactions between protein and DNA molecules include electrostatics and by an excluded volume term of $V(r) = \frac{\varepsilon_0}{r^{12}}$, where r is the distance between a protein bead and a DNA bead, and ϵ_0 is a constant of 1.6264×10^{-3} kJ mol^{-1} nm^{12}. All simulations were performed using the GROMACS simulation package[362].

Simulations followed the slab methodology for determining critical temperature[294, 363]. To model effects seen experimentally, a ratio of 4DNA:1H1 was preserved for all simulations. The ssDNA simulations included 160 DNA molecules and 40 H1 molecules, while those with dsDNA were performed with 320 DNA molecules and 80 H1 molecules. To check for finite-size effects, we also performed simulations with 160 DNA molecules and 40 H1 molecules for dsDNA and obtained similar results. These molecules were initially placed in a large simulation box of size 100 nm × 100 nm × 100 nm. We then performed the steepest descent energy minimization. This step was followed by an NPT simulation for 0.1 μs at 150 K and 1 bar, using a Parrinello-Rahman isotropic barostat and time coupling constant of 1 ps. The NPT simulation collapses the molecules into one dense phase. The z-dimension was then expanded by ~20 times the original size, resulting in a droplet with a dilute phase on either side. We then performed an NVT simulation for 0.1 μs with a time coupling constant of 100 ps. During the simulation, we raised the temperature from 150 K to the desired temperature. The resulting equilibrated system was run for 2 μs in the NVT ensemble, with the first 1 μs discarded for analysis. For simulations where the dsDNA and ssDNA droplets were merged, both dense phases from NPT simulations were placed in the same simulation box. The total molecules in the simulation were 80 H1, 160

ssDNA, and 160 dsDNA. We then performed the standard slab technique described above to this new system.

The critical temperatures were determined using the difference in density in the high density and low-density phase, as described previously[294, 363]. MDAnalysis was used to help with analysis[364].

7.10 Phase separation methods

A more detailed description can be found in Chapter 2.

Solution turbidity experiments

Phase separation experiments were performed in 20 μL volumes in the following buffer: 20 mM Tris-HCl pH 7.5, 200 mM NaCl, 10% PEG8000, at DNA concentrations of 10 μM and linker histone concentrations of 2.5 μM unless otherwise specified. Absorption measurements were taken on a Nano drop instrument at 350 nm after incubating for 10 min.

Droplet imaging using optical traps, bright field imaging, and confocal scanning

As described above, single droplets were captured using the LUMICKS C-trap. Droplets were prepared as described above for the phase separation assay and loaded onto a home-made flow cell, sandwiched between a glass cover slip and slide using double-sided tape and sealed with clear nail polish. One droplet was captured per trap at a trapping laser power of 5%. Droplets were imaged using 532 nm and 649 nm lasers. For bright field imaging, images were acquired using the bright field camera.

Fluorescence recovery after photobleaching (FRAP)

An automated FRAP script, "FRAP droplet imaging" was written to provide consistent results (https://harbor.lumicks.com/single-script/3a796fac-dbb3-4fe1-8ce7-8b0cf8c25ad9). Droplets were prepared as described above and imaged with a 2D confocal scan at 10% laser power. Then droplets were partially bleached at 100% laser power with an 800-nm radius point scan for 2 s. Then the droplets were imaged with a 2D confocal scan every 5 s. Images were uploaded to FIJI as a stack and analyzed by the FRAP Profiler plugin. Normalized FRAP intensity was extracted and plotted in Prism.

Controlled droplet fusion experiments

Droplets were prepared as described above and captured in dual traps and imaged with a continuous 2D confocal scan every 5 s. One trap was manually stepped towards the other trap in 200 nm intervals after every image scan until droplets were fused (or were pushed together but unable to fuse). The force for each trap was recorded concurrently. Analysis of fusion times of H1 separated droplets under the control of the optical trap experimental set-up were calculated first by an approximate manual selection of the timepoints of droplet fusion initiation and termination. The manual selections are then computationally refined in a local search of 40—200 time points. Droplet fusion timepoints were refined to the point of maximum rate of force change (for droplet fusion initiation) or minimum rate of force change (for droplet fusion termination).

Phase separation data analysis

Developed by John Watters.

Force and fluorescence data from .h5 files generated were analyzed using the tools in the lumicks.pylake Python library supplemented with other Python modules (Numpy, Matplotlib, Pandas)[365] in a custom GUI Python script titled "C-Trap .h5 File Visualization GUI" (https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f). This script was used to extract confocal images and fusion traces from droplet formation, fusion, and FRAP experiments.

To analyze the fusion of H1 separated droplets using the optical trap experimental set-up, a sigmoidal curve was fit to the high frequency force-time measurements (78 kHz) of the total, normalized force from the stationary optical trap. The sigmoidal fit was then used to calculate τ, droplet fusion time, as defined by the difference between the timepoints for 0.3 and 0.8 of the normalized force values. Manual selection around the approximate fusion time window was applied to improve the curve fits and the ability of the SciPy curve fitting algorithm to converge on a solution.

7.11 Live cell imaging

Developed by Wola Osunsade.

Human embryonic kidney 293T cells were transfected with either eGFP-H1 (N-terminally labeled) constructs alone or in combination with RPA1-mCherry (C-terminally labeled) and cultured in Dulbecco's modified Eagle's medium media with 10% fetal bovine serum and 4 mM Glutamine at 37 °C in 5% CO_2. All live-cell confocal imaging was conducted under cell culture conditions on a Leica SP8 microscope with inverted stand. For experiments investigating replication stress, cells were subjected to either mock (1% PBS and 0.1% DMSO) or treatment conditions. Treatment involved exposure to 2 mM hydroxyurea (Sigma-Aldrich) and 20 μM ceralasertib (MedChemExpress) for 12 hours before imaging. Each experimental condition was measure in at least 10 different cells. Puncta were counted using the 3D Object Counter Plugin in FIJI.

7.12 Cryo-electron tomography procedures

Developed in conjunction with Ayala Carl and Santiago Espinosa.

PRC2 samples were prepared for cryo-electron tomography technique as follows. Grids of various type were tested for ideal sample adherence, including 200 Au C-flat and 300 Au lacey carbon. These grids were plasma cleaned for 7 seconds with O2/H2 mixture, then spotted with 3 or 10 uL of sample (300 nM PRC2, 300 nM 12-mer array in 25mM HEPES pH 7.9, 50 mM KCl, 1 mM TCEP) and incubated for either 30 seconds or 1 minute. After incubation the grids were blotted for 5 seconds, then immediately plunged using Leica EM Grid Plunger machine. Grids were clipped and transferred to Talos Arctica CryoTEM (FEI) operating at 200 kV for collection of tomograms. Tomograms were collected using a dose symmetric scheme from -60 to +60 degrees with angular increments of 3 degrees, a total dose of 100 e/A^2, and a defocus range between -1.5 μ M and -4 μ M. The acquisition magnification was 17,000 times with a pixel size of 2.4 Å/pix.

Tomograms were reconstructed using the software appion-protomo. Images were aligned, and dose weighted accordingly, then reconstructed using Simultaneous Iterative Reconstruction Technique (SIRT). Reconstructions were binned by 4 resulting in a final pixel size of 9.6 Å. Reconstructed tomograms were visualized using 3DMOD software.

CHAPTER 8. Appendix

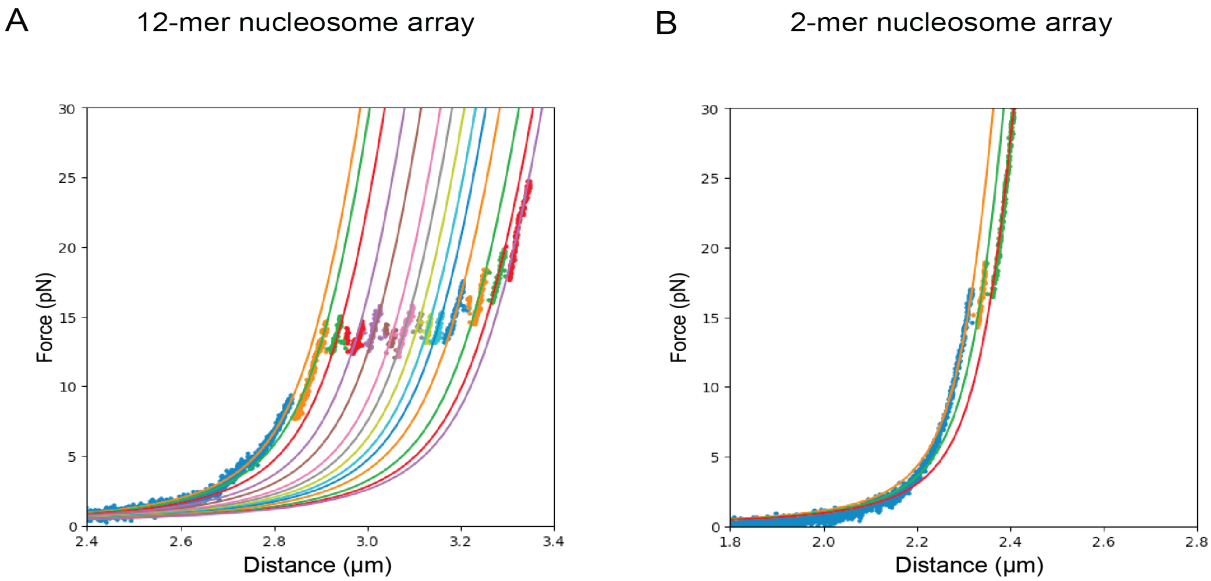
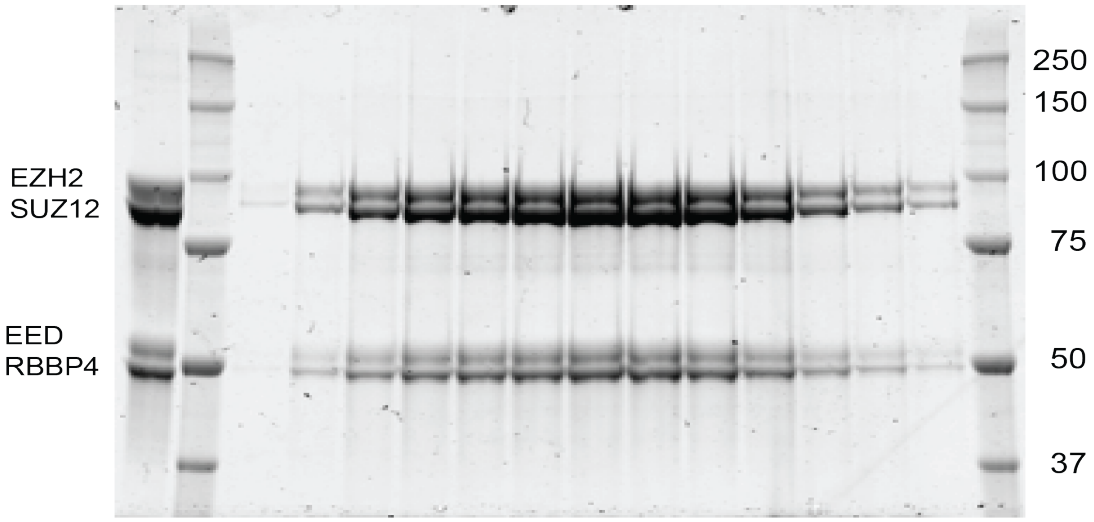


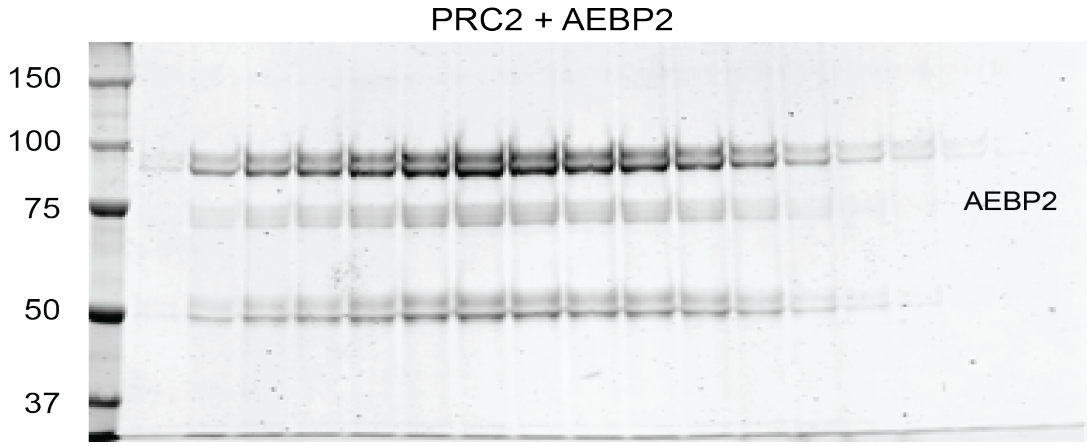
Figure 8.1. Segmentation of force-extension curves. (**A**) A representative force-extension curve and its fitted segments for a 12-mer nucleosome array. See Materials and Methods for details of the segmentation analysis. (**B**) A representative force-extension curve and its fitted segments for a dinucleosome substrate.

Figure 8.2. SDS-PAGE analysis of PRC2 complexes. SDS-PAGE gels showing the fractions of PRC2 core (**A**), PRC2-AEBP2 (**B**), and PRC2-JARID2 (**C**) complexes from size-exclusion chromatography.

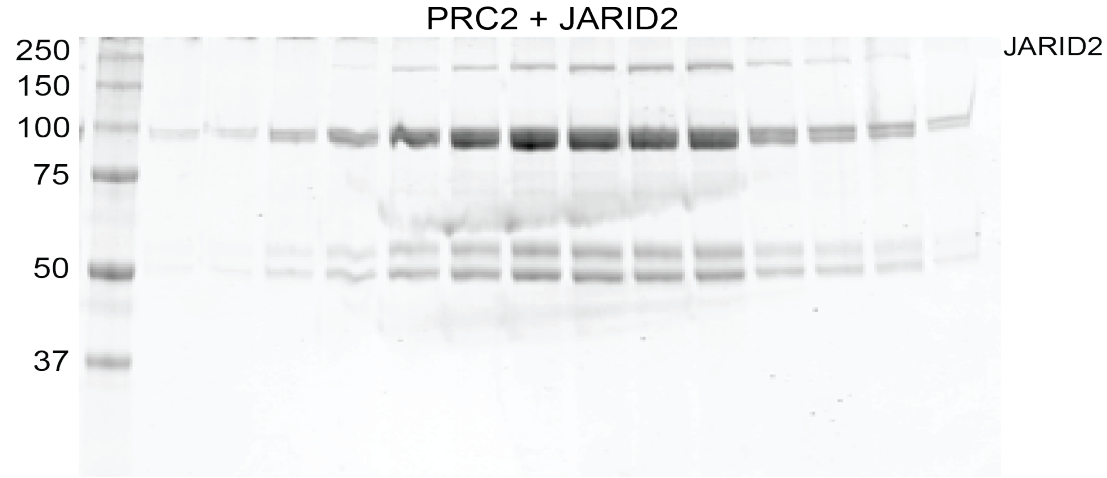
A


PRC2 Core Complex
EZH2
SUZ12
EED
RBBP4
250
150
100
75
50
37

B


PRC2 + AEBP2
150
100
75
50
37
AEBP2

C


PRC2 + JARID2
250
150
100
75
50
37
JARID2

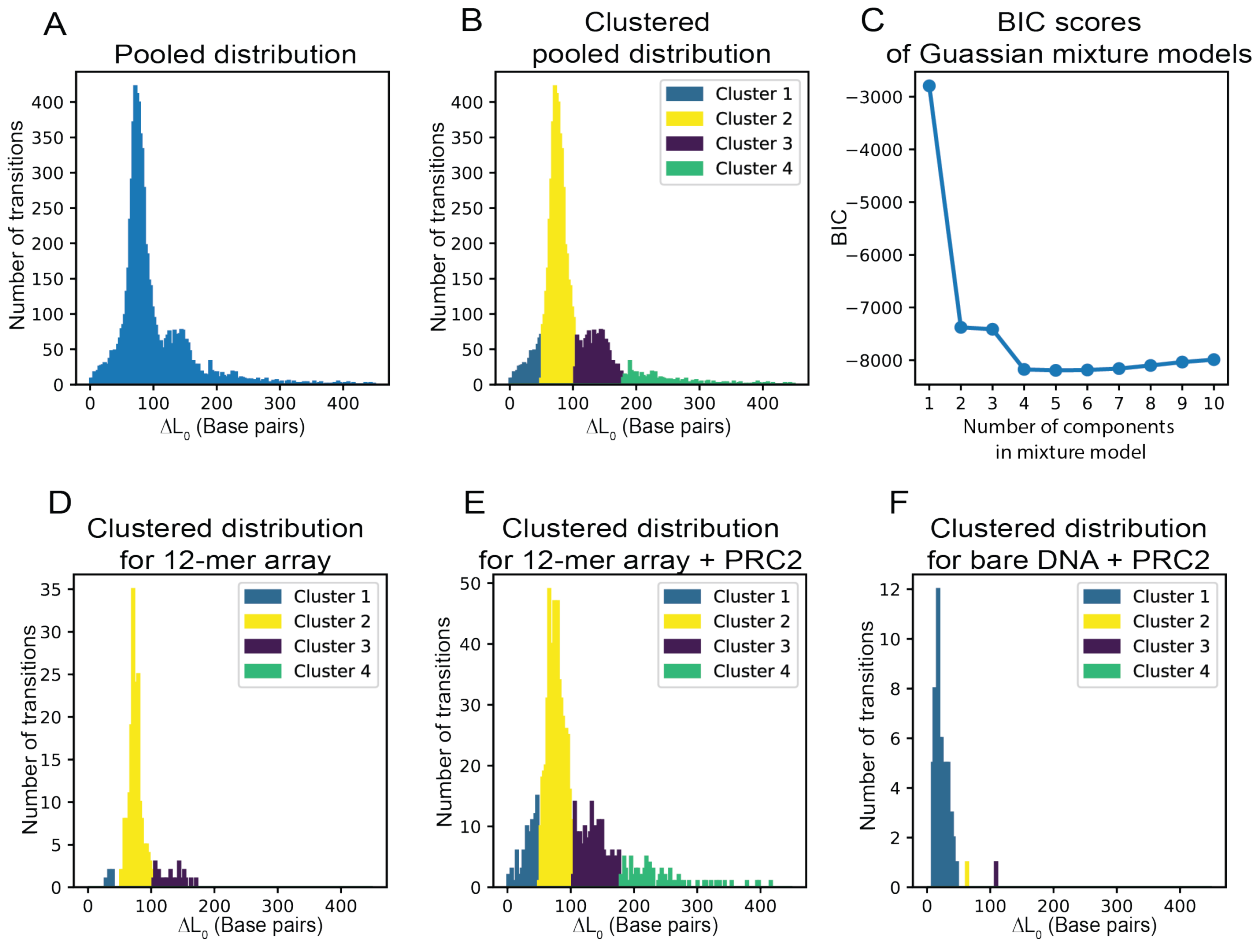
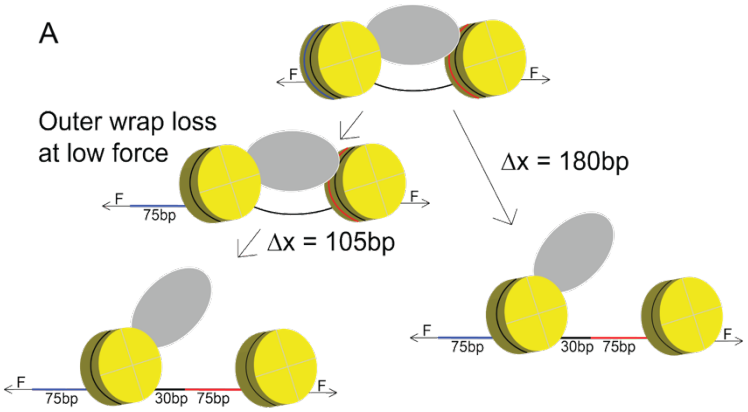
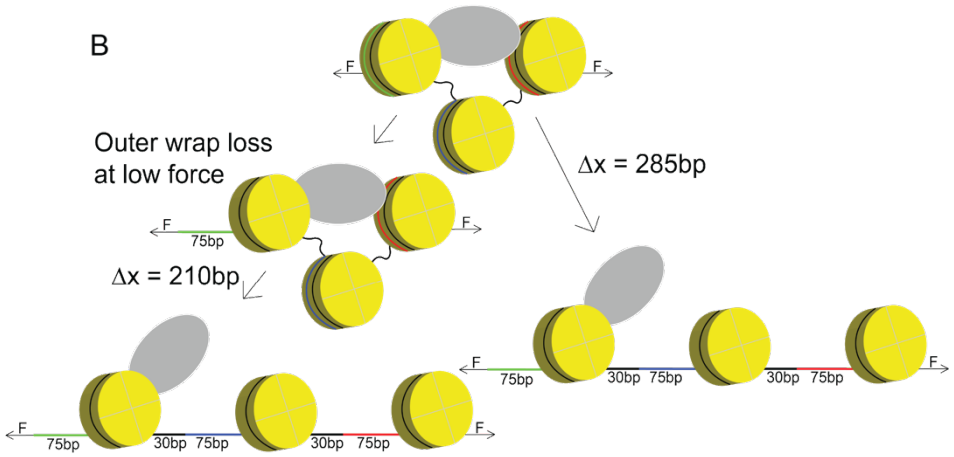
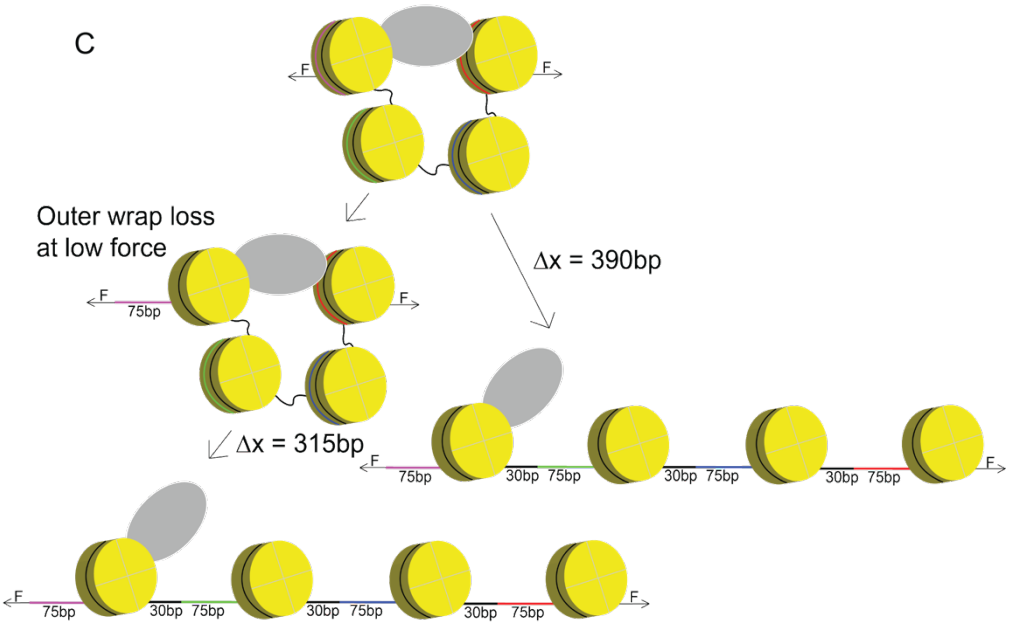


Figure 8.3. Clustering of transitions identified from force-extension curves. (**A**) Histogram of contour length changes (ΔL_0) of transitions derived from the force-extension curves pooled from all conditions. (**B**) The same histogram as in (**A**) is grouped into four clusters based on a four-component Gaussian mixture model. The cluster boundaries are determined using a bootstrapping procedure. In black is the estimated kernel density of the underlying distribution. See *Materials and Methods* for details of cluster assignment, robustness evaluation, and kernel density estimation. (**C**) Bayesian information criterion (BIC) scores of *M*-component Gaussian mixture models (*M* ranges from 1 to 10) to fit the pooled experimental data of ΔL_0. A four-component model was selected for all cluster assignments in this study because the BIC score ceased to improve when more components were added to the model. (**D-F**) Clustered histograms of ΔL_0 distribution for the 12-mer array data (**D**), 12-mer + PRC2 data (**E**), and bare DNA + PRC2 data (**F**). Kernel density estimates are shown as black contours. *N* denotes the total number of transitions within each histogram.

Figure 8.4. Schematic illustration of the amount of DNA released after disruption of PRC2-mediated bridging of nucleosome pairs. (**A**) Scenarios for force-induced disengagement of PRC2 initially engaged with two adjacent nucleosomes (Nuc1-2 mode). In the left pathway, the outer wrap of the first nucleosome (blue segment) is not sequestered by PRC2, thus undone at low forces. Upon PRC2 unbinding that occurs at a higher force, the outer wrap of the second nucleosome (red segment, ~75 bp) and the 30-bp linker DNA—with a total length of ~105 bp—are released. The inner wraps of the nucleosomes have similar stabilities to PRC2 engagement. Therefore, they are expected to unravel later as independent transitions. Alternatively, as depicted in the right pathway, PRC2 sequesters both outer wraps. In this scenario, PRC2 disengagement would release two outer wraps (~150 bp) plus one linker DNA, totaling ~180 bp. If PRC2 sequesters part of the first outer wrap, a number between 105 and 180 bp is expected for the amount of DNA released. (**B**) Scenarios for force-induced disengagement of PRC2 bound to two nucleosomes that are separated by one spacer nucleosome (Nuc1-3 mode). A total between 210 and 285 bp of DNA is expected to be released upon PRC2 disengagement. (**C**) Scenarios for force-induced disengagement of PRC2 bridging a pair of nucleosomes that are separated by two spacer nucleosomes (Nuc1-4 mode). A total between 315 and 390 bp of DNA is expected to be released upon PRC2 disengagement.


A
Outer wrap loss
at low force
Δx = 180bp
Δx = 105bp
F
75bp
30bp
75bp


B
Outer wrap loss
at low force
Δx = 285bp
Δx = 210bp
F
75bp
30bp
75bp


C
Outer wrap loss
at low force
Δx = 390bp
Δx = 315bp
F
75bp
30bp
75bp

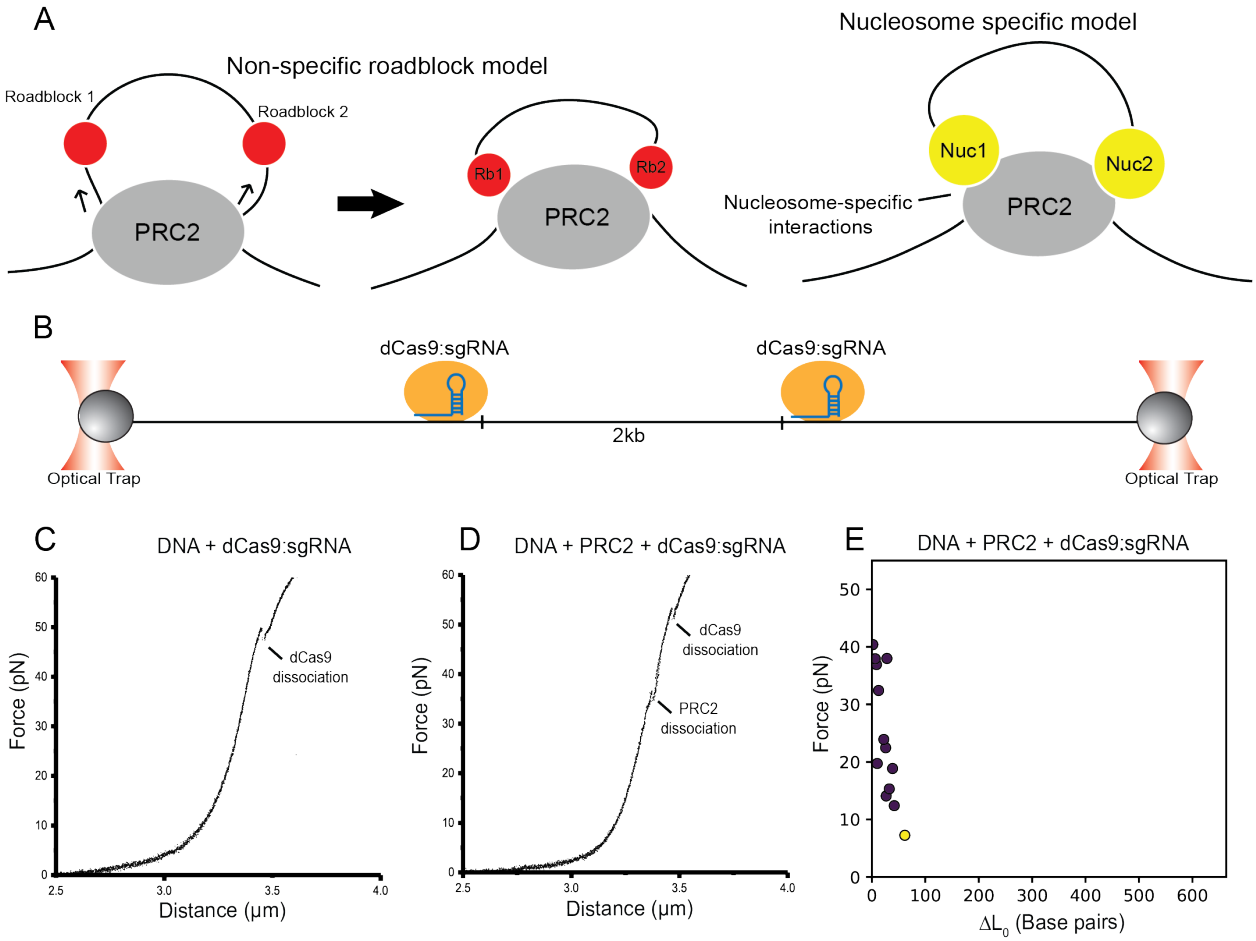
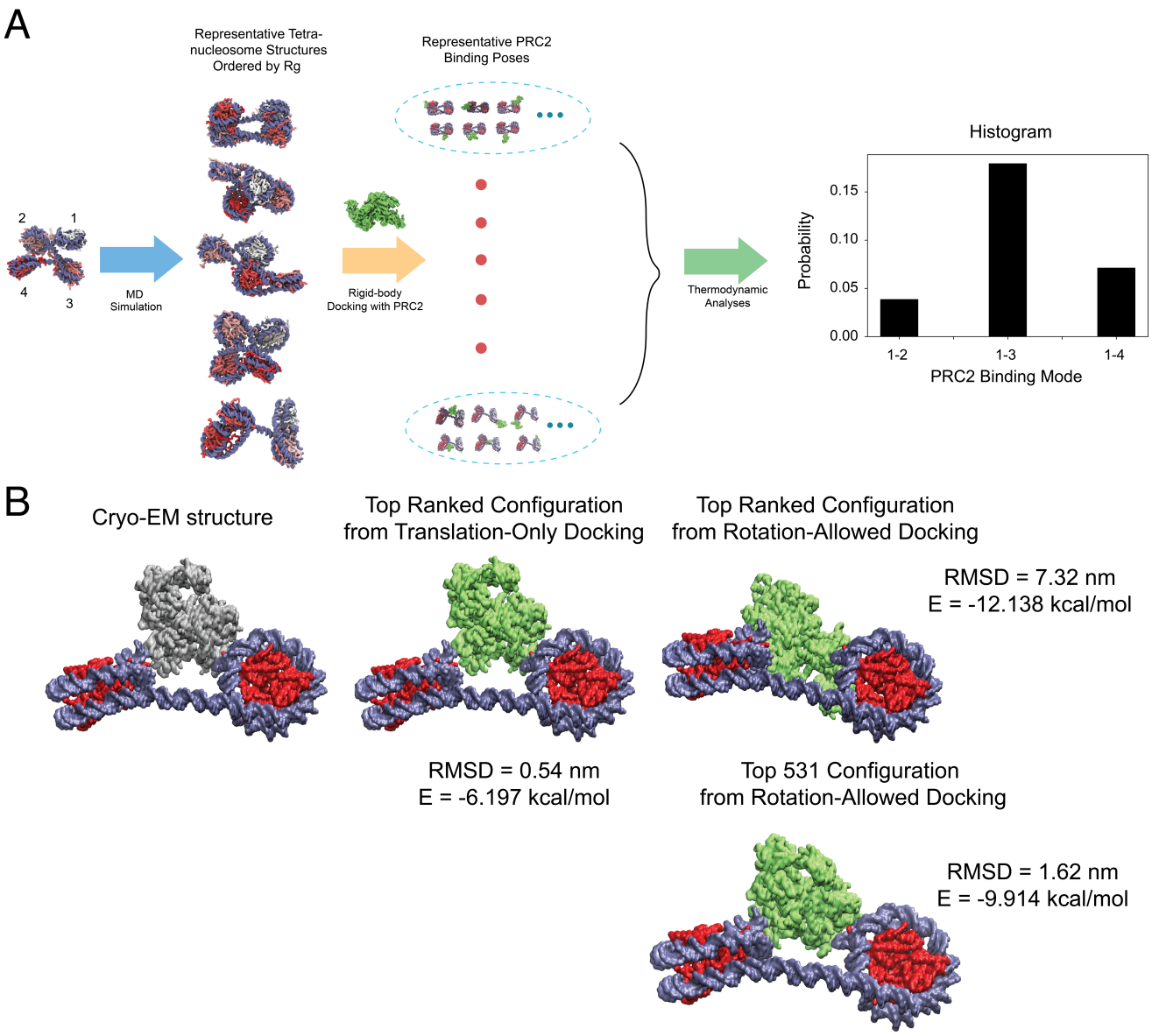


Figure 8.5. Differentiating between alternative models for PRC2-mediated chromatin looping. (**A**) (Left) Schematic of a non-specific roadblock model in which PRC2 engages with two distal segments of DNA and slides on them—either via an inherent one-dimensional sliding activity of PRC2 or through force-induced threading—until running into a pair of steric blocks. (Right) Schematic of a nucleosome-specific model in which PRC2 makes specific contacts with a pair of nucleosomes. Both models could explain the observed changes in the end-to-end distance of DNA when PRC2 is mechanically dissociated by pulling the DNA. (**B**) Experimental design of two roadblocks engineered into a piece of tethered DNA based on dCas9:sgRNA targeting to specific DNA sites. The two dCas9 complexes are separated by ~2,000 bp. The same DNA was used to make 12-mer nucleosome arrays. (**C**) A representative force-extension curve for dCas9:sgRNA loaded DNA. Force-induced dCas9 dissociation can be observed at ~50 pN, indicating stable dCas9 binding. (**D**) A representative force-extension curve for dCas9:sgRNA loaded DNA incubated with PRC2. PRC2 disengagement events occurred at lower forces compared to dCas9 dissociation events. Importantly, no transition with a size close to 2,000 bp was observed, thus disfavoring the non-specific roadblock model. (**E**) Cluster analysis of PRC2-mediated transitions observed in the force-extension curves of PRC2-bound bare DNA in the presence of dCas9:sgRNA.

Figure 8.6. Coarse-grained modeling of PRC2–chromatin interactions. (**A**) Flowchart of the algorithm used to quantify the various binding modes in which PRC2 engages with a tetranucleosome. See *Supplemental Methods* for details of the modeling. (**B**) Comparison between experimental and simulated PRC2 binding poses. (*Left*) Cryo-EM structure of the PRC2-dinucleosome complex determined by Ref. [181]. (*Middle*) Lowest energy configuration predicted from rigid docking using a coarse-grained force field. See *Supplemental Methods* for details of the force field. The PRC2 structure used for docking includes SUZ12, EZH2, EED, and RBBP4 residues resolved in the cryo-EM structure. Only translational motion was allowed in these docking simulations and PRC2 was fixed in the same orientation as that found in the cryo-EM structure. The Cα RMSD between docked and cryo-EM structures is 0.54 nm. (*Right*) The lowest energy (*Top*) and a low RMSD (*Bottom*) configurations from docking simulations that allow both translational and rotational motions. In the lowest energy configuration, PRC2 is again juxtaposed between the two nucleosomes as found in cryo-EM, but with a slightly tilted orientation. The difference in orientation between simulated and experimental configurations could be due to the omission of disordered regions in PRC2 and histones during docking. The accuracy of the force field could also potentially impact the PRC2 orientation. We found that the top 531st ranked configuration, though with an energy difference of less than 3 kcal/mol from the lowest energy configuration, is in much better agreement with the cryo-EM structure with a Cα RMSD of 1.62 nm. These results with the dinucleosome, therefore, support the use of the coarse-grained model for studying PRC2–chromatin binding modes.


A
Representative Tetra-nucleosome Structures Ordered by Rg
Representative PRC2 Binding Poses
2
1
4
3
MD Simulation
Rigid-body Docking with PRC2
Thermodynamic Analyses
Histogram
Probability
0.15
0.10
0.05
0.00
1-2
1-3
1-4
PRC2 Binding Mode
B
Cryo-EM structure
Top Ranked Configuration from Translation-Only Docking
Top Ranked Configuration from Rotation-Allowed Docking
RMSD = 7.32 nm
E = -12.138 kcal/mol
RMSD = 0.54 nm
E = -6.197 kcal/mol
Top 531 Configuration from Rotation-Allowed Docking
RMSD = 1.62 nm
E = -9.914 kcal/mol

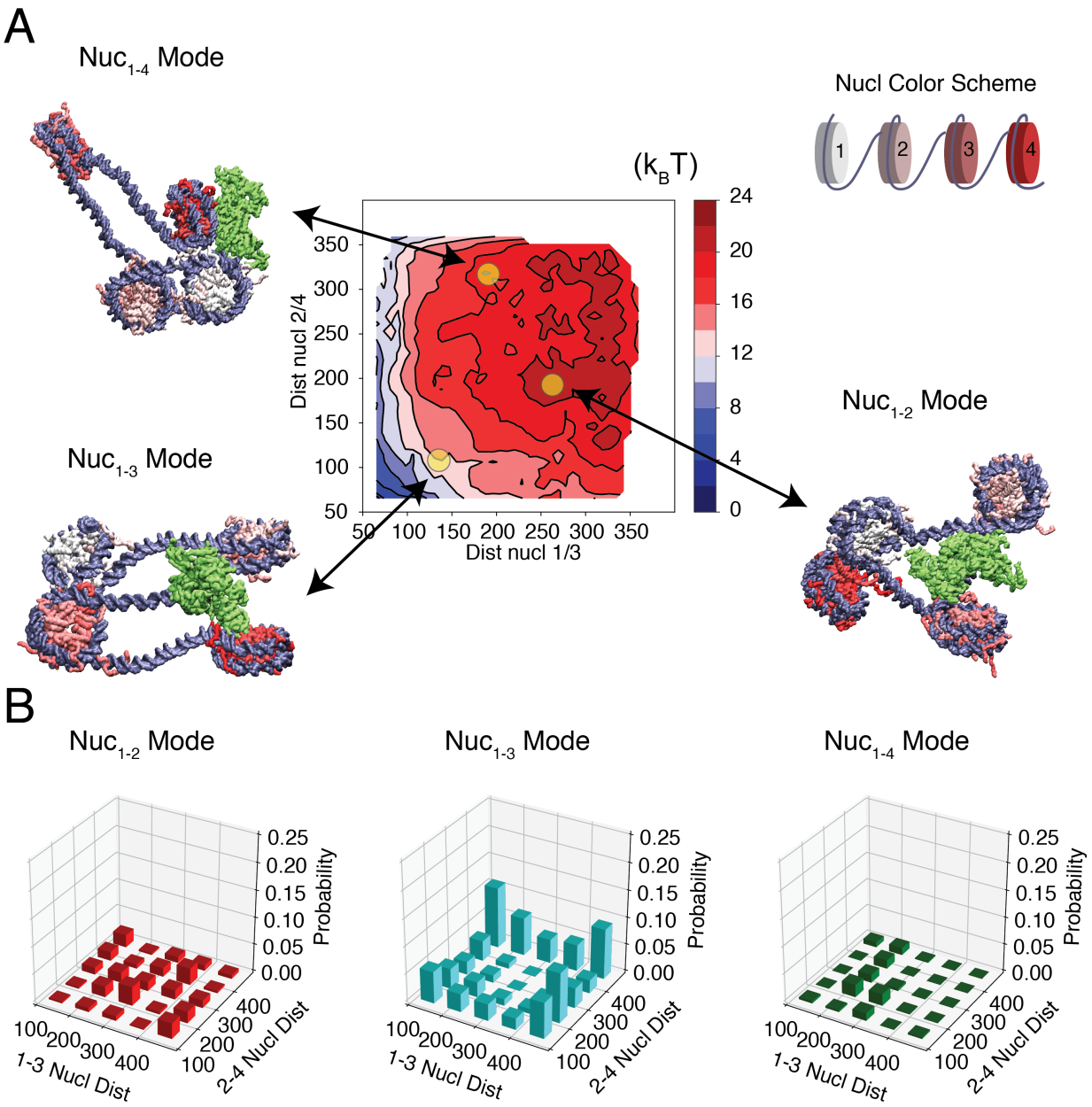


Figure 8.7. In silico analysis of the interaction between PRC2 and tetranucleosomes with 50-bp linker. (**A**) Thermodynamic stability (free energy) for the tetranucleosome with a 50-bp-long linker DNA as a function of the spatial distances between the 1-3 and 2-4 nucleosome pairs. The impact of PRC2 was not included when computing the free energies. Example chromatin configurations are shown on the side, with DNA in purple and histone colors varying from white to red as the nucleosome index increases. PRC2 is colored in green and shown in its lowest energy pose in the three examples. (**B**) Fraction of PRC2 engaging in different nucleosome-bridging modes (Nuc_{1-2}, Nuc_{1-3}, and Nuc_{1-4}) at given tetranucleosome configurations grouped by its inter-nucleosome distances.

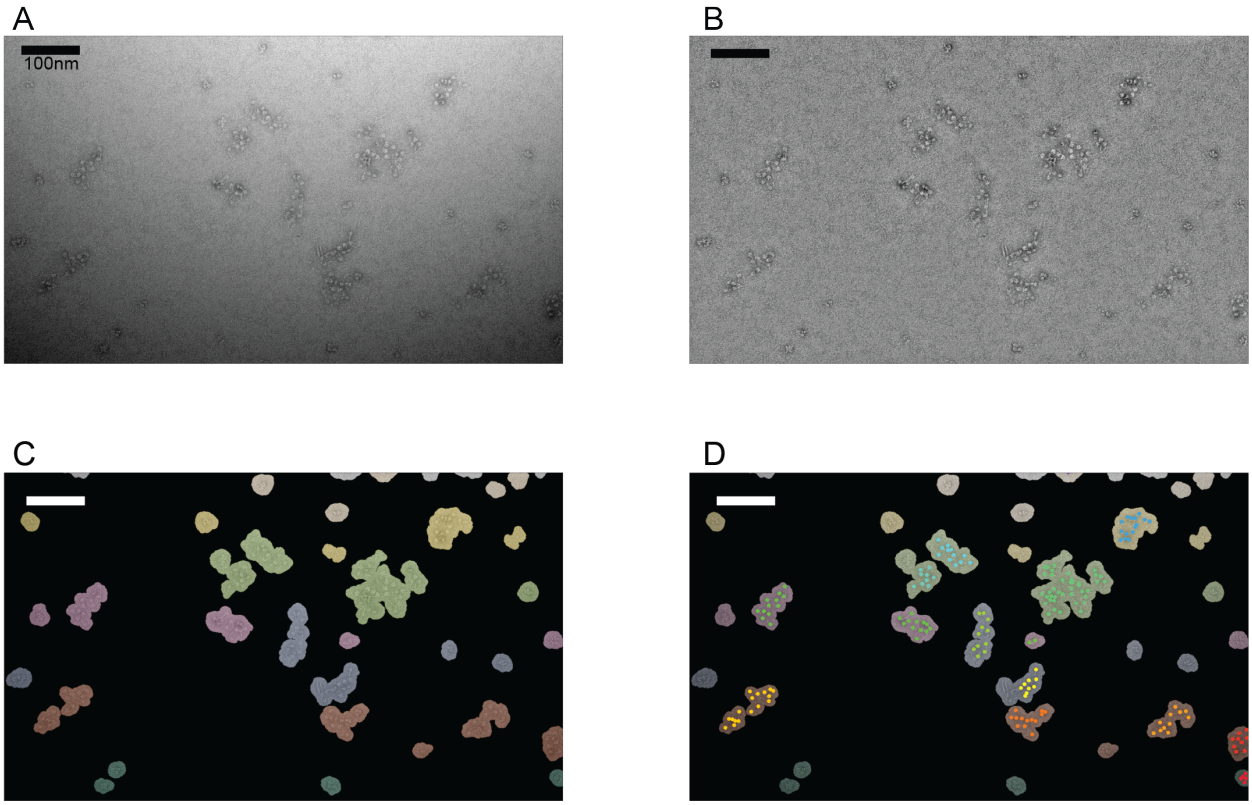


Figure 8.8. Workflow for analyzing the electron micrographs of nucleosome arrays. (**A**) A representative negative-stain EM micrograph of 12-mer nucleosome arrays mixed with PRC2 at a 1:1 ratio (150 nM each). (**B**) The same micrograph after illumination correction to achieve a uniform background. (**C**) Binary masks of individualized nucleosome arrays. (**D**) Manually picked nucleosomes (colored dots) inside their corresponding masks. See *Materials and Methods* for details.

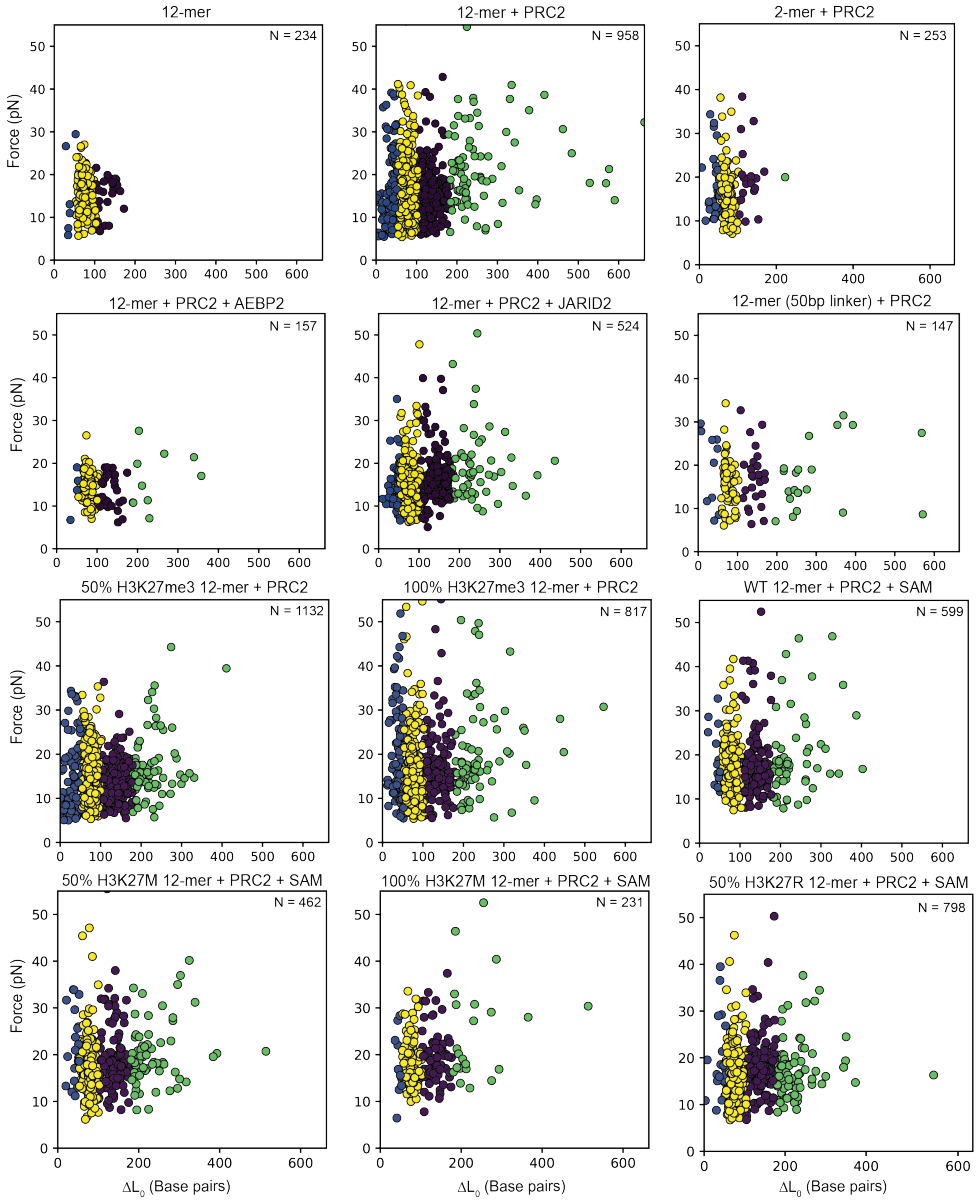


Figure 8.9. Cluster analysis of force-induced transitions under different experimental conditions. X-axis and y-axis represent the contour length change and the rupture force of each transition, respectively. Blue, yellow, purple, and green symbols represent Cluster 1, Cluster 2, Cluster 3, and Cluster 4 transitions, respectively. *N* denotes the number of transitions collected under each condition.

Table 8.1. Statistics of force-induced transitions collected with different PRC2–chromatin assemblies. The nucleosome arrays all have 30-bp-long linker DNA unless noted otherwise. Force values are presented as mean ± SEM. Individual data points are plotted in **Figure 8.9**.

Assembly	# of transitions	Average transition force (pN)	**Cluster 1** transitions		**Cluster 2** transitions		**Cluster 3** transitions		**Cluster 4** transitions	
			#	Transition force (pN)	#	Transition force (pN)	#	Transition force (pN)	#	Transition force (pN)
12-mer	234	16.0 ± 0.3	6	15.6 ± 4.1	203	16.2 ± 0.3	25	14.7 ± 0.9	0	-
12-mer + PRC2	958	17.0 ± 0.3	119	16.8 ± 0.8	541	16.6 ± 0.3	212	16.3 ± 0.4	86	21.4 ± 1.0
2-mer + PRC2	253	17.5 ± 0.3	36	17.4 ± 1.0	198	17.3 ± 0.3	18	20.1 ± 1.8	1	-
12-mer + PRC2 + AEBP2	157	14.5 ± 0.3	4	13.8 ± 2.7	112	14.7 ± 0.3	31	13.3 ± 0.7	10	16.3 ± 2.1
12-mer + PRC2 + JARID2	524	16.7 ± 0.3	43	15.3 ± 0.9	277	16.1 ± 0.3	149	17.2 ± 0.5	55	19.0 ± 1.1
12-mer (50-bp linker) + PRC2	147	17.1 ± 0.6	10	19.4 ± 2.7	88	16.1 ± 0.5	27	17.3 ± 1.3	22	19.6 ± 2.9
50% H3K27me3 12-mer + PRC2	1132	15.0 ± 0.1	112	13.7 ± 0.7	801	15.1 ± 0.1	316	14.6 ± 0.2	83	16.9 ± 0.8
100% H3K27me3 12-mer + PRC2	817	17.3 ± 0.4	103	20.3 ± 1.3	475	16.2 ± 0.4	163	17.2 ± 0.7	76	21.0 ± 1.3
WT 12-mer + PRC2 + SAM	599	17.8 ± 0.3	17	18.5 ± 1.6	350	17.4 ± 0.3	175	17.8 ± 0.5	57	19.8 ± 1.2
50% H3K27M 12-mer + PRC2 + SAM	462	18.3 ± 0.3	22	21.5 ± 1.6	252	17.2 ± 0.4	126	19.1 ± 0.6	62	20.3 ± 0.9
100% H3K27M 12-mer + PRC2 + SAM	231	19.6 ± 0.5	10	18.1 ± 2.2	126	18.5 ± 0.4	71	19.6 ± 0.7	24	26.1 ± 2.6
100% H3K27R 12-mer + PRC2 + SAM	798	17.2 ± 0.3	23	19.3 ± 1.7	495	16.8 ± 0.2	198	18.0 ± 0.4	82	17.7 ± 0.7

Table 8.2. Oligonucleotide sequences for histone H1 condensate formation

Name	Length (nt/bp)	GC%	Sequence
$ssDNA_{16}$	16	50	TGT ACT TCT AGA GGC C
$ssDNA_{30}$	30	50	GCA TGT ATT GAA GTA CTT CTA GAG GCC GCC
$dsDNA_{30}$	30	50	same as above
RNA_{30}	30	50	rGrCrA rUrGrU rArUrU rGrArA rGrUrA rCrUrU rCrUrA rGrArG rGrCrC rGrCrC
$ssDNA_{70}$	70	50	GCA TGT ATT GAA GTA CTT CTA GAG GCC GCC GCA TGT ATT GAA GTA CTT CTA GAG GCC GCC TGC CTC TAG A
$dsDNA_{70}$	70	50	same as above
$ssDNA_{75}$	75	33	CCA AAA CTA ATC TTT ATA TAA ATG GGA GAC TCA CTA TCA TGG CTA CTA CGA CTA CTA AAC GAG GTT AAA ATG AAT
$dsDNA_{75}$	75	33	same as above
$ssDNA_{147}$	147	56	CT GGA GAA TCC CGG TGC CGA GGC CGC TCA ATT GGT CGT AGA CAG CTC TAG CAC CGC TTA AAC GCA CGT ACG CGC TGT CCC CCG CGT TTT AAC CGC CAA GGG GAT TAC TCC CTA GTC TCC AGG CAC GTG TCA GAT ATA TAC ATC CTG T

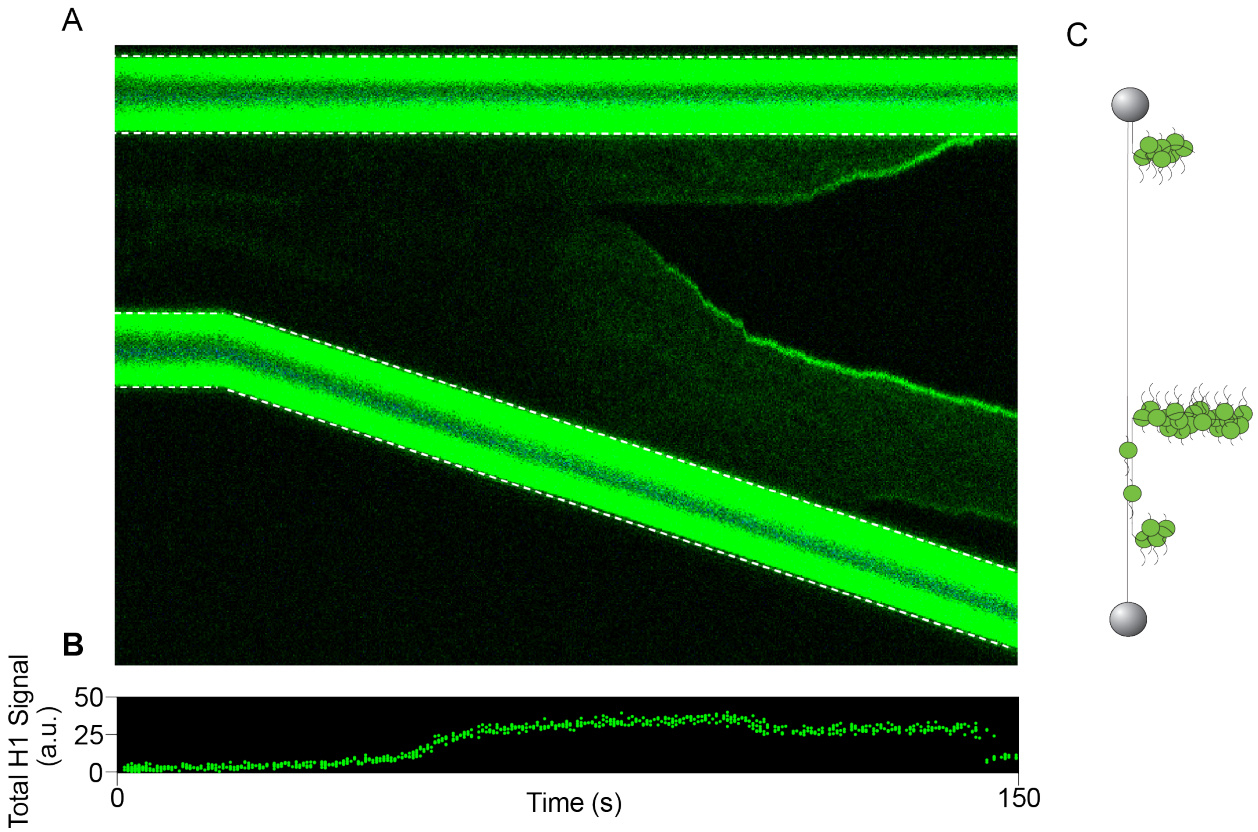


Figure 8.10. Additional H1 kymograph. (**A**) A representative kymograph of Cy3-H1 binding to DNA over time as tension is increased. (**B**) Normalized plot of Cy3 intensity over time for kymograph in **A**. (**C**) Schematic of Cy3-H1 binding to ssDNA.

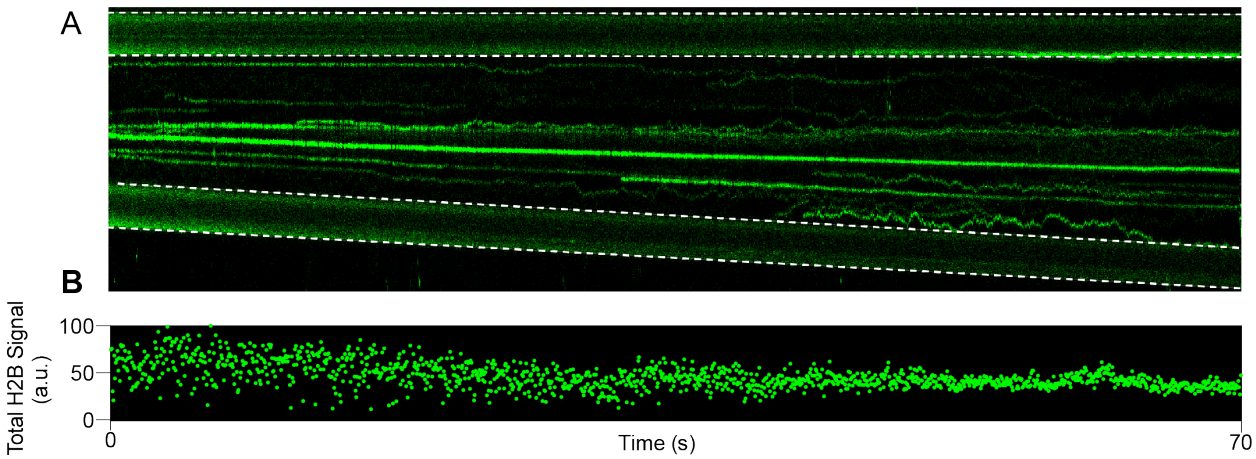


Figure 8.11. Example H2B kymograph. (**A**) A representative kymograph of Cy3-H2B binding to DNA over time as tension is increased. (**A**) Normalized plot of Cy3 intensity over time for kymograph in **A**.

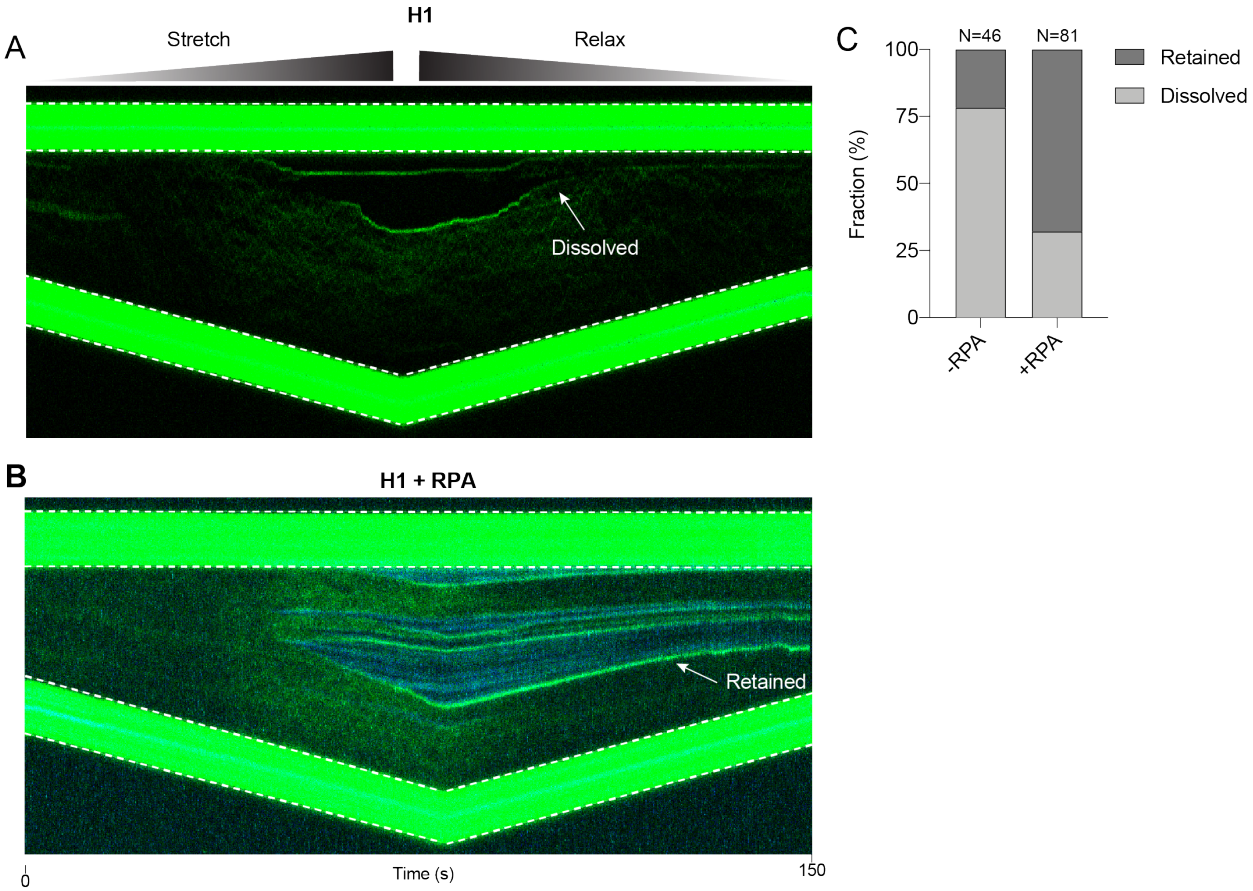


Figure 8.12. Reversibility o H1:ssDNA condensates +/- RPA. (**A**) Reversible formation and dissolution of Cy3-H1 condensates on DNA during stretching and relaxing. (**B**) Formation of Cy3-H1 condensates on DNA in the presence of RPA during stretching and relaxing. Condensates remain formed on DNA during relaxation. (**C**) Fraction of condensates dissolved for Cy3-H1 and Cy3-H1 plus RPA.

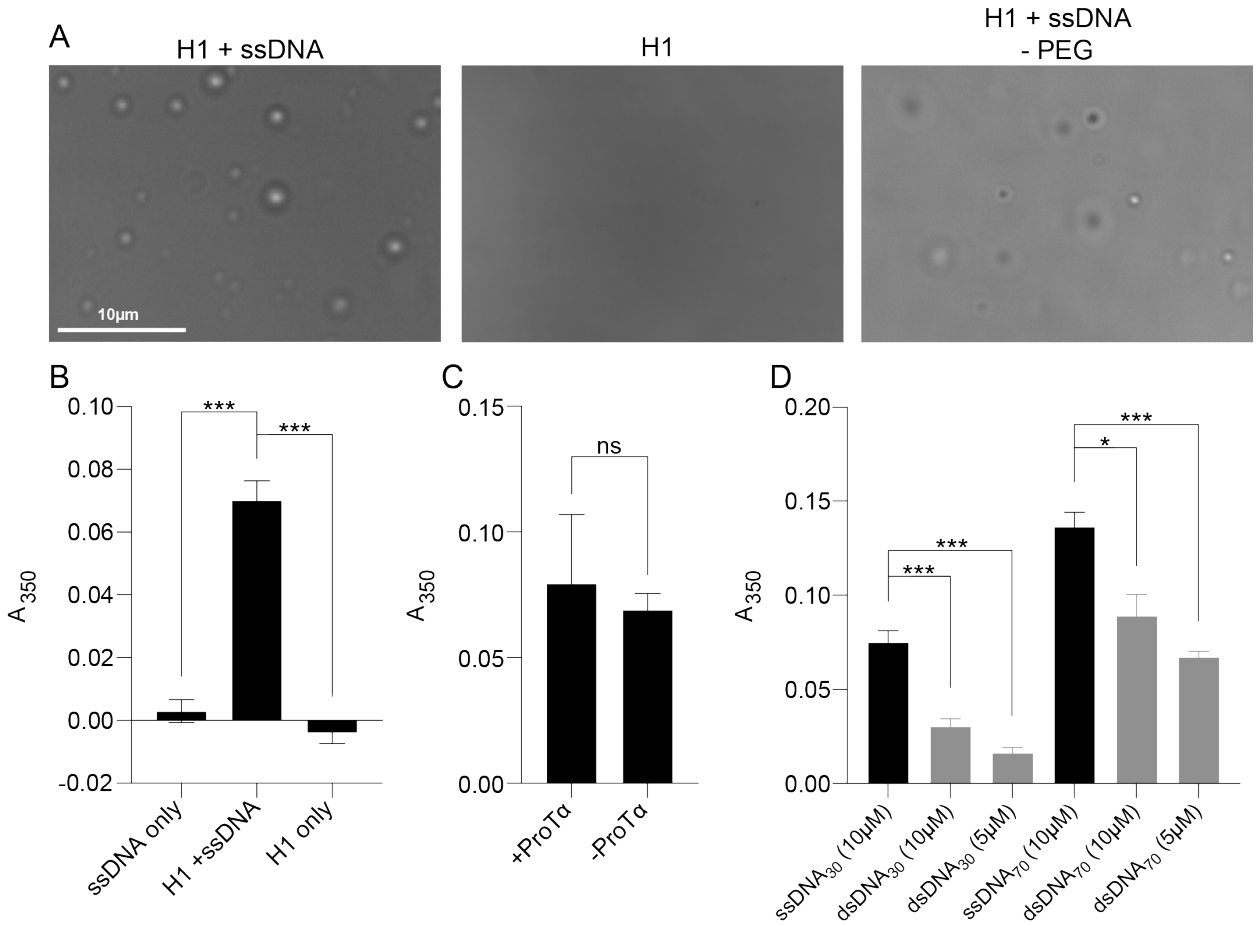


Figure 8.13. Brightfield images and additional A_{350} results. (**A**) Brightfield images of H1 (2.5 μM) and $ssDNA_{75}$ (10 μM) droplets with (left), without DNA (middle), and without PEG (right). All droplet experiments performed with 2.5 μM H1 with the exception of image with PEG, which contains 10 μM H1. (**B**) Comparison between average A_{350} turbidity values without H1 or 10 μM $ssDNA_{75}$. (**C**) Comparison between average turbidity values for H1:$ssDNA_{75}$ (10 μM) with or without chaperone ProT-α. (**D**) Comparison between average turbidity values for 30 and 70 nt 10 μM ssDNA, 10 μM dsDNA, and 5 μM dsDNA (corrected for the same number of bases).

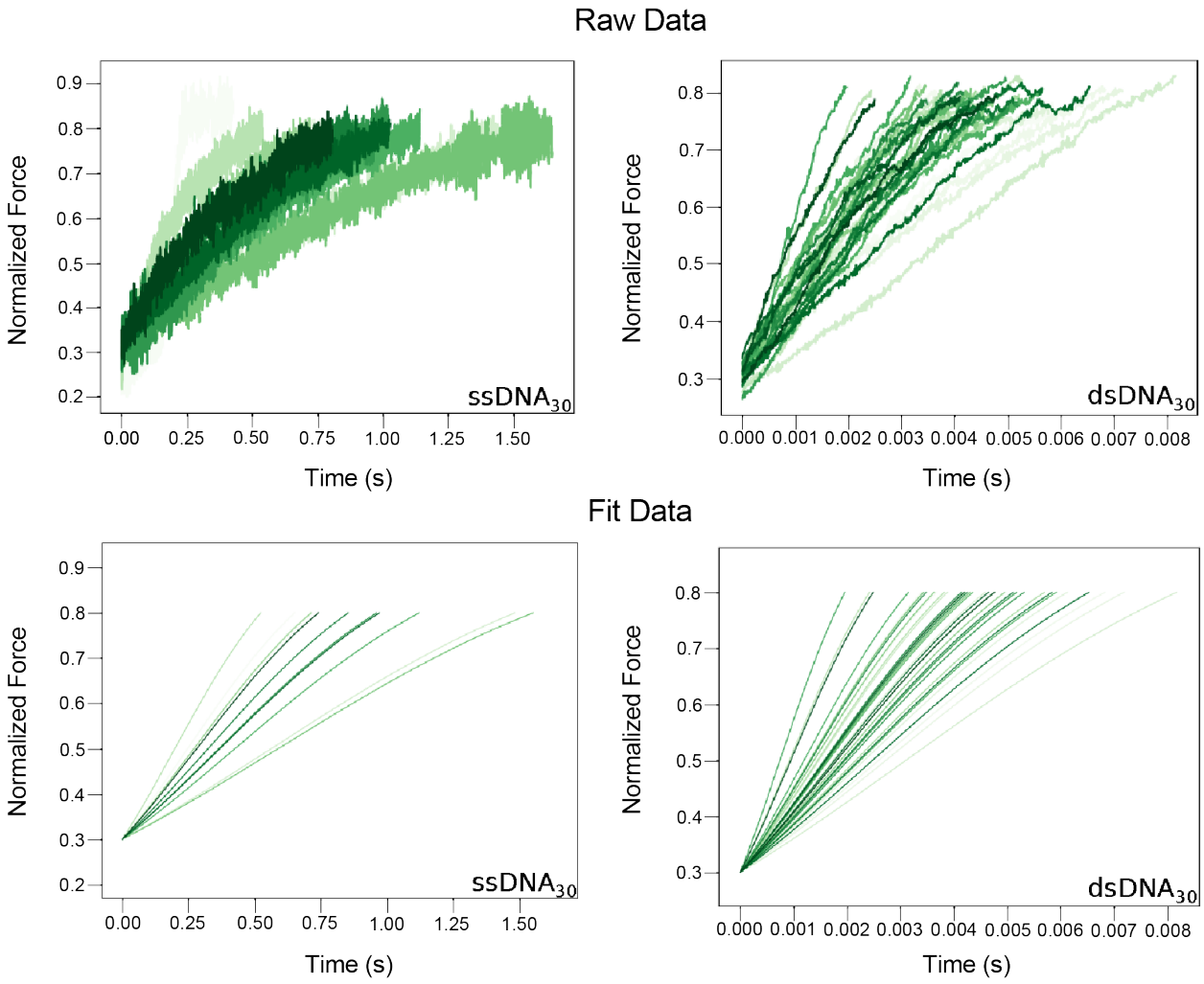


Figure 8.14. Additional force profiles for H1:ssDNA and H1:dsDNA droplets. Plots of the raw, high-frequency data between the interpolated points of droplet fusion (normalized force values of 0.3 and 0.8) from the sigmoidal fit for 30 bases of ssDNA (top left) and dsDNA (top right). Plots showing the calculated sigmoidal fits in the relevant time ranges for droplet fusion: 30 ssDNA (bottom left) and (bottom right).

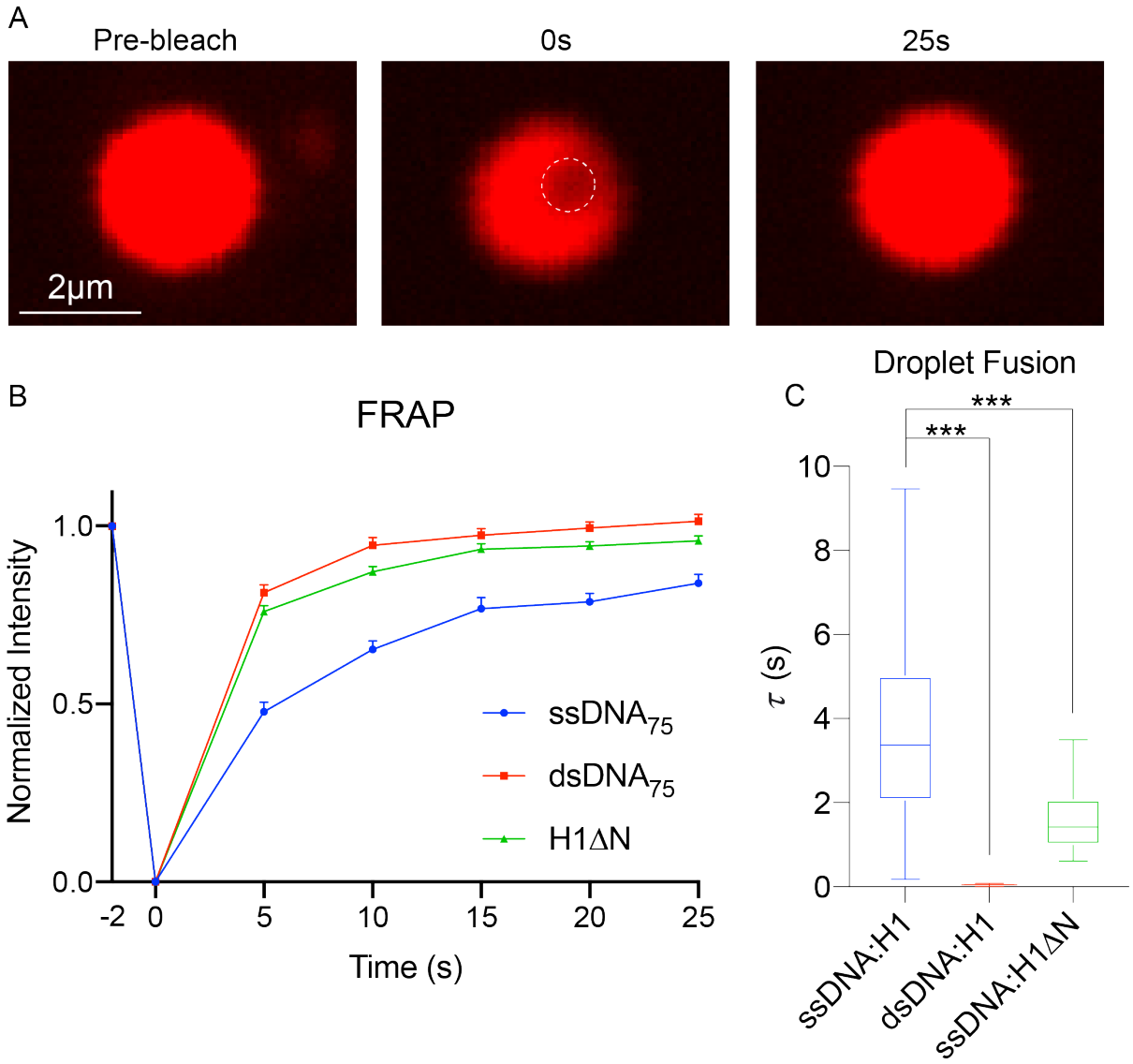


Figure 8.15. FRAP results. (**A**) Representative images of droplet photobleaching and recovery for Cy5-ssDNA$_{75}$ (10 μM, 10% labeled) and 2.5 μM H1 droplets. All droplet experiments performed with 2.5 μM H1 unless otherwise noted. (**B**) Quantification of fluorescence recovery over time for H1:Cy5-ssDNA$_{75}$ (10 μM), H1:Cy5-dsDNA$_{75}$ (10 μM), and dNTD H1:ssDNA$_{75}$ (10 μM) droplets (all DNA are 10% labeled). (**C**) Average fusion time for H1 droplets with ssDNA$_{75}$, dsDNA$_{75}$, and dNTD H1 with ssDNA$_{75}$. At least 12 fusion events were quantified for each condition.

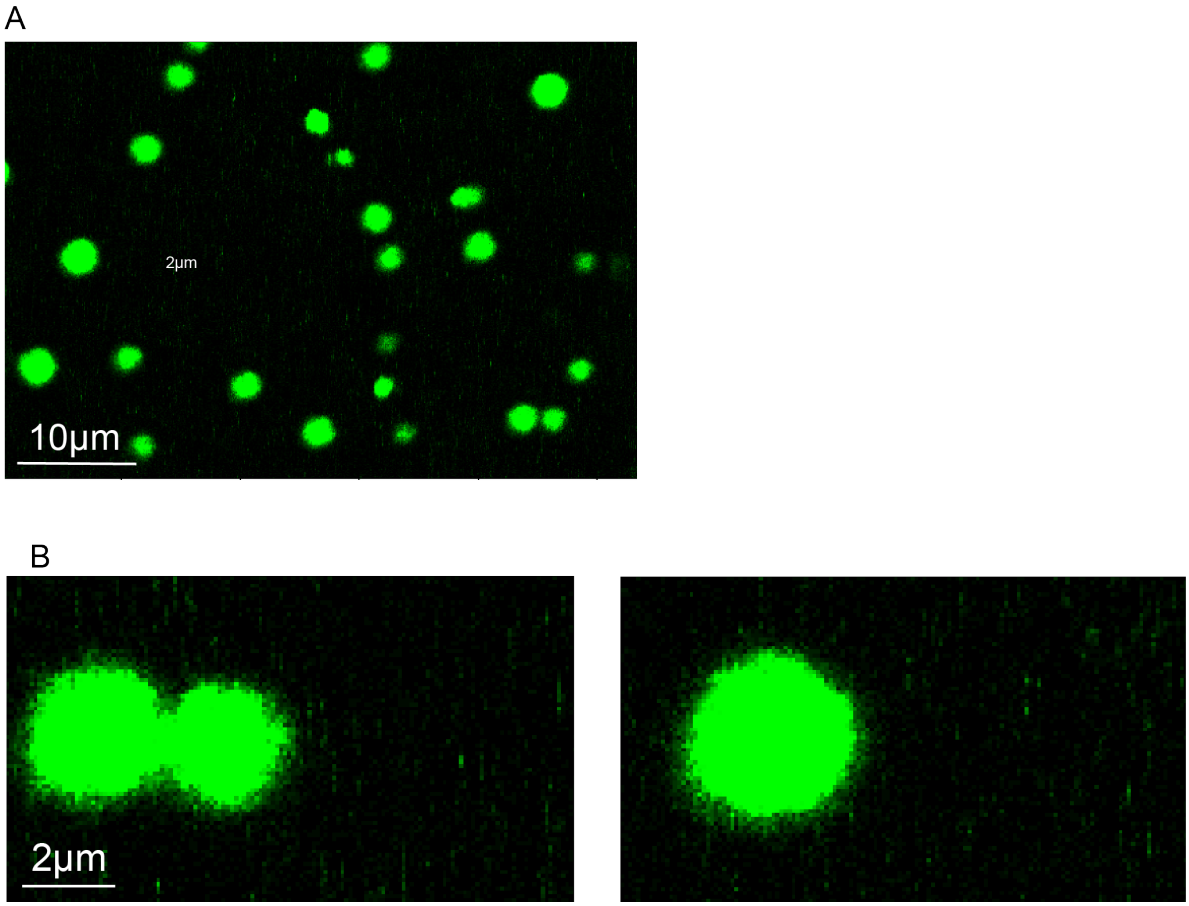


Figure 8.16. eGFP-H1 droplets. (**A**) Sample image of 2.5 μM H1-GFP droplets *in vitro* with 10 μM ssDNA$_{75}$. (**B**) H1-GFP and 10 μM ssDNA$_{75}$ fusion images.

CHAPTER 9. References

1. Muller, H.J. and E. Altenburg, *The Frequency of Translocations Produced by X-Rays in Drosophila.* Genetics, 1930. **15**(4): p. 283-311.

2. Lyon, M.F., *Gene Action in the X-chromosome of the Mouse (Mus musculus L.).* Nature, 1961. **190**(4773): p. 372-373.

3. Passarge, E., *Emil Heitz and the concept of heterochromatin: longitudinal chromosome differentiation was recognized fifty years ago.* American journal of human genetics, 1979. **31**(2): p. 106-115.

4. Brown, S.W., *Heterochromatin.* Science, 1966. **151**(3709): p. 417.

5. Trojer, P. and D. Reinberg, *Facultative Heterochromatin: Is There a Distinctive Molecular Signature?* Molecular Cell, 2007. **28**(1): p. 1-13.

6. Allis, C.D. and T. Jenuwein, *The molecular hallmarks of epigenetic control.* Nat Rev Genet, 2016. **17**(8): p. 487-500.

7. Luger, K., et al., *Crystal structure of the nucleosome core particle at 2.8A resolution.* Nature, 1997. **389**(6648): p. 251-260.

8. Olins, D.E. and A.L. Olins, *Chromatin history: our view from the bridge.* Nature Reviews Molecular Cell Biology, 2003. **4**(10): p. 809-814.

9. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS.* Proceedings of the National Academy of Sciences of the United States of America, 1964. **51**(5): p. 786-794.

10. Brownell, J.E., et al., *Tetrahymena Histone Acetyltransferase A: A Homolog to Yeast Gcn5p Linking Histone Acetylation to Gene Activation.* Cell, 1996. **84**(6): p. 843-851.

11. Raurell-Vila, H., et al., *An HP1 isoform-specific feedback mechanism regulates Suv39h1 activity under stress conditions.* Epigenetics, 2017. **12**(2): p. 166-175.

12. Aagaard, L., et al., *Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31.* The EMBO journal, 1999. **18**(7): p. 1923-1938.

13. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-45.

14. Margueron, R. and D. Reinberg, *Chromatin structure and the inheritance of epigenetic information.* Nature reviews. Genetics, 2010. **11**(4): p. 285-296.

15. Ransom, M., B.K. Dennehey, and J.K. Tyler, *Chaperoning Histones during DNA Replication and Repair.* Cell, 2010. **140**(2): p. 183-195.

16. Allfrey, V.G., R. Faulkner, and A.E. Mirsky, *ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS.* Proceedings of the National Academy of Sciences, 1964. **51**(5): p. 786.

17. Marmorstein, R. and M.-M. Zhou, *Writers and Readers of Histone Acetylation: Structure, Mechanism, and Inhibition.* Cold Spring Harbor Perspectives in Biology, 2014. **6**(7).

18. Tschiersch, B., et al., *The protein encoded by the Drosophila position-effect variegation suppressor gene Su(var)3-9 combines domains of antagonistic regulators of homeotic gene complexes.* The EMBO Journal, 1994. **13**(16): p. 3822-3831.

19. Muller, M.M., et al., *A two-state activation mechanism controls the histone methyltransferase Suv39h1.* Nat Chem Biol, 2016. **12**(3): p. 188-193.

20. Rea, S., et al., *Regulation of chromatin structure by site-specific histone H3 methyltransferases.* Nature, 2000. **406**(6796): p. 593-599.

21. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-120.

22. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-124.

23. Milne, T.A., et al., *MLL Targets SET Domain Methyltransferase Activity to Hox Gene Promoters.* Molecular Cell, 2002. **10**(5): p. 1107-1117.

24. Dou, Y., et al., *Regulation of MLL1 H3K4 methyltransferase activity by its core components.* Nature Structural & Molecular Biology, 2006. **13**(8): p. 713-719.

25. Yokoyama, A., et al., *Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin To Regulate Hox Gene Expression.* Molecular and Cellular Biology, 2004. **24**(13): p. 5639.

26. Müller, J., et al., *Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex.* Cell, 2002. **111**(2): p. 197-208.

27. Cao, R., et al., *Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing.* Science, 2002. **298**(5595): p. 1039.

28. Willcockson, M.A., et al., *H1 histones control the epigenetic landscape by local chromatin compaction.* Nature, 2020.

29. Risca, V.I., et al., *Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping.* Nature, 2017. **541**(7636): p. 237-241.

30. Allis, C.D. and T.W. Muir, *Spreading Chromatin into Chemical Biology.* ChemBioChem, 2011. **12**(2): p. 264-279.

31. Daban, J.-R., *Electron microscopy and atomic force microscopy studies of chromatin and metaphase chromosome structure.* Micron, 2011. **42**(8): p. 733-750.

32. Grigoryev, S.A., *Nucleosome spacing and chromatin higher-order folding.* Nucleus (Austin, Tex.), 2012. **3**(6): p. 493-499.

33. Bartolome, S., A. Bermudez, and J.R. Daban, *Internal structure of the 30 nm chromatin fiber.* Journal of Cell Science, 1994. **107**(11): p. 2983.

34. Song, F., et al., *Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units.* Science, 2014. **344**(6182): p. 376.

35. Ou, H.D., et al., *ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells.* Science, 2017. **357**(6349): p. eaag0025.

36. Szabo, Q., F. Bantignies, and G. Cavalli, *Principles of genome folding into topologically associating domains.* Science Advances, 2019. **5**(4): p. eaaw1668.

37. Dixon, J.R., D.U. Gorkin, and B. Ren, *Chromatin Domains: The Unit of Chromosome Organization.* Molecular cell, 2016. **62**(5): p. 668-680.

38. Cattoni, D.I., et al., *Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions.* Nature Communications, 2017. **8**(1): p. 1753.

39. Cuvier, O. and B. Fierz, *Dynamic chromatin technologies: from individual molecules to epigenomic regulation in cells.* Nature Reviews Genetics, 2017. **18**(8): p. 457-472.

40. Ashkin, A., et al., *Observation of a single-beam gradient force optical trap for dielectric particles.* Optics Letters, 1986. **11**(5): p. 288-290.

41. Neuman, K.C. and A. Nagy, *Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy.* Nature methods, 2008. **5**(6): p. 491-505.

42. Block, S.M., D.F. Blair, and H.C. Berg, *Compliance of bacterial flagella measured with optical tweezers.* Nature, 1989. **338**(6215): p. 514-518.

43. Block, S.M., L.S.B. Goldstein, and B.J. Schnapp, *Bead movement by single kinesin molecules studied with optical tweezers.* Nature, 1990. **348**(6299): p. 348-352.

44. Abbondanzieri, E.A., et al., *Direct observation of base-pair stepping by RNA polymerase.* Nature, 2005. **438**(7067): p. 460-465.

45. Smith, D.E., et al., *The bacteriophage φ29 portal motor can package DNA against a large internal force.* Nature, 2001. **413**(6857): p. 748-752.

46. Liu, S., G. Chistol, and C. Bustamante, *Mechanical operation and intersubunit coordination of ring-shaped molecular motors: insights from single-molecule studies.* Biophysical journal, 2014. **106**(9): p. 1844-1858.

47. Hao, Y., et al., *Integrated Method to Attach DNA Handles and Functionally Select Proteins to Study Folding and Protein-Ligand Interactions with Optical Tweezers.* Scientific Reports, 2017. **7**(1): p. 10843.

48. Bustamante, C., et al., *Entropic elasticity of lambda-phage DNA.* Science, 1994. **265**(5178): p. 1599.

49. Bouchiat, C., et al., *Estimating the persistence length of a worm-like chain molecule from force-extension measurements.* Biophysical journal, 1999. **76**(1 Pt 1): p. 409-413.

50. Smith, S.B., Y. Cui, and C. Bustamante, *Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules.* Science, 1996. **271**(5250): p. 795.

51. Bustamante, C., W. Cheng, and Y.X. Mejia, *Revisiting the Central Dogma One Molecule at a Time.* Cell, 2011. **144**(4): p. 480-497.

52. Killian, J.L., F. Ye, and M.D. Wang, *Optical Tweezers: A Force to Be Reckoned With.* Cell, 2018. **175**(6): p. 1445-1448.

53. Zhou, K., G. Gaullier, and K. Luger, *Nucleosome structure and dynamics are coming of age.* Nature structural & molecular biology, 2019. **26**(1): p. 3-13.

54. Fierz, B. and M.G. Poirier, *Biophysics of Chromatin Dynamics.* Annual Review of Biophysics, 2019. **48**(1): p. 321-345.

55. Brower-Toland, B.D., et al., *Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA.* Proceedings of the National Academy of Sciences, 2002. **99**(4): p. 1960.

56. Cui, Y. and C. Bustamante, *Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure.* Proceedings of the National Academy of Sciences, 2000. **97**(1): p. 127.

57. Mihardja, S., et al., *Effect of force on mononucleosomal dynamics.* Proceedings of the National Academy of Sciences, 2006. **103**(43): p. 15871.

58. Pope, L.H., et al., *Single Chromatin Fiber Stretching Reveals Physically Distinct Populations of Disassembly Events.* Biophysical Journal, 2005. **88**(5): p. 3572-3583.

59. Mack, A.H., et al., *Kinetics and thermodynamics of phenotype: unwinding and rewinding the nucleosome.* Journal of molecular biology, 2012. **423**(5): p. 687-701.

60. Kaczmarczyk, A., et al., *Single-molecule force spectroscopy on histone H4 tail-cross-linked chromatin reveals fiber folding.* The Journal of biological chemistry, 2017. **292**(42): p. 17506-17513.

61. Bennink, M.L., et al., *Unfolding individual nucleosomes by stretching single chromatin fibers with optical tweezers.* Nature Structural Biology, 2001. **8**(7): p. 606-610.

62. Poirier, M.G. and J.F. Marko, *Mitotic chromosomes are chromatin networks without a mechanically contiguous protein scaffold.* Proceedings of the National Academy of Sciences, 2002. **99**(24): p. 15393.

63. Hall, M.A., et al., *High-resolution dynamic mapping of histone-DNA interactions in a nucleosome.* Nature Structural & Molecular Biology, 2009. **16**(2): p. 124-129.

64. Kruithof, M., et al., *Single-molecule force spectroscopy reveals a highly compliant helical folding for the 30-nm chromatin fiber.* Nature Structural & Molecular Biology, 2009. **16**(5): p. 534-540.

65. Vlijm, R., et al., *The supercoiling state of DNA determines the handedness of both H3 and CENP-A nucleosomes.* Nanoscale, 2017. **9**(5): p. 1862-1870.

66. Ordu, O., A. Lusser, and N.H. Dekker, *Recent insights from in vitro single-molecule studies into nucleosome structure and dynamics.* Biophysical reviews, 2016. **8**(Suppl 1): p. 33-49.

67. Hashemi Shabestari, M., et al., *Chapter Four - Recent Advances in Biological Single-Molecule Applications of Optical Tweezers and Fluorescence Microscopy*, in *Methods in Enzymology*, M. Spies and Y.R. Chemla, Editors. 2017, Academic Press. p. 85-119.

68. Wasserman, M.R., et al., *Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase.* Cell, 2019. **178**(3): p. 600-611.e16.

69. Gutierrez-Escribano, P., et al., *A conserved ATP- and Scc2/4-dependent activity for cohesin in tethering DNA molecules.* Science Advances, 2019. **5**(11): p. eaay6804.

70. Newton, M.D., et al., *DNA stretching induces Cas9 off-target activity.* Nature Structural & Molecular Biology, 2019. **26**(3): p. 185-192.

71. Hyman, A.A., C.A. Weber, and F. Jülicher, *Liquid-Liquid Phase Separation in Biology.* Annual Review of Cell and Developmental Biology, 2014. **30**(1): p. 39-58.

72. Flory, P.J., *Thermodynamics of High Polymer Solutions.* The Journal of Chemical Physics, 1942. **10**(1): p. 51-61.

73. Boeynaems, S., et al., *Protein Phase Separation: A New Phase in Cell Biology.* Trends in Cell Biology, 2018. **28**(6): p. 420-435.

74. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry.* Nature Reviews Molecular Cell Biology, 2017. **18**(5): p. 285-298.

75. Sheu-Gruttadauria, J. and I.J. MacRae, *Phase Transitions in the Assembly and Function of Human miRISC.* Cell, 2018. **173**(4): p. 946-957.e16.

76. Feric, M., et al., *Coexisting Liquid Phases Underlie Nucleolar Subcompartments.* Cell, 2016. **165**(7): p. 1686-1697.

77. Schmidt, H.B. and D. Görlich, *Transport Selectivity of Nuclear Pores, Phase Separation, and Membraneless Organelles.* Trends Biochem Sci, 2016. **41**(1): p. 46-61.

78. Bergeron-Sandoval, L.-P., et al., *Endocytosis caused by liquid-liquid phase separation of proteins.* bioRxiv, 2018: p. 145664.

79. Patel, A., et al., *A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.* Cell, 2015. **162**(5): p. 1066-1077.

80. Larson, A.G. and G.J. Narlikar, *The Role of Phase Separation in Heterochromatin Formation, Function, and Regulation.* Biochemistry, 2018. **57**(17): p. 2540-2548.

81. Yoshizawa, T., et al., *Biological phase separation: cell biology meets biophysics.* Biophysical Reviews, 2020. **12**(2): p. 519-539.

82. Strom, A.R., et al., *Phase separation drives heterochromatin domain formation.* Nature, 2017. **547**(7662): p. 241-245.

83. Keenen, M.M., et al., *HP1 proteins compact DNA into mechanically and positionally stable phase separated domains.* bioRxiv, 2020: p. 2020.10.30.362772.

84. Sabari, B.R., et al., *Coactivator condensation at super-enhancers links phase separation and gene control.* Science, 2018. **361**(6400): p. eaar3958.

85. Larson, A.G., et al., *Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin.* Nature, 2017. **547**(7662): p. 236-240.

86. Sanulli, S., et al., *HP1 reshapes nucleosome core to promote phase separation of heterochromatin.* Nature, 2019. **575**(7782): p. 390-394.

87. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484.e21.

88. Wang, L., et al., *Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism.* Molecular Cell, 2019. **76**(4): p. 646-659.e6.

89. Erdel, F. and K. Rippe, *Formation of Chromatin Subcompartments by Phase Separation.* Biophysical Journal, 2018. **114**(10): p. 2262-2270.

90. Brackley, C.A., et al., *Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization.* Proceedings of the National Academy of Sciences, 2013. **110**(38): p. E3605.

91. Strickfaden, H., et al., *Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells.* Cell.

92. Kornberg, R.D., *Chromatin Structure: A Repeating Unit of Histones and DNA.* Science, 1974. **184**(4139): p. 868.

93. Zaret, K.S., J. Lerner, and M. Iwafuchi-Doi, *Chromatin Scanning by Dynamic Binding of Pioneer Factors.* Molecular cell, 2016. **62**(5): p. 665-667.

94. Clapier, C.R., et al., *Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes.* Nature Reviews Molecular Cell Biology, 2017. **18**(7): p. 407-422.

95. Margueron, R. and D. Reinberg, *The Polycomb complex PRC2 and its mark in life.* Nature, 2011. **469**(7330): p. 343-349.

96. Leicher, R., et al., *Single-molecule and in silico dissection of the interaction between Polycomb repressive complex 2 and chromatin.* Proceedings of the National Academy of Sciences, 2020. **117**(48): p. 30465.

97. Ngo, T.T.M., et al., *Asymmetric unwrapping of nucleosomes under tension directed by DNA local flexibility.* Cell, 2015. **160**(6): p. 1135-1144.

98. Ge, E.J., et al., *Nucleation and Propagation of Heterochromatin by the Histone Methyltransferase PRC2: Geometric Constraints and Impact of the Regulatory Subunit JARID2.* Journal of the American Chemical Society, 2019. **141**(38): p. 15029-15039.

99. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning1 1Edited by T. Richmond.* Journal of Molecular Biology, 1998. **276**(1): p. 19-42.

100. Lusser, A. and J.T. Kadonaga, *Strategies for the reconstitution of chromatin.* Nature Methods, 2004. **1**(1): p. 19-26.

101. Wang, M.D., et al., *Stretching DNA with optical tweezers.* Biophysical journal, 1997. **72**(3): p. 1335-1346.

102. Wasserman, M.R. and S. Liu, *A Tour de Force on the Double Helix: Exploiting DNA Mechanics To Study DNA-Based Molecular Machines.* Biochemistry, 2019. **58**(47): p. 4667-4676.

103. Müller, M.M. and T.W. Muir, *Histones: At the Crossroads of Peptide and Protein Chemistry.* Chemical Reviews, 2015. **115**(6): p. 2296-2349.

104. Marko, J.F. and E.D. Siggia, *Stretching DNA.* Macromolecules, 1995. **28**(26): p. 8759-8770.

105. Wang, Z., G. Zhang, and H. Zhang, *Protocol for analyzing protein liquid–liquid phase separation.* Biophysics Reports, 2019. **5**(1): p. 1-9.

106. Weber, S.C., *Sequence-encoded material properties dictate the structure and function of nuclear bodies.* Curr Opin Cell Biol, 2017. **46**: p. 62-71.

107. Quail, T., et al., *Capillary forces drive pioneer transcription factor-mediated DNA condensation.* bioRxiv, 2020: p. 2020.09.17.302299.

108. Alshareedah, I., et al., *Phase transition of RNA−protein complexes into ordered hollow condensates.* Proceedings of the National Academy of Sciences, 2020. **117**(27): p. 15650.

109. Saha, S., et al., *Polar Positioning of Phase-Separated Liquid Compartments in Cells Regulated by an mRNA Competition Mechanism.* Cell, 2016. **166**(6): p. 1572-1584.e16.

110. Jawerth, L., et al., *Protein condensates as aging Maxwell fluids.* Science, 2020. **370**(6522): p. 1317.

111. Ghosh, A. and H.-X. Zhou, *Determinants for Fusion Speed of Biomolecular Droplets.* Angewandte Chemie International Edition, 2020. **59**(47): p. 20837-20840.

112. Ph, L., *Pc: Polycomb.* Drosophila information service., 1949. **21**: p. 69.

113. Lewis, E.B., *A gene complex controlling segmentation in Drosophila.* Nature, 1978. **276**(5688): p. 565-570.

114. Schuettengruber, B. and G. Cavalli, *Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice.* Development, 2009. **136**(21): p. 3531.

115. Klymenko, T., et al., *A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities.* Genes & development, 2006. **20**(9): p. 1110-1122.

116. Simon, J.A. and R.E. Kingston, *Mechanisms of Polycomb gene silencing: knowns and unknowns.* Nature Reviews Molecular Cell Biology, 2009. **10**(10): p. 697-708.

117. Müller, J. and J.A. Kassis, *Polycomb response elements and targeting of Polycomb group proteins in Drosophila.* Current Opinion in Genetics & Development, 2006. **16**(5): p. 476-484.

118. Whitcomb, S.J., et al., *Polycomb Group proteins: an evolutionary perspective.* Trends in Genetics, 2007. **23**(10): p. 494-502.

119. Margueron, R., et al., *Role of the polycomb protein EED in the propagation of repressive histone marks.* Nature, 2009. **461**(7265): p. 762-767.

120. Hansen, K.H., et al., *A model for transmission of the H3K27me3 epigenetic mark.* Nature Cell Biology, 2008. **10**(11): p. 1291-1300.

121. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nature Structural & Molecular Biology, 2018. **25**(2): p. 154-162.

122. Justin, N., et al., *Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2.* Nature Communications, 2016. **7**: p. 11316.

123. Kasinath, V., S. Poepsel, and E. Nogales, *Recent Structural Insights into Polycomb Repressive Complex 2 Regulation and Substrate Binding.* Biochemistry, 2019. **58**(5): p. 346-354.

124. Ciferri, C., et al., *Molecular architecture of human polycomb repressive complex 2.* eLife, 2012. **1**: p. e00005.

125. Kasinath, V., et al., *JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications.* Science, 2021. **371**(6527): p. eabc3393.

126. Wang, X., et al., *Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA.* Nature Structural & Molecular Biology, 2017. **24**(12): p. 1028-1038.

127. Lee, C.-H., et al., *Distinct Stimulatory Mechanisms Regulate the Catalytic Activity of Polycomb Repressive Complex 2.* Molecular Cell, 2018. **70**(3): p. 435-448.e5.

128. Grau, D., et al., *Structures of monomeric and dimeric PRC2:EZH1 reveal flexible modules involved in chromatin compaction.* Nature Communications, 2021. **12**(1): p. 714.

129. Kasinath, V., et al., *Structures of human PRC2 with its cofactors AEBP2 and JARID2.* Science (New York, N.Y.), 2018. **359**(6378): p. 940-944.

130. Grijzenhout, A., et al., *Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs.* Development, 2016. **143**(15): p. 2716.

131. Li, G., et al., *Jarid2 and PRC2, partners in regulating gene expression.* Genes & Development, 2010. **24**(4): p. 368-380.

132. Pasini, D., et al., *JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells.* Nature, 2010. **464**(7286): p. 306-310.

133. Sanulli, S., et al., *Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation.* Molecular Cell, 2015. **57**(5): p. 769-783.

134. Wang, S., G.P. Robertson, and J. Zhu, *A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers.* Gene, 2004. **343**(1): p. 69-78.

135. Nekrasov, M., et al., *Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes.* The EMBO journal, 2007. **26**(18): p. 4078-4088.

136. Yuan, W., et al., *Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation.* Science, 2012. **337**(6097): p. 971.

137. Heenan, P.R., et al., *Bending and looping of long DNA by Polycomb repressive complex 2 revealed by AFM imaging in liquid.* Nucleic acids research, 2020. **48**(6): p. 2969-2981.

138. Pengelly, A.R., et al., *A Histone Mutant Reproduces the Phenotype Caused by Loss of Histone-Modifying Factor Polycomb.* Science, 2013. **339**(6120): p. 698.

139. Bernstein, B.E., et al., *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.* Cell, 2006. **125**(2): p. 315-326.

140. Gaydos, L.J., W. Wang, and S. Strome, *H3K27me and PRC2 transmit a memory of repression across generations and during development.* Science, 2014. **345**(6203): p. 1515.

141. Zee, B.M., et al., *In Vivo Residue-specific Histone Methylation Dynamics.* Journal of Biological Chemistry, 2010. **285**(5): p. 3341-3350.

142. Peters, A.H.F.M., et al., *Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin.* Molecular Cell, 2003. **12**(6): p. 1577-1589.

143. Barski, A., et al., *High-Resolution Profiling of Histone Methylations in the Human Genome.* Cell, 2007. **129**(4): p. 823-837.

144. Lee, T.I., et al., *Control of Developmental Regulators by Polycomb in Human Embryonic Stem Cells.* Cell, 2006. **125**(2): p. 301-313.

145. Bracken, A.P., et al., *Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions.* Genes & Development, 2006. **20**(9): p. 1123-1136.

146. Al-Sady, B., Hiten D. Madhani, and Geeta J. Narlikar, *Division of Labor between the Chromodomains of HP1 and Suv39 Methylase Enables Coordination of Heterochromatin Spread.* Molecular Cell, 2013. **51**(1): p. 80-91.

147. Bryan, L.C., et al., *Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions.* Nucleic Acids Research, 2017. **45**(18): p. 10504-10517.

148. Kilic, S., et al., *Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1α.* Nature Communications, 2018. **9**(1): p. 235.

149. Canzio, D., et al., *Chromodomain-Mediated Oligomerization of HP1 Suggests a Nucleosome-Bridging Mechanism for Heterochromatin Assembly.* Molecular Cell, 2011. **41**(1): p. 67-81.

150. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proceedings of the National Academy of Sciences, 2010. **107**(50): p. 21931.

151. Mikkelsen, T.S., et al., *Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.* Nature, 2007. **448**(7153): p. 553-560.

152. Jani, K.S., et al., *Histone H3 tail binds a unique sensing pocket in EZH2 to activate the PRC2 methyltransferase.* Proceedings of the National Academy of Sciences, 2019. **116**(17): p. 8295.

153. Voigt, P., et al., *Asymmetrically Modified Nucleosomes.* Cell, 2012. **151**(1): p. 181-193.

154. Schmitges, Frank W., et al., *Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks.* Molecular Cell, 2011. **42**(3): p. 330-341.

155. Schoeftner, S., et al., *Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing.* The EMBO Journal, 2006. **25**(13): p. 3110-3122.

156. Grau, D.J., et al., *Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge.* Genes & Development, 2011. **25**(20): p. 2210-2221.

157. Eskeland, R., et al., *Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination.* Molecular cell, 2010. **38**(3): p. 452-464.

158. Lau, M.S., et al., *Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning.* Science, 2017. **355**(6329): p. 1081.

159. Fischle, W., et al., *Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains.* Genes & Development, 2003. **17**(15): p. 1870-1881.

160. Min, J., Y. Zhang, and R.-M. Xu, *Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27.* Genes & Development, 2003. **17**(15): p. 1823-1828.

161. Mujtaba, S., et al., *Epigenetic transcriptional repression of cellular genes by a viral SET protein.* Nature Cell Biology, 2008. **10**(9): p. 1114-1122.

162. Simon, J.A. and R.E. Kingston, *Mechanisms of Polycomb gene silencing: knowns and unknowns.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 697-708.

163. Vallot, C., et al., *PRC2-independent chromatin compaction and transcriptional repression in cancer.* Oncogene, 2015. **34**(6): p. 741-751.

164. Varambally, S., et al., *The polycomb group protein EZH2 is involved in progression of prostate cancer.* Nature, 2002. **419**(6907): p. 624-629.

165. Kleer, C.G., et al., *EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells.* Proceedings of the National Academy of Sciences, 2003. **100**(20): p. 11606.

166. Kim, K.H. and C.W.M. Roberts, *Targeting EZH2 in cancer.* Nat Med, 2016. **22**(2): p. 128-134.

167. Conway, E., E. Healy, and A.P. Bracken, *PRC2 mediated H3K27 methylations in cellular identity and cancer.* Current Opinion in Cell Biology, 2015. **37**: p. 42-48.

168. Lewis, P.W., et al., *Inhibition of PRC2 Activity by a Gain-of-Function H3 Mutation Found in Pediatric Glioblastoma.* Science, 2013. **340**(6134): p. 857.

169. Diehl, K.L., et al., *PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor.* Proceedings of the National Academy of Sciences, 2019. **116**(44): p. 22152.

170. Martin, M.C., et al., *Small Molecule Approaches for Targeting the Polycomb Repressive Complex 2 (PRC2) in Cancer.* Journal of Medicinal Chemistry, 2020.

171. Khanna, A., et al., *Design, Synthesis, and Pharmacological Evaluation of Second Generation EZH2 Inhibitors with Long Residence Time.* ACS Medicinal Chemistry Letters, 2020. **11**(6): p. 1205-1212.

172. He, Y., et al., *The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex.* Nat Chem Biol, 2017. **13**(4): p. 389-395.

173. Qi, W., et al., *An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED.* Nature Chemical Biology, 2017. **13**(4): p. 381-388.

174. Dong, H., et al., *An Allosteric PRC2 Inhibitor Targeting EED Suppresses Tumor Progression by Modulating the Immune Response.* Cancer Research, 2019. **79**(21): p. 5587.

175. Danishuddin, et al., *Polycomb repressive complex 2 inhibitors: emerging epigenetic modulators.* Drug Discovery Today, 2019. **24**(1): p. 179-188.

176. Brower-Toland, B.D., et al., *Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA.* Proc Natl Acad Sci U S A, 2002. **99**(4): p. 1960-5.

177. Hall, M.A., et al., *High-resolution dynamic mapping of histone-DNA interactions in a nucleosome.* Nat Struct Mol Biol, 2009. **16**(2): p. 124-9.

178. Mihardja, S., et al., *Effect of force on mononucleosomal dynamics.* Proc Natl Acad Sci U S A, 2006. **103**(43): p. 15871-6.

179. Wang, X., et al., *Molecular analysis of PRC2 recruitment to DNA in chromatin and its inhibition by RNA.* Nat Struct Mol Biol, 2017. **24**(12): p. 1028-1038.

180. Heenan, P.R., et al., *Bending and looping of long DNA by Polycomb repressive complex 2 revealed by AFM imaging in liquid.* Nucleic Acids Res, 2020.

181. Poepsel, S., V. Kasinath, and E. Nogales, *Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes.* Nat Struct Mol Biol, 2018. **25**(2): p. 154-162.

182. Noel, J.K., et al., *SMOG@ctbp: simplified deployment of structure-based models in GROMACS.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W657-61.

183. Noel, J.K., et al., *SMOG 2: A Versatile Software Package for Generating Structure-Based Models.* PLoS Comput Biol, 2016. **12**(3): p. e1004794.

184. Freeman, G.S., et al., *Coarse-grained modeling of DNA curvature.* J Chem Phys, 2014. **141**(16): p. 165103.

185. Lequieu, J., et al., *Tension-Dependent Free Energies of Nucleosome Unwrapping.* ACS Cent Sci, 2016. **2**(9): p. 660-666.

186. Parsons, T. and B. Zhang, *Critical role of histone tail entropy in nucleosome unwinding.* J Chem Phys, 2019. **150**(18): p. 185103.

187. Moller, J., J. Lequieu, and J.J. de Pablo, *The Free Energy Landscape of Internucleosome Interactions and Its Relation to Chromatin Fiber Structure.* ACS Cent Sci, 2019. **5**(2): p. 341-348.

188. Chen, S., et al., *Unique Structural Platforms of Suz12 Dictate Distinct Classes of PRC2 for Chromatin Binding.* Mol Cell, 2018. **69**(5): p. 840-852 e5.

189. Justin, N., et al., *Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2.* Nat Commun, 2016. **7**: p. 11316.

190. Kasinath, V., et al., *Structures of human PRC2 with its cofactors AEBP2 and JARID2.* Science, 2018. **359**(6378): p. 940-944.

191. Eswar, N., et al., *Comparative protein structure modeling using MODELLER.* Curr Protoc Protein Sci, 2007. **Chapter 2**: p. Unit 2 9.

192. Hojfeldt, J.W., et al., *Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2.* Nat Struct Mol Biol, 2018. **25**(3): p. 225-232.

193. Sanulli, S., et al., *Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation.* Mol Cell, 2015. **57**(5): p. 769-783.

194. Son, J., et al., *Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin.* Genes Dev, 2013. **27**(24): p. 2663-77.

195. Ge, E.J., et al., *Nucleation and Propagation of Heterochromatin by the Histone Methyltransferase PRC2: Geometric Constraints and Impact of the Regulatory Subunit JARID2.* J Am Chem Soc, 2019. **141**(38): p. 15029-15039.

196. Cooper, S., et al., *Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2.* Nat Commun, 2016. **7**: p. 13661.

197. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma.* Science, 2013. **340**(6134): p. 857-61.

198. Stafford, J.M., et al., *Multiple modes of PRC2 inhibition elicit global chromatin alterations in H3K27M pediatric glioma.* Sci Adv, 2018. **4**(10): p. eaau5935.

199. Diehl, K.L., et al., *PRC2 engages a bivalent H3K27M-H3K27me3 dinucleosome inhibitor.* Proc Natl Acad Sci U S A, 2019. **116**(44): p. 22152-22157.

200. Tatavosian, R., et al., *Live-cell single-molecule dynamics of PcG proteins imposed by the DIPG H3.3K27M mutation.* Nat Commun, 2018. **9**(1): p. 2080.

201. Margueron, R., et al., *Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms.* Mol Cell, 2008. **32**(4): p. 503-18.

202. Comet, I., et al., *Maintaining cell identity: PRC2-mediated regulation of transcription and cancer.* Nat Rev Cancer, 2016. **16**(12): p. 803-810.

203. Grigoryev, S.A., et al., *Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes.* Proc Natl Acad Sci U S A, 2016. **113**(5): p. 1238-43.

204. Dodd, I.B., et al., *Theoretical analysis of epigenetic cell memory by nucleosome modification.* Cell, 2007. **129**(4): p. 813-22.

205. Erdel, F. and E.C. Greene, *Generalized nucleation and looping model for epigenetic memory of histone modifications.* Proc Natl Acad Sci U S A, 2016. **113**(29): p. E4180-9.

206. Xie, W.J. and B. Zhang, *Learning the Formation Mechanism of Domain-Level Chromatin States with Epigenomics Data.* Biophys J, 2019. **116**(10): p. 2047-2056.

207. Simon, J.A. and R.E. Kingston, *Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put.* Mol Cell, 2013. **49**(5): p. 808-24.

208. Oksuz, O., et al., *Capturing the Onset of PRC2-Mediated Repressive Domain Formation.* Mol Cell, 2018. **70**(6): p. 1149-1162 e5.

209. Davidovich, C., et al., *A dimeric state for PRC2.* Nucleic Acids Res, 2014. **42**(14): p. 9236-48.

210. Chen, S., et al., *A Dimeric Structural Scaffold for PRC2-PCL Targeting to CpG Island Chromatin.* Mol Cell, 2020. **77**(6): p. 1265-1278 e7.

211. Kinkade, J.M. and R.D. Cole, *The Resolution of Four Lysine-rich Histones Derived from Calf Thymus.* Journal of Biological Chemistry, 1966. **241**(24): p. 5790-5797.

212. Hergeth, S.P. and R. Schneider, *The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle.* EMBO reports, 2015. **16**(11): p. 1439-1453.

213. Brockers, K. and R. Schneider, *Histone H1, the forgotten histone.* Epigenomics, 2019. **11**(4): p. 363-366.

214. Simpson, R.T., *Structure of the chromatosome, a chromatin particle containing 160 base pairs of DNA and all the histones.* Biochemistry, 1978. **17**(25): p. 5524-5531.

215. Fyodorov, D.V., et al., *Emerging roles of linker histones in regulating chromatin structure and function.* Nature reviews. Molecular cell biology, 2018. **19**(3): p. 192-206.

216. White, A.E., A.R. Hieb, and K. Luger, *A quantitative investigation of linker histone interactions with nucleosomes and chromatin.* Scientific Reports, 2016. **6**(1): p. 19122.

217. Yusufova, N., et al., *Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture.* Nature, 2020.

218. Zhou, B.-R., et al., *Structural Mechanisms of Nucleosome Recognition by Linker Histones.* Molecular Cell, 2015. **59**(4): p. 628-638.

219. Zhou, B.-R., et al., *Distinct Structures and Dynamics of Chromatosomes with Different Human Linker Histone Isoforms.* Molecular Cell, 2020.

220. Clark, D.J. and T. Kimura, *Electrostatic mechanism of chromatin folding.* Journal of Molecular Biology, 1990. **211**(4): p. 883-896.

221. Misteli, T., et al., *Dynamic binding of histone H1 to chromatin in living cells.* Nature, 2000. **408**(6814): p. 877-881.

222. Catez, F., T. Ueda, and M. Bustin, *Determinants of histone H1 mobility and chromatin binding in living cells.* Nature Structural & Molecular Biology, 2006. **13**(4): p. 305-310.

223. Karetsou, Z., et al., *Prothymosin alpha modulates the interaction of histone H1 with chromatin.* Nucleic acids research, 1998. **26**(13): p. 3111-3118.

224. Karetsou, Z., et al., *Prothymosin α associates with the oncoprotein SET and is involved in chromatin decondensation.* FEBS Letters, 2004. **577**(3): p. 496-500.

225. Gomez-Marquez, J. and P. Rodríguez, *Prothymosin alpha is a chromatin-remodelling protein in mammalian cells.* The Biochemical journal, 1998. **333 (Pt 1)**(Pt 1): p. 1-3.

226. Boán, F., et al., *Prothymosin α, a mammalian c-myc-regulated acidic nuclear protein, provokes the decondensation of human chromosomes in vitro.* Cytogenetic and Genome Research, 2001. **93**(3-4): p. 171-174.

227. George, E.M. and D.T. Brown, *Prothymosin alpha is a component of a linker histone chaperone.* FEBS letters, 2010. **584**(13): p. 2833-2836.

228. Elsässer, S.J. and S. D'Arcy, *Towards a mechanism for histone chaperones.* Biochimica et biophysica acta, 2013. **1819**(3-4): p. 211-221.

229. Torres, C.M., et al., *The linker histone H1.0 generates epigenetic and functional intratumor heterogeneity.* Science, 2016. **353**(6307): p. aaf1644.

230. Medrzycki, M., et al., *Profiling of linker histone variants in ovarian cancer.* Frontiers in bioscience (Landmark edition), 2012. **17**: p. 396-406.

231. Khachaturov, V., et al., *Histone H1.5, a novel prostatic cancer marker: an immunohistochemical study.* Human Pathology, 2014. **45**(10): p. 2115-2119.

232. Ye, X., et al., *Linker Histone in Diseases.* International journal of biological sciences, 2017. **13**(8): p. 1008-1018.

233. Hansen, J.C., *Conformational Dynamics of the Chromatin Fiber in Solution: Determinants, Mechanisms, and Functions.* Annual Review of Biophysics and Biomolecular Structure, 2002. **31**(1): p. 361-392.

234. Shen, X., et al., *Linker histories are not essential and affect chromatin condensation in vivo.* Cell, 1995. **82**(1): p. 47-56.

235. Lin, R., R.G. Cook, and C.D. Allis, *Proteolytic removal of core histone amino termini and dephosphorylation of histone H1 correlate with the formation of condensed chromatin and transcriptional silencing during Tetrahymena macronuclear development.* Genes Dev, 1991. **5**(9): p. 1601-10.

236. Lu, M.J., et al., *Generation and characterization of novel antibodies highly selective for phosphorylated linker histone H1 in Tetrahymena and HeLa cells.* Chromosoma, 1994. **103**(2): p. 111-21.

237. Halmer, L. and C. Gruss, *Effects of Cell Cycle Dependent Histone H1 Phosphorylation on Chromatin Structure and Chromatin Replication.* Nucleic Acids Research, 1996. **24**(8): p. 1420-1427.

238. Boggs, B.A., C.D. Allis, and A.C. Chinault, *Immunofluorescent studies of human chromosomes with antibodies against phosphorylated H1 histone.* Chromosoma, 2000. **108**(8): p. 485-490.

239. Maresca, T.J., B.S. Freedman, and R. Heald, *Histone H1 is essential for mitotic chromosome architecture and segregation in Xenopus laevis egg extracts.* The Journal of cell biology, 2005. **169**(6): p. 859-869.

240. Fan, Y., et al., *Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation.* Cell, 2005. **123**(7): p. 1199-1212.

241. Geeven, G., et al., *Local compartment changes and regulatory landscape alterations in histone H1-depleted cells.* Genome Biology, 2015. **16**(1): p. 289.

242. Izzo, A., et al., *Dynamic changes in H1 subtype composition during epigenetic reprogramming.* The Journal of cell biology, 2017. **216**(10): p. 3017-3028.

243. Harshman, S.W., et al., *H1 histones: current perspectives and challenges.* Nucleic acids research, 2013. **41**(21): p. 9593-9609.

244. Izzo, A. and R. Schneider, *The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics.* Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2016. **1859**(3): p. 486-495.

245. Cheung, P., C.D. Allis, and P. Sassone-Corsi, *Signaling to Chromatin through Histone Modifications.* Cell, 2000. **103**(2): p. 263-271.

246. Kamieniarz, K., et al., *A dual role of linker histone H1.4 Lys 34 acetylation in transcriptional activation.* Genes & Development, 2012. **26**(8): p. 797-802.

247. Saha, A., et al., *Site-Specific Phosphorylation of Histone H1.4 Is Associated with Transcription Activation.* Int J Mol Sci, 2020. **21**(22).

248. Hale, T.K., et al., *Phosphorylation of the Linker Histone H1 by CDK Regulates Its Binding to HP1α.* Molecular Cell, 2006. **22**(5): p. 693-699.

249. Roque, A., et al., *Phosphorylation of the carboxy-terminal domain of histone H1: effects on secondary structure and DNA condensation.* Nucleic acids research, 2008. **36**(14): p. 4719-4726.

250. Fischle, W., Y. Wang, and C. David Allis, *Binary switches and modification cassettes in histone biology and beyond.* Nature, 2003. **425**(6957): p. 475-479.

251. Terme, J.-M., et al., *Dynamics and dispensability of variant-specific histone H1 Lys-26/Ser-27 and Thr-165 post-translational modifications.* FEBS Letters, 2014. **588**(14): p. 2353-2362.

252. Aubin R Fau - Frechette, A., et al., *Nucleosomal poly(ADP-ribose) polymerase: properties and relaxation of the chromatin structure.*

253. Krishnakumar, R., et al., *Reciprocal Binding of PARP-1 and Histone H1 at Promoters Specifies Transcriptional Outcomes.* Science, 2008. **319**(5864): p. 819.

254. Weiss, T., et al., *Histone H1 variant-specific lysine methylation by G9a/KMT1C and Glp1/KMT1D.* Epigenetics & chromatin, 2010. **3**(1): p. 7-7.

255. Lu, X., et al., *Linker histone H1 is essential for Drosophila development, the establishment of pericentric heterochromatin, and a normal polytene chromosome structure.* Genes & development, 2009. **23**(4): p. 452-465.

256. Lu, X., et al., *Drosophila H1 Regulates the Genetic Activity of Heterochromatin by Recruitment of Su(var)3-9.* Science, 2013. **340**(6128): p. 78.

257. Nielsen, A.L., et al., *Heterochromatin Formation in Mammalian Cells: Interaction between Histones and HP1 Proteins.* Molecular Cell, 2001. **7**(4): p. 729-739.

258. Popova, E.Y., et al., *Developmentally regulated linker histone H1c promotes heterochromatin condensation and mediates structural integrity of rod photoreceptors in mouse retina.* The Journal of biological chemistry, 2013. **288**(24): p. 17895-17907.

259. Kim, K., et al., *Linker Histone H1.2 cooperates with Cul4A and PAF1 to drive H4K31 ubiquitylation-mediated transactivation.* Cell reports, 2013. **5**(6): p. 1690-1703.

260. Yang, S.-M., et al., *H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation.* Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(5): p. 1708-1713.

261. Soria, G., Sophie E. Polo, and G. Almouzni, *Prime, Repair, Restore: The Active Role of Chromatin in the DNA Damage Response.* Molecular Cell, 2012. **46**(6): p. 722-734.

262. Hashimoto, H., et al., *Histone H1 variant, H1R is involved in DNA damage response.* DNA Repair, 2007. **6**(11): p. 1584-1595.

263. Murga, M., et al., *Global chromatin compaction limits the strength of the DNA damage response.* The Journal of cell biology, 2007. **178**(7): p. 1101-1108.

264. Peng, J.C. and G.H. Karpen, *Epigenetic regulation of heterochromatic DNA stability.* Current opinion in genetics & development, 2008. **18**(2): p. 204-211.

265. Luijsterburg, M.S., et al., *Heterochromatin protein 1 is recruited to various types of DNA damage.* The Journal of cell biology, 2009. **185**(4): p. 577-586.

266. Hong, Z., et al., *A polycomb group protein, PHF1, is involved in the response to DNA double-strand breaks in human cell.* Nucleic acids research, 2008. **36**(9): p. 2939-2947.

267. Kalashnikova, A.A., et al., *Linker histone H1.0 interacts with an extensive network of proteins found in the nucleolus.* Nucleic Acids Research, 2013. **41**(7): p. 4026-4035.

268. Thorslund, T., et al., *Histone H1 couples initiation and amplification of ubiquitin signalling after DNA damage.* Nature, 2015. **527**(7578): p. 389-393.

269. Downs, J.A., et al., *Suppression of Homologous Recombination by the Saccharomyces cerevisiae Linker Histone.* Molecular Cell, 2003. **11**(6): p. 1685-1692.

270. Jennifer E. Chubb, S.R., *Core and Linker HIstone Modifications Involved in the DNA Damage Response*. Subcellular Biochemistry. Vol. Genome Stability and Human Diseases. 2010: Springer, Dordrecht.

271. Lu, Z.H., et al., *Histone H1 modulates DNA replication through multiple pathways in Xenopus egg extract.* Journal of Cell Science, 1997. **110**(21): p. 2745.

272. Andreyeva, E.N., et al., *Regulatory functions and chromatin loading dynamics of linker histone H1 during endoreplication in Drosophila.* Genes & Development, 2017. **31**(6): p. 603-616.

273. Zou, Y., et al., *Functions of human replication protein A (RPA): from DNA replication to DNA damage and stress responses.* Journal of cellular physiology, 2006. **208**(2): p. 267-273.

274. Wold, M.S., *REPLICATION PROTEIN A: A Heterotrimeric, Single-Stranded DNA-Binding Protein Required for Eukaryotic DNA Metabolism.* Annual Review of Biochemistry, 1997. **66**(1): p. 61-92.

275. Liu, V.F. and D.T. Weaver, *The ionizing radiation-induced replication protein A phosphorylation response differs between ataxia telangiectasia and normal human cells.* Molecular and cellular biology, 1993. **13**(12): p. 7222-7231.

276. Binz, S.K., A.M. Sheehan, and M.S. Wold, *Replication Protein A phosphorylation and the cellular response to DNA damage.* DNA Repair, 2004. **3**(8): p. 1015-1024.

277. de Laat, W.L., et al., *DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair.* Genes & development, 1998. **12**(16): p. 2598-2609.

278. Kim, C., R.O. Snyder, and M.S. Wold, *Binding properties of replication protein A from human and yeast cells.* Molecular and cellular biology, 1992. **12**(7): p. 3050-3059.

279. Qin, Z., et al., *Human RPA activates BLM's bidirectional DNA unwinding from a nick.* eLife, 2020. **9**: p. e54098.

280. Osunsade, A., et al., *A Robust Method for the Purification and Characterization of Recombinant Human Histone H1 Variants.* Biochemistry, 2019. **58**(3): p. 171-176.

281. Th'ng, J.P.H., et al., *H1 Family Histones in the Nucleus: CONTROL OF BINDING AND LOCALIZATION BY THE C-TERMINAL DOMAIN.* Journal of Biological Chemistry, 2005. **280**(30): p. 27809-27814.

282. Öberg, C., et al., *Linker Histone Subtypes Differ in Their Effect on Nucleosomal Spacing In Vivo.* Journal of Molecular Biology, 2012. **419**(3): p. 183-197.

283. Kalashnikova, A.A., R.A. Rogge, and J.C. Hansen, *Linker histone H1 and protein-protein interactions.* Biochimica et biophysica acta, 2016. **1859**(3): p. 455-461.

284. Borgia, A., et al., *Extreme disorder in an ultrahigh-affinity protein complex.* Nature, 2018. **555**(7694): p. 61-66.

285. Segade, F. and J. Gómez-Márquez, *Prothymosin α.* The International Journal of Biochemistry & Cell Biology, 1999. **31**(11): p. 1243-1248.

286. Shakya, A., et al., *Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization.* Biophysical Journal, 2020. **118**(3): p. 753-764.

287. Turner, A.L., et al., *Highly disordered histone H1−DNA model complexes and their condensates.* Proceedings of the National Academy of Sciences, 2018. **115**(47): p. 11964.

288. Smith, S.B., Y. Cui, and C. Bustamante, *Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules.* Science, 1996. **271**(5250): p. 795-9.

289. Hergeth, S.P. and R. Schneider, *The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle.* EMBO Rep, 2015. **16**(11): p. 1439-53.

290. van Mameren, J., et al., *Unraveling the structure of DNA during overstretching by using multicolor, single-molecule fluorescence imaging.* Proc Natl Acad Sci U S A, 2009. **106**(43): p. 18231-6.

291. Shakya, A., et al., *Liquid-Liquid Phase Separation of Histone Proteins in Cells: Role in Chromatin Organization.* Biophys J, 2020. **118**(3): p. 753-764.

292. Gibson, B.A., et al., *Organization of Chromatin by Intrinsic and Regulated Phase Separation.* Cell, 2019. **179**(2): p. 470-484 e21.

293. Savelyev, A. and G.A. Papoian, *Chemically accurate coarse graining of double-stranded DNA.* Proc Natl Acad Sci U S A, 2010. **107**(47): p. 20340-5.

294. Latham, A.P. and B. Zhang, *Consistent Force Field Captures Homolog Resolved HP1 Phase Separation.* bioRxiv: p. 2021.01.06.425600.

295. Xu, B., et al., *Rigidity enhances a magic-number effect in polymer phase separation.* Nat Commun, 2020. **11**(1): p. 1561.

296. Espinosa, J.R., et al., *Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components.* Proc Natl Acad Sci U S A, 2020. **117**(24): p. 13238-13247.

297. Alshareedah, I., et al., *Interplay between Short-Range Attraction and Long-Range Repulsion Controls Reentrant Liquid Condensation of Ribonucleoprotein-RNA Complexes.* J Am Chem Soc, 2019. **141**(37): p. 14593-14602.

298. Rhine, K., et al., *ALS/FTLD-Linked Mutations in FUS Glycine Residues Cause Accelerated Gelation and Reduced Interactions with Wild-Type FUS.* Mol Cell, 2020. **80**(4): p. 666-681 e8.

299. Hendzel, M.J., et al., *The C-terminal domain is the primary determinant of histone H1 binding to chromatin in vivo.* J Biol Chem, 2004. **279**(19): p. 20028-34.

300. Fan, Y., et al., *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.* Cell, 2005. **123**(7): p. 1199-212.

301. Jawerth, L., et al., *Protein condensates as aging Maxwell fluids.* Science, 2020. **370**(6522): p. 1317-1323.

302. Downs, J.A., et al., *Suppression of homologous recombination by the Saccharomyces cerevisiae linker histone.* Mol Cell, 2003. **11**(6): p. 1685-92.

303. Hashimoto, H., et al., *Histone H1 variant, H1R is involved in DNA damage response.* DNA Repair (Amst), 2007. **6**(11): p. 1584-95.

304. Zheng, Q., et al., *Reversible histone glycation is associated with disease-related changes in chromatin architecture.* Nature Communications, 2019. **10**(1): p. 1289.

305. Hellwig, M. and T. Henle, *Baking, Ageing, Diabetes: A Short History of the Maillard Reaction.* Angewandte Chemie International Edition, 2014. **53**(39): p. 10316-10329.

306. Rabbani, N. and P.J. Thornalley, *Measurement of methylglyoxal by stable isotopic dilution analysis LC-MS/MS with corroborative prediction in physiological samples.* Nature Protocols, 2014. **9**: p. 1969.

307. Lo, T.W., et al., *Binding and modification of proteins by methylglyoxal under physiological conditions. A kinetic and mechanistic study with N alpha-acetylarginine, N alpha-acetylcysteine, and N alpha-acetyllysine, and bovine serum albumin.* Journal of Biological Chemistry, 1994. **269**(51): p. 32299-32305.

308. Schwartzentruber, J., et al., *Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma.* Nature, 2012. **482**(7384): p. 226-31.

309. Wu, G., et al., *Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas.* Nature genetics, 2012. **44**(3): p. 251-253.

310. Nacev, B.A., et al., *The expanding landscape of 'oncohistone' mutations in human cancers.* Nature, 2019. **567**(7749): p. 473-478.

311. Arimura, Y., et al., *Cancer-associated mutations of histones H2B, H3.1 and H2A.Z.1 affect the structure and stability of the nucleosome.* Nucleic Acids Research, 2018. **46**(19): p. 10007-10018.

312. Wan, L., et al., *ENL links histone acetylation to oncogenic gene expression in acute myeloid leukaemia.* Nature, 2017. **543**(7644): p. 265-269.

313. Klein, B.J., et al., *Structural insights into the π-π-π stacking mechanism and DNA-binding activity of the YEATS domain.* Nature Communications, 2018. **9**(1): p. 4574.

314. Wan, L., et al., *Impaired cell fate through gain-of-function mutations in a chromatin reader.* Nature, 2020. **577**(7788): p. 121-126.

315. Bernstein, B.E., et al., *A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells.* Cell. **125**(2): p. 315-326.

316. Burkardt, D.D., et al., *HIST1H1E heterozygous protein-truncating variants cause a recognizable syndrome with intellectual disability and distinctive facial gestalt: A study to clarify the HIST1H1E syndrome phenotype in 30 individuals.* Am J Med Genet A, 2019. **179**(10): p. 2049-2055.

317. Owen, I. and F. Shewmaker, *The Role of Post-Translational Modifications in the Phase Transitions of Intrinsically Disordered Proteins.* Int J Mol Sci, 2019. **20**(21).

318. Kim, J.-M., et al., *Linker histone H1.2 establishes chromatin compaction and gene silencing through recognition of H3K27me3.* Scientific Reports, 2015. **5**(1): p. 16714.

319. Kuzmichev, A., et al., *Different Ezh2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3.* Molecular Cell, 2004. **14**(2): p. 183-193.

320. Swaroop, A., et al., *An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia.* Oncogene, 2019. **38**(5): p. 671-686.

321. Flintoft, L., *Human cells clear foreign DNA.* Nature Reviews Genetics, 2010. **11**(3): p. 172-172.

322. Fox, A.H., et al., *Paraspeckles: Where Long Noncoding RNA Meets Phase Separation.* Trends in Biochemical Sciences, 2018. **43**(2): p. 124-135.

323. Rashid, F., A. Shah, and G. Shan, *Long Non-coding RNAs in the Cytoplasm.* Genomics, Proteomics & Bioinformatics, 2016. **14**(2): p. 73-80.

324. Hajjari, M. and A. Salavaty, *HOTAIR: an oncogenic long non-coding RNA in different cancers.* Cancer biology & medicine, 2015. **12**(1): p. 1-9.

325. Muthunayake, N.S., et al., *Phase-separated bacterial ribonucleoprotein bodies organize mRNA decay.* Wiley Interdiscip Rev RNA, 2020. **11**(6): p. e1599.

326. Fay, M.M. and P.J. Anderson, *The Role of RNA in Biological Phase Separations.* Journal of molecular biology, 2018. **430**(23): p. 4685-4701.

327. Mei, L., et al., *Molecular mechanism for direct actin force-sensing by α-catenin.* Elife, 2020. **9**.

328. Ma, J. and M.D. Wang, *RNA polymerase is a powerful torsional motor.* Cell cycle (Georgetown, Tex.), 2014. **13**(3): p. 337-338.

329. Wang, H.-Y., et al., *Force Generation in RNA Polymerase.* Biophysical Journal, 1998. **74**(3): p. 1186-1202.

330. Arias-Gonzalez, J.R., *A DNA-centered explanation of the DNA polymerase translocation mechanism.* Scientific Reports, 2017. **7**(1): p. 7566.

331. Maier, B., D. Bensimon, and V. Croquette, *Replication by a single DNA polymerase of a stretched single-stranded DNA.* Proceedings of the National Academy of Sciences, 2000. **97**(22): p. 12002.

332. Alam, S., et al., *Nuclear forces and cell mechanosensing.* Progress in molecular biology and translational science, 2014. **126**: p. 205-215.

333. Poh, Y.-C., et al., *Dynamic force-induced direct dissociation of protein complexes in a nuclear body in living cells.* Nature Communications, 2012. **3**(1): p. 866.

334. Wang, N., J.D. Tytell, and D.E. Ingber, *Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus.* Nature Reviews Molecular Cell Biology, 2009. **10**(1): p. 75-82.

335. Stamenović, D. and D.E. Ingber, *Tensegrity-guided self assembly: from molecules to living cells.* Soft Matter, 2009. **5**(6): p. 1137-1145.

336. Stein, G.S., et al., *Organization of transcriptional regulatory machinery in nuclear microenvironments: implications for biological control and cancer.* Advances in enzyme regulation, 2007. **47**: p. 242-250.

337. Thomas, C.H., et al., *Engineering gene expression and protein synthesis by modulation of nuclear shape.* Proceedings of the National Academy of Sciences, 2002. **99**(4): p. 1972.

338. Klein, I.A., et al., *Partitioning of cancer therapeutics in nuclear condensates.* Science, 2020. **368**(6497): p. 1386.

339. Berger, I., et al., *The multiBac protein complex production platform at the EMBL.* J. Vis. Exp., 2013(77): p. e50159.

340. Vary, J.C., Jr., T.G. Fazzio, and T. Tsukiyama, *Assembly of yeast chromatin using ISWI complexes.* Methods Enzymol, 2004. **375**: p. 88-102.

341. Lowary, P.T. and J. Widom, *New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning.* J Mol Biol, 1998. **276**(1): p. 19-42.

342. Fierz, B., et al., *Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction.* Nat Chem Biol, 2011. **7**(2): p. 113-9.

343. Harada, B.T., et al., *Stepwise nucleosome translocation by RSC remodeling complexes.* eLife, 2016. **5**: p. e10051.

344. Hashemi Shabestari, M., et al., *Recent Advances in Biological Single-Molecule Applications of Optical Tweezers and Fluorescence Microscopy.* Methods Enzymol, 2017. **582**: p. 85-119.

345. Wasserman, M.R., et al., *Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase.* Cell, 2019. **178**(3): p. 600-611 e16.

346. Wang, M.D., et al., *Stretching DNA with optical tweezers.* Biophys J, 1997. **72**(3): p. 1335-46.

347. Marko, J.F. and E.D. Siggia, *Stretching DNA.* Macromolecules, 1995. **28**: p. 8759-8770.

348. Pedregosa, F., et al., *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.

349. Sbalzarini, I.F. and P. Koumoutsakos, *Feature point tracking and trajectory analysis for video imaging in cell biology.* J Struct Biol, 2005. **151**(2): p. 182-95.

350. Mangeol, P., B. Prevo, and E.J. Peterman, *KymographClear and KymographDirect: two tools for the automated quantitative analysis of molecular and cellular dynamics using kymographs.* Mol Biol Cell, 2016. **27**(12): p. 1948-57.

351. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.

352. Jinek, M., et al., *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.

353. Newton, M.D., et al., *DNA stretching induces Cas9 off-target activity.* Nat Struct Mol Biol, 2019. **26**(3): p. 185-192.

354. Lu, X.J. and W.K. Olson, *3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures.* Nucleic Acids Res, 2003. **31**(17): p. 5108-21.

355. Clementi, C., H. Nymeyer, and J.N. Onuchic, *Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins.* J Mol Biol, 2000. **298**(5): p. 937-53.

356. Noel, J.K., P.C. Whitford, and J.N. Onuchic, *The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function.* J Phys Chem B, 2012. **116**(29): p. 8692-702.

357. Zhang, B., et al., *Exploring the Free Energy Landscape of Nucleosomes.* J Am Chem Soc, 2016. **138**(26): p. 8126-33.

358. Lequieu, J., D.C. Schwartz, and J.J. de Pablo, *In silico evidence for sequence-dependent nucleosome sliding.* Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9197-E9205.

359. Kumar, S., et al., *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* J. Comput. Chem., 1992. **13**(8): p. 1011-1021.

360. Murphy, M.C., et al., *Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy.* Biophys J, 2004. **86**(4): p. 2530-7.

361. Rizzo, V. and J. Schellman, *Flow dichroism of T7 DNA as a function of salt concentration.* Biopolymers, 1981. **20**(10): p. 2143-63.

362. Berendsen, H.J.C., D. van der Spoel, and R. van Drunen, *GROMACS: A message-passing parallel molecular dynamics implementation.* Comput Phys Commun, 1995. **91**(1-3): p. 43-56.

363. Dignon, G.L., et al., *Sequence determinants of protein phase behavior from a coarse-grained model.* PLoS Comput Biol, 2018. **14**(1): p. e1005941.

364. Michaud-Agrawal, N., et al., *MDAnalysis: a toolkit for the analysis of molecular dynamics simulations.* J Comput Chem, 2011. **32**(10): p. 2319-27.

365. Harris, C.R., et al., *Array programming with NumPy.* Nature, 2020. **585**(7825): p. 357-362.