

Rockefeller University

Digital Commons @ RU

Student Theses and Dissertations

2023

Mechanical Manipulation of Eukaryotic Chromatin by DNA-Binding Proteins

Tuan Nguyen

Follow this and additional works at: https://digitalcommons.rockefeller.edu/student_theses_and_dissertations



Part of the Life Sciences Commons



MECHANICAL MANIPULATION OF EUKARYOTIC CHROMATIN BY DNA-BINDING PROTEINS

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

by
Tuan Nguyen
June 2023

MECHANICAL MANIPULATION OF EUKARYOTIC CHROMATIN BY DNA-BINDING PROTEINS

Tuan Nguyen, Ph.D.

The Rockefeller University 2023

The eukaryotic genome is organized in many length scales, reflecting the intricacy associated with evolution of complex biological processes. This organization serves to exert spatiotemporal control of many DNA-transacting processes such as gene expression. Despite emerging progress, the biophysical mechanism underpinning eukaryotic genome organization remains an outstanding question in the field. In this thesis, I describe mechanistic insights on genome organization and its regulation through leveraging single-molecule biophysical techniques.

In Chapter 2, I characterize the dynamic interplay of Sox2 and H1 DNA binding activity. Both families constitute large classes of chromatin and DNA binding proteins that have been historically thought to be antagonistic regulators of each other, but the underlying mechanism is not well understood. Using single-molecule fluorescence-based approach, I show that Sox2 and H1 regulate each other's loading rate on bare DNA and nucleosomes in a concentration-dependent fashion. In particular, H1 promotes the Sox2's loading rate at low concentration but inhibits Sox2's loading rate at higher concentration. Together, these findings highlight the potential importance of tuning protein concentrations in the regulation of gene expression.

In Chapter 3, I characterize the mechanical effects on DNA from biomolecular condensation, which has recently emerged as an important mechanism of gene regulation. In particular, I investigate how Sox2, which constitutes an important pioneer factor implicated in the maintenance of pluripotency, forms co-condensates with DNA and chromatin components. The described results present three conceptual advances to the field: 1) protein:DNA co-condensation can generate high forces, up to ~ 7 pN, comparable to other reported cellular forces, 2) the intrinsically disordered regions (IDRs) are dispensable for condensate formation but necessary for high force generation, and lastly, 3) chromatin components, such as nucleosomes and linker histone H1, attenuate the force generating capacity of Sox2 condensates and reduce their mechanical effects on DNA via colocalization. The results add to the growing body of studies that the chromatin architecture can function as a mechanical sink that regulates cellular forces.

In Chapter 4, I visualize the DNA compaction activity of the structural maintenance of chromosome (SMC) complex 5/6, an important ATPase implicated in regulating DNA repair and replication. Despite emerging insights on the SMC5/6 complex's cellular function, the molecular mechanism behind the complex's DNA binding activity is not well understood. Using single-molecule fluorescence method, I present data showing the SMC5/6 complex can compact DNA in a tether-like mechanism without the requirement for ATP hydrolysis. Thus, this work adds a novel perspective towards understanding the molecular mechanism of the SMC5/6 complex.

Together, the thesis below contributes novel mechanistic insights towards understanding genome organization and regulation. I reveal unique modes of DNA compaction spanning from

transcription factors to ATPase molecular motor as well as associated regulatory mechanism. Due to the implications in diverse molecular pathways, aberrant regulation of genome organization underpins many disease processes. Thus, these findings help establish a molecular basis towards understanding many disease mechanisms, which can be potentially exploited for therapeutic avenues.

To my friends, family, and many others who have supported me along the way.

ACKNOWLEDGMENTS

Many people made this work possible. I want to express my most sincere gratitude to my advisor Shixin Liu for all of his support and guidance. Shixin possesses an unparalleled dedication, enthusiasm, and vision for science. Working with someone with such scientific dedication and trenchant insight has truly been an intellectual joy. Most importantly, Shixin is a kind human being. During this journey, there were times when things were hard and stormy, when the ship was swaying with no land in sight, when I was barely holding on to the helm, but Shixin was there, supporting me along the way and making his own sacrifices. I will always be grateful for having Shixin as mentor.

Much of the work here is made possible with the help and companionship of many members in the Liu Lab. First, I would like to thank Sai Li and Michael Wasserman for all the technical help and guidance throughout my time here. I owe it to them to get many of my important experiments running and troubleshoot them along the way. Second, I am grateful for unmeasurable contributions of John Watters, Jeremy Chang, Htet Ng, and Wola Osunsade for the project in Chapter 3. Lastly, I would like to thank the rest of the Liu lab for their companionship, their insightful advice, and their technical help throughout my time here, including Ling Wang, Xiangwu Ju, Rochelle Shih, Dayi Li, Gabriella Chua, and Rachel Leicher. Additionally, I want to thank many rotation students who I had great pleasure working with, including Matt Reynolds, Brandon Malone, Sarah Cai, and Joshua Chandanani. Everyone from the Liu lab has helped me at some point that cannot be simply recounted within the narrow margins of this page – thank you.

I would like to further thank Xiaolan Zhao and members of her laboratory, notably, Shibai Li and Cory Haluska, for the insightful discussion and technical support regarding the complex biology of SMC5/6. Further, I sincerely thank members of the Alushin, Funabiki, and Birsoy lab for their time and letting me temporarily use their equipment at times. In particular, thank you Yasu Arimura for his invaluable advice and expertise on histones. Thank you to Robbie and many others in Birsoy lab for the jovial company and all the help running Western blots early on. Furthermore, I thank the advice of many in the Bio-imaging resource center with regards to TIRFM data analysis, Jim Petrillo from the Precision Instrument Technologies, for his valuable time in helping me make the TIRFM slides and double-sided tapes, and Brian Miller from Olympus, for his technical support of the TIRFM. Additionally, much of the committee meeting and progress have been facilitated by the Dean's Office; thank you in particular to Cris Rosario, Kristen Cullen, and Sid Strickland. Lastly, I would like to express my gratitude towards the Tri-Institutional MD/PhD office for all their support and advice during the entire journey up to this point, especially Olaf Andersen, Catharine Boothroyd, Ruth Gotian. Thank you all for everything.

In addition, I also want to express my sincere gratitude to Hiro Funabiki, Yael David, and Sandy Simon for their thoughtful guidance and advice as members of my thesis committee. The critical and constructive advice has been instrumental in improving the science conducted in this thesis. Thank you, Hiro, for being a critical scientist and all the insightful conversations, for allowing me to see my work in a different light. Thank you, Yael, for your expertise in chromatin biology and all the knowledge on linker histone. Thank you, Sandy, for your witty humor, expertise on the TIRFM, and the all science that I could ever learn from you. I also thank

Sebastian Deindl for lending his valuable time to serve as my external referee. Thank you all for making me a better scientist.

Lastly, I am grateful for all the love and support from my friends and family. My parents, Minh and Ha have navigated through the obstacles of immigration so that I would have the opportunities like this today. Thank you for all your sacrifices, for all your hard work, and importantly, for being there with me through all the tough and joyful times. Additionally, thank you to my partner Lan for staying by my side for this stretch of the journey. I am forever grateful.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iv
TABLE OF CONTENTS	vi
LIST OF FIGURES	ix
CHAPTER 1. Introduction	1
1.1 Eukaryotic genome organization	1
1.2 Chromatin organization	2
1.3 Phase separated compartmentalization	2
1.3.1 Mechano-regulation by transcriptional condensates	3
1.4 Chromatin loops.....	4
1.4.1 Genome organization via the SMC complex	4
1.5 Techniques used to study genome organization	7
1.6 Single-molecule techniques	7
1.6.1 Single-molecule fluorescence approach	8
1.6.2 Single-molecule force-based approach.....	9
1.6.3 Combining single-molecule fluorescence and force-based approaches	10
1.7 Overview and significance of findings	10
CHAPTER 2. Dynamic interplay between Sox2 and linker histone H1 binding activity 12	
2.1 The HMG-box protein family	12
2.1.1 Transcriptional regulation via HMG-box protein	12
2.2 Linker histone H1	13
2.2.1 Linker histone H1 and gene regulation	14
2.3 Dynamic interplay between H1 and HMG-box protein.....	15
2.4 Single-molecule study of Sox2 and H1 binding to nucleosomes and DNA.....	16
2.4.1 H1 does not alter Sox2 binding dwell time	19
2.4.2 H1 displays a concentration-dependent regulation of Sox2 loading	20
2.4.3 Deletion of H1.4's C-terminal tail promotes Sox2 loading at higher concentration	21
2.4.4 Sox2 promotes H1 loading on DNA and nucleosomes.....	22
2.5 Conclusion.....	23
CHAPTER 3. Chromatin sequesters pioneer factor Sox2 from exerting force on DNA 24	
3.1 Pioneer transcription factor	24
3.2 Sox2: an introduction	24
3.2.1 Sox2 and pluripotency.....	24
3.2.2 Sox2's biochemical properties	25
3.3 TF-mediated transcriptional condensates.....	27

3.4	Single-molecule findings of Sox2:DNA co-condensate formation.....	28
3.4.1	Sox2 forms co-condensates with DNA	29
3.4.2	Sox2:DNA co-condensation exerts mechanical stress on DNA	33
3.4.3	IDRs of Sox2 are dispensable for condensate formation	40
3.4.4	Quantification of the force mediated by Sox2:DNA co-condensation	43
3.4.5	Nucleosomes attenuate the mechanical effect of Sox2:DNA co-condensation	46
3.4.6	Quantification of force generated by Sox2 condensates on nucleosomes	49
3.4.7	Linker histone H1.4 suppresses Sox2-induced stress on DNA through colocalization.....	50
3.5	Conclusion.....	51
CHAPTER 4. Single molecule visualization of DNA compaction by the SMC5/6 complex		52
4.1	SMC5/6: the yet-to-be named complex of the SMC complex family.....	52
4.1.1	SMC5/6 in DNA repair	53
4.1.2	SMC5/6 in DNA replication	54
4.1.3	Recently discovered miscellaneous functions of SMC5/6.....	54
4.1.4	Biochemical activity underlying SMC5/6's function.....	55
4.2	Visualization of SMC5/6 complex's DNA compacting capacity	56
4.2.1	SMC5/6 complex compacts DNA without ATP via a tethering-like mechanism	56
4.2.2	SMC5/6 complex mediates DNA bridging without requiring ATP.....	58
4.2.3	DNA curvature appears to regulate SMC5/6's DNA compaction activity .	58
4.2.4	ATP appears to inhibit SMC5/6's DNA compaction activity	58
4.3	Conclusion.....	59
CHAPTER 5. Discussion.....		60
5.1	Force-generation by biomolecular condensates.....	60
5.2	Biophysical nature of biomolecular condensates	62
5.3	Chromatin component as a force-regulating mechanical sink.....	63
5.4	Mechanism underpinning the regulatory interplay between H1 and Sox2	64
5.4.1	Biomolecular condensate model	64
5.4.2	Other alternative models.....	65
5.5	Cellular implications of HMGB and H1 regulation	65
5.6	DNA-tethering by the SMC5/6 complex.....	66
5.7	Loop extrusion vs DNA tethering	66
5.8	ATP-dependent decrease of SMC5/6 compaction: a paradox	67
5.9	Cellular implications of SMC5/6 compaction modes	67
5.10	Concluding remarks	68
CHAPTER 6. Materials and methods.....		69

6.1	Protein purification and labeling.....	69
6.2	DNA construct preparation	70
6.3	Single-molecule TIRFM experiments.....	70
7.3.1	TIRFM data analysis	71
6.4	Optical tweezers experiments	73
6.5	Electrophoretic mobility shift assay (EMSA)	74
6.6	Statistical analysis	74
CHAPTER 7.	References.....	75

LIST OF FIGURES

Figure 1. The eukaryotic genome organization	1
Figure 2. Force-exertion via protein-DNA co-condensation	4
Figure 3. Structural organization of the SMC complex family	5
Figure 4. Schematics of loop extrusion vs tethering model	6
Figure 5. Schematics of single-molecule TIRFM	9
Figure 6. Schematics of force-fluorescence microscopy in studying condensates.....	10
Figure 7. Structural delineation of linker histone H1.....	14
Figure 8. Overview of DNA and nucleosome constructs used in this study	16
Figure 9. Depiction of colocalization single-molecule spectroscopy (CoSMoS)	19
Figure 10. H1 does not modulate the dwell time of Sox2 on DNA and nucleosomes	20
Figure 11. H1 modulates Sox2's loading rate onto DNA and nucleosomes.....	21
Figure 12. Modulation of H1 loading rate on DNA and nucleosomes by Sox2 and Pro α 22	
Figure 13. DNA-binding activity and domain organization of Sox2	26
Figure 14. Schematics of TF-mediated transcriptional condensate	28
Figure 15. Sox2 forms condensates on DNA.....	30
Figure 16. Purification and labeling of recombinant human Sox2.	31
Figure 17. Distribution of Sox2 binding motifs on λ DNA.	31
Figure 18. Behavior of Sox2 condensates on DNA at different time points	33
Figure 19. Evaluation of the mechanical effects of Sox2:DNA co-condensation using unlabeled Sox2.	34
Figure 20. Sox2:DNA co-condensation exerts intra- and inter-strand mechanical stress	36
Figure 21. Evaluation of Sox2-mediated DNA breakage under different TIRFM experimental conditions	38
Figure 22. DNA condensation by linker histone H1	39
Figure 23. DNA condensation mediated by Sox2-HMGB constructs	42
Figure 24. EMSA for the binding of different Sox2 constructs to DNA	42
Figure 25. Optical tweezers assays for quantitative measurements of the force generated by Sox2:DNA co-condensation	45
Figure 26. Nucleosomes colocalize with Sox2 condensates and attenuate their mechanical effects on DNA	48
Figure 27. Sox2 binding and condensation on nucleosomal DNA	49
Figure 28. Nucleosomes attenuate the condensation force exerted by Sox2 on DNA ...	50
Figure 29. H1 suppresses the Sox2-induced stress on DNA through colocalization.....	51

Figure 30. Schematics of the SMC5/6 complex53
Figure 31. Schematics depicting the discussed cellular functions of the SMC5/6 complex55
Figure 32. SMC5/6 complex compacts DNA without ATP requirement via a tethering-like mechanism57
Figure 33. DNA bridging interaction in the presence of SMC5/658
Figure 34. Regulation of DNA compaction by ATP and end-to-end tether length59

CHAPTER 1. Introduction

1.1 Eukaryotic genome organization

The sheer length of genetic materials embedded within the eukaryotic genome gives rise to the need for efficient packaging within a confined nuclear space. Apart from this, there exists the necessity for organization to satisfy the biological complexity linked to many DNA-transacting processes in eukaryotes. As such, eukaryotic genomes are organized in multiple length scales (**Figure 1**).^{1,2} Understanding the genome organizational pattern and more importantly, its underlying molecular mechanism, is the overarching theme of this thesis. Below I discuss several themes pertaining to the eukaryotic genome organization, highlight particular aspects that this thesis addresses, and discuss the major experimental approaches that are covered.

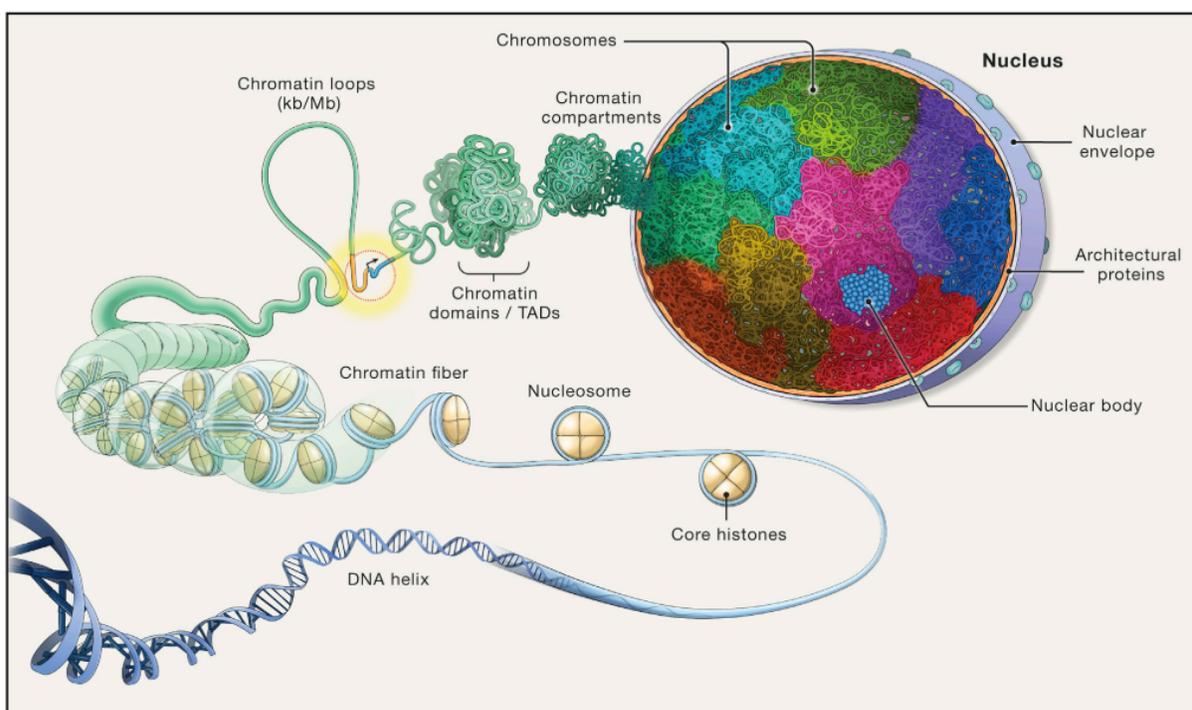


Figure 1. The eukaryotic genome organization

Adapted from Misteli.¹ Eukaryotes organize their genetic materials into many length scales, as outlined in this figure. In short, DNA is packaged by the nucleosomes, consisting of an octamer of histone proteins. Nucleosomes organize into the chromatin fiber structure, which can form loops of various sizes. The chromatin fiber can fold into topological associating domains, which further form various chromatin compartments. The layers of genome organization and its underlying mechanism form the conceptual basis of this thesis work.

1.2 Chromatin organization

One fundamental organizational element of eukaryotic genome is the chromatin fiber (**Figure 1**).³ At the most basic level, the chromatin fiber comprises of nucleosomes, which consists of ~147 base pairs of DNA wrapped around an octamer of core histones with a central (H3-H4)₂ tetramer and two H2A-H2B dimers. Apart from the nucleosome core particle, the next fundamental packaging unit of the chromatin consists of a linker H1 histone bound to ~20 bp of entry/exit DNA from core nucleosome, comprising the chromatosome structure.⁴ Together, the chromatin physically compartmentalizes eukaryotic genetic materials, regulating DNA-dependent transactions such as replication, DNA repair, and transcription through physically restricting nuclear proteins that execute cellular tasks.^{5,6} One important nuclear process through which the chromatin plays a central role is transcriptional regulation, which entails transcription factors (TF) accessing and reading genomic DNA sequence.^{7,8} The mechanism through which TF navigates the complex chromatin environment to execute gene expression programs is an important question that remains poorly understood.

In addition to its role in packaging genomic DNA, the chromatin has emerged as an important mechano-regulator of important intra-nuclear processes.⁹⁻¹² Early characterization of the chromatin's intrinsic mechanical property reveals that it resembles a spring that resists outward pulling force by kinesins during cell division.¹³ Accumulated studies suggest that the chromatin can deform in the presence of an extrinsic force and serve as a potential sensor to mediate transcriptional changes in response to extrinsic forces.^{11,12,14} Given the recent finding of the chromatin's solid-like properties,¹⁵ it has been envisioned to serve as a structural scaffold for other nuclear components that display liquid-like state. Recent studies, using single-molecule approaches, have presented evidence that the chromatin structure can buffer against torsional stress and thereby facilitate DNA replication.¹⁶ In light of these findings, the fashion through which the chromatin architecture regulates intra and extra-nuclear forces remain not well understood.

1.3 Phase separated compartmentalization

In addition to the classic view of genome organization via the chromatin structure, recent studies have revealed a novel paradigm for nuclear compartmentalization: phase separation.¹⁷⁻²⁰ Phase separation, at its most rudimentary level, describes the creation of two distinct phases, one dense and one dilute phase, from a well-mixed solution, akin to the immiscible behavior between oil and water. Although well-studied within the context of membraneless cellular organelles,²¹ phase separation's application to biomacromolecules, particularly within the context of gene regulation, is a recent development.

One concept that has been invoked to explain many aspects of biomacromolecular phase separation is biomolecular condensation, defined as molecular-scale, membraneless assemblies that possess the ability to concentrate biomacromolecules in the cells.²⁰ Physically, biomolecular condensates display material properties spanning from a more liquid-like to a more gel-like or solid state.^{20,22,23} Notably, some condensates exhibit maturing behavior, in which the incipient liquid state transitions into a more solid-like state over time.^{20,22,24} Biomolecular condensates, often driven by multivalent, weak interactions by

modular domains and intrinsically disordered regions (IDRs),²⁰ have been noted to enhance or suppress biochemical reactions,^{25–27} buffer protein concentration,^{20,28} and notably, exert mechanical forces^{29,30} etc. Due to their potentially vast implications in diverse physiological processes, aberrant regulation of biomolecular condensates can lead to pathological states such as cancer^{31,32} and neurodegenerative diseases.^{22,33,34}

Biomolecular condensation provides a fresh perspective towards understanding gene regulation. For example, it has been postulated to explain many physiological features associated spatiotemporal control of transcription via physically bringing together multiple distal enhancer sites.³⁵ In support of this notion, many transcriptional machineries such as mediator,^{36,37} RNA polymerase II,^{38,39} and TFs^{40–42} assembled into liquid-like puncta *in vivo*, termed transcriptional condensates,³⁵ which can also be modulated by the presence RNA.⁴³ Through its intrinsic physicochemical milieu, transcriptional condensates have been shown to enhance transcription and explain many of its properties such as the formation of super-enhancers, their bursting behavior, and their simultaneous activation of multiple genes.³⁵ Despite emerging progress made in understanding the function and physical nature of condensates, we still lack a full understanding of its biological function. Importantly, the mechanism through which condensates form at a molecular level and its physical nature remains poorly understood.

Furthermore, biomolecular condensation has been implicated in the formation of the chromatin.^{44–46} Accumulated data suggest that chromatin components, such as nucleosome arrays,⁴⁷ heterochromatin protein 1 (HP1),^{48,49} and linker histone H1 display liquid-like condensates,^{47,50–52} and recent studies suggest that the chromatin behaves more solid-like at a mesoscale level.¹⁵ Given that transcription needs to traverse the complex chromatin landscape,⁵³ how biomolecular condensate components interact and regulate each other remain an outstanding question in light of these new paradigm shifts.

1.3.1 Mechano-regulation by transcriptional condensates

In principle, formation of protein:DNA co-condensates can generate forces, driven by surface tension,⁵⁴ that pull onto non-condensed DNA to perform mechanical work, physically bringing distal enhancer elements that are necessary for gene regulation (**Figure 2**). Consistent with this notion, liquid-like condensates formed through optogenetic manipulation appear to exert mechanical forces driving genomic rearrangement *in vivo*.²⁹ The force generated by protein:DNA co-condensation, as demonstrated via FoxA1, akin to capillary forces that arise between liquid-surface interaction, was estimated to in the order of sub-pN,³⁰ placing them among the weakest known nuclear forces alongside with loop extruding molecular motors.⁵⁵ Given its low force, how does force generation from protein-DNA condensate contribute to gene regulation in the crowded nuclear environment is unclear. Moreover, the mechanism through which chromatin organization regulates protein:DNA co-condensation is an important question that arises from these emerging findings. In particular, many IDR-containing proteins such as BRD4, FUS, and TAF4 display affinity towards nucleosome-depleted region and exclude chromatin as they grow.²⁹ In contrast, others like HP1 α exhibit an affinity towards chromatin rich regions and facilitate their compaction.^{48,49} Hence, how the

chromatin organization impacts TF condensation and its force-generating capacity remains an unaddressed question. In this thesis, I discuss how protein:DNA co-condensation can generate high force, how the chromatin organization can modulate biomolecular condensates' force-generating capacity (Chapter 3), and potential implications in gene regulation (Chapter 5).

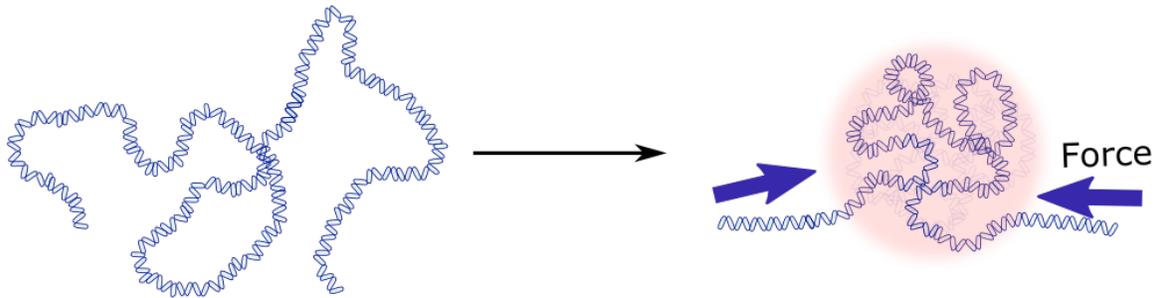


Figure 2. Force-exertion via protein-DNA co-condensation.

Co-condensation between protein and DNA can generate forces, driven surface tension of the droplet, that can pull onto non-condensed DNA, as shown in recent theory and experiment.^{30,54}

1.4 Chromatin loops

The chromatin can fold back into itself to form loops, which can vary in sizes, structure, and function (**Figure 1**).^{1,56,57} Small-scale loops, ranging up to hundreds of kb in lengths, often form promoters to regulate transcription via mediating interactions between promoters and enhancers.^{56,57} Larger sized loops, ranging up to several Mbs in length, contribute to higher order functional yet insulated units in interphase called topologically associating domains (TADs), which can be leveraged to exert spatiotemporal control of gene clusters. During mitosis, larger sized loop formation is also involved in the compaction of DNA into the chromosomal structure.⁵⁶

1.4.1 Genome organization via the SMC complex

In eukaryotes, larger sized chromatin loops, such as those used for mitotic compaction or TADs, are thought to be mediated by members of the structural maintenance of chromosomes (SMC) protein family, which expends energy in the form of adenosine triphosphate (ATP) to drive DNA conformational change.^{55,58} Accumulated literature over decades of literature reveals that the SMC complex is implicated beyond their initial roles in maintaining chromosomal shape and segregation, spanning from regulating DNA replication to transcription.⁵⁹

A distinguishing structural feature of the SMC complex is ring-like shape, consisting of two coiled-coil “arm” subunits and a single connecting kleisin subunit that form the core components (**Figure 3**).⁵⁸ The one end of the coiled-coil “arm” subunit is connected to

another via a stable dimerization interface known as the “hinge,” while the other opposite end forms the transient “head” domain, which are ATP-binding cassette (ABC) family of ATPase domains. The “head,” in particular, is held together by ATP-dependent dimerization in conjunction with the kleisin subunit. Apart from the core components, accessory subunits unique to each SMC complex family assemble around the kleisin and ATPase “head” region (Figure 3).⁵⁸

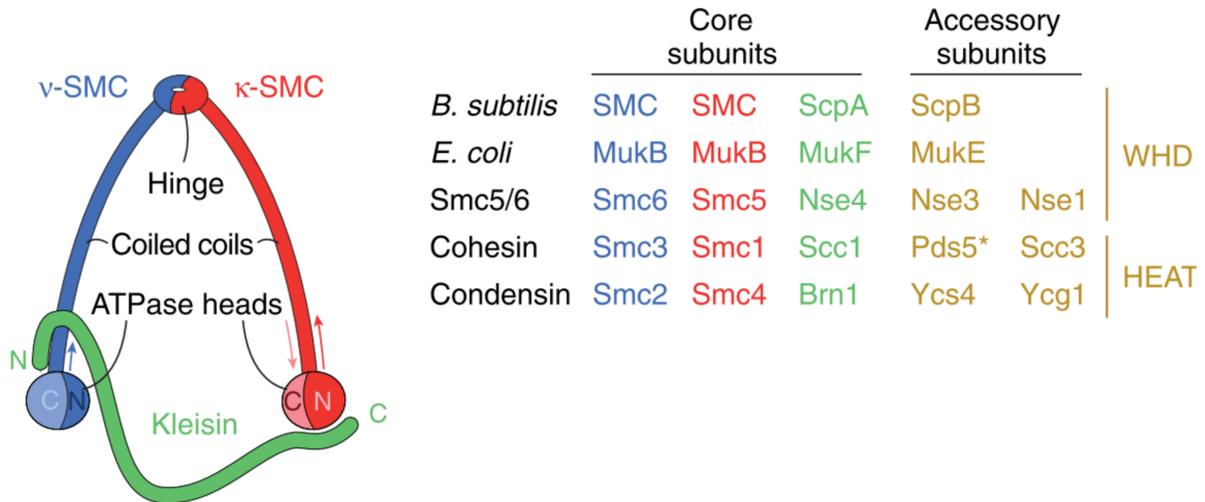


Figure 3. Structural organization of the SMC complex family.

Adapted from Hassler et al.⁵⁸ (Left) Cartoon depiction of the ring-like structure created by the assembly of the core and accessory SMC subunits. (Right) Corresponding subunit composition of prokaryotic and eukaryotic SMC

Evolutionarily conserved, the SMC complex is present in all domains of life. In prokaryotes, the SMC complex forms homodimers making up the MukBEF and SMC-ScpAB complexes that play important roles in bacterial nucleoid organization.^{58,60} In eukaryotes, SMC complexes form heterodimers of different SMC subunits, making up the condensin, cohesin, the dosage compensation complex, and SMC 5/6 complexes with each having its own unique function. Condensin plays a direct role in chromosome condensation.⁶¹ Cohesin plays a role in sister chromatid cohesion.⁶² The dosage compensation complex contributes to heterochromatin formation and X chromosome silencing.⁶³ The SMC5/6 complex, on the other hand, does not appear affect the chromosome structure or sister chromatid cohesion, but rather, is implicated in DNA replication and repair.⁶⁴ In contrast to other SMC complex members, SMC5/6 complex is yet to be named, owing to the comparatively lack of understanding of its *in vitro* and *in vivo* functions. Later in this thesis, I discuss in detail the function of SMC5/6 and its activity on DNA (Chapter 4).

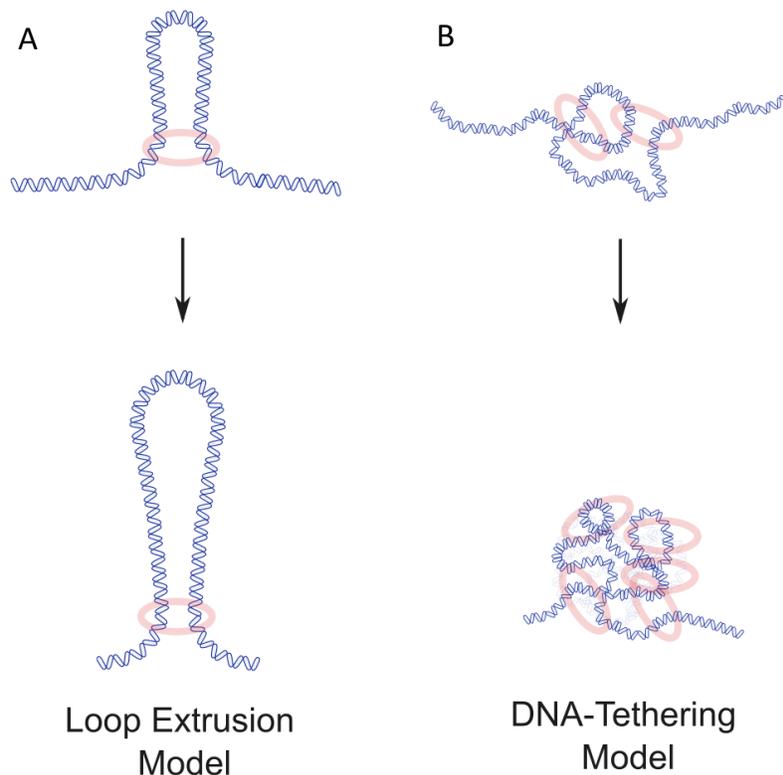


Figure 4. Schematics of loop extrusion and DNA-tethering model.

Loop extrusion model postulates topological entrapment of DNA (shown in cartoon), which does not entail opening of the ring, or pseudotopological entrapment, which does not entail opening of the ring. DNA-tethering model entails the ring structure to bridge two neighboring DNA segments together. This model also implies the cooperative action of multiple complexes to compact DNA.

The SMC complexes are thought organize the genome by physically linking distal chromosomal loci,⁵⁹ but the exact fashion through which this is carried out is not well established. An increasing amount of evidence, supported from *in vivo* observations to *in vitro* single-molecule visualization, suggests the ability of SMC complex to topologically entrap and enlarge DNA loops in what is known as loop extrusion (**Figure 4A**).^{55,58} In recent years, loop extrusion has emerged as a prominent hypothesis towards a unified SMC complex function, supported by direct visualization through single-molecule techniques.^{65–67} Apart from loop extrusion, many SMC complex displays ATP-dependent translocase activity,^{67,68} but it remains unclear how this is coupled to loop extrusion. Additionally, given the low stalling force of loop extrusion, which is typically below 1 pN,^{65–67} it is unclear how loop extrusion navigates the crowded chromatin environment, which would encounter other cellular forces such as those exerted molecular motors in the nucleus. Moreover, other modes of

compaction, such as DNA tethering^{69–71} (**Figure 4B**) and phase-separation may be at play, as suggested by recent findings.^{72,73} However, it is entirely unclear if they are mutually exclusive or not with respect to loop extrusion. As such, clarifying the mechanism through which the SMC complex compacts the chromosome structure remains to be an outstanding question in the field. In chapter 4, I describe how using single-molecule platform developed, as outlined below (see Chapter 1.6.1), yields insights into the SMC5/6 complex's DNA compaction mechanism.

1.5 Techniques used to study genome organization

Two major classes of experimental techniques are currently used to interrogate genome organization *in vivo*: sequencing-based and microscopy-based approaches.^{74,75} Chromosome conformation capture (3C) and its derivative, based on a nuclear ligation assay combined with PCR, have led to the discovery of (TADs) and chromatin loops.⁷⁶ Microscopy-based approaches, such as STORM and FISH-based methods,⁷⁴ have contributed important insights into the chromatin folding in the nucleus, such as the visualization of “nucleosomal clutches.”⁷⁷ Despite the unique strengths of these methods in inferring and visualizing genome organization *in vivo*, the mechanistic contribution of individual molecular component remains difficult to tease out.

In vitro, biochemical reconstitution of a pre-defined biological system facilitates the characterization of individual molecular component in organizing the genome. For example, the material properties of *in vitro* phase separation within the context of reconstituted chromatin arrays and associated components have been characterized via fluorescence recovery after photobleaching (FRAP).⁷⁸ Structural biology approaches, particularly those coming from cryo-electron microscopy,⁷⁹ have revealed the structures of basic chromatin repeating unit, the tetranucleosome, and many DNA-transacting machineries that directly act on chromatin components.^{80–82}

Despite the unique advantages conferred by these techniques, the molecular mechanism underpinning genome organization, particular those pertaining to dynamic conformational changes,⁶ remains a major question in the field. In this thesis, I describe how *in vitro* single-molecule approaches can reveal unique insights into how chromatin component regulates TF binding (Chapter 2), how force-generation by biomolecular condensates is regulated (Chapter 3), and how ATPase molecular machines compact DNA (Chapter 4).

1.6 Single-molecule techniques

Ensemble biochemical approaches provide a population-averaged description of the studied system. As such, they are prone to miss important dynamic information that occur at a shorter time scales (ranging from millisecond to seconds) such as transitional states and nucleation events that often become masked during averaging. This intrinsic disadvantage is particularly apparent if the biological system being studied is asynchronous at a population level. Address these limitations, single-molecule techniques enable direct observation of individual molecules at a high temporal resolution, allowing the detailed interrogation of more transient and stochastic biological events. Moreover, the real-time resolution allows

direct visualization of assembly/disassembly of macromolecular complexes, and therefore, can reveal the temporal order and stoichiometry of molecular events that cannot be discerned using other techniques. Below I describe two major classes of single-molecule approaches employed in this thesis: a fluorescence-based and force-based method.

1.6.1 Single-molecule fluorescence approach

Standard fluorescence microscopy has been hampered by its low spatial resolution, in part, due to contributions by out-of-focus fluorescence that increases background noises. Total internal reflection fluorescence microscopy (TIRFM), exploiting an evanescent field spanning hundreds of nanometers that resulted from total-internal-reflection, facilitates the excitation of fluorophores near the surface (**Figure 5**).⁸³ As such, TIRFM allows for visualization of fluorescence at a high signal-noise ratio and the consequent single molecule resolution.

Applications of single-molecule TIRFM towards the study of genome organization have revealed, using Forster resonance energy transfer (FRET) detection,⁸⁴ that the nucleosomes and chromatin display intrinsic “breathing” dynamics, which can be exploited by PFs and chromatin binding proteins.⁶ Additionally, colocalization single-molecule spectroscopy (CoSMoS) takes advantage of real-time imaging of multiple biomolecules that are conjugated to spectrally separated fluorescent dyes, which colocalize to the same location if a complex is formed.⁸⁵ CoSMoS has shed light on the recruitment and temporal assembly of transcription machineries,⁸⁶ such as cooperativity of transcription factor binding,⁸⁷ and DNA replication initiation.⁸⁸ I present in Chapter 2 data in how CoSMoS can shed light on the binding dynamics of Sox2 PF and linker histone H1.

Recent adaptations of single-molecule TIRFM, particularly through the use of DNA curtains,⁸⁹ have revealed how DNA is compacted by many nuclear proteins spanning from structural proteins such as HP1 α ⁴⁸ to ATP-consuming molecular motor such as condensin.^{65,90} Single-molecule TIRFM confers unique advantages via its high signal to noise ratio, temporal resolution, and throughputs, thereby allowing direct observation of many heretofore elusive molecular mechanism. For example, direct visualization of loop extrusion was achieved for the condensin and cohesin by leveraging TIRFM.⁶⁵⁻⁶⁷ More recently, application of TIRFM led to the formulation that that force can be generated via protein:DNA co-condensation.³⁰ In Chapter 3 and 4, I describe how using a similar adaptation can gain molecular insights in the mechanism behind DNA organization by biomolecular condensates and the SMC5/6 complex.

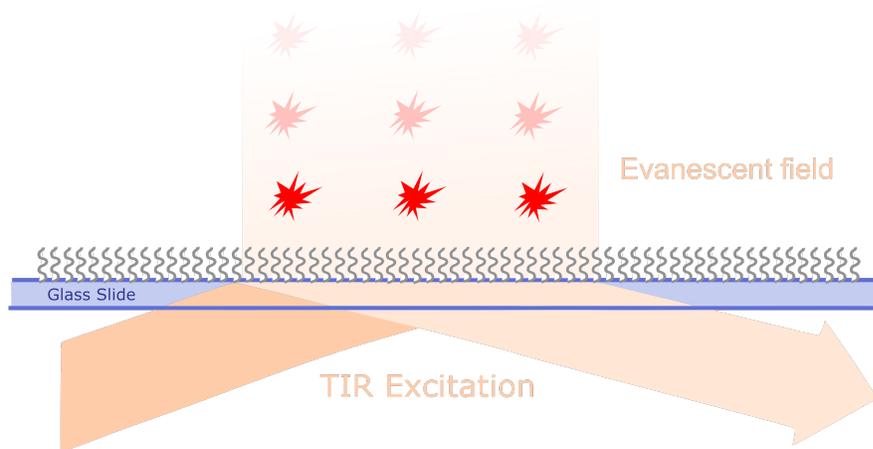


Figure 5. Schematics of single-molecule TIRFM.

TIRFM exploits the induced and decaying evanescent field created by total-internal-reflection of a laser source. The net effect results in excitation of fluorophores that are localized near the surface, reducing background signals. The high signal-noise ratio fluorescence visualization enabled by TIRFM allows for single-molecule detection.

1.6.2 Single-molecule force-based approach

Force microscopy is a powerful class of technique to investigate biological molecules. Optical traps, born out of the discovery that highly focused light source can be used to hold and maintain microscopic particles,⁹¹ confers the ability to study force-mediated biological processes at a high spatiotemporal resolution.⁹² In comparison to fluorescence-based approach such as TIRFM, optical traps allow direct force measurement and physical manipulation of the substrate of interest. The unique technical advantages of optical traps have contributed to the current mechanistic understanding of motor proteins, protein-nucleic interactions, protein folding⁹² and chromatin biology⁶ etc.

In light of the recent paradigm shifts in biomolecular condensate-mediated gene regulation, optical traps have been leveraged to interrogate the physical nature of phase separation at the molecular level. This has led to the formulation of a polymer-surface-mediated condensation model^{93,94} and the notion that protein:DNA co-condensation can generate forces.³⁰ In Chapter 3, I discuss the application of optical traps, performed in collaboration with Jeremy Chang, to study the biophysical properties of Sox2 PF condensates and their force-generating ability.

1.6.3 Combining single-molecule fluorescence- and force-based approaches

Technological advances enabled the combination of fluorescence and optical tweezers, allowing simultaneous visualization and force manipulation. Fluorescence-trap combined instruments range from wide-field epifluorescence to confocal and TIRF microscopy.⁹² In particular, confocal microscopy combined with optical traps⁹⁵ have recently been leveraged to tackle fundamental questions in DNA replication⁹⁶, biomolecular condensation,^{93,94} and heterochromatin formation⁴⁹ etc. In Chapter 3, I describe how combining force and fluorescence can be applied to gain insights into the innerworkings of biomolecular condensates (**Figure 6**).

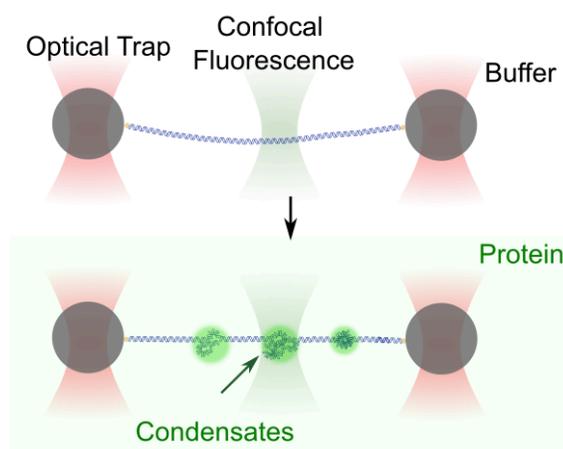


Figure 6. Schematics of force-fluorescence microscopy in studying condensates.

A cartoon depicting a representative DNA tether that is held at a fixed distance by the two optical traps. The trap positions can be moved into a different channel containing protein (shown in light green), where condensates can form. Confocal scanning lasers (shown in dark green) are used to visualize fluorescently labeled proteins bound to DNA tether.

1.7 Overview and significance of findings

Using single-molecule techniques, here, I interrogate the molecular mechanism of eukaryotic genome organization. In Chapter 2, I present data suggestive of a concentration-dependent mode of regulation for H1's and Sox2's binding dynamic. Although the mechanism that underlies such observation remains to be clarified, this surprising observation underscores the importance of tuning the concentration of molecular components to regulate many DNA-binding processes. In Chapter 3, I characterized how TF can form co-condensates with DNA, using Sox2 PF as a model system⁹⁷. I found that Sox2:DNA co-condensation can generate high

forces which can be attenuated by the presence of nucleosomes and linker histone H1, thus implicating the role of chromatin component as a mechanical sink to regulate biomolecular condensation. In Chapter 4, I present data on the ATP-independent compaction activity by the SMC5/6 complex, thus highlighting how an ATPase motor protein can potentially adopt a distinct ATP-independent function to organize genetic materials.

CHAPTER 2. Dynamic interplay between linker histone H1 and Sox2 binding

2.1 The high mobility group protein family

The High Mobility Group (HMG) protein superfamily, named after the high mobility in polyacrylamide gel after its first isolation, consists of a large group of chromosomal proteins that regulate many DNA-transacting processes spanning from transcription to DNA replication. Composed of three distinct families, HMGN, HMGA, and HMGB protein, each displays distinct DNA and chromatin binding sequence motif.⁹⁸ The HMGB family, in particular, is distinguished by its conserved DNA-binding HMG-box. Members of the HMGB family contains single (e.g. Sox, Tcf/Lef) or multiple HMG boxes (e.g. HMGB1-2, mitochondrial transcription factor A (TFAM)), and display sequence specific (e.g. Sox, Tcf/Lef) or non-specific DNA binding activity (e.g. HMGB1-2, FACT, protein TFAM).⁹⁹ It is unclear how the sequence specificity of the HMG-box domain has evolved.

The HMG-box's ability to bend DNA underlies its functional role.⁹⁹ Structurally, the HMG-box binds to the minor groove of the DNA, forming an L shaped structure from three alpha helices that widens the minor groove via multiple electrostatic and hydrophobic interactions. The ability to bend DNA is thought to be dependent on multiple factors: 1) HMG-box's sequence composition, 2) DNA target's sequence composition, and 3) the flanking disordered region.⁹⁹ Given that the HMG-box protein members displaying an affinity towards bent, single stranded, supercoiled DNA,¹⁰⁰ and even RNA hairpins,¹⁰¹ it is curious to ponder if HMG-box proteins display an intrinsic preference for distorted nucleic acid structures, which may explain many of HMG-box member's affinity towards chromatinized DNA.

In comparison to other HMG-box protein, Sox2 contains a single HMG box and displays sequence specificity (see Chapter 3.2 for a more detailed Sox2 review). Thematically, while the overall protein folds of non-specific HMG-box are similar to Sox2, most typically contain multiple HMG boxes, with each having two non-polar intercalating residues instead of a single one in Sox2.^{99,102} Despite emerging progress, much of the structural understanding comes from the HMG-box, and the mechanistic contributions from IDRs, which have been shown to affect DNA binding,¹⁰³ have remained elusive to date.

2.1.1 Transcriptional regulation via HMG-box protein

Due to its unique DNA-bending properties, the HMG-box family is long thought to involve extensively in chromatin remodeling, but the molecular basis for such role is not well established. Initial model for HMG-box's activity postulates it can loosen histone-DNA contact and assist the recruitment of chromatin remodeler.¹⁰⁴ In support of this notion, HMGB1 has been shown to directly bend nucleosomal DNA¹⁰⁵ and assist the recruitment of SWI/SNF complex, BAF complex, and histone chaperone FACT.⁹⁸ Additionally, recent structures of Sox2 bound to nucleosome reveal partial distortion of DNA structure that may render it more amenable for remodeling.^{106,107}

Although the HMG-box family is often associated the opening of closed chromatin, it has been also associated with transcriptional repression. For example, the Tcf/Lef TF has been shown to repress Wnt target genes via directly binding to its promoter.¹⁰⁸ Additional role of the HMG-box as a repressor has been demonstrated in Sox17 and Sox3, which forms a complex with β -catenin and repress Wnt pathway.¹⁰⁹ Together, these findings of HMG-box protein as a repressor of Wnt pathway underscores its multifaceted role in transcription.

Recent paradigm shifts of gene regulation (see Chapter 1.3 and 2.3) point to the idea of phase separation as a potential unifying mechanism. In particular, several studies have demonstrated that many HMG-box members form phase separated compartments. For example, TFAM, a mitochondrial TF, has been shown to phase separate with mitochondrial DNA to drive nucleoid self-assembly, which further recruits many components of transcriptional machineries.^{42,110} Sox2, in particular, has been shown to phase separate with and without the presence of a Mediator.^{36,111} Given the explosion of findings surrounding phase separation, it is curious to ponder that it may underlie an important mechanism through which HMG-box protein regulate transcription, and altering the material properties of condensates may confer a regulatory role to both activate and repress transcription.

2.2 Linker histone H1

Apart from the core nucleosome particle, the next recurring packaging unit of the chromatin consists of linker histone H1 bound to the nucleosome structure.⁴ H1 is a lysine-rich and highly basic protein comprised of a globular domain flanked by N- and C-terminal IDR. Compared to core histones, H1 exhibits much higher sequence diversity among different species. In human, 11 variants of H1 genes have been described, which fall into three major categories: 1) somatic, replication-dependent (e.g. H1.1-H1.5), 2) somatic, replication-independent (e.g. H1.0 and H1x), and 3) germline specific variants (e.g. testis-specific H1t, H1T2, and HILS1, and oocyte-specific H1oo).^{4,112} Among paralogous H1 family members, the highest sequence similarity lies in the globular domain, while the IDRs display highest variance. Curiously, among orthologues, the IDRs display strong conservation, suggesting that the functional differences among different H1 variants mainly lie in the tail regions.¹¹³ The explanation for H1's sheer diversity in variants¹¹⁴ and their evolutionary conservation is a major unanswered question in the field.

H1 is subjected to various modes of post-translational modifications (PTMs) in the globular and IDRs to regulate its association to other proteins. For example, phosphorylation somatic variants H1.2-H1.5 has been extensively shown to alter its cellular localization and chromatin association.¹¹⁵ Additionally, methylation and acetylation of H1.4's N-terminal tail were shown to contribute to heterochromatin formation.^{4,115} Citrullination of H1's globular domain is known to promote its dissociation of chromatin, thereby promoting an open chromatin state.¹¹⁶ Given the differential sequence features associated with each H1's variants in the IDRs, it is thought that PTMs in these regions underlie the functional divergence of H1 variants.⁴

Structurally, H1 histone bound to ~20 bp of linker DNA, forming a chromatosome (**Figure 7A and B**).⁴ Accumulated biochemical and structural data indicate that the globular domain of

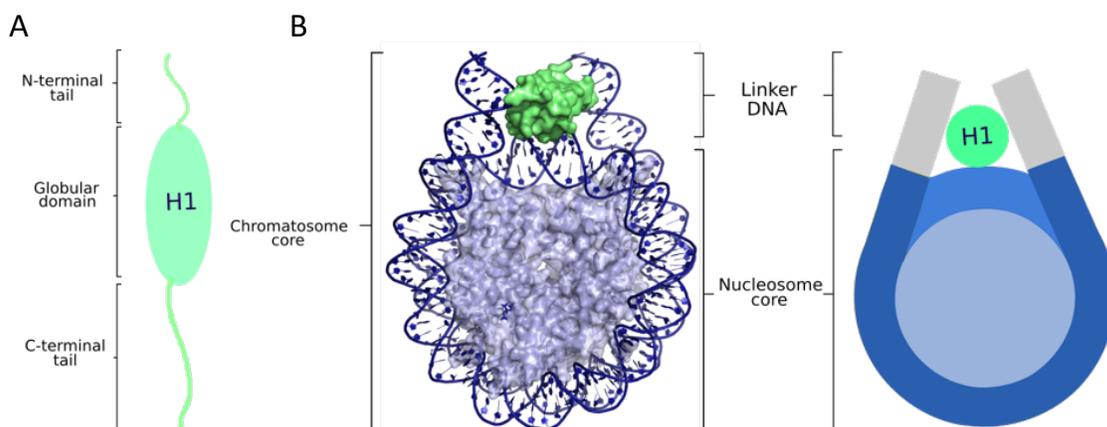


Figure 7. Structural delineation of linker histone H1. A) Outlines the functional organization of H1. B) Depicts the schematics of a chromosome core (PDBID: 4QLC).

linker H1 histone binds to the nucleosome in the vicinity of the dyad region, but the fashion through which the disordered tails of H1 interact with linker DNA is less well known. For a long time, H1 has been shown to contribute to condensed chromatin structure from the individual chromosome to the nucleosome arrays.⁴ However, the structural features of H1 within the chromatin context have remained less well understood. Recent structures of H1 with nucleosome arrays have suggested the role of H1 in contributing to the twisted and ladder-like configuration,^{4,117,118} which is a subject of debates in the field.

H1 displays dynamic bindings to chromatin targets *in vivo*, which contrasts its high, picomolar binding affinity to DNA and chromatin *in vitro*.¹¹⁹ In the cells, assembly and disassembly of H1 are accomplished by chaperones. Many core histone chaperones have been shown to facilitate the loading and unloading of H1 to the chromatin, such as NASP¹²⁰ and Nap1.¹²¹ One well described histone chaperone is prothymosin α (i.e. ProT α), which has been shown to increase the exchange of H1 onto native chromatin substrates.^{122,123} Mechanistically, ProT α displays high affinity towards the C-terminal tail of H1, displacing H1 via a competitive substitution mechanism.¹²⁴ In this section, I present data on the role of ProT α in modulating H1 binding dynamics.

2.2.1 Linker histone H1 and gene regulation

Despite its role in chromatin compaction, H1 has been noted to serve as both a positive and negative regulator of transcription. Early studies in *Tetrahymena* demonstrated that H1 does not have a major effect on global transcription.¹²⁵ In ES cells and mouse models, knockout of H1 demonstrated only a set of specific genes that are positively and negatively regulated.¹²⁶ These findings and among many others suggest that direct regulation of the chromatin structure is not the only means through which H1 can affect transcription.^{127,128}

The scaffolding role of H1, which enables H1's recruitment of transcriptional activators and repressors, has emerged to be a critical factor in H1's gene-specific mode of regulation. In this fashion, recruitment of activators facilitates transcription. For example, H1.2 was shown to recruit Cul4A ubiquitin ligase and PAF elongation complex to the phosphorylated C-terminal tail of RNA polymerase II, and the ubiquitin ligase activity of Cul4A was shown to be necessary for H3K4me3 and H3K79me2, both marks are associated with active transcription.¹²⁹ Conversely, H1 can mediate recruitment of molecular machineries associated with gene repression. For example, H1 murine-specific variant recruits DNA methyltransferase DNMT1 and 3B, leading to DNA methylation and transcriptional silencing.¹³⁰ Additionally, H1.2's scaffolding function mediates p53's and MSX's transcriptional repression activity.^{131,132}

The more intuitive interpretation of H1's role in transcriptional repression is through H1's well defined role in chromatin compaction. The template obstruction model posits that the increased steric hindrance brought about by increased chromatin compaction would inhibit TFs and transcription machineries from interacting with DNA.⁴ This model is consistent with many observations, such as the competition between H1 and PF FoxA1, which shares a similar tertiary structure.¹³³ Additionally, this notion is consistent with the observation that H1 contributes directly via its biochemical properties and indirectly via its scaffolding function to the formation of heterochromatin.⁴ However, increased chromatin folding driven by H1 may serve as favorable substrate for many DNA-binding proteins, such as PF, and histone-modifying enzymes such as PRC2.¹³⁴

Phase separation provides a new perspective towards understanding and interpreting H1's cellular function. In recent years, many studies have implicated the role of H1 in phase separation in the context of double-stranded DNA (dsDNA),⁵² single-stranded DNA (ssDNA),⁵¹ and nucleosome arrays,⁴⁷ driven by its basic IDRs. Therefore, organization of DNA into distinct phases from liquid to solid-like may underlie the ability of H1 to both enhance and repress transcription. Nonetheless, the molecular mechanism behind H1's role in gene regulation is not well understood.

2.3 Dynamic interplay between H1 and HMG-box proteins

Among the first proteins known to alter the properties of the chromatin fiber, HMG-box proteins were demonstrated to de-compact chromatin structure, while H1 condenses it. Consequently, members of the HMG-box, most notably illustrated via HMGB1, and H1 were long thought to act in competing fashion with each other. Both H1 and many characterized HMG-box proteins display similar binding properties near the DNA dyad, albeit each protein results in protection of micrococcal nuclease (MN) digestion on the opposite end of the linker arm.^{135,136}

Several lines of evidence support direct competition between HMG-box family and H1. *In vivo* studies using FRAP demonstrated that HMGB1, when injected into cells expressing GFP-H1, resulted in increased GFP-H1 exchange rate with chromatin, vice versa.^{137,138} The effect appears to be dose-dependent and contingent on the DNA-binding activity of HMGB1. Additionally, direct HMGB1-H1 interactions appear to be important for the increased turnover of H1 *in vivo*.¹⁰⁰ Further *in vivo* evidence from extracted mouse myeloma nuclei and *Xenopus* eggs suggests that H1 and HMG-box protein levels are inversely correlated with each

other.¹⁰⁰

What is the mechanism for the interplay between H1 and HMG-box proteins? H1-HMGB1 have been demonstrated to directly interact with each other *in vitro*, driven by the oppositely charged IDRs.¹³⁹ Further, *in vitro* experiments demonstrated direct competition between HMB1-2 and H1 within the context of four-way junctions and cisplatin-modified DNA.¹⁴⁰ Given the proximity of their binding location in the nucleosome, it is likely that many HMG-box proteins and H1 simply directly compete each other via steric hindrances (i.e. via template obstruction model). Nonetheless, the molecular basis for the competition between HMG-box proteins and H1 is unknown, particularly in the context of the nucleosome structure. Moreover, it is not known if the competition model holds between H1 and all HMG-box family members.

2.4 Single-molecule study of Sox2 and H1 binding to nucleosomes and DNA.

Using single-molecule TIRFM and CoSMoS, I seek to characterize the dynamic binding interplay between H1 and HMG-box protein, using Sox2 as a model system. The presented data reveal a surprising role for the concentration-dependent mode of regulation Sox2 and H1 binding to chromatin. In this work, H1 purification and labeling was performed by Wola Osunsade under the supervision of Yael David. The cloning of the DNA constructs and Sox2 was performed by Sai Li.

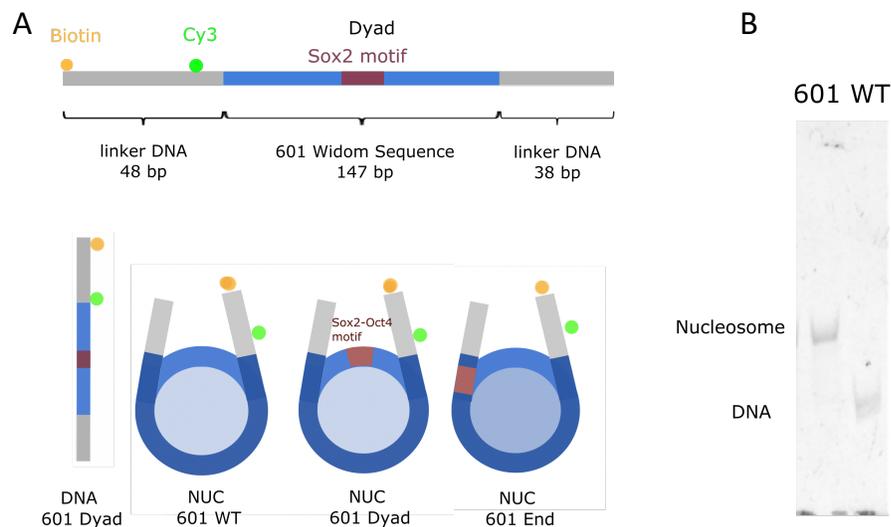


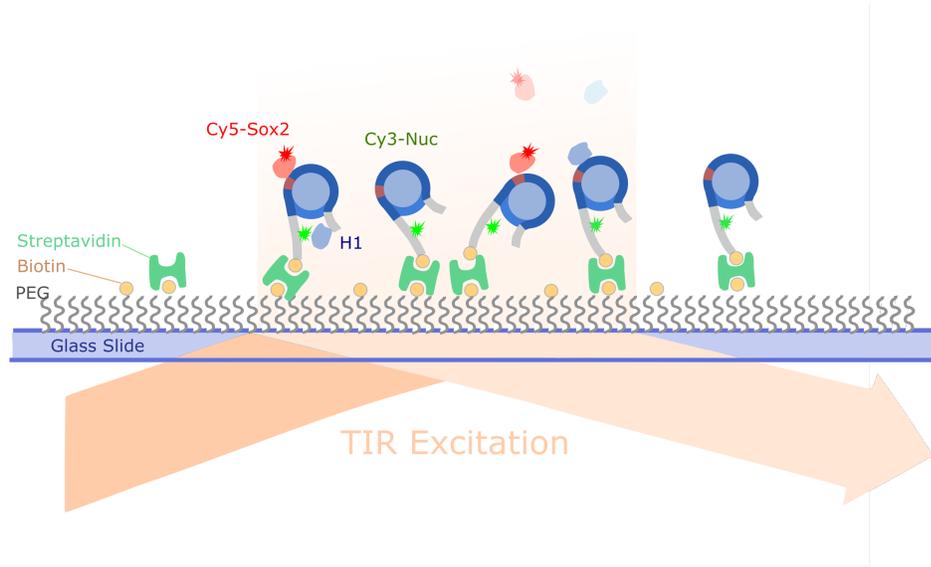
Figure 8. Overview of DNA and nucleosome constructs used in this study.

A) Schematics of biotinylated DNA and reconstituted mononucleosome constructs containing Sox2 binding motif used in this study.

B) Representative native PAGE gel showing reconstituted mononucleosome construct used in this study. Image was taken via a Cy3-scan.

I constructed a DNA template containing the 601 Widom strong nucleosome positioning sequence (i.e. DNA 601 WT) and 601 Widom sequence with an engineered Sox2 TFBS (i.e. CTTTGTT) located in the entry and dyad region, which was previously described⁸⁷ (hereafter named DNA 601 Dyad and 601 End) (**Figure 8A**). The construct is biotinylated and Cy3-labeled. I further reconstituted nucleosomes, as described previously, using this sequence and wild-type 601 Widom sequence (i.e. NUC 601 WT, Dyad, and End, respectively) (**Figure 8B**). The Cy3-labeled bare DNA template and nucleosome constructs were respectively immobilized onto a glass coverslip and their locations visualized via TIRFM in each experiment. I then injected 2 nM of Cy5-labeled Sox2 into the microfluidic chamber containing immobilized DNA/nucleosome substrate. The binding of Sox2 to DNA or nucleosomes was visualized in real time (**Figure 9A**). Single-molecule fluorescence trajectories, extracted via colocalization (see Methods in Chapter 6.3), were further analyzed to assign binding dwell time and binding events ratio (**Figure 9B and C**).

A



B



C

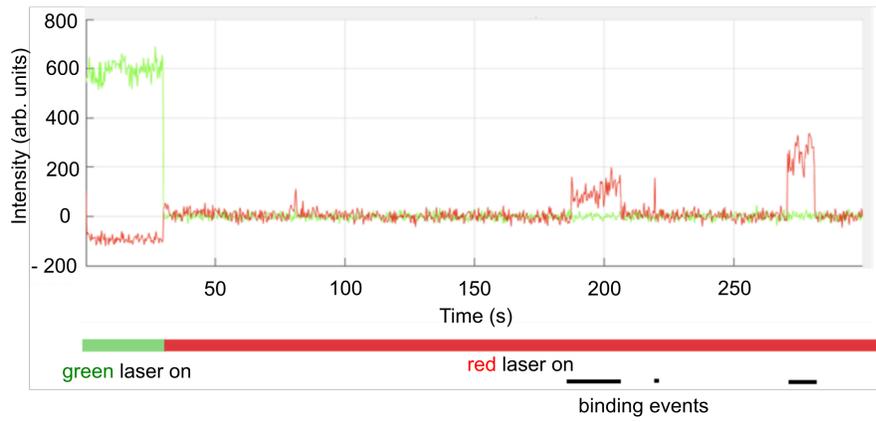


Figure 9. Depiction of colocalization single-molecule spectroscopy (CoSMoS).

- A) Schematics of a representative experiment using the TIRFM setup in this study. Nucleosomes/DNA substrates are biotinylated to a passivated glass surface. Labeled molecules were then flowed into the channel.
- B) Schematics describing the general colocalization scheme used to extract the traces of bound Sox2 molecules to a given DNA/nucleosome construct.
- C) Representative trace, corresponding to single “dot” in B). Note that the green laser (532 nm, corresponding to Cy3) was turned off after 100 frames. After, the red laser (640 nm, corresponding to Cy5), was turned on. Representative colocalization of Cy5-Sox2 to a given Cy3-labeled substrate is noted by black underline, corresponding to a binding event.

2.4.1 H1 does not alter Sox2 binding dwell time

I visualized the binding dynamics of Cy5-labeled Sox2 on Cy3-labeled DNA 601 Dyad, NUC WT Dyad, and NUC 601 WT in the presence of increasing H1.0 and H1.4 concentrations. I calculated the average dwell time for Cy5-labeled Sox2, revealing that Sox2 dynamically binds to DNA and nucleosome with an average dwell time around 6-8 seconds. I observed no statistically-significant differences of Sox2 dwell time among different DNA/nucleosome constructs (**Figure 10A and B**). Furthermore, I discerned no statistical significances of Sox2 dwell time among different H1 concentrations (**Figure 10A and B**), suggesting that H1 does not regulate the binding time of Sox2 to DNA and nucleosome constructs. To rule out the effect of Cy5 photobleaching, I characterized Cy5-Sox2 binding on DNA 601 Dyad under different 640 nM of laser intensities and observed no statistically significant differences in average dwell time of Sox2. This control suggests that photobleaching due to the employed laser power does not account for the dynamic residence time of Sox2 (**Figure 10C**).

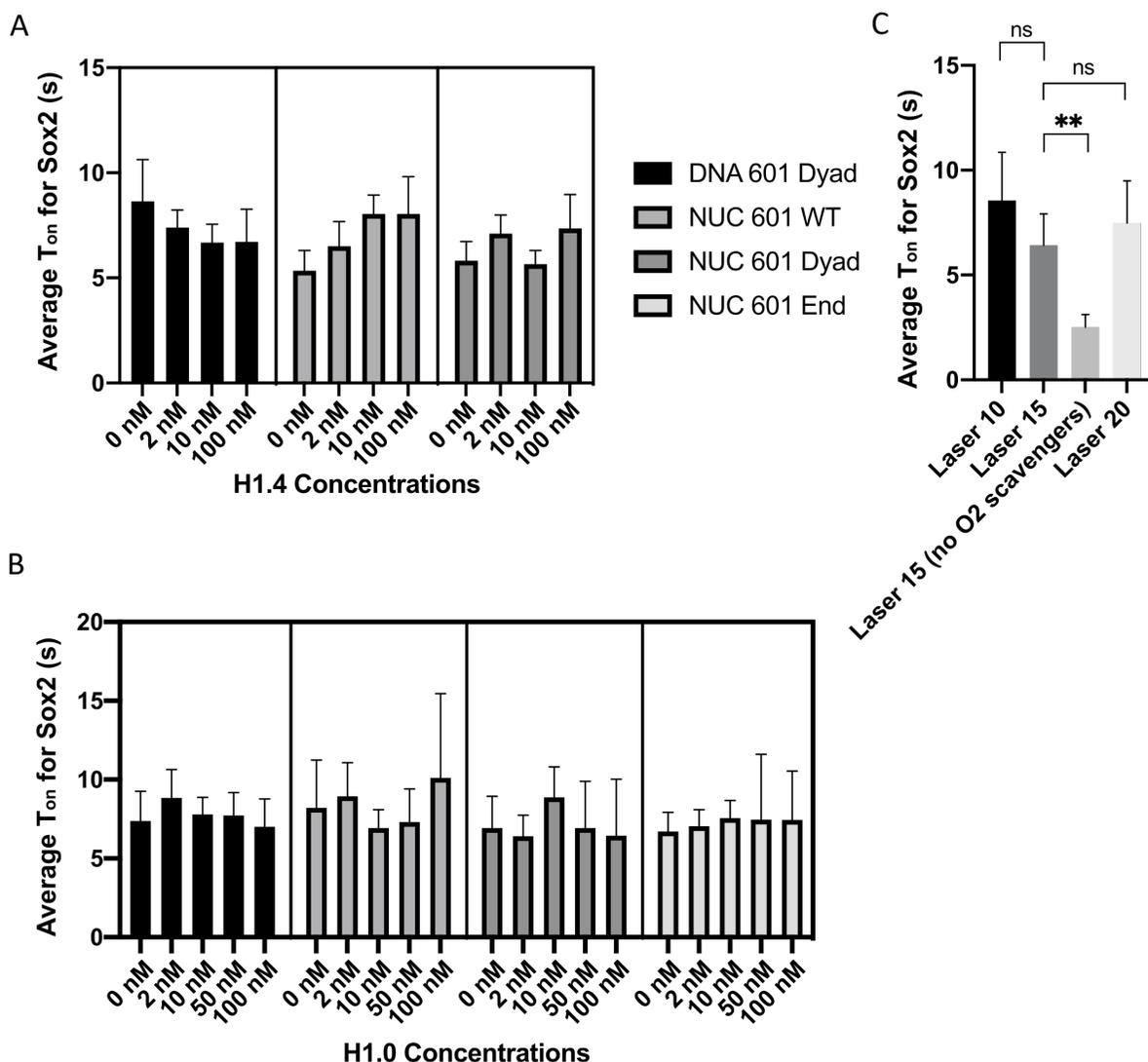


Figure 10. H1 does not modulate the dwell time of Sox2 on DNA and nucleosomes.

Calculated average dwell time (T_{on}) of Sox2 for a given DNA/mononucleosome construct with A) H1.4 titration and B) H1.0 titration. C) Shows the T_{on} of Sox2 on DNA 601 Dyad under different laser conditions. Laser 15 is the power used in all experiments. Laser 15 (no O₂ scavengers) corresponds to a negative control wherein oxygen scavengers are not present, which would result in fast photobleaching.

2.4.2 H1 displays a concentration-dependent regulation of Sox2 loading

I further analyzed the loading rate, measured via the binding events ratio (see Methods in Chapter 6.3), of Sox2 on the described DNA and nucleosomes (Figure 8). The described

analysis reveals that H1.0 and H1.4 both facilitate the loading of Sox2 on DNA and nucleosomes at low concentrations (i.e. at 2 nM and 10 nM). Curiously, higher concentrations of H1.0 and H1.4 (i.e. beyond 50 nM) decrease the loading rate of Sox2 on both DNA and nucleosomes (**Figure 11A and B**). Together, these data suggest that H1 regulates Sox2 loading rate on DNA and nucleosomes in a concentration-dependent fashion.

2.4.3 Deletion of H1.4's C-terminal tail promotes Sox2 loading at higher concentration

I asked what are the molecular determinants driving H1's regulation of Sox2 binding. The C-terminal tail of H1 contains a long stretch of intrinsically disordered and lysine-rich stretch of residues that was previously known to bind to HMG-box protein.¹⁴¹ I hypothesized that it may play a role. I measured Sox2's loading rate on DNA and nucleosome substrates with increasing concentration of H1.4 with its C-terminal tail truncated (i.e. H1.4 CTDdel, obtained via David

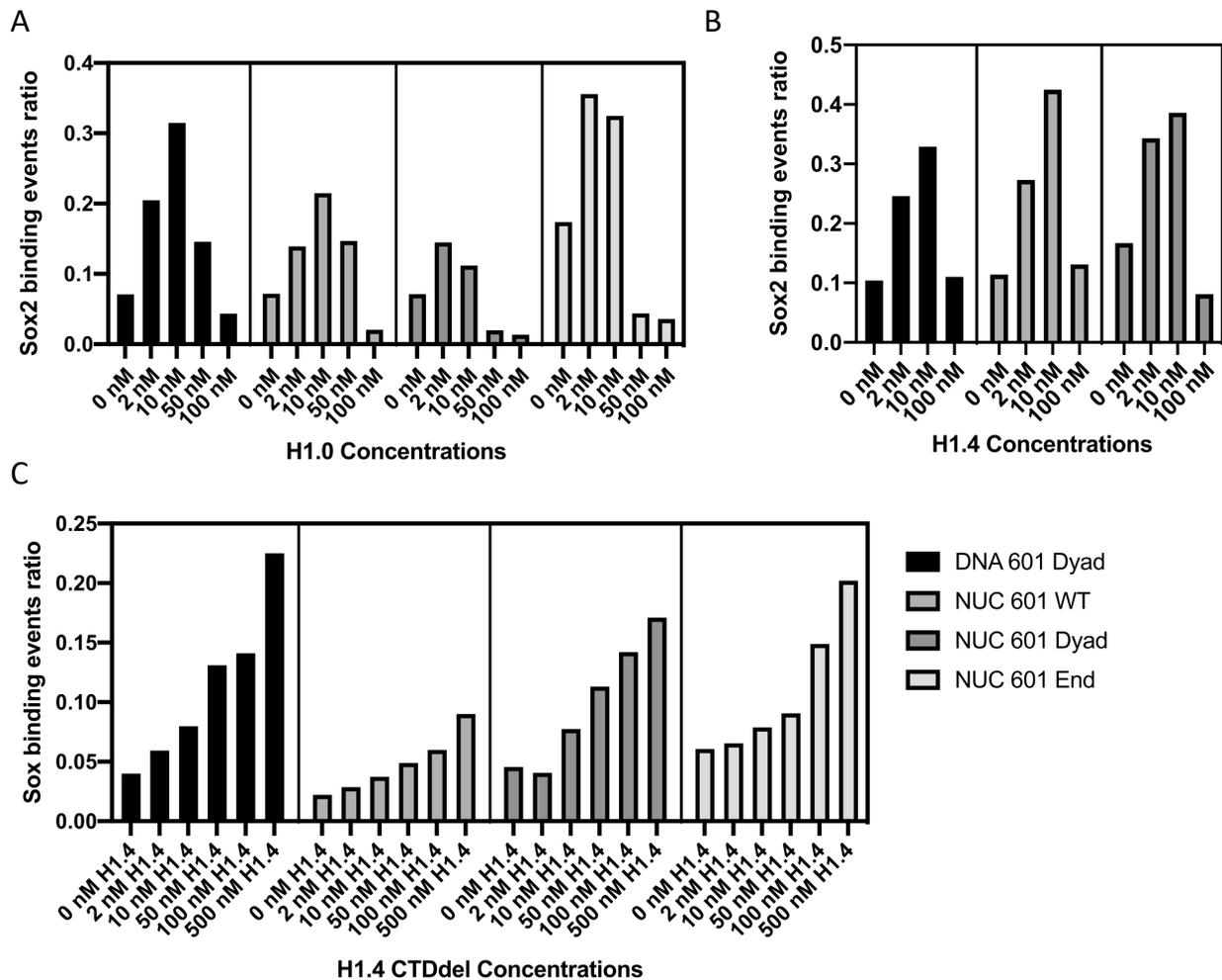


Figure 11. H1 modulates Sox2's loading rate onto DNA and nucleosomes.

Bar graph showing the Sox2's binding events ratio on DNA and nucleosome constructs (see Figure 8) in the presence of increasing A) H1.0, B) H1.4, and C) H1.4 CTDdel concentrations.

lab). I observed that increasing H1.4 CTDdel concentration further promotes the loading rate of Sox2 on DNA and nucleosomes (**Figure 11C**).

2.4.4 Sox2 promotes H1 loading on DNA and nucleosomes

I visualized the binding of H1.4 on Cy5-labeled DNA and nucleosomes (NUC 601 WT and DNA 601 Dyad) by flowing in 10 pM of Cy3-labeled H1.4 (obtained via David lab). In particular, I asked how Sox2 regulates H1 binding on DNA and nucleosomes. I visualized the binding of Cy3-labeled H1.4 on Cy5-labeled DNA and nucleosome in the presence of increasing unlabeled Sox2 concentration. My analysis reveals that higher concentrations of Sox2 promote H1.4 loading, as measured via the binding events ratio, on both DNA and nucleosomes (**Figure 12A**). Interestingly, I observed a large spike of increased Cy3-H1.4 loading for both DNA and nucleosome under high 500 nM Sox2 concentrations. H1 loading *in vivo* is mediated in the presence of chaperones.¹⁴² As a control, I measured the loading rate of H1 with increasing concentration of ProT α , a known chaperone of H1, promotes the loading of H1.4 on nucleosomes but on DNA (**Figure 12B**).

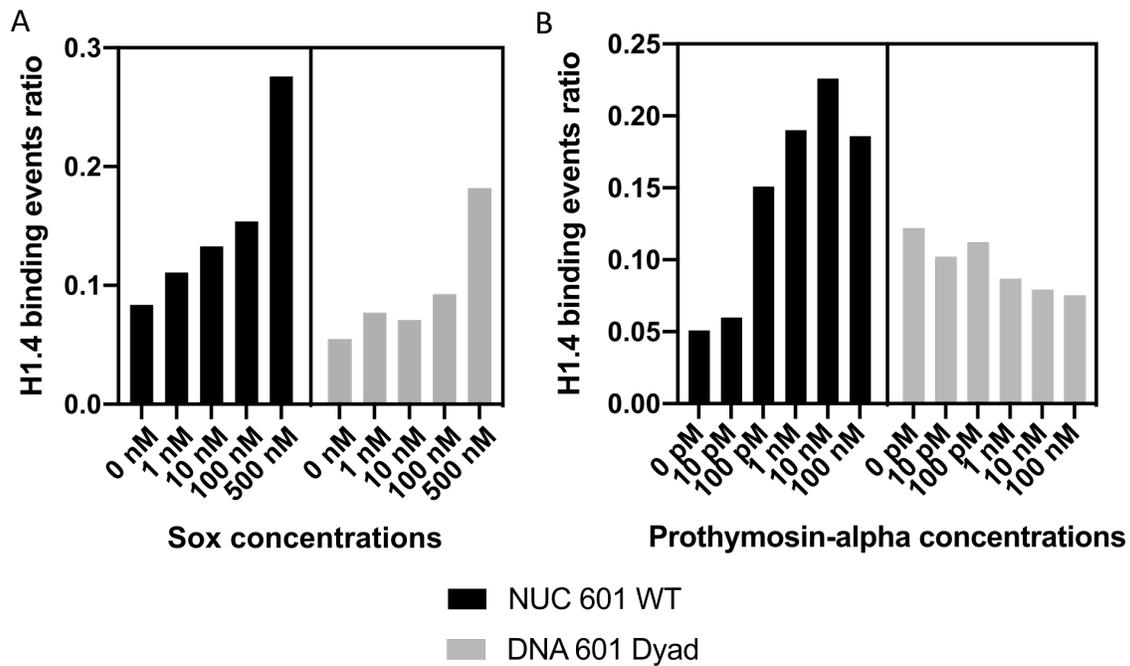


Figure 12. Modulation of H1 loading rate on DNA and nucleosome by Sox2 and ProT α

H1.4 binding events ratio on selected DNA and nucleosome constructs in the presence of increasing A) Sox2 and B) ProT α concentrations.

2.5 Conclusion

Taken together, the presented data reveals that Sox2 and H1 can both promote each other's loading on DNA and nucleosomes, as demonstrated via increased binding events ratio. H1 can inhibit Sox2 binding on both nucleosomes and DNA at higher concentrations beyond 50 nM, and this phenomenon appears to be mediated via the intrinsically disordered tail.

Additionally, H1 does not alter Sox2's binding dwell time on DNA and nucleosomes. Lastly, I show that ProT α can promote H1 loading on nucleosomes but not on bare DNA. Below, I further discuss these data in the context known literature and speculate on the underlying mechanism in Chapter 5.

CHAPTER 3. Chromatin sequesters pioneer factor Sox2 from exerting force on DNA.

3.1 Pioneer transcription factor

Given the complex chromatin organization, a subset of TFs, known as pioneer factors (PFs), possesses the ability to access silent chromatin through their intrinsic affinity for nucleosomal DNA, thereby further recruiting transcriptional machineries to initiate cell fate transition and reprogramming events.¹⁴³ Thus, PFs are deemed to be among the first to engage and open up closed chromatin for subsequent TFs' and chromatin remodeling machineries' activity. Recent insights into the structures and functions of PFs reveal that they recognize and further induce distorted DNA structures embedded within the nucleosome structures.^{106,144,145} Within the chromatin fiber, PFs have been shown, using single-molecule techniques, to physically open the chromatin structure via preventing nucleosome stacking.¹⁴⁶

Despite these recent insights into the nucleosome-binding activities of PFs, many outstanding questions remain unanswered. PF activity appears to be constrained by diverse factors *in vivo*, such as chromatin states and co-factor bindings, etc.^{147–150} Notably, the intrinsic nucleosomal structure appears to be a constraint even among PFs.^{148,151} Thus, TF pioneering activity is likely conditional on intrinsic and extrinsic factors specific to chromatin states that await further clarifications. Moreover, PFs are long thought to induce local changes in chromatin and facilitate recruitment of transcriptional machineries to facilitate gene regulation, but exactly how this process is orchestrated has not been well understood.

3.2 Sox2: an introduction

One prime example of a pioneer factor shaping gene expression program is Sox2, which belongs to the original Yamanaka TF cocktail (i.e. Oct4, Klf4, and c-Myc) that enables reprogramming of somatic cells to induced pluripotent stem cells (iPSCs).¹⁵² At its most basic level, Sox2 is a cell-fate determining TF that is implicated in the maintenance and induction of pluripotency with over decades of literature.¹⁵² Due to its critical role in development, Sox2 has been extensively studied in multiple contexts from development to cancer.^{153,154} Below, I limit my review towards its role in pluripotency and biochemical properties, particularly in relation to transcription.

3.2.1 Sox2 and pluripotency

Underscoring its role in pluripotency, high level of Sox2 expression is required for the maintenance of embryonic stem cells (ES), as its deletion drives ES cell differentiation.¹⁵⁵ In the setting of iPSCs, endogenous expression of Sox2, driven by multiple feedback loops, lying at the center of the gene network regulating pluripotency.¹⁵⁶ Chromatin immunoprecipitation sequencing (ChIP-seq) studies revealed that Sox2 occupies cis-regulatory elements of a large number of genes necessary for pluripotency.^{150,157,158} In diseased states, aberrant Sox2

expression has been linked to many forms of cancer stem cells and their endothelial to mesenchymal-transition.¹⁵³ Taken together, while the cellular implications of Sox2 on pluripotency have been relatively well defined with decades of literature, the molecular mechanism underlying Sox2 regulation, particularly within the context of transcription, is not as well understood.

3.2.2 Sox2's biochemical properties

Sox2 belongs to the High Mobility Group (HMG) box superfamily of proteins, containing an evolutionarily conserved HMG box that binds to the minor groove of DNA (see Chapter 2.1 for HMG-box protein review). Sox2 displays a specific affinity for the consensus sequence [C/A][A/T]TTGT in human with the central TTGT motif being the common motif among all Sox proteins.^{102,153} Given the existence of several thousands of copies of Sox2's consensus motifs in the human genome, Sox2 appears to be a rather promiscuous factor, a notion supported by its short *in vivo* dwelling time at a specific target DNA.¹⁵⁹ Recent findings also suggest that Sox2 is capable of binding to RNA *in vitro* and *in vivo*, although the functional significance of this remains to be clarified.^{101,160} Given the complexity of its binding modes, Sox2's specific binding to its cognate site may be contingent on extrinsic factors such as the chromatin state and co-factor binding, as discussed below.

Much of the current understanding of Sox2 binding activity on DNA comes from the HMG box (see Chapter 2.1), which displays the minimum sufficiency for Sox2's DNA binding activity. Structurally the HMG box folds into an L-shaped structure consisted of three alpha helices (**Figure 13A**), which makes extensive contact with the consensus motif via base-specific interactions.¹⁰² The HMG box, through insights from circular permutation assays and structural studies, displays the ability to distort DNA (**Figure 13B**), which is thought to contribute to Sox2's pioneer activity.^{161,162} In support of this notion, the degree of HMG box-induced DNA bending is well correlated with Sox2's capacity to drive transcriptional output.¹⁶³ However, the molecular mechanism through which the DNA bending activity of Sox2 and many others in the HMG box family regulates transcription remains unclear.

Besides the HMG domain, Sox2 contains two flanking N- and C-terminal domains that remain largely disordered and not as well characterized. The C-terminal tail spans a length of ~200 residues, containing a stretch of serine-rich residues, a dimerization domain, a nuclear localization signal (NLS), a conserved Trans-Activation Domain (TAD) (**Figure 13C**), and many reported sites of post-translational modification.^{102,153,164} The N-terminal segment, on the other hand, spans a shorter stretch of ~40 residues that contain a minor site for NLS, a stretch of poly-glycine residues, and overall is less characterized compared to the C-terminal tail (**Figure 13C**).^{165,166}

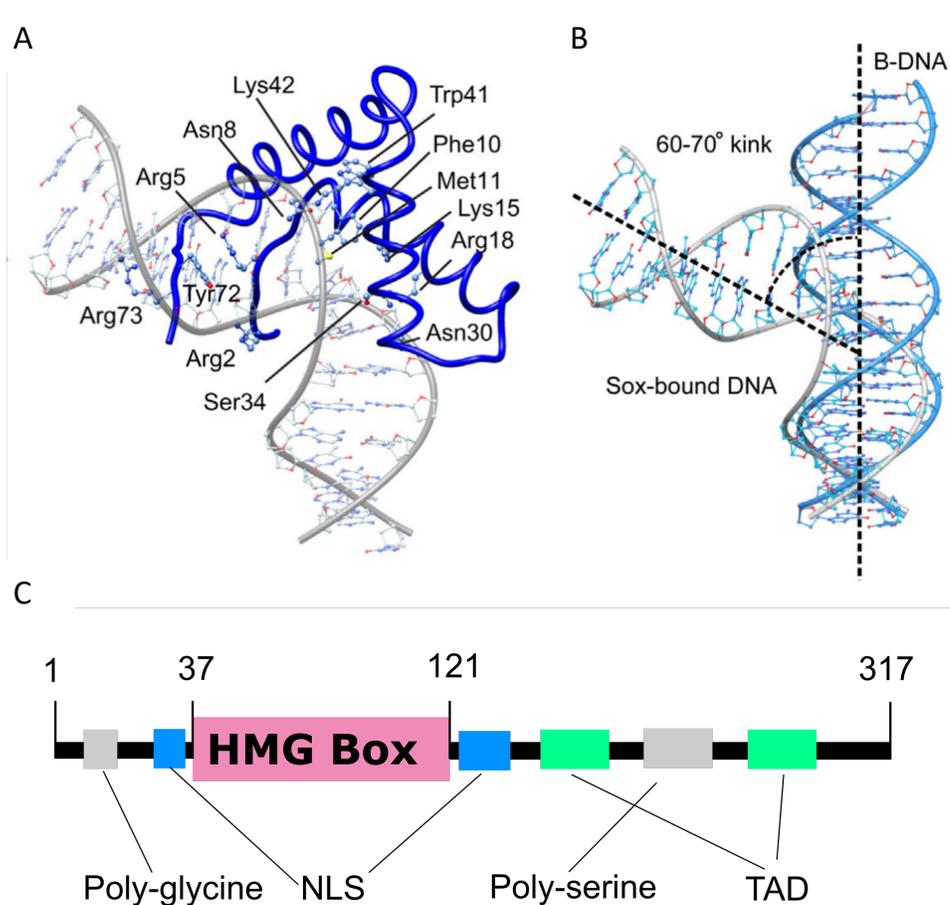


Figure 13. DNA-binding activity and domain organization of Sox2.

- A) Adapted from Hou et al.¹⁰⁴, ball-and-sticks representation showing the contact DNA:protein interface between Sox2 and the CATTGT binding motif. Methionine 11 is a key hydrophobic intercalating residue that is structurally conserved among all HMG-box protein.
- B) Superimposition of ideal B-DNA and Sox2-bound DNA, showing the bending angle induced by Sox2.
- C) Domain organization schematics of Sox2 as discussed in the text.

Mounting evidence points to the idea that Sox2 is PF, given its intrinsic affinity to nucleosomes *in vitro* and DNase-resistant, silent chromatin *in vivo*.¹⁴⁹ Recent studies suggested that PF's activity may be constrained by the chromatin state, and in fact, by the molecular architecture of the nucleosome itself.^{149,151,167} For example, Sox2's engagement to chromatin is dependent on Sox2's TF binding site (TFBS) orientation (i.e. rotational setting) and location (i.e. translational setting) relative to the histone octamer, where it prefers to bind to the nucleosomal dyad location.^{87,151} Recent structures of Sox2 bound to nucleosomes revealed that its recognition of distorted DNA structure in the nucleosomal architecture, providing a possible explanation for the bypass of a more stringent consensus TFBS requirement *in vivo*.¹⁴⁹ Despite the emerging evidence that supports the idea of Sox2 as PFs,

there is also counterevidence. Namely, *in vivo* studies that measure the ability of TF to open up chromatin, assayed via dynamic changes in DNase I hypersensitivity footprint sites during ES cells' differentiation, suggested that Sox2 more likely resembles a "migrant TFs" that bind sporadically to their chromatin sites.¹⁶⁸ Thus, despite the wealth of information brought about by new findings in the field, we still lack a coherent picture of how the chromatin regulates Sox2's pioneer activity, particularly in light of new conceptual advances delineated below.

3.3 TF-mediated transcriptional condensates

Given the paradigm shift involving phase separation (See Chapter 1.3), an emerging view behind PF-mediated gene-regulation involves the formation of TF-driven transcriptional condensates (**Figure 14**).³⁵ In this model, enhancer elements are viewed as nucleation sites for formation of condensates, which further recruit transcription machineries to execute gene-regulation. Consistent with this notion, increasing the density of TF-binding site facilitates phase separation in bulk and biomolecular condensate formation at the molecular level.^{169,170} Furthermore, TF-mediated condensates have been shown to directly recruit RNA-polymerase II and enhance transcriptional output in multiple settings.¹⁷⁰⁻¹⁷² Given the crowded chromatin environment, it remains unclear how transcriptional condensates are regulated.

The connection of TF-mediated to condensates to disease is readily made apparent in light of recent findings implicating mutations, fusion, and disease-repeating expansion in TF to condensate formation.^{41,170,173} Importantly, TF-driven condensates also impact the partitioning of cancer therapeutics, fueling many start-up efforts.^{174,175} Therefore, further basic understanding condensates will support drug discovery efforts and therapeutic strategies to address unmet clinical needs.

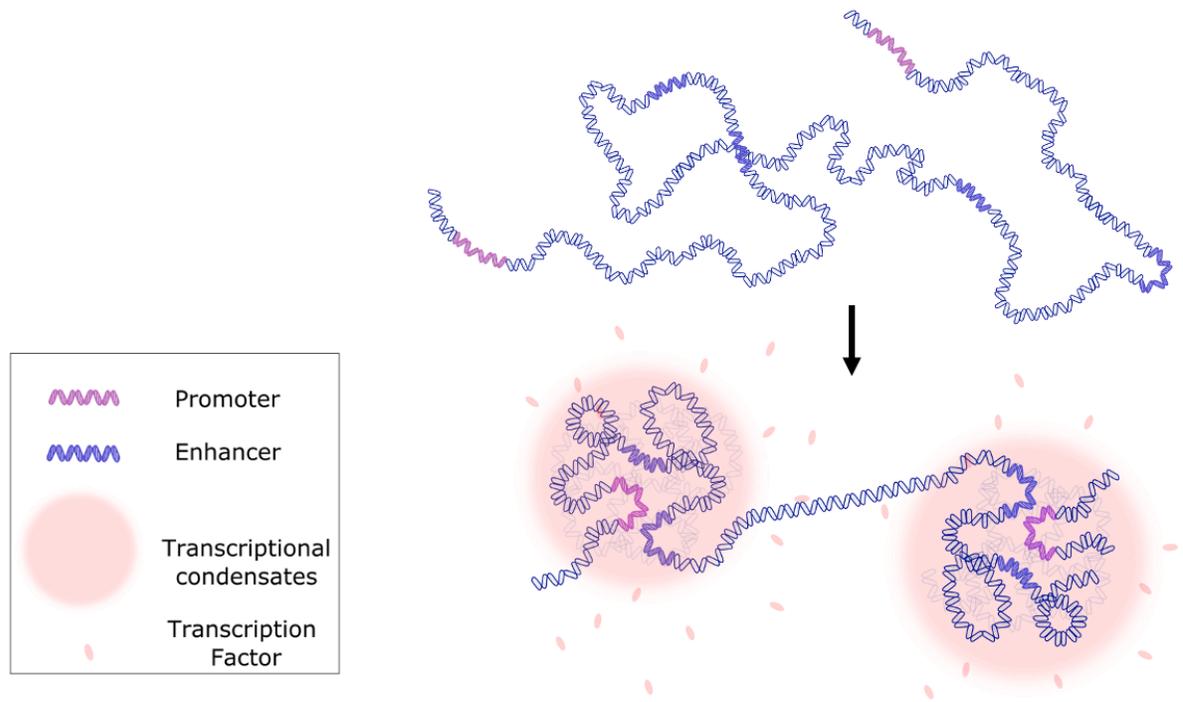


Figure 14. Schematics of TF-mediated transcriptional condensate.

The transcriptional condensate model posits the phase-separation of transcriptional component, which would facilitate bringing together regulatory elements such as promoter and enhancer together. The model explains many properties related to enhancer-related gene regulation, as noted in Chapter 1.3, particularly with respect to super-enhancers.

3.4 Single-molecule findings of Sox2:DNA co-condensate formation

Below I show the findings that are adapted from Nguyen et al.,⁹⁷ in which I leveraged single-molecule TIRFM and optical-tweezer to characterize how Sox2 forms co-condensates with DNA and chromatin. The results below present three conceptual advances to the field: 1) protein:DNA co-condensation can generate high forces, up to ~ 7 pN comparable to other reported cellular forces, 2) the intrinsically disordered regions (IDRs) are dispensable for condensate formation but necessary for high force generation, and lastly, 3) chromatin component, such as nucleosome and linker histone H1, attenuates the force generating capacity of Sox2 condensates, and reduce their mechanical effects on DNA. The presented findings add to the growing body of studies that the chromatin can function as a mechanical sink that regulates the cellular forces.

In this work, the cloning of Sox2 constructs and many of its mutants was performed by Sai Li. Optical tweezer experiment was performed by Jeremy Chang. Computational analysis was aided via automation with John Watter's help. EMSA experiment was performed with the

help of Htet Ng. Linker histone H1 was purified and labeled by Adewola Osunsade of Yael David's lab.

3.4.1 Sox2 forms co-condensates with DNA

I used the bacteriophage λ genomic DNA (λ DNA) as a model DNA substrate for this study. Individual λ DNA molecules were immobilized on a glass surface via biotin-streptavidin linkage, stained with the YOPRO1 dye that intercalates into the DNA backbone, and imaged with total-internal-reflection fluorescence microscopy (TIRFM) (**Figure 15A**).⁸⁷ Double-tethered λ DNA molecules exhibited a distribution of end-to-end distances due to heterogeneous anchoring of the two ends. Molecules with short end-to-end distances displayed larger transversal fluctuations—due to more slacks—than those with long end-to-end distances (**Figure 15A**). After flowing in Cy5-labeled recombinant full-length human Sox2 (**Figure 16**), I observed the formation of Sox2 foci on the DNA (**Figure 15B**), which contains numerous Sox2 binding motifs across its native sequence (**Figure 17**). These foci displayed mobility on the DNA as well as fusion and splitting events (**Figure 18A**), indicating liquid-like properties.^{36,111} Upon Sox2 binding and foci formation, I also observed that the fluorescence signal of the DNA

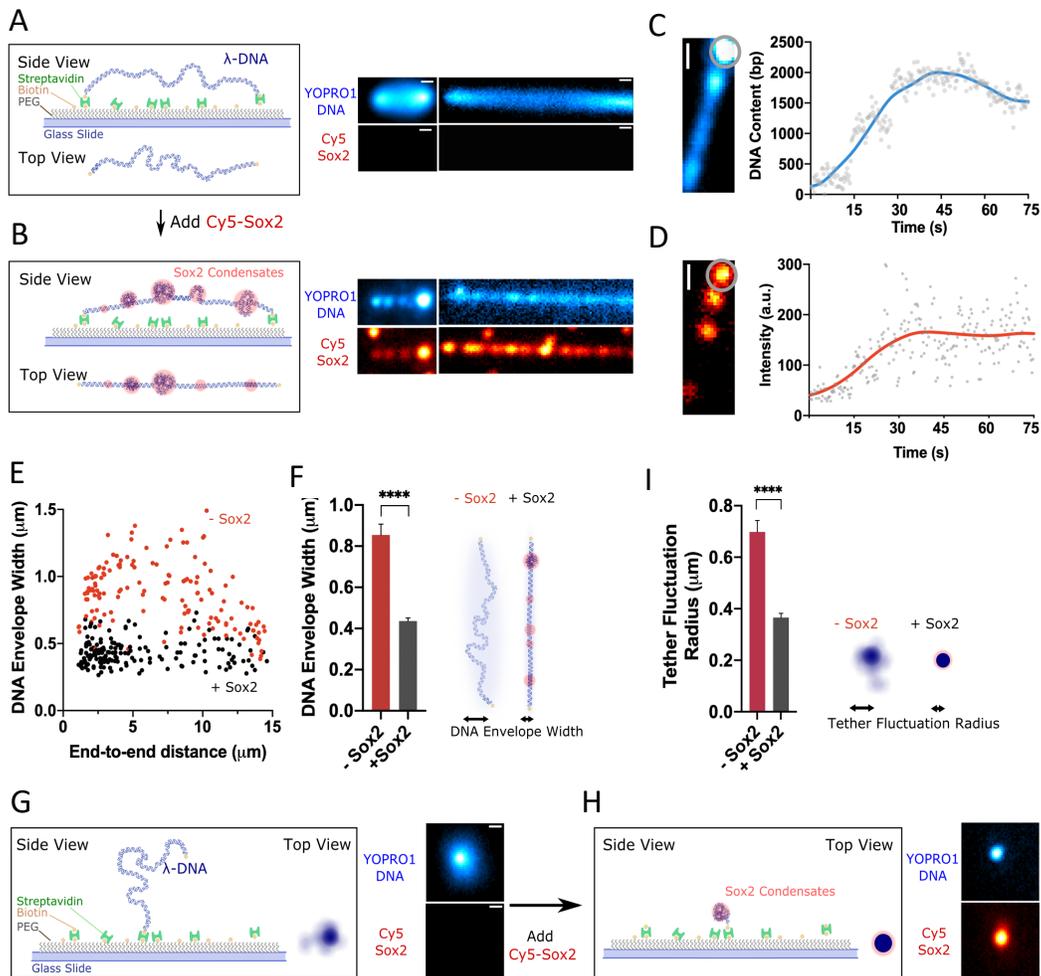


Figure 15. Sox2 forms condensates on DNA.

A) (Left) Schematic of a double-tethered λ DNA. (Right) Two example images of double-tethered λ DNA molecules with different end-to-end distances (among 4 independent experiments). DNA was stained with YOPRO1 (20 nM) and imaged by TIRFM. Shown are time-averaged projections over a 75-s period. Scale bar, 0.5 μ m.

B) Schematic and time-averaged projections of the same two λ DNA molecules as in a (among 4 independent experiments) when incubated with 10 nM Cy5-labeled Sox2. Scale bar, 0.5 μ m.

C) Real-time tracking of the DNA content (YOPRO1 fluorescence intensity converted to the amount of DNA base pairs) within a condensate (circled region) on a double-tethered λ DNA (among 4 independent experiments). Scale bar, 0.5 μ m.

D) Corresponding changes in the Sox2 intensity within the same circled region as in c (among 4 independent experiments). Scale bar, 0.5 μ m.

E) DNA envelope width as a function of the end-to-end distance of double-tethered DNA measured in the absence ($n = 147$) or presence of Sox2 ($n = 162$), where n represents the number of DNA molecules analyzed.

F) Bar graph and cartoon showing a reduction in the mean DNA envelope width averaged over all the molecules shown in e ($n = 147$) upon Sox2-mediated co-condensation ($n = 162$). Error bars denote 95% CI. Significance was obtained using an unpaired two-sample t test (**** $P < 0.0001$).

G) Schematic and time-averaged projection of a single-tethered λ DNA stained with YOPRO1 displaying random fluctuations. Scale bar, 0.5 μ m.

H) Schematic and time-averaged projection of the same single-tethered λ DNA as in g showing Sox2-mediated condensation. Scale bar, 0.5 μ m.

I) Bar graph and cartoon showing the mean fluctuation radius of single-tethered DNA molecules in the absence ($n = 38$) or presence of Sox2 ($n = 37$). Error bars denote 95% CI. Significance was obtained using an unpaired two-sample t test (**** $P < 0.0001$).

transitioned from a relatively uniform distribution to a few clusters that colocalized with the Sox2 foci (**Figure 15B**). This was particularly apparent in the DNA strands with a short end-to-end distance.

Once nucleated, the Sox2 foci on DNA were long-lived, and the fluorescence intensities of both Sox2 and DNA at the foci increased with time until reaching a steady state (**Figure 15C and D**). Interestingly, I observed a loss of the fluctuating motion in the DNA concurrent with Sox2 foci formation (**Figure 15A and B**). Indeed, the average DNA envelope width—a measure for the degree of transversal fluctuations—was significantly reduced in the presence of Sox2 (**Figure 15E and F**). Even though the DNA envelope is wider for molecules with shorter end-to-end distances in the absence of Sox2 as expected, the addition of Sox2 reduced the envelope

width for all double-tethered molecules to the same level (**Figure 15E**). These findings can be

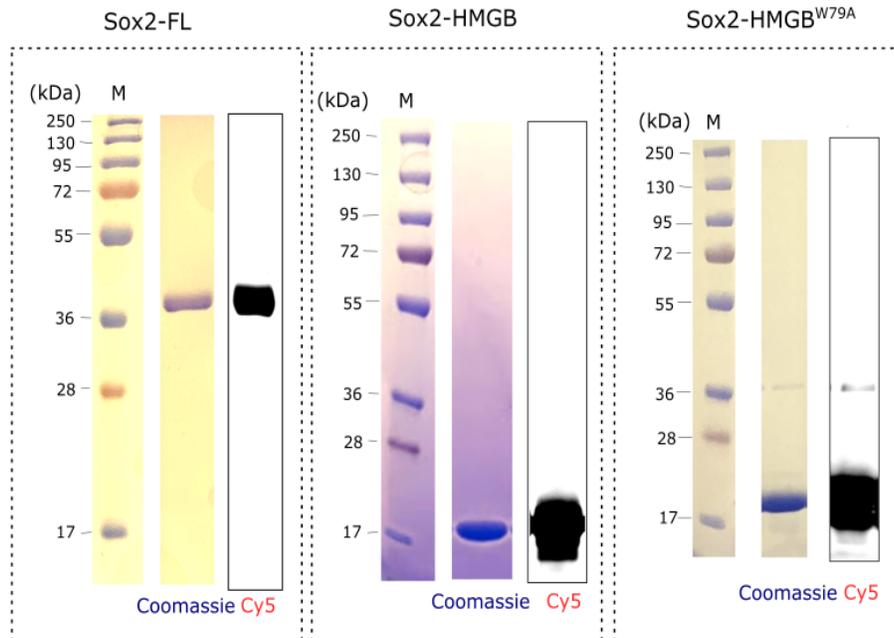


Figure 16. Purification and labeling of recombinant human Sox2.

Coomassie stain and Cy5 fluorescence scan of Sox2-FL, Sox2-HMGB, and Sox2-HMGB^{W79A} proteins (among 3 independent preparations).

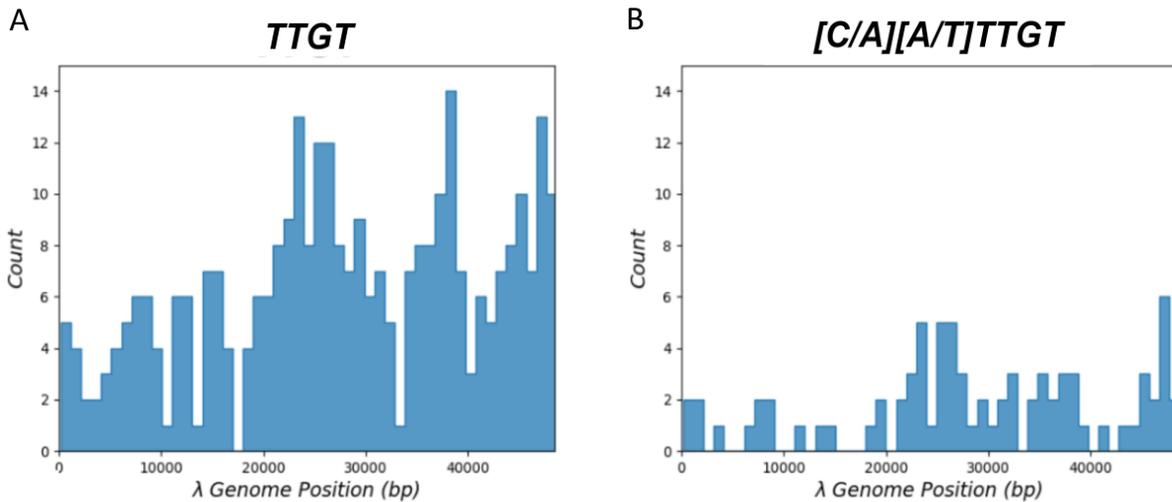


Figure 17. Distribution of Sox2 binding motifs on λDNA.

A) Histogram displaying the occurrence of the canonical Sox2 motif TTGT along the λDNA genomic sequence.

B) Histogram displaying the occurrence of the extended Sox2 motif [C/A][A/T]TTGT.

Bin size in the histograms is 1 kb.

rationalized by an ability of Sox2 to form co-condensates with DNA. As more DNA being pulled into the condensates, the previously slacked DNA transitioned into a tensed state. In addition, I observed a fraction of λ DNA molecules that were tethered to the surface at only one end (**Figure 15G**), likely because the other biotinylated end did not find a streptavidin to bind during flow injection. Without Sox2, these single-tethered DNA molecules displayed random fluctuations characterized by a measurable radius (**Figure 15G and I**). The addition of Sox2 again visibly suppressed such fluctuations —most likely due to co-condensation with DNA—resulting in a significantly decreased average fluctuation radius (**Figure 15H and I**). Altogether, these results demonstrate that Sox2 and DNA form co-condensates wherein proteins and DNA accumulate, reducing the amount of free DNA outside the condensates.

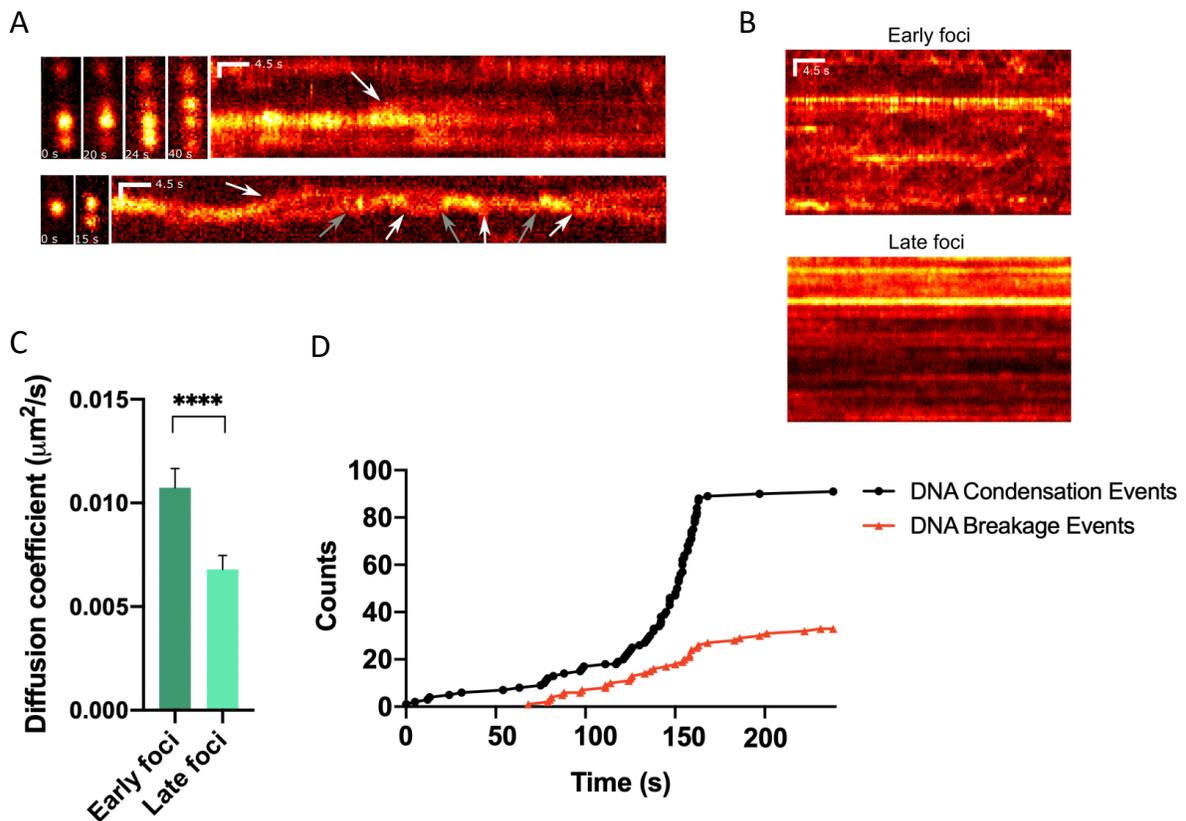


Figure 18. Behavior of Sox2 condensates on DNA at different time points.

A) Snapshots and kymographs of two representative double-tethered λ DNA molecules displaying fusion (gray arrows) and splitting (white arrows) of Cy5-Sox2 foci. Vertical scale bar denotes 0.5 μm .

B) Kymograph of a representative double-tethered λ DNA molecule displaying decreased Sox2 foci mobility over time. Early foci were recorded ~ 5 min after Sox2 injection, and late foci were recorded ~ 15 min after injection. Vertical scale bar denotes 1 μm .

C) Bar graph displaying the diffusion coefficient of early Sox2 foci ($n = 179$), recorded at ~ 5 mins after Sox2 injection, and late Sox2 foci ($n = 104$), recorded at ~ 15 mins after Sox2 injection, where n denotes the number of kymograph traces tracked. Error bars denote 95% CI. Significance test was calculated using an unpaired two-sample t tests (**** $P < 0.0001$).

D) Cumulative incidence of Sox2-mediated DNA condensation and breakage events as a function of time in a representative field of view.

3.4.2 Sox2:DNA co-condensation exerts mechanical stress on DNA

The loss of fluctuations in both single- and double-tethered DNA suggests that Sox2-induced condensation generates mechanical tension within the DNA. These effects were recapitulated using unlabeled wild-type Sox2, ruling out the possibility of labeling artifact (**Figure 19A and B**). In accordance with this notion but nonetheless unexpectedly, I observed that a significant population of double-tethered DNA underwent sudden breakage after losing slacks (**Figure 20A, B and Figure 19C**). The breakage was accompanied by a rapid collapse of the Sox2 and DNA fluorescence signals into the two tethered ends. Notably, these breakage events occurred over a time window that coincided with the formation of Sox2:DNA co-condensates and became much less frequent as the mobility of Sox2 foci decreased over time (**Figure 18B and D**). In contrast, virtually no DNA breakage was observed in the absence of Sox2 (**Figure 20B**), or for single-tethered λ DNA, where the tension can be released from the free end, even after the addition of Sox2 (**Figure 21A**).

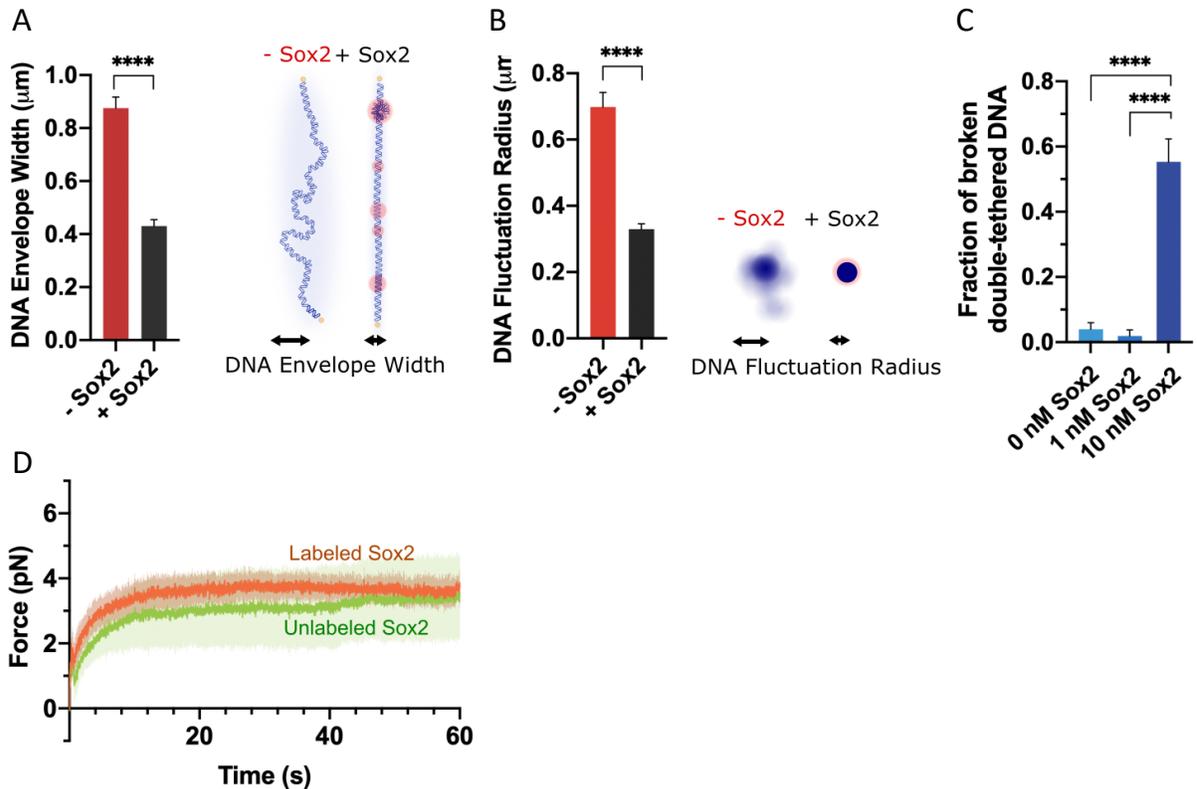


Figure 19. Evaluation of the mechanical effects of Sox2:DNA co-condensation using unlabeled Sox2.

A) Bar graph and cartoon showing the DNA envelope width of double-tethered DNA measured in the absence ($n = 147$) or presence of unlabeled Sox2 ($n = 32$). Error bars denote 95% CI. Significance test was calculated using an unpaired two-sample t tests (**** $P < 0.0001$).

B) Bar graph and cartoon showing the fluctuation radius of single-tethered DNA molecules in the absence ($n = 38$) or presence of unlabeled Sox2 ($n = 32$). Error bars denote 95% CI. Significance test was calculated using an unpaired two-sample t tests (**** $P < 0.0001$).

C) Fraction of double-tethered λ DNA molecules that broke after 15 min without Sox2 ($n = 251$), with 1 nM unlabeled Sox2 ($n = 131$), or with 10 nM unlabeled Sox2 ($n = 250$). Data are collected from at least three fields of view. Results shown in panels a-c are from TIRFM experiments. Error bars denote standard deviation. Significance test was calculated using a one-way ANOVA with Dunnett's test for multiple comparisons (**** $P < 0.0001$).

D) Force measurements as a function of time with Cy3-labeled Sox2 (orange, $n = 4$) or unlabeled Sox2 (green, $n = 5$) made by the optical tweezers assay. The shades correspond to standard deviation. The Sox2 concentration in these experiments was 100 nM.

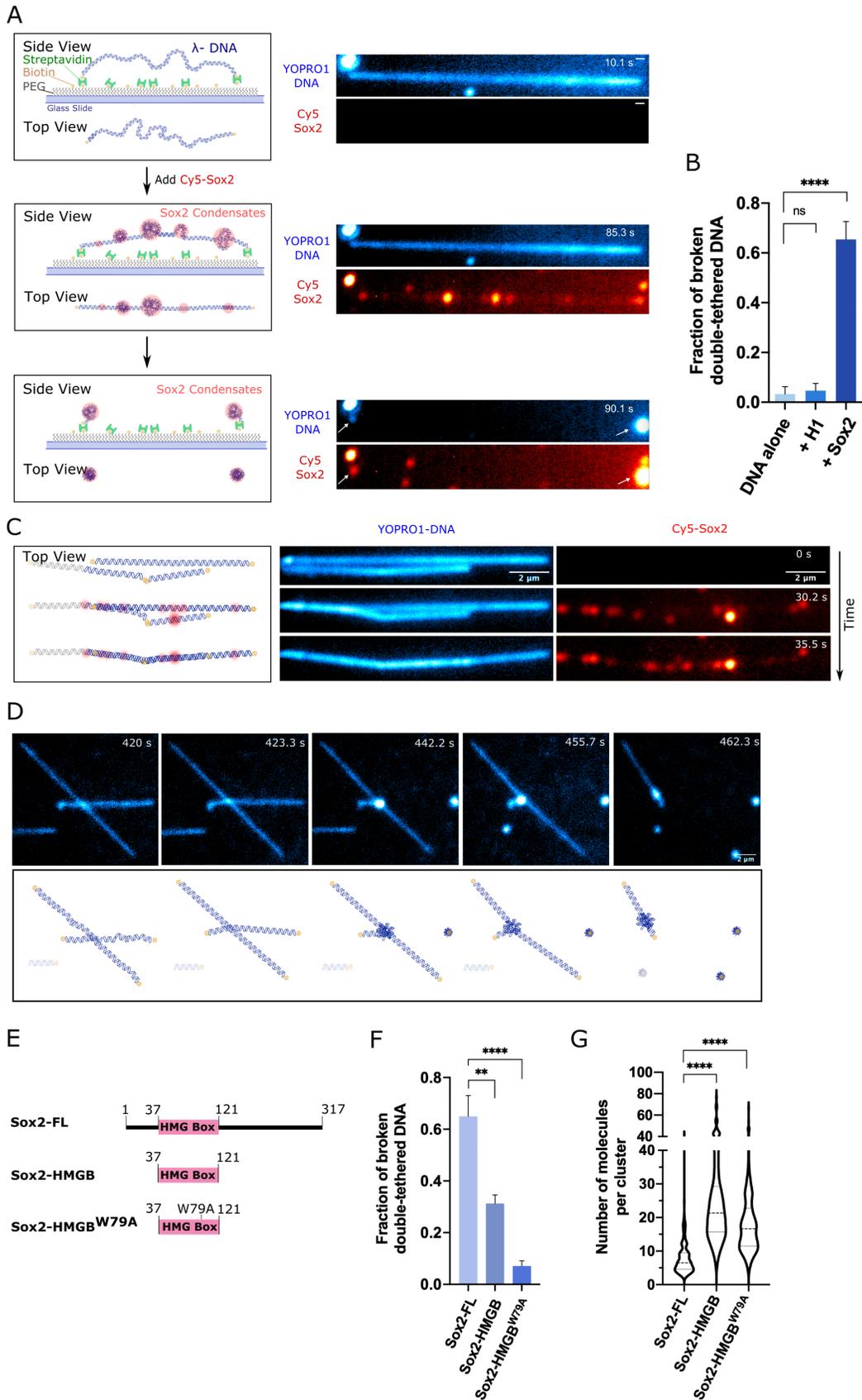


Figure 20. Sox2:DNA co-condensation exerts intra- and inter-strand mechanical stress.

A) Schematic (left) and time-lapse snapshots (right) showing Sox2 condensate formation on a double-tethered λ DNA and the subsequent breakage event upon which both DNA and Sox2 signals collapsed to the two tethered ends (white arrows) (among 4 independent experiments). Scale bar, 0.5 μ m.

B) Fraction of double-tethered λ DNA molecules that broke after 15 min without any protein ($n = 251$), with 10 nM H1 ($n = 150$), or with 10 nM Sox2 ($n = 379$). Error bars denote standard deviation. Data are collected from at least three fields of view. Significance was obtained using a one-way ANOVA with Dunnett's test for multiple comparisons (ns, $P = 0.9267$; **** $P < 0.0001$).

C) Schematic (left) and time-lapse snapshots (right) showing multiple adjacent DNA strands (among 4 independent experiments) being joined upon Sox2 condensate formation.

D) Time-lapse snapshots (top) and cartoon illustrations (bottom) showing a series of DNA breaking and joining events occurring among multiple λ DNA strands in the presence of Sox2.

E) Schematic of different Sox2 constructs used in this study.

F) Fraction of double-tethered λ DNA molecules that broke after 15 min of incubation with Sox2-FL ($n = 379$), Sox2-HMGB ($n = 357$), or Sox2-HMGB^{W79A} ($n = 297$). Error bars denote standard deviation. Significance was obtained using a one-way ANOVA with Dunnett's test for multiple comparisons (** $P < 0.001$; **** $P < 0.0001$).

G) Violin plot showing the distribution of the number of Sox2 molecules within each cluster for Sox2-FL ($n = 167$), Sox2-HMGB ($n = 168$), or Sox2-HMGB^{W79A} ($n = 155$), where n represents the number of clusters analyzed. Significance was obtained using a one-way ANOVA with Dunnett's test for multiple comparisons (**** $P < 0.0001$).

I then explored other factors besides condensation-induced tension that could contribute to the tether breakage. I found that the fraction of broken DNA tethers was not significantly affected by the duration of laser exposure in my experiments (**Figure 21B**). Given the known effect of DNA intercalating dyes on the mechanical properties of DNA,^{176,177} I washed out YOPRO1 prior to the addition of Sox2 and observed a lower fraction of ruptured DNA (**Figure 21C**). To evaluate whether nicks that inevitably exist in these λ DNA samples played a role in tether breakage, I treated the DNA with T4 ligase and observed fewer breakage events upon Sox2 condensate formation (**Figure 21D**). Based on these results, I speculate that the DNA breakage observed in the TIRFM experiments resulted from a combination of tension generated by Sox2:DNA co-condensation and mechanical instability in the DNA substrate due to nicks and the intercalating dye. Nonetheless, the breakage fraction is still a useful proxy for the magnitude of mechanical tension to compare different proteins and substrates if the same imaging conditions and DNA batch are used.

To examine whether other DNA-binding proteins can exert the same level of tension on DNA, I repeated the above TIRFM assay with another abundant nuclear protein, the human linker

histone H1.4 (referred to as H1 hereafter). H1 is known to form co-condensates with DNA.^{30,51,52} I found that H1:DNA co-condensation also reduced the double-tethered DNA envelope width and single-tethered DNA fluctuation radius (**Figure 22**). However, H1-mediated DNA condensation resulted in much fewer DNA breakage events compared to Sox2-mediated condensation (**Figure 20B**), suggesting that H1 generates a lower force on DNA. I then sought to examine whether Sox2:DNA co-condensation can generate inter-strand tension. In the absence of Sox2, the neighboring λ DNA strands immobilized in proximity of each other fluctuated independently. Upon the addition of Sox2, these strands lost slack and joined one another through the fusion of Sox2 foci (**Figure 20C**). In some cases, I observed successive severing and joining of DNA located nearby (**Figure 20D**). Together, these results suggest that Sox2 condensates exert force on DNA both within the same strand (when both ends are anchored) and between multiple strands.

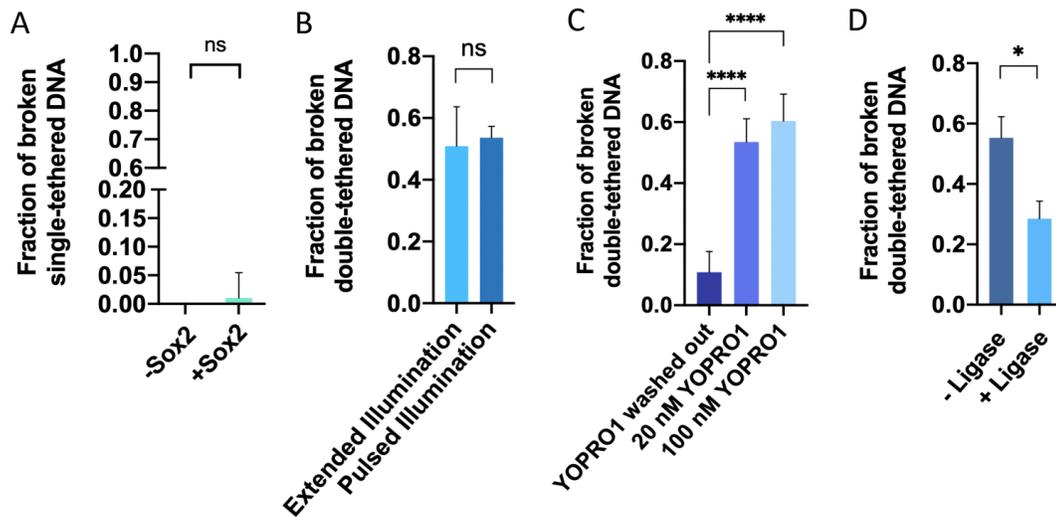


Figure 21. Evaluation of Sox2-mediated DNA breakage under different TIRFM experimental conditions.

A) Fraction of single-tethered λ DNA molecules that broke after 15 min of imaging in the absence of Sox2 ($n = 165$) or in the presence of Sox2 ($n = 306$). Error bars denote standard deviation. Significance test was calculated using an unpaired two-sample t tests (ns $P=0.37$).

B) Fraction of double-tethered λ DNA molecules in the presence of Sox2 that broke after 15 min under different laser illumination schemes. Under “Extended Illumination”, 7.5 min of continuous 488-nm laser illumination was applied ($n = 55$). Under “Pulsed Illumination”, a single 300-ms pulse of 488-nm laser at the same power was applied ($n = 703$). Error bars denote standard deviation. Significance test was calculated using an unpaired two-sample t tests (ns $P=0.92$).

C) Fraction of double-tethered λ DNA molecules that broke after 15 min of incubation with Sox2 in the absence of YOPRO1 (washed out with 500 μ L of buffer after initial imaging of the DNA) ($n = 161$), or in the presence of 20 nM YOPRO1 ($n = 157$) or 100 nM YOPRO1 ($n = 106$). Error bars denote standard deviation. Significance test was calculated using a one-way ANOVA with Dunnett’s test for multiple comparisons (**** $P<0.0001$).

D) Fraction of double-tethered λ DNA molecules without ($n = 250$) or with a T4 ligase pre-treatment ($n = 184$) that broke after 15 min of incubation with Sox2. Error bars denote standard deviation. Significance test was calculated using an unpaired two-sample t tests (** $P=0.0057$).

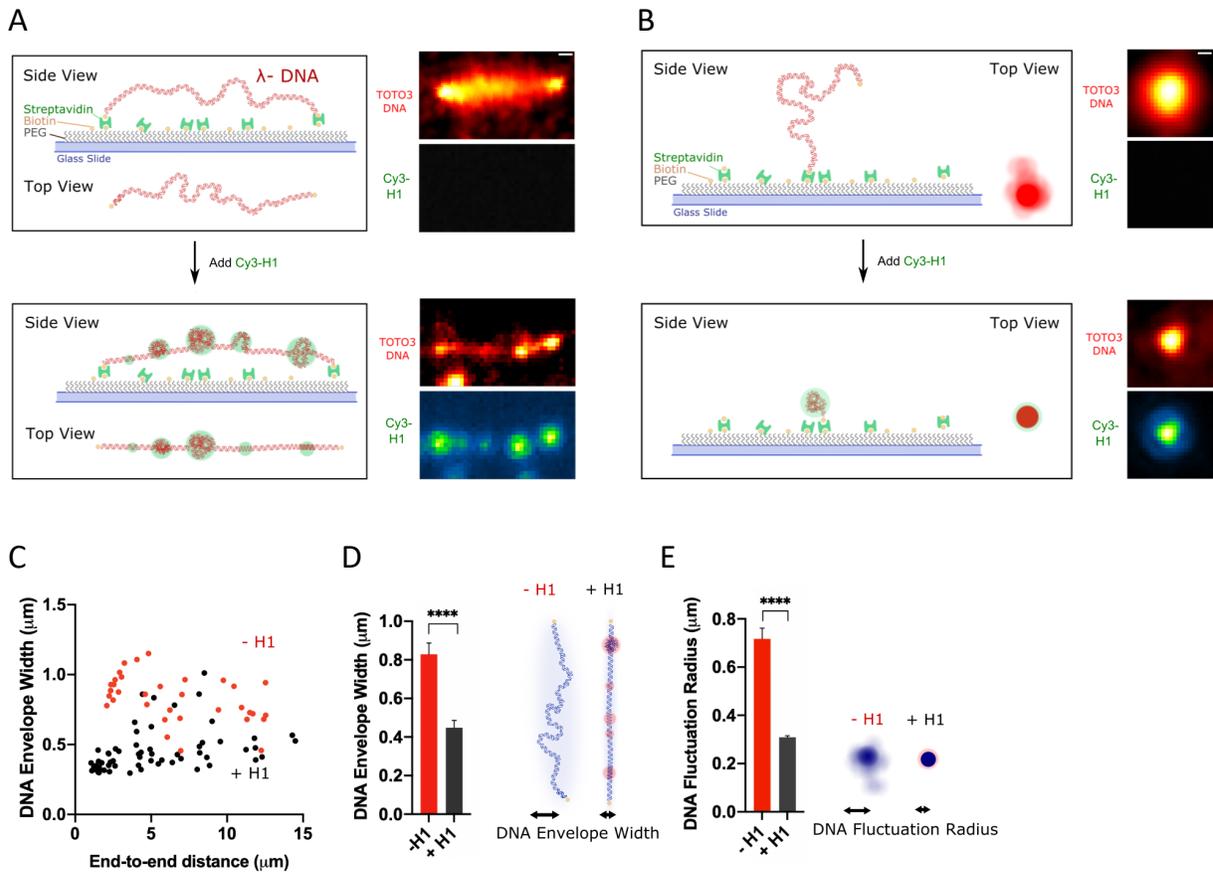


Figure 22. DNA condensation by linker histone H1

A) Schematic (left) and time-averaged projection (right) of a double-tethered λ DNA molecule (among 3 independent experiments) stained with 30 nM of the TOTO3 dye and incubated with 150 pM of Cy3-labeled H1. Scale bar, 0.5 μm .

B) Schematic (left) and time-averaged projection (right) of a single-tethered λ DNA molecule (among 3 independent experiments) stained with TOTO3 and incubated with Cy3-H1. Scale bar, 0.5 μm .

C) Double-tethered DNA envelope width as a function of the end-to-end distance measured in the absence ($n = 35$) or presence of H1 ($n = 63$).

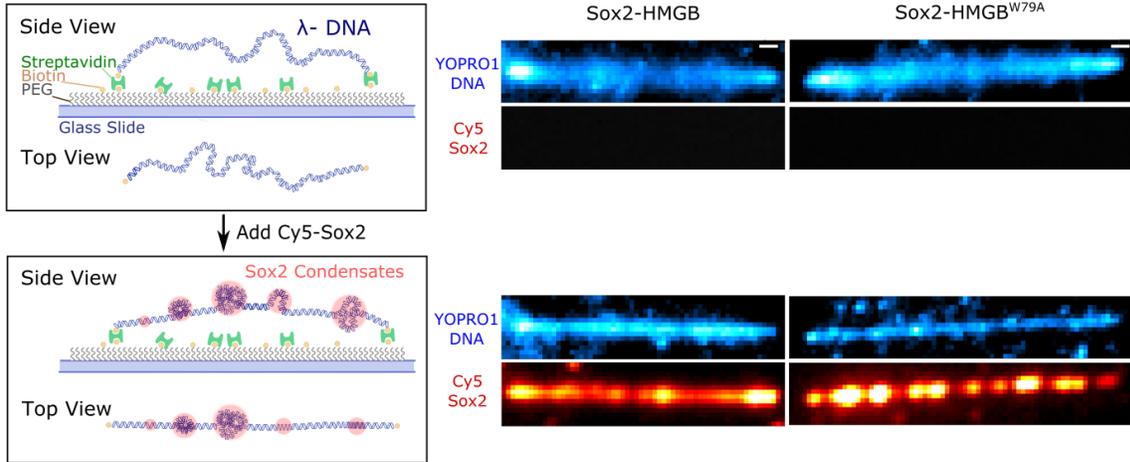
D) Bar graph and cartoon showing a reduction in the average DNA envelope width of double-tethered λ DNA molecules in c upon H1-mediated DNA condensation. Error bars denote 95% CI. Significance test was calculated using an unpaired two-sample t tests (**** $P < 0.0001$).

E) Bar graph and cartoon showing the fluctuation radius of single-tethered DNA molecules in the absence ($n = 34$) or presence of H1 ($n = 33$). Error bars denote 95% CI. Significance test was calculated using an unpaired two-sample t tests (**** $P < 0.0001$).

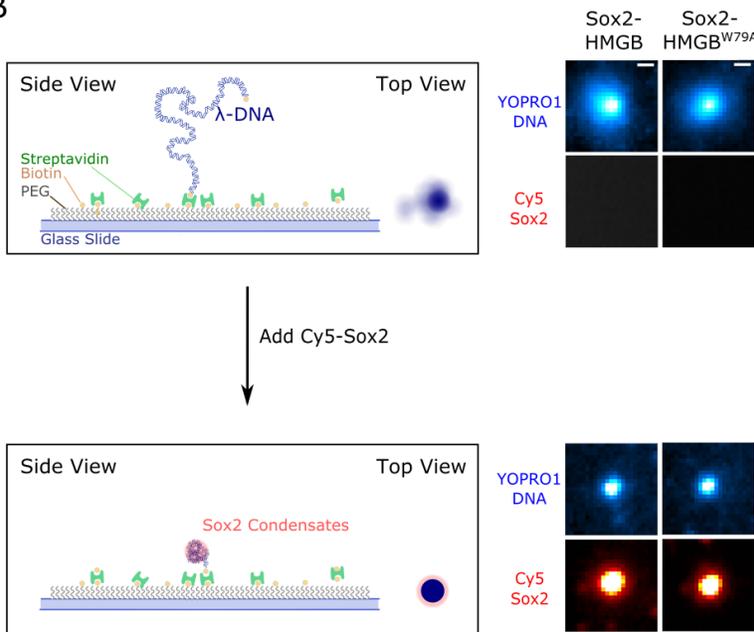
3.4.3 IDRs of Sox2 are dispensable for condensate formation but required for force exertion

Sox2 contains N- and C-terminal IDRs flanking the globular DNA-binding HMGB domain.¹⁷⁸ To gain insight into the molecular mechanism that underlies the capacity of Sox2 to form co-condensates with DNA, I generated and fluorescently labeled Sox2 truncations (**Figure 20E** and **Figure 16**). I first examined a Sox2 construct that contains only its HMGB domain without the IDRs (Sox2-HMGB). Somewhat unexpectedly, similar to the full-length Sox2 (Sox2-FL), Sox2-HMGB also formed foci on λ DNA strands—both doubly and singly tethered—along with a concomitant loss of DNA fluctuations (**Figure 23**). This observation indicates that the IDRs of Sox2 are not required for its co-condensation with DNA. However, Sox2-HMGB took a much longer time to form the same amount of DNA condensation (measured through the loss of fluctuations) ($T_{\text{condense}} = 184 \pm 45$ s) compared to Sox2-FL ($T_{\text{condense}} = 30 \pm 4$ s). Sox2-HMGB:DNA co-condensation also resulted in significantly fewer DNA breakage events (**Figure 20F**). I next introduced a single-residue mutation W79A to Sox2-HMGB, which was previously shown markedly impair DNA binding by Sox2,¹⁰¹ generating Sox2-HMGB^{W79A}. Consistent with previous results,¹⁰¹ Sox2-HMGB^{W79A} displayed diminished DNA binding activity (**Figure 24**). Nevertheless, it still retained the ability to form co-condensates with DNA (**Figure 23**), albeit with much slower condensation kinetics ($T_{\text{condense}} = 251 \pm 59$ s) compared to Sox2-FL. This point mutation further attenuates the condensation-dependent mechanical tension exerted on DNA (quantified by the fraction of broken double-tethered DNA) compared to both Sox2-FL and Sox2-HMGB (**Figure 20F**). Notably, the mechanical effect of Sox2:DNA co-condensates is not directly correlated with their size, as both Sox2-HMGB and Sox2-HMGB^{W79A} foci on average contained more Sox2 molecules—estimated from the brightness of the fluorescent foci—than Sox2-FL foci (**Figure 20G**). Together, these results demonstrate that the HMGB domain alone can mediate Sox2:DNA co-condensation, but the high mechanical stress on DNA is largely driven by the IDRs of Sox2.

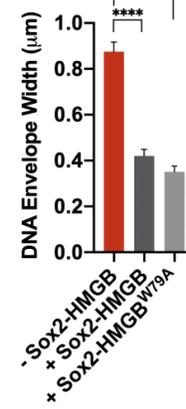
A



B



C



D

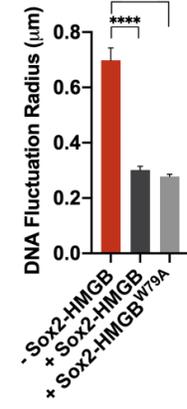


Figure 23. DNA condensation mediated by Sox2-HMGB constructs.

A) Schematic and time-averaged projections of double-tethered λ DNA stained with YO-PRO1 (among 3 independent experiments) and incubated with 10 nM of Cy5-labeled Sox2-HMGB or Sox2-HMGB^{W79A}. Scale bar, 0.5 μ m.

B) Schematic and time-averaged projections of single-tethered λ DNA stained with YO-PRO1 (among 3 independent experiments) and incubated with Cy5-labeled Sox2-HMGB or Sox2-HMGB^{W79A}. Scale bar, 0.5 μ m.

C) Bar graph showing the DNA envelope width of double-tethered λ DNA molecules without Sox2-HMGB ($n = 147$), or after incubation with Sox2-HMGB ($n = 34$) or with Sox2-HMGB^{W79A} ($n = 36$). Error bars denote 95% CI. Significance test was calculated using a one-way ANOVA with Dunnett's test for multiple comparisons (**** $P < 0.0001$).

D) Bar graph showing the fluctuation radius of single-tethered λ DNA molecules without Sox2-HMGB ($n = 38$), or after incubation with Sox2-HMGB ($n = 33$) or with Sox2-HMGB^{W79A} ($n = 32$). Error bars denote 95% CI. Significance test was calculated using a one-way ANOVA with Dunnett's test for multiple comparisons (**** $P < 0.0001$).

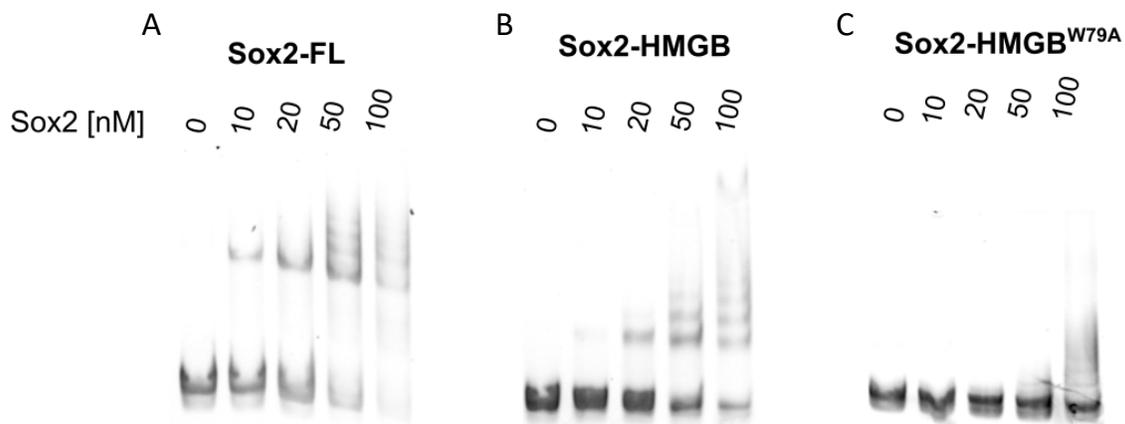


Figure 24. EMSA for the binding of different Sox2 constructs to DNA.

A) SYBR-stained gel results for full-length Sox2 (Sox2-FL) binding to a 233-bp DNA that contains a Sox2 binding motif (CTTTGTT).

B) Gel results for Sox2-HMGB binding to the same DNA substrate as in A.

C) Gel results for Sox2-HMGB^{W79A} binding to the same DNA substrate as in A.

3.4.4 Quantification of the force generated by Sox2:DNA co-condensation

In collaboration with Jeremy Chang.

Next, we sought to quantitatively measure the force exerted by Sox2:DNA co-condensates on the DNA strand. Using optical tweezers combined with scanning confocal microscopy,⁹⁶ we tethered a single λ DNA molecule between two optically trapped beads, moved the tether in its relaxed form (i.e. zero applied force) to a channel containing Cy3-labeled Sox2, and monitored the force on DNA as a function of time. We first conducted experiments in a passive mode by keeping the trap positions fixed (**Figure 25A**). As Sox2 foci appeared and accumulated on the DNA tether, the force reading concurrently increased. Both fluorescence and force values reached a plateau after 10-20 sec (**Figure 25B**). Force generation requires the presence of Sox2, and the force plateau level is dependent on the concentration of Sox2 in solution (**Figure 25C**) with a maximum value of ~ 7 pN measured at the highest Sox2 concentration tested (500 nM).

We then conducted force-clamp experiments in which the tethered DNA was incubated with Sox2 at a constant force by adjusting the trap separation (**Figure 25D**). We observed that, with a force clamp set at 0.5 pN, Sox2 and DNA underwent continued condensation, reducing the length of free DNA and bringing the two beads closer to each other (**Figure 25E**). In contrast, a 10-pN force clamp largely abolished the condensation process (**Figure 25F**), consistent with the above passive-mode results reporting a 7-pN maximum force that Sox2:DNA co-condensation can generate.

Next, we asked how much force is required to dissolve Sox2:DNA co-condensates. To address this question, we first formed Sox2 foci on a DNA tether under low force (0.5 pN) and then gradually pulled the two beads apart, thereby increasing the force applied to the tether (**Figure 25G**). From the resultant force-extension curve, we found that the extension of a Sox2-bound tether was much shorter than that of a bare DNA tether, indicating significant DNA accumulation inside the condensates (**Figure 25H**). Some transitions were observed in the force-extension curve, which likely correspond to force-induced condensate dissolution events (**Figure 25H**). Nonetheless, a significant fraction of condensates persisted even when the force reached the DNA overstretching regime (~ 65 pN), as reflected by the shorter extension at high forces compared to the bare DNA (**Figure 25H**). Concomitant fluorescence imaging confirmed that the Sox2 foci remained intact during pulling (**Figure 25I**). These results demonstrate that Sox2:DNA co-condensates are stable and resistant to high disruptive forces.

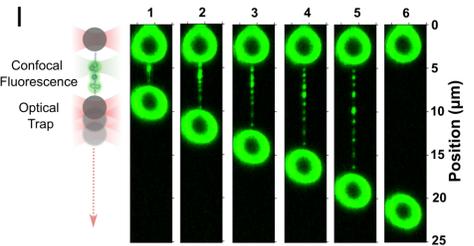
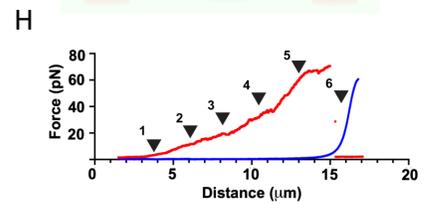
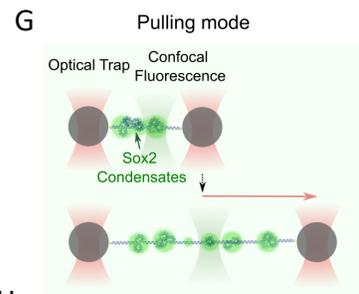
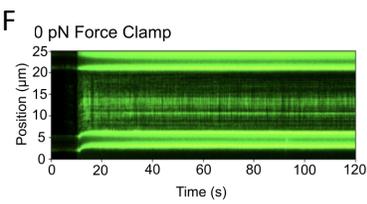
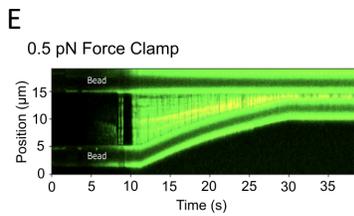
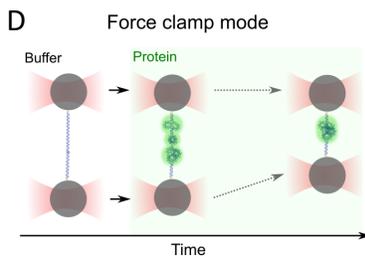
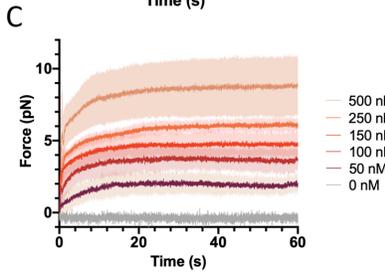
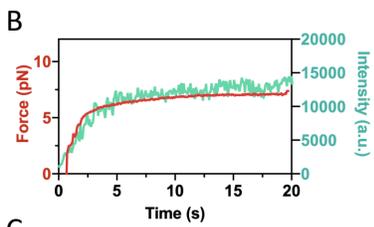
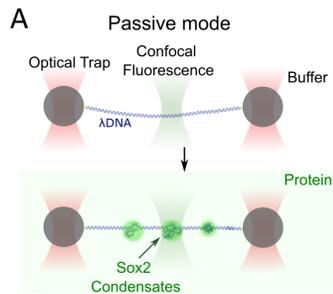


Figure 25. Optical tweezers assays for quantitative measurements of the force generated by Sox2:DNA co-condensation.

- A) Schematic of the optical tweezers assay that measures forces generated by Sox2-mediated DNA condensation. The trap positions were held fixed in this experiment and moved from a buffer to protein channel.
- B) Force measurements (red) and the corresponding fluorescence intensities (green) as a function of time for the assay depicted in A). Data are averaged from 4 representative tethers.
- C) Force measurements made with different Sox2 concentrations. The colored lines correspond to the mean force as a function of time averaged from multiple DNA tethers ($n = 4$ or 5). The shades correspond to standard deviation.
- D) Schematic of force-clamp experiments. The force applied to the tether was kept at a constant value via feedback such that DNA condensation would result in shortening of the tether.
- E) A representative kymograph showing significant tether contraction and Sox2 condensate formation under a 0.5-pN force clamp.
- F) A representative kymograph showing suppressed tether contraction under a 10-pN force clamp.
- G) Schematic of pulling experiments. Sox2 condensates first formed on tethered λ DNA under a low force (~ 0.5 pN). The tether was then subjected to mechanical pulling by gradually separating the two traps apart.
- H) A representative force-distance curve from pulling a λ DNA tether harboring Sox2 condensates (red) in comparison to a representative curve from pulling a bare λ DNA (blue). The black arrowheads denote selected time points imaged in I).
- I) Two-dimensional fluorescence scan of the same tether (red) as in H) at selected time points during pulling showing that Sox2 condensates persisted under forces up to 60 pN (time points #1-5) until tether rupture (time point #6).

3.4.5 Nucleosomes attenuate the mechanical stress that Sox2 condensation exerts on DNA

Given that Sox2 is a nucleosome-binding pioneer TF,¹⁴⁹ I asked how the mechanical stress exerted by Sox2:DNA co-condensation on DNA may be regulated by nucleosome wrapping and chromatin organization. To this end, I loaded histone octamers containing Cy3-labeled H2B onto surface-immobilized λ DNA in the TIRFM setup (**Figure 26A**) and then added Cy5-Sox2 to bind the nucleosomal DNA (**Figure 26B**). As expected, I observed that Sox2 foci nucleate around nucleosome locations (**Figure 26C and D**). Sox2 foci preferentially colocalized with nucleosomes over bare DNA sites (**Figure 26E and F**). The majority of Sox2:nucleosome foci contained multiple Sox2 molecules based on the Cy5 fluorescence intensity, similar to those on bare DNA (**Figure 27A**).

Strikingly, I detected drastically fewer DNA breakage events upon the formation of Sox2 foci on nucleosomal DNA than on bare DNA (**Figure 26G**). In the few examples in which nucleosomal DNA breakage did occur, the tether ruptured at one of the anchor positions, and the full DNA contour was sustained and underwent rigid-body-like fluctuations (**Figure 27B**). This is in contrast to the breakage events observed on bare DNA where the tether broke in the middle and the Sox2/DNA signals abruptly collapsed into the two anchor positions (**Figure 20A**). I also analyzed the single-tethered nucleosomal λ DNA molecules and found that the addition of Sox2 did not significantly suppress their fluctuating motions (**Figure 26H-J**), again in contrast to the bare DNA results (**Figure 15I**).

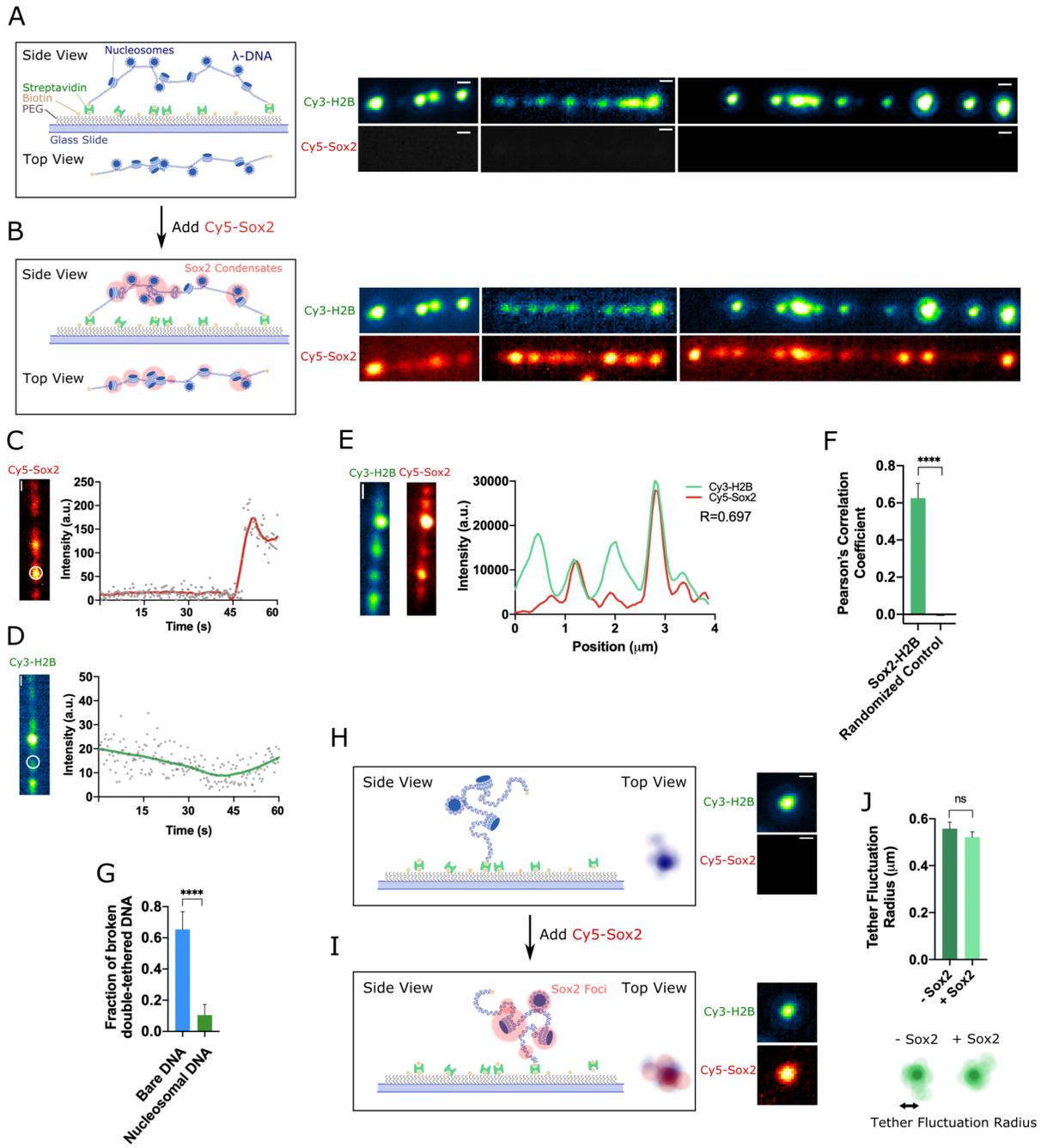


Figure 26. Nucleosomes colocalize with Sox2 condensates and attenuate their mechanical effects on DNA.

A) (Left) Schematic of double-tethered λ DNA loaded with nucleosomes. (Right) Time-averaged projections of three double-tethered nucleosomal DNA molecules with different end-to-end distances (among 5 independent experiments). Nucleosomes were visualized by Cy3-labeled histone H2B. Scale bar, 0.5 μ m.

B) Schematic and time-averaged projections of the same three nucleosomal DNA molecules as in a (among 5 independent experiments) when incubated with 10 nM Cy5-labeled Sox2. Scale bar, 0.5 μ m.

C) Real-time tracking of Cy5-Sox2 intensities at a nucleosome position (circled region) on a double-tethered nucleosomal λ DNA. Scale bar, 0.5 μ m.

D) Corresponding Cy3-H2B intensities within the same circled region as in c. Scale bar, 0.5 μ m.

E) (Left) Snapshot of a double-tethered DNA harboring Cy3-H2B nucleosomes and Cy5-Sox2 condensates. (Right) Intensity profiles of Cy3-H2B (green) and Cy5-Sox2 (red) along the length of the same DNA molecule. R value represents Pearson's correlation coefficient. Scale bar, 0.5 μ m.

F) Pearson's correlation coefficients averaged from all aligned Cy3-H2B and Cy5-Sox2 intensity profiles and from Costes' randomized control ($n = 158$). Error bars denote 95% CI. Significance was obtained using an unpaired two-sample t test (**** $P < 0.0001$).

G) Fraction of double-tethered bare DNA ($n = 379$) versus nucleosomal DNA molecules ($n = 303$) that broke after 15 min of incubation with 10 nM Sox2. Data are averaged from at least three fields of view. Error bars denote standard deviation. Significance was obtained using an unpaired two-sample t test (**** $P < 0.0001$).

H) Schematic and time-averaged projection of a single-tethered nucleosomal λ DNA visualized by Cy3-H2B fluorescence. Scale bar, 0.5 μ m.

I) Schematic and time-averaged projection of the same nucleosomal DNA molecule as in h when incubated with 10 nM Cy5-Sox2. Scale bar, 0.5 μ m.

J) Average fluctuation radius of single-tethered nucleosomal λ DNA in the absence or presence of Sox2 ($n = 42$). Error bars denote 95% CI. Significance was obtained using an unpaired two-sample t test (ns, $P = 0.28$).

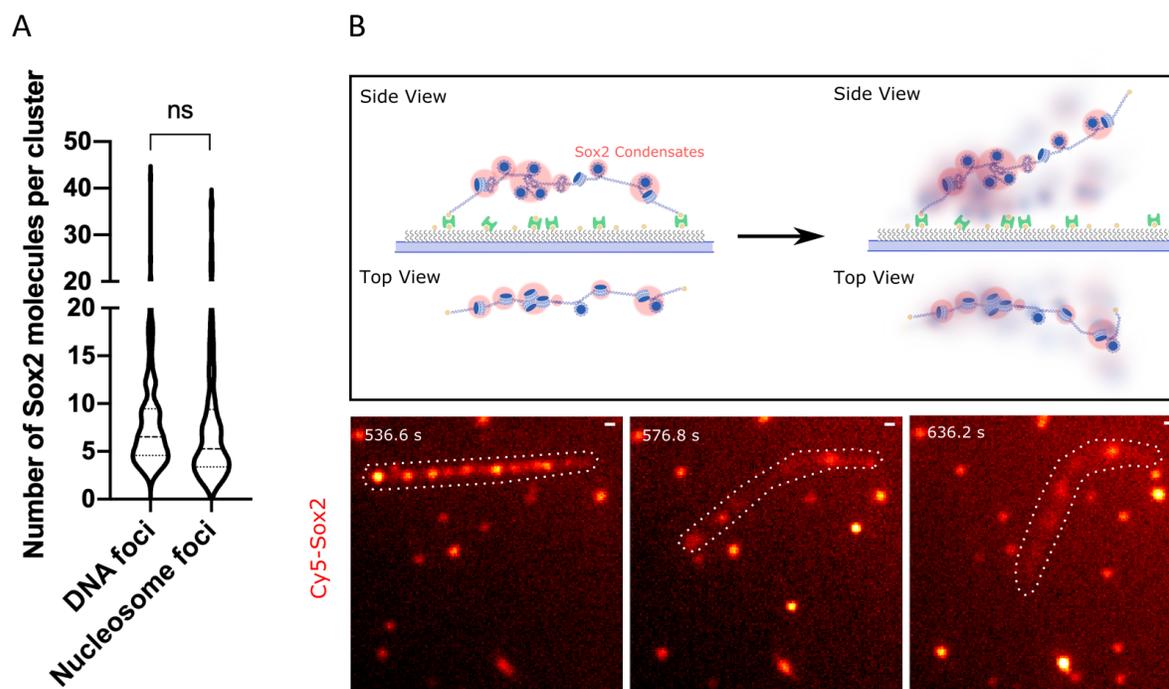


Figure 27. Sox2 binding and condensation on nucleosomal DNA.

A) Violin plot showing the distribution of the number of Sox2 molecules within bare DNA foci ($n = 167$) and the distribution within nucleosome foci ($n = 150$), where n represents the number of foci analyzed. Significance test was calculated using an unpaired two-sample t tests (ns $P=0.96$).

B) Schematic (top) and snapshots (bottom) showing Sox2 condensates on a double-tethered nucleosomal DNA molecule. The nucleosomal DNA contour (dashed line) remained in an extended configuration and, when detaching from one anchored end, underwent rigid-body-like fluctuations. Scale bar, $0.5 \mu\text{m}$.

3.4.6 Quantification of force generated by Sox2 condensates on nucleosomes

In collaboration with Jeremy Chang.

We performed optical tweezers experiments to directly measure the force that Sox2 condensates exert on nucleosomal DNA. We assembled histone octamers containing AlexaFluor488-labeled H2A onto a λ DNA tether, moved the tether to a channel containing Cy3-Sox2, and monitored the force reading in the passive mode (**Figure 28A**). Satisfyingly, we observed that, even though the Sox2 foci predominantly colocalized with nucleosomes (**Figure 28B**), their formation hardly caused any increase in force, in contrast to the significant force increase on bare DNA (**Figure 28C**). These results corroborate the above TIRFM data, together suggesting that nucleosomes attenuate the force exerted by Sox2 condensates on DNA through colocalization with Sox2.

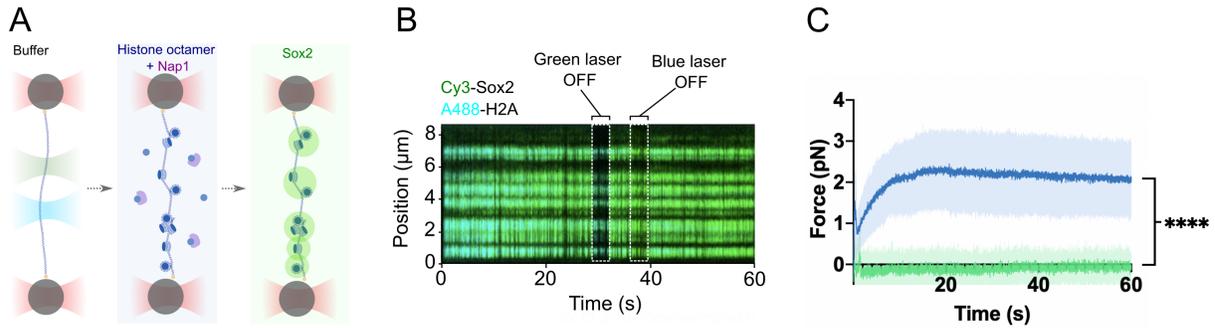


Figure 28. Nucleosomes attenuate the condensation force exerted by Sox2 on DNA.

A) Schematic of *in situ* nucleosome assembly and Sox2 condensate formation on a λ DNA tether in an optical tweezers assay.

B) A representative kymograph showing the colocalization of Sox2 condensates (green) with nucleosomes (cyan) on a λ DNA tether.

C) Force measurements on nucleosomal DNA as a function of time (green line) for the assay depicted in A). Data are averaged from 7 representative tethers. Force measurements on bare DNA tethers are shown in blue (averaged from 14 tethers). The dark colored lines correspond to average force trajectories. The shades correspond to standard deviation. The Sox2 concentration in these experiments was 75 nM. Significance was obtained using an unpaired two-sample *t* test (**** $P < 0.0001$).

3.4.7 Linker histone suppresses Sox2-induced stress on DNA through colocalization

Finally, I examined how linker histone H1 may affect the mechanical effect of Sox2:DNA co-condensation. I then added both Cy3-H1 and Cy5-Sox2 to the DNA and curiously observed a high degree of colocalization of H1 and Sox2 signals (**Figure 29A and B**). Strikingly, the presence of H1 in the Sox2 foci dramatically decreased the occurrence of DNA breakage events on double-tethered DNA compared to Sox2 alone (**Figure 29C**). This result suggests that, like the nucleosome core particle, linker histone H1 also attenuates the mechanical effects of Sox2:DNA co-condensates on DNA through colocalization with Sox2.

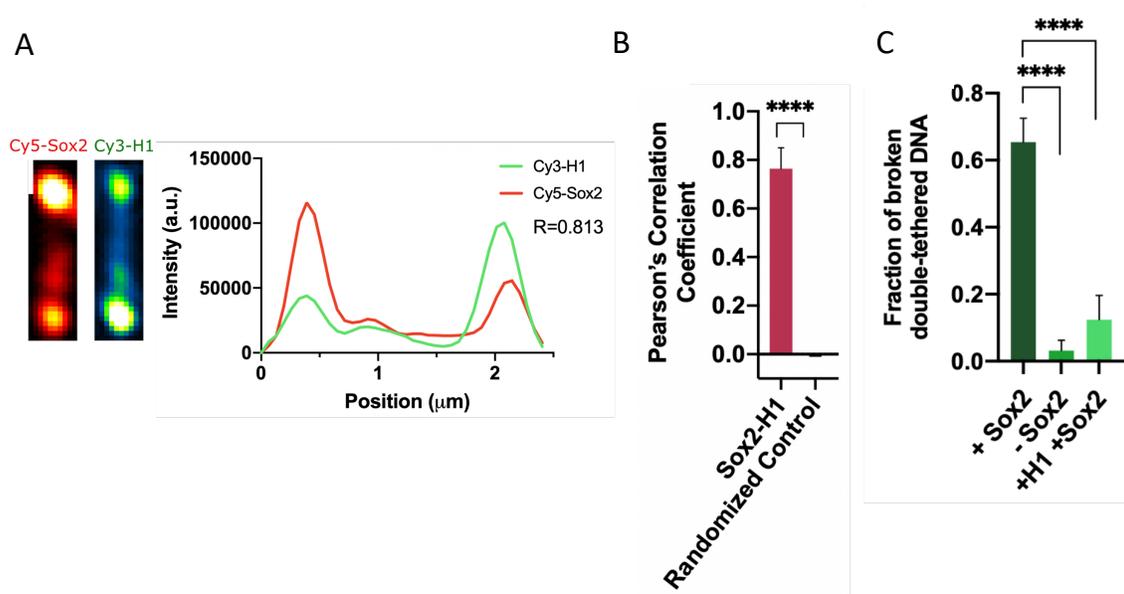


Figure 29. H1 suppresses Sox2-induced stress on DNA through colocalization.

A) (Left) Snapshot of a double tethered DNA harboring Cy5-Sox2 condensates and Cy3-H1 condensates. (Right) Intensity profiles of Cy5-Sox2 (red) and Cy3-H1 (green) along the length of the same DNA molecule. R value represents Pearson's correlation coefficient. Scale bar, $0.5 \mu\text{m}$.

B) Pearson's correlation coefficients averaged from all aligned Cy3-H2B and Cy5-Sox2 intensity profiles and from Costes' randomized control ($n = 83$). Error bars denote 95% CI. Significance was obtained using an unpaired two-sample t test (**** $P < 0.0001$).

C) Fraction of double-tethered λ DNA molecules that broke after 15 min with 10 nM Sox2 ($n = 379$), without any protein ($n = 251$), with 100 pM H1 and 10 nM Sox2 ($n = 54$), or. Error bars denote standard deviation. Significance was obtained using a one-way ANOVA with Dunnett's test for multiple comparisons (**** $P < 0.0001$).

3.5 Conclusion

Together, these results reveal that formation of protein:DNA co-condensates can generate high forces, up to 7 pN, similar to other cellular forces.¹⁷⁹ Moreover, I show that Sox2 condensates are mechanically stable, being able to resist high forces. I further show that IDRs are dispensable for condensate formation, but not required for force-generation per se. Lastly, I demonstrate that nucleosomes and linker histone H1 attenuate the force-generation capacity by Sox2:DNA co-condensates, suggesting a novel role of the chromatin architecture as a mechanical sink. I further discuss these findings and future directions in Chapters 5.

CHAPTER 4. Single-molecule visualization of DNA compaction by the SMC5/6 complex

4.1 SMC5/6: the yet-to-be named complex of the SMC complex family.

The SMC complexes mediate important processes in genome organization. In eukaryotes, four SMC complexes have been characterized so far: condensin, cohesin, the dosage compensation complex, and the SMC5/6 complex.⁵⁹ In contrast its counterparts, SMC5/6 complex has not been shown to directly modulate the chromosome structure and organization.^{180,181} Rather, the SMC5/6 complex appears to be more strongly implicated in DNA repair and replication processes.⁶⁴ Recent studies, as discussed below, have uncovered new surprising functional roles for SMC5/6 (see Chapter 4.1.3) within the context of transcription¹⁸² and developmental regulation.¹⁸³

Structurally, the SMC5/6 complex deviates from its counterparts through variances in the SMC5 and SMC6 arm region, which doesn't fold back on themselves.¹⁸⁴ Importantly, the complex is distinguished by a set up six accessory subunits, aptly named non-SMC elements (Nse) 1-6 (**Figure 30**). Curiously, Nse subunits belong to the winged helix domain (WHD) protein family, which more closely resembles the regulatory subunits of many prokaryotic SMCs than cohesin and condensin.¹⁸⁵ In particular, Nse1 has been recently shown to contain ubiquitin ligase activity towards Nse4 kleisin subunit.¹⁸⁶ Nse2 displays a SUMO (Small Ubiquitin-like modifier) E3 ligase activity on many characterized SMC5/6 substrates and subunits.¹⁸⁷ The Nse3 subunit has recently been reported to display dsDNA binding activity,¹⁸⁸ and Nse4 forms the connecting kleisin core subunit and is thought to regulate ATP binding.¹⁸⁹ Lastly, Nse5/6 has been shown to function as negative regulator of SMC5/6's ATPase and loop extrusion activity.¹⁹⁰⁻¹⁹² Together, Nse1, 3, 4-6 are thought to mediate the scaffolding function of the SMC5/6 complex that has been not been well characterized.⁶⁴

In comparison to cohesin and condensin, the SMC5/6 complex displays a lack of elbow bending and a much shorter coiled-coil arm region, which makes extensive contacts with its unique subunits Nse2 and Nse6.¹⁸⁴ The Nse5/6 subunits in SMC5/6, in contrast to the HEAT repeating units in cohesin and condensin, don't display DNA-binding activity, suggesting a potentially distinct molecular function. Although displaying a fundamentally similar DNA-clamping mechanism to its counterparts, SMC5/6 displays subtle differences in DNA-binding activity.¹⁹³ Importantly, how the structural variations contribute to SMC5/6 complex's functional divergence are not well understood. Below I discuss the known functional roles of SMC5/6 complex, its underlying molecular mechanism, and open questions in the field that are relevant to this chapter.

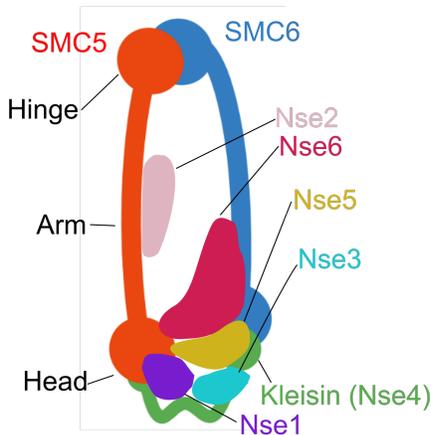


Figure 30. Schematics of the SMC5/6 complex.

Cartoon depiction of the SMC5/6 complex and its accessory subunits.

4.1.1 SMC5/6 in DNA repair

The most well-known functional role of SMC5/6 is its involvement in the repair of DNA double-strand breaks (DSB) via promoting homologous recombination (HR) between sister chromatids^{64,183,194} (**Figure 31A**). Originally discovered via genetic screens for mutations that sensitize *S. pombe* to genotoxic stress,¹⁹⁵ SMC5/6 complex was later implicated in HR repair in multiple organisms.¹⁸³ In support of this notion, SMC5/6 complex inactivation was shown to cause defect in HR repair via reducing sister-chromatid recombination.¹⁹⁶ Mechanistically, early studies in multiple organisms, using ChIP, demonstrated that the SMC5/6 complex is recruited to DSB regions to promote sister chromatid recombination along with cohesin.¹⁹⁴ Mounting evidence suggests a functional cooperativity between the SMC5/6 complex and cohesin in HR-repair: 1) epistasis analysis of their mutants suggest that both complexes work in the same pathway,^{197,198} 2) SMC5/6 complex was shown to directly recruit cohesin to DSB sites via its Nse2 subunit's SUMOylation activity.¹⁹⁷

Several working hypotheses delineate the SMC5/6's mechanistic contribution towards HR repair. Apart from its scaffolding function, the SMC5/6 complex is hypothesized to align the two DNA strands along with cohesin to promote HR.^{194,199} Additionally, the SUMOylation activity, mediated by its Nse2 subunit of SMC5/6 complex, may enhance the stability of cohesin and coordinate the recruitment of DNA repair factors.¹⁹⁴ Whether this model holds true is an open question. Moreover, *in vitro* work suggests that the SMC5/6's ATPase activity is coupled to its SUMOylation activity,²⁰⁰ but the underlying molecular mechanism is not well understood.

In addition, the SMC5/6 complex has been demonstrated to negatively regulate HR in repetitive sequences via its SUMOylation activity, particularly in rDNA, centromeres, and telomeres.¹⁹⁴ For example, SUMOylation of multiple telomeric proteins such as TRF1-2 and RAP1 appears to be important for telomere maintenance.²⁰¹ Despite SMC5/6's extensive

implication in HR-repair and its association with the Nse2's SUMOylation activity, these observations don't fully explain the ring-like architecture or the ATPase activity of SMC5/6.

4.1.2 SMC5/6 in DNA replication

In addition to its role in promoting HR-repair, the SMC5/6 complex is important for maintaining replication fork (RF) stability and removal of toxic DNA structures associated with replication stress (**Figure 31C**). In cells that replicate in the presence of DNA damage agents, HR-recombination intermediates accumulate in the form of an X-shaped DNA structure, which can lead to stalled RF.¹⁹⁴ Early genetic studies have shown that SMC5/6 mutants are synthetically lethal with mutants in genes promoting recovery of stalled RF, suggesting its involvement in the pathway.^{196,202} In support of this view, further studies in *S. cerevisiae* showed that SMC5/6 localizes to collapsed RF, and mutants abrogating SMC5/6 activity resulted in the toxic X-shaped recombination intermediate structure.^{202,203} Further studies implicated the role of SMC5/6 SUMO ligase activity, mediated by its Nse2 subunit, in resolving recombination intermediates.²⁰⁴ In human, mutations in the Nse2 subunit have been shown to lead to genetic disorders characterized by increased chromosomal fragility.²⁰⁵

Despite its functional importance, the SUMO ligase activity of Nse2 subunit appears to be not essential for cell viability.²⁰⁶ Recent studies have revealed a role for SMC5/6's scaffolding function in coordinating the RecQ helicase complex Sgs-Top3-Rmi1 (STR) in resolving recombination intermediate structures.^{64,207} Thus, an emerging view of SMC5/6 complex's function involves its appropriate balance between the accumulation and removal of HR-recombination intermediates via exploiting the scaffolding and SUMO ligase function. Given this, it is not clear how SMC5/6's intrinsic DNA-binding activity contribute to this process.

Besides its role in regulating stressed replication fork, studies have implicated SMC5/6's involvement in normal DNA replication such as relief of DNA supercoiling²⁰⁸ (**Figure 31B**). In *S. cerevisiae*, SMC5/6 complex is present at replication initiation (autosomal replication sites (ARS)) sites during the S phase and changes to a cohesin-like distribution in G2/M phase, suggesting that the complex may directly follow RF progression and dissociates during the M phase.¹⁹⁴ Given these findings, it remains unclear why SMC5/6 and cohesin colocalize to the same chromosomal site and what functional cooperativity exists between the two complexes mechanistically. Additionally, it is unknown if SMC5/6 displays a direct functional role in molding the chromatin structure.

4.1.3 Recently discovered miscellaneous functions of SMC5/6

In addition to these discussed functions above, novel surprising roles have been ascribed to SMC5/6 in recent years. In particular, SMC5/6 complex's involvement as a restriction factor (RF), suppressing the transcription of HBV genome in human cells, has been suggested (**Figure 31D**). Notably, hepatitis B virus (HBV) has been shown to hijack ubiquitin ligases from human hosts to degrade SMC5/6, promoting transcription of viral genes.¹⁸² Nonetheless, the mechanism through which SMC5/6 acts as a RF is unknown. Interestingly, SMC5/6 was noted to directly interact with microtubules, suggesting a potential role in connecting spindle microtubules to chromosome during cell division.²⁰⁹ In plants, SMC5/6 has been implicated in

regulating the developmental processes, which appear to be linked to its SUMO ligase activity.¹⁸³ Together, these studies and among many others noted earlier in this chapter have highlighted a functional context for SMC5/6's unique acquisition of its SUMO ligase activity. Nonetheless, the contribution of SMC5/6's extensive accessory subunits and ATPase activity to its cellular function has remained thus far less understood.

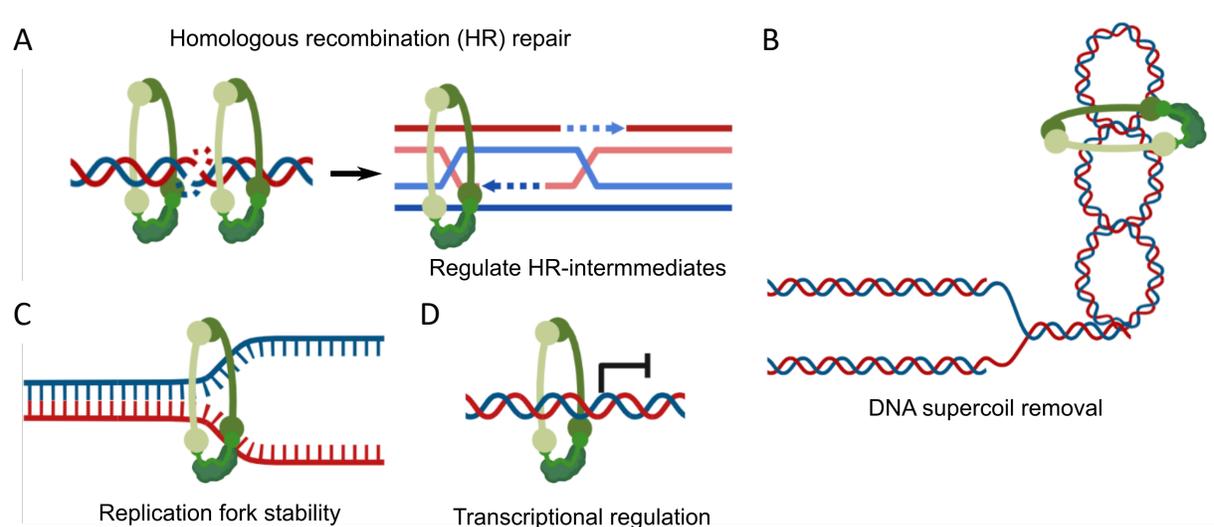


Figure 31. Schematics depicting the discussed cellular functions of the SMC5/6 complex.

4.1.4 Biochemical activity underlying SMC5/6's function

In comparison to its cellular function, the molecular mechanism of the SMC5/6 complex's action on DNA is not well understood. Moreover, how it relates to SMC5/6's cellular function is even less clear. *In vitro* studies have established that the complex can bind to ssDNA and dsDNA via exploiting its hinge and head region, respectively.^{210–212} Additionally, chromatin association *in vivo* appears to be dependent on the dsDNA-binding head region and ATPase activity.²¹³ The ssDNA-binding activity appears to mediate HR-recombination, as mutation in the hinge region leads to chromosomal rearrangement during replication restart.²¹³

What are the molecular modes of SMC5/6 action on DNA? Several models have been proposed, including DNA-tethering,⁷⁰ stabilization of supercoiled DNA,^{211,212} and loop extrusion model.²¹² A recent study, using single molecule TIRFM, has shown that the SMC5/6 complex is a loop extruding motor.¹⁹² Curiously, loop extrusion is driven by a dimer, in contrast to condensin, while monomeric form of the SMC5/6 complex translocates unidirectionally along DNA. Additionally, the Nse5-6 subunits act as negative regulator of loop extrusion by inhibiting ATPase activity of SMC5/6.¹⁹² Despite recent advances, it remains enigmatic how SMC5/6's loop extrusion activity plays out in the context of its cellular function and why SMC5/6's ATPase activity needs to be tightly regulated by the Nse5-6 complex. Importantly, although loop extrusion appears to establish a conserved mechanism among all SMC complexes, it is unclear if it is a predominant mechanism through which the SMC5/6 acts

on DNA (i.e. translocase vs loop extrusion vs others etc), which may be context dependent *in vivo*. To provide additional insights, I report below data on SMC5/6's DNA compaction activity, showing that the complex can compact DNA in an ATP independent fashion.

4.2 Visualization of SMC5/6 complex's DNA compacting capacity.

In the chapters below, I discuss data detailing the SMC5/6 complex's DNA-compaction activity. My findings suggest that the SMC5/6 complex can compact DNA and bridge neighboring DNA without the requirement for ATP hydrolysis via a tethering-like mechanism. Moreover, I observed that the DNA curvature appears to regulate SMC5/6's DNA compaction activity. Paradoxically, I observed that that ATP hydrolysis appears to inhibit SMC5/6's DNA compaction activity. All SMC5/6 protein purification work from the work below was performed by Shibai Li, Thane Than, Cory Haluska, and Dany Guan of Xiaolan Zhao's lab.

4.2.1 SMC5/6 complex compacts DNA without ATP requirement via a tethering-like mechanism

I tethered both ends of λ DNA molecules onto a passivated glass slide, stained them with 100 nM Sytox Orange (SxO), and visualized them using TIRFM, similarly to that discussed in Chapter 2.4.1 (**Figure 32A**). I then applied 10 nM of SMC5/6 holocomplex (i.e. containing SMC5/6 and Nse1-6) into the channel containing immobilized λ DNA. I observed a clusters of high intensity DNA that formed in the presence or absence of ATP, suggesting that SMC5/6 can mediate DNA compaction independently from ATP (**Figure 32C**).

To gain additional insights into the underlying mechanism, I visualized DNA compaction activity in the presence of continuous flow ($\sim 100 \mu\text{L}/\text{sec}$), during which I flowed in 10 nM of SMC5/6 into the channel orthogonally with respect to the direction of immobilized λ DNA (**Figure 32B**). I observed the accumulation of fluorescence intensity along the tip of the DNA arch formed from the hydrodynamic flow. Notably the fluorescence intensity gradually grows and moves against the applied hydrodynamic flow along with an accompanied decrease in the tether's contour length (**Figure 32B**). Furthermore, I observed single-tethered DNA compaction in the presence of flow with 10 nM of SMC5/6. Single-tether compaction entails a high intensity DNA cluster forming at the tip of the tether, which appears to reel in DNA over time against the flow (**Figure 32D**).

I attempted to reproduce the loop extrusion activity, as recently reported by Pradhan et al., under my buffer condition and at lower flow rate ($\sim 10 \mu\text{L}/\text{sec}$) but failed to observe any loop extrusion. Rather, I observed a qualitatively similar mode of compaction: a predominant DNA cluster forming at the top of the DNA arch that reeled in DNA (similar to **Figure 32B and C**). Together, these observations suggest that, under my imaging condition, SMC5/6 drives local compaction of DNA via a tethering-like mechanism without the need for ATP binding or hydrolysis.

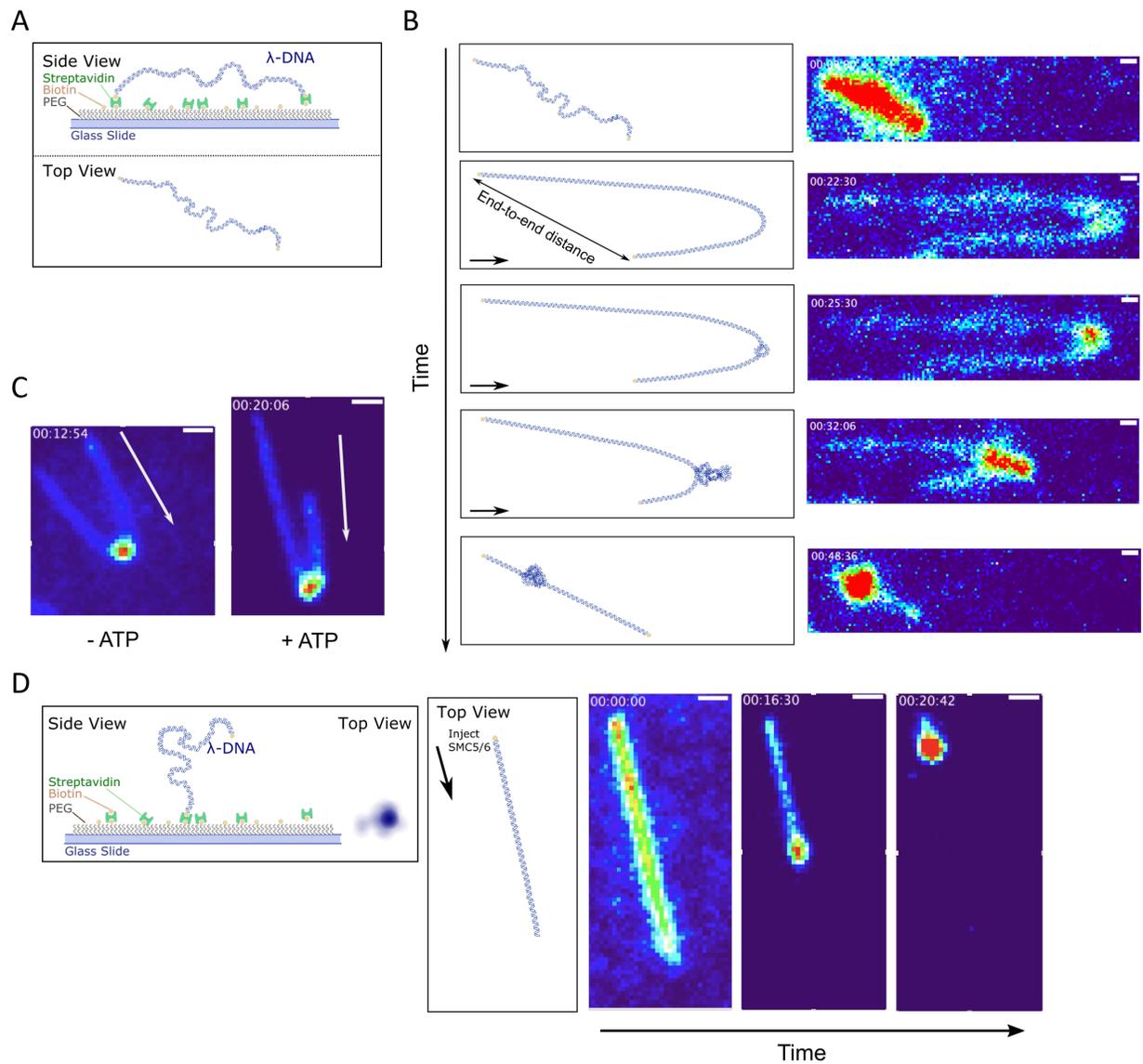


Figure 32. SMC5/6 complex compacts DNA without ATP requirement via a tethering-like mechanism.

A) Schematic of a double-tethered λ DNA molecule immobilized on a glass coverslip.

B) (Left) Schematic of λ DNA under flow in the presence of unlabeled SMC5/6. Flow direction was applied from left to right in the direction of the arrow at the bottom left of the panel. (Right) An example snapshot series of a double-tethered λ DNA molecule under flow in the presence of SMC5/6. DNA was stained with SxO (100 nM) and imaged by TIRFM. Scale bar, 0.5 μ m.

C) High intensity cluster formation on λ DNA in the presence of SMC5/6 flow with or without ATP. Direction of the flow is denoted by the white arrow. Scale bar, 1 μ m.

D) (Left) Schematic of a single-tethered λ DNA molecule immobilized on a glass coverslip. (Right) Snapshots of single-tether λ DNA compaction in the presence of flow with SMC5/6.

4.2.2 SMC5/6 complex mediates DNA bridging without requiring ATP

Consistent with the observation above, I discerned the bridging of DNA segments that are in close proximity to each other in the presence of SMC5/6 (**Figure 33**). This phenomenon happens in the presence or absence of ATP, again consistent with its ATP-independent mode of compaction. The bridging of DNA segments is mediated by newly formed DNA clusters from neighboring strands, which can mediate different variations of zipping and bridging interactions.

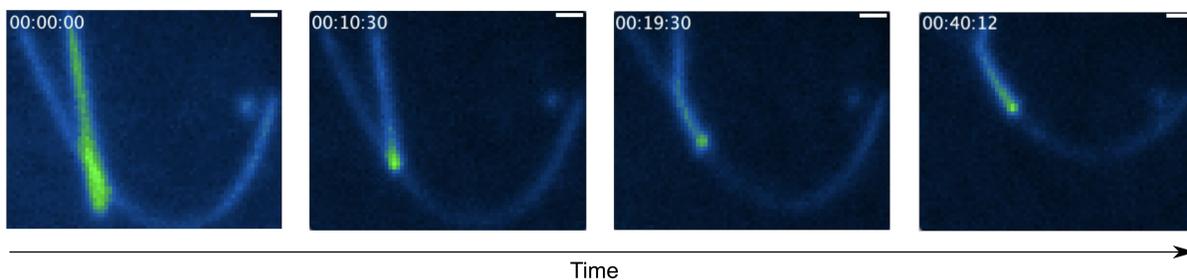


Figure 33. DNA bridging interaction in the presence of SMC5/6.

A representative snapshot series showing two DNA segments in close proximity that are bridged and zipped together in the presence of buffer flow with 10 nM SMC5/6. DNA is stained with Sytox Orange (SxO). Scale bar, 1 μm .

4.2.3 DNA curvature appears to regulate SMC5/6's DNA compaction activity

Further, I observed that λ DNA tethers that have large end-to-end anchoring distances display lesser extent of compaction (**Figure 34A and B**) and slower compaction kinetics (**Figure 34C**) compared to those with lower end-to-end distances. In general, λ DNA tethers with large end-to-end distances become more efficiently compacted if the flow is applied at a sharper angle relative to the line connecting the two anchored ends. This observation can be rationalized by a narrower DNA curvature induced by the shorter anchored end-to-end distances, allowing for closer proximity of DNA for SMC5/6 to bridge together (see Chapter 5.6).

4.2.4 ATP appears to inhibit SMC5/6's DNA compaction activity

Upon flowing in 10 nM of SMC5/6 and 5 mM ATP, I similarly observed the formation of a prominent DNA cluster at the tip of the double-tethered DNA arch and single-tethered free end (**Figure 32B-D**) that becomes compacted over time. I noticed that the presence of ATP slows down the rate of compaction for both double- and single-tethered λ DNA (**Figure 34C and D**) but does not appear to affect the maximal extent of compaction (**Figure 34B**). In the presence of AMPPNP, a non-hydrolyzable ATP analogue, SMC5/6 exhibits compaction kinetics that more resemble those without ATP (**Figure 34C and D**). Together, these data suggest that the ATPase activity of SMC5/6 appears to inhibit its DNA compaction activity.

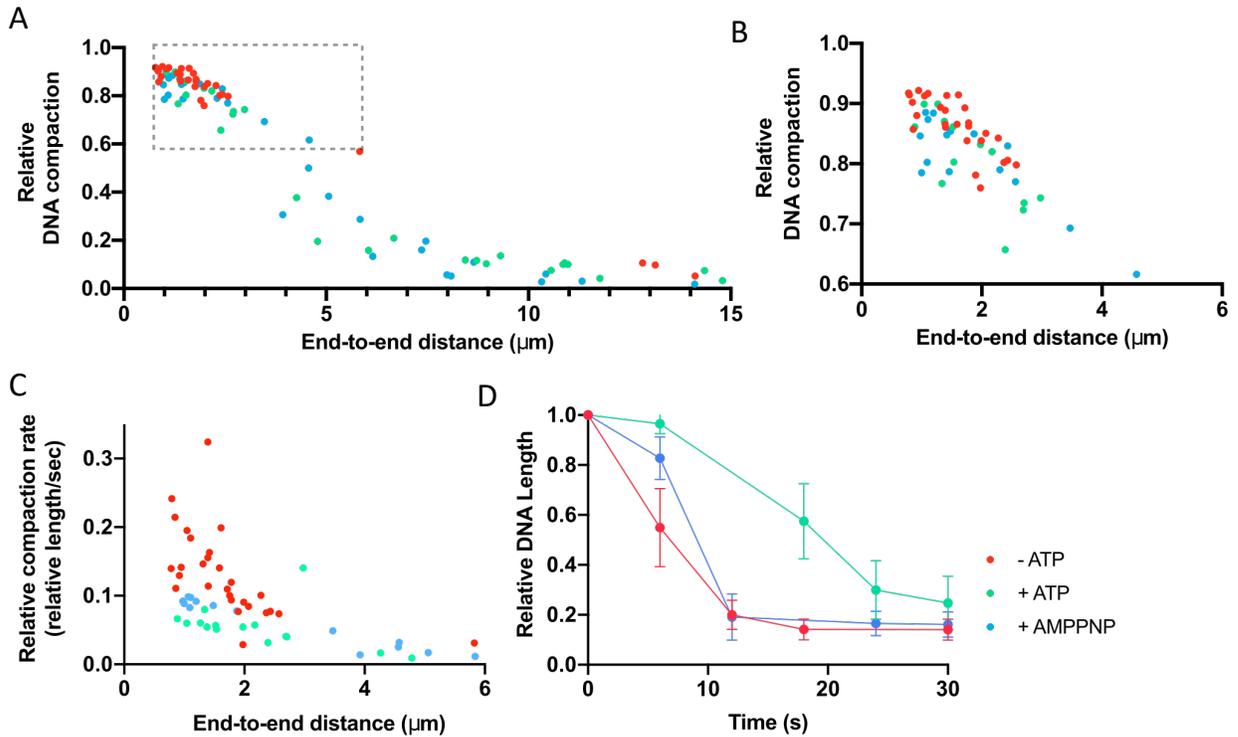


Figure 34. Regulation of DNA compaction by ATP and end-to-end tether length.

A) Shows the relative DNA compaction, corresponding to the proportion of the original DNA length that becomes compacted in the presence of SMC5/6 under different conditions, as a function of the λ DNA tether's end-to-end distances (see methods in Chapter 6.3.1). The gray inset corresponds to the zoomed-in portion shown in B).

B) Shows the zoomed in portion of A), as denoted by the gray inset.

C) Shows the relative compaction rate, which measures the rate at which the relative length of the λ DNA tether becomes compacted in the presence of SMC5/6 under different conditions (see methods in Chapter 6.3.1), as a function of the λ DNA tether's end-to-end distances.

D) Shows the relative single-tether DNA length as a function of time in the presence of SMC5/6 under different conditions.

4.3 Conclusion

In summary, the presented data suggest that the SMC5/6 complex can compact and bridge neighboring DNA without the need for ATP binding or hydrolysis. Further, I show that the geometry of the DNA, such as its curvature, regulates DNA compaction by the complex. Lastly, I present data suggestive of the notion that ATP hydrolysis paradoxically inhibits DNA compaction by SMC5/6 complex. I discuss these results, implications, and future directions in the context of known literature in Chapter 5.

CHAPTER 5. Discussion

Eukaryotic genome organization is subject to regulatory diversity to satisfy the functional needs arising from the evolution of complex biological processes. Despite emerging progress in technological development, I still lack a fundamental molecular understanding of many biological processes underpinning genome organization.

As delineated in the previous chapters, I demonstrate the application of single-molecule biophysical techniques in addressing important biological problems in the setting of eukaryotic genome organization. The delineated contributions confer novel perspectives towards understanding biomolecular condensates, chromatin organization, transcriptional regulation, and motor proteins etc. In this chapter, I discuss these findings in the wealth of existing literature, my interpretations, caveats, and future outstanding questions that arise from the presented work.

5.1 Force-generation by biomolecular condensates

The concept of biomolecular condensates has emerged as an attractive model to explain many properties related to gene regulation (see Chapter 1.3). Although condensates have been studied within the context of compartmentalization, their force-generating aspect, which is in theory possible,⁵⁴ has only been recently visited.³⁰ Below, I discuss some outstanding questions arising from the presented data in Chapter 3.

What mechanical effects can result from force-generation by biomolecular condensates? Protein:DNA co-condensation has been likened to interactions between liquids and surface, which in theory can generate forces.^{30,54} Experimentally, the forces generated by co-condensation between DNA and proteins—such as FoxA1 and PARP1—were estimated to be on the order of sub-pN,^{30,214} placing them among the weakest nuclear forces alongside those generated by loop-extruding SMC complexes.²¹⁵ I show that Sox2, an abundant TF central to pluripotency and embryogenesis, can actively generate condensation forces up to 7 pN, one order of magnitude higher than previously reported values. The cellular Sox2 concentration is estimated to be in the low micromolar range.^{159,216} Therefore, I speculate that the forces generated by Sox2 *in vivo* are at least comparable to those measured in my *in vitro* experiments. It is worth noting that Klf4, another pluripotency TF, can also form condensates on DNA against a relatively high force (~8 pN).⁹⁴ These findings are of significance because they show that forces exerted by certain protein:DNA co-condensates are comparable to other cellular forces such as those generated by molecular motors.¹⁷⁹

What are the molecular features driving force generation by biomolecular condensates? The critical force below which a protein-rich condensate is able to pull DNA inside likely depends on the physicochemical properties of the condensate, such as its surface tension,⁵⁴ which in turn are determined by the characteristics of the TF including its charge distribution and intrinsic disorder. I show that the ability of Sox2 to generate high forces through co-condensation with DNA critically relies on its IDRs. Compositionally, Sox2's N-terminal tail is characterized by having clusters of negatively charged residues and clusters of small, hydrophobic residues consisting of glycine and alanine stretch near the globular HMGB. On

the other hand, the C-terminal tail consists of predominantly of negatively charged residues apart from a small cluster of glycine rich segment near the HMGB. Given that the sequence features dictating phase separation is protein specific, it remains difficult to speculate what combination of sequence features drive the high force generation in Sox2. Given its incipient stage, the full capacity of biomolecular condensates as a force-generator is not known. Thus, further studies using force measurement techniques would be useful for investigating the underlying biological mechanism.

What is the mechanism through which the HMGB may mediate condensate formation? In Chapter 3, I show that the DNA-binding HMGB domain of Sox2 alone is sufficient for forming co-condensates with DNA, reminiscent of recent findings with Klf4 and SMN proteins.^{40,217} It has been reported that Sox2 can form a dimer on DNA that requires a motif located at the C-terminus of HMGB.²¹⁸ Hence the dimerization activity of Sox2-HMGB may underlie its ability to form co-condensates with DNA, whereas the multivalent interaction mediated by Sox2's IDRs is likely responsible for its force generation effect. Given that Sox2's HMGB domain's DNA-bending propensity and alternative putative nucleic acid-binding site, it is conceivable that HMGB may involve in multiple bridging interactions, facilitated by its DNA-bending activity, to drive biomolecular condensation. This model is reminiscent to what was described in bacterial SMC protein, in which both DNA bridging and bending contributes to DNA compaction mechanism.⁷¹

What are the cellular implications behind force-generation by biomolecular condensates? I surmise that the range of forces displayed by different protein-DNA co-condensates represent factor-specific modes of gene regulation that can be further tuned *in vivo* to achieve spatiotemporal control. In the cellular milieu, many DNA-binding factors and motor proteins, which display tension-dependent binding and stalling force *in vitro*, will be affected. Recent findings suggest that applied mechanical tension can directly regulate gene expression *in vivo*.^{14,219,220} Thus, condensate-driven mechanical tension may serve as an initial cue to trigger and tune gene expression machineries, either via cooperatively enhancing or stalling force-dependent motor protein activity.¹² This represents an additional, but not mutually exclusive, mechanism for gene regulation besides the canonical sequence-specific TF-DNA interaction paradigm.

One caveat with the interpretation of the presented data in Chapter 3 is the DNA breakage events that was observed in TIRFM. It is worth noting that a force on the order of 7 pN by itself is not sufficient to break intact DNA. Indeed, I observed few breakage events in the optical tweezers' experiments. This discrepancy can be attributed to 1) the fact that I did not use any intercalating dye to stain DNA in the optical tweezers experiments; and 2) different illumination geometries between the TIRFM and optical tweezers assays, which may render DNA in the latter assay less susceptible to nicks and other types of photodamage. Hence, it should be noted that the force values obtained from the optical tweezers assay represent a more direct and accurate measure of the mechanical tension that Sox2 condensates exert on DNA. Another limitation from the presented findings is that it remains difficult to discern mechanical stress *in vivo*, which remains experimentally challenging. Additionally, given the crowded the nuclear environment, eukaryotic DNA would be subjected to various external forces from molecular motor to extra-nuclear forces. Thus, it is likely that in the presence of

other external forces present in the nucleus, the mechanical properties of Sox2 condensates may be different from the simple tether that is observed in the discussed experimental setup.

5.2 Biophysical nature of biomolecular condensates

In Chapter 3, I show that Sox2:DNA co-condensates are extremely stable, resisting against pulling forces sufficient to overstretch B-form DNA. In comparison, a fraction of the condensates formed by DNA and Heterochromatin Protein 1 α can resist disruptive forces of up to 40 pN.²²¹ Recent studies demonstrate that at a more mesoscale level, liquid and gel-like droplets can resist high forces from 20 to 40 pN before fusing.⁵¹ Together, these observations highlight that the mechanical properties of biomolecular condensates can potentially be tailored towards their specific biological roles. For example, phase separation of heterochromatin has been suggested to perform mechanical work via resisting extra-nuclear deformation.²²² In a similar fashion, the stable mechano-properties of Sox2:DNA condensates may ensure fidelity of gene expression, given that mechanical stress driven by extra nuclear forces and motor proteins exist.^{10,12}

Although early studies of phase separated condensates show that they have liquid-like properties, a more nuanced picture emerges in which condensates display properties ranging from liquid to more solid-like. In particular, condensates can exhibit maturation, where they physically transition from liquid to solid-state as they age.^{20,22} For example, FUS has widely been shown to phase separate into liquid-like condensates which transition into a more solid-like state through time,²²³ an observation similarly reported in other protein families.²⁰ Several mechanisms explaining condensate maturation have been proposed, ranging from gelation to polymerization, which would require further testing. In Chapter 3, I observe that Sox2 condensates undergo what is akin to maturation as the condensate mobility decreases over time. Anecdotally, I observe that the DNA-breakage that happens at later time points exhibit rigid body-like fluctuation in a manner reminiscent to a solid-like material. Curiously, I also observe that high force generation by Sox2 condensates, signified by the DNA breakage events, predominantly occurs early but diminishes as the condensates' mobility decreases over time. This observation indicates that the maturation of Sox2:DNA co-condensates from a liquid-like form to a solid-like one—akin to what was described in other systems²⁰—attenuates their force-generating capacity. Together, these observations highlight the relationship between condensate's force generating capacity and its material properties, which would entail further investigation.

I further show in Chapter 3 that although the HMGB domains can form condensates that appear to exert less force compared to full length Sox2 (FL-Sox2), their condensate sizes appear to be on average larger than FL-Sox2. Thus, it is interesting to ponder the relationship between the condensate size limit and its force-generating capacity. Given that Sox2 forms relatively small visualizable droplets in bulk, it is conceivable that the surface tension of Sox2 condensates is relatively low compared to other known condensates. Additionally, the high force-generating ability, dictated by the physicochemical properties of a given protein, may dictate size limit of a condensate at a given concentration.

5.3 Chromatin component as a force-regulating mechanical sink

The work in Chapter 3 adds to the mechanical regulatory role of chromatin by showing that nucleosomes and linker histone H1 sequester TFs from exerting high force on DNA via co-condensation. Together, these data suggest a role of the chromatin as a mechanical sink that buffers stress within genomic DNA.

How does the chromatin structure attenuate force generation by biomolecular condensates? One possible scenario is that nucleosomes, through the contact of DNA and histone octamer, present additional barrier for Sox2 condensates to reel in additional DNA. Thus, by effectively reducing Sox2 concentration bound to bare DNA via sequestration, nucleosomes can alleviate the average force that is exerted on bare DNA by Sox2 condensates. An alternative scenario is that the intrinsic biochemical properties of condensates on nucleosomes and bare DNA are dissimilar, rendering the differential force-generation effects. Also, given that nucleosome arrays can phase separate and form biomolecular condensates,^{46,47} it is interesting to ponder how the mixing of condensates can potentially tune their biochemical properties and mechanics. I show that H1 can form condensates that colocalize with Sox2 and attenuate its mechanical effect on DNA. This observation suggests that the mechanical properties of condensates can be potentially regulated by tuning the condensate's composition, and given its alterations in diseased states,⁴¹ it would be intriguing to ascertain how the potentially altered mechanics of condensates affect gene regulation. Taken together, it can be envisioned that the chromatin landscape—shaped by many factors and altered during development and disease—is directly related to the force field in the nucleus. Further studies are warranted to elucidate this relationship, which will improve the understanding of how chromatin mechanics influence genome architecture and gene expression.

What is the interplay between nucleosomes and sequence motif features as potential nucleation point for condensate formation? Emerging studies suggest condensates sizes correlate with TF sequence motif location on bare DNA.⁹⁴ Given the colocalization of Sox2 condensates and nucleosomes, it is interesting to ponder how nucleosomes can potentially dampen Sox2's dependency on its sequence motif, vice versa. Previous study using ChIP-seq suggests that Sox2 displays a more promiscuous sequence motif in nucleosome position, likely driven by the intrinsically distorted geometry dictated by the nucleosomal architecture.¹⁴⁹ However, how this is relevant in biomolecular condensation is not known.

Are the sequestering effects limited to only PF or other chromatin-binding factors? Given the diversity of nucleosome-binding factors characterized thus far, it can be envisioned that chromatin condensate formation is generalizable to many proteins outside of PF. HP1 α , for example, has been demonstrated form condensates on chromatin.^{48,49} Nonetheless, it remains difficult to speculate on the spectrum of sequestering effect by the chromatin structure and the biological sequelae. Given that the chromatin displays properties resembling condensates with a range of material properties, it remains to be seen how these novel characteristics are relevant towards regulation of DNA-binding factors. Taken together, a nuanced understanding of the chromatin and its regulatory role in condensate formation will facilitate the understanding of diverse pathologies that have ties to phase separation, thereby facilitating novel avenues of therapeutic intervention.

What are the cellular implications of chromatin's role as a mechanical sink, particularly within the context of biomolecular condensates? It is conceivable the excess force on DNA by condensates may be deleterious, provided that it can potentially affect other force-dependent cellular processes such as DNA-binding activity and motor protein processivity. Therefore, the presence of nucleosomes may serve to tune condensates by raising the threshold for force-generation. In this way, higher concentration of proteins would be necessary to exert mechanical stress on DNA.

5.4 Mechanism underpinning the regulatory interplay between H1 and Sox2

The presented data in Chapter 2 show that H1 and Sox2 display a concentration-dependent mode of regulation. Indeed, higher Sox2 concentration promotes H1 loading on both DNA and nucleosomes, while higher H1 concentration promotes Sox2 loading rate up to a certain level (i.e. up to 50 nM), but inhibits Sox2 loading at higher concentration (i.e. at 100 nM and above). Conceptually, these observations are not expected, given that the HMGB, to which Sox2 belongs, and H1 are historically long thought to be antagonistic regulators of each other.¹³⁹ What mechanism may underlie H1's concentration-dependent regulation of Sox2 binding to nucleosome and DNA?

5.4.1 Biomolecular condensates model

One unifying model that can be invoked to explain this phenomenon is through the formation of biomolecular condensates. In Chapter 3, I show that H1 can form condensates at 100 pM on λ DNA in TIRFM. Similarly, Sox2 was shown to form condensates on λ DNA at 10 nM. Importantly, I show that H1 and Sox2 condensates colocalize with each other. At the lowest concentration of tested H1, namely, at 2 nM, the immobilized DNA/nucleosome, which was at picomolar concentration, would be well saturated. As such, the presence of existing H1 condensate may recruit additional Sox2 binding onto DNA and nucleosomes, thereby resulting in increased loading rate. At higher concentrations of H1, Sox2 loading rate onto DNA and nucleosome is decreased. This curious observation can reflect the condensate's changing material properties at higher H1 concentrations, which can become more solid-like via crosslinking of the IDRs. In support of this model, the phenomenon is abolished via the C-terminal tail truncation (CTDdel) of H1.4, which displays continually increased Sox2 loading at higher H1.4 CTDdel concentrations. Similarly, biomolecular condensate formation can be used to explain Sox2's concentration dependent recruitment of H1 onto DNA and nucleosome.

Several weaknesses are present in the proposed model: 1) I cannot directly visualize condensate formation, if it exists, for the substrate tested, unlike the case for λ DNA in Chapter 3) the residence time of Sox2 on DNA and nucleosomes remains unchanged, suggesting that Sox2 is rapidly dissociating after being loaded, albeit an alternative explanation can involve diffusion within a condensate to outside the TIRF field, 3) the CTDdel of H1.4 abolishes its bulk phase separation properties, but it remains unknown how it behaves at a molecular level, which may behave, as recently suggested, in a distinct fashion,⁹⁴ and 4) excess non-specific adherence of H1 at high concentration on the TIRFM surface may artificially recruit Sox2 binding.

5.4.2 Other alternative models

In vivo, H1s have a much more dynamic dwell time compared to that *in vitro*, likely as a result of competing factors and chaperones, which facilitates the exchanges of H1s.²²⁴ For example, Nap1 was found to remove non-specifically bound H1s from both DNA and mononucleosome.¹²¹ Hence, it is conceivable that in the presence of histone chaperones, H1's propensity to form biomolecular condensates can change.²²⁵ The presented data suggest that Prot α , a long acidic peptide that tightly binds to C-terminal tail of H1,²²⁶ promotes the loading of H1.4 on mononucleosome but not on DNA. It is entirely possible that at higher concentrations of H1, molecular aggregates may accumulate. Thus, the aggregated mode of binding can explain H1 inhibition of Sox2 loading rate at higher concentrations, and CTDdel simply reduces the aggregated mode of binding. H1 may directly interact with Sox2 for enhanced loading effect. This model is consistent with some bulk experimental observations, such as H1's trapping at the loading well in an EMSA at high, supersaturated concentrations. Further experiments involving the use of histone chaperones and DNA substrates of differing lengths would be useful to further clarify the underlying mechanism.

Additionally, it is also worth noting that Sox2 binding dwell time appears to be unchanged across all DNA and nucleosome substrate used, which is contrast to that previously reported.⁸⁷ One potential explanation for this discrepancy lies in the different buffer condition used, wherein I used 150 mM NaCl compared to 50 mM NaCl in a previous study.⁸⁷ I also used labeled DNA constructs for my nucleosomes and thus, I cannot rule out nucleosome dissociation in TIRM condition.^{227,228} Further repeat and control experiments are necessary to gain additional insights into these observations.

5.5 Cellular implications of HMGB and H1 regulation

The findings in Chapter 2 suggest a surprising notion that H1 and Sox2, a HMGB protein member, can both positively promote each other's loading onto DNA and nucleosome. This observation is important because it implicates H1 in transcriptional activation processes mediated by Sox2, thereby further expanding the role of H1 to more than just an architectural protein. Additionally, the findings of Sox2 as a facilitator of H1 loading onto DNA and nucleosome implicates Sox2 as a possible context-dependent chaperone, potentially expand Sox2's role more than just a transcription factor. Together, the presented data highlight a potentially complex interplay between the HMG protein family and H1 in the cellular milieu, in which positive and negative cooperativity is contingent on the relative concentration of each molecular component.

In addition, it should be noted that the concentration of labeled Sox2 and H1 are in the low nanomolar and picomolar concentrations, which would be below the known measured concentrations *in vivo*.^{4,159,216} Therefore, the described setup using TIRFM cannot capture the binding activity at higher protein concentrations observed *in vivo*. Thus, a further understanding of Sox2 and H1 binding dynamics at a higher concentration would entail further follow-up studies.

5.6 DNA-tethering by the SMC5/6 complex

In Chapter 4, I visualized a tethering-like mode of compaction from the SMC5/6 complex that does not require ATP hydrolysis. Under the presence of continuous flow, I observed accumulation of DNA into a distinct cluster located at the tip of the DNA arch, which gradually reels in DNA. The compaction appears to be dependent on a cluster of multiple SMC5/6 complexes. Together, these observations suggest a distinct mechanism involving cooperative action of multiple SMC5/6 complexes to compact DNA.

Qualitatively, the compaction mediated by multiple SMC5/6 complexes resemble a tethering-like model,⁷⁰ in which multiple complexes can oligomerize and bridge neighboring DNA. This model is consistent with the observation of DNA bridging, which is mediated by multiple SMC5/6 located in compacted DNA foci. Importantly, this model explains why the DNA geometry appears to dictate the efficiency of compaction mediated by the SMC5/6 complex: the narrow DNA arch allowed by shorter end-to-end anchoring distances facilitates higher probability of DNA to be tethered by the complex at the opposite ends. The tethering model is reminiscent of bacterial SMC complex, which also displays DNA tethering.^{71,229} Given the recent structural findings of the SMC5/6 complex, it is likely that the binding mode may involve an alternative binding site distinct from the cleft interaction noted in the head domain, which entails ATP binding.¹⁹³

The described observation of SMC5/6's mode of compaction strikingly resembles that of *S. cerevisiae* cohesin, which also appears to display DNA compaction without the requirement for ATP hydrolysis.⁷³ Phase separation has been invoked to explain this observation in *S. cerevisiae* cohesin, along with many other reports in bacterial SMC loading factor ParB and cohesin loader Scc2/4.^{72,230} In comparison, SMC5/6 complex does not contain an equivalent loader, and although I anecdotally observed fusion of DNA and SMC5/6 foci, it remains unclear if the complex can phase separate, which would entail further investigation.

5.7 Loop extrusion vs DNA tethering

Whether loop extrusion is a common unifying mechanism among all SMC complex is an open question. Recent preprint published during the writing of this thesis demonstrated that SMC5/6 is a loop extruding motor, mediated by a dimeric form. On the other hand, the monomeric form of SMC5/6 complex displays unidirectional DNA translocase activity.¹⁹² As described in Chapter 4, I attempted to reproduce the observed loop extrusion in my buffer condition under low flow but was thus far unsuccessful. I note that the different buffer conditions employed between my experiments can potentially explain the observed discrepancies. I used low salt at ~50 mM K₂Glu, similar to what was used in loop extrusion assay in condensin and cohesin.⁶⁵⁻⁶⁷ The higher salt used in Pradhan et al. (i.e. 100 mM NaCl)¹⁹² may enable the complex to adopt more monomeric/dimeric forms instead of higher order clusters at higher protein concentration. Reports from condensin and cohesin noted that loop extrusion activity is sensitive to salt and hydrodynamic flow,⁶⁵⁻⁶⁷ and thus, careful calibration of these parameters is warranted in future investigations.

How the SMC5/6's DNA compaction activity, either via loop extrusion or tethering mechanism, functions in the context of its cellular function is not well understood. Given the known colocalization with cohesin *in vivo*, it is possible that SMC5/6 may act in a synergistic fashion with cohesin to facilitate HR-repair processes via loop extrusion. Recent *in vivo* findings suggest that SMC5/6's ATPase activity, which enables its loop extrusion, facilitates its chromatin-binding activity.²¹³ However, the exact molecular mechanism, particularly with respect to loop extrusion, is unknown. Besides loop extrusion, DNA compaction by SMC5/6 complex via the tethering-model, as suggested by the discussed findings, may promote physical proximity of DNA molecules in HR DNA repair to promote more efficient interaction. In a similar fashion, local DNA compaction may explain SMC5/6's ability to inhibit HBV transcription. Given SMC5/6's involvement in complex molecular pathways, such as those related to HBV proliferation, further biochemical studies would be helpful to dissect the molecular properties of SMC5/6 underpinning its cellular function. Novel therapeutic strategies may arise from such understanding, and due to the high prevalence of HBV infection and related consequences such as cirrhosis and cancer, understanding the SMC5/6 complex's molecular function is an important endeavor that can be leveraged to address many unmet clinical needs.

5.8 ATP-dependent decrease of SMC5/6 compaction: a paradox

The presented data in Chapter 4 suggests how ATP binding and hydrolysis activity can potentially decrease SMC5/6's DNA compaction. Below I present possible mechanistic explanations for this paradoxical activity.

One possible explanation for this observation is that ATP binding renders the SMC5/6 complex more dynamic, and as such, the effective concentration, particularly at the cluster of the DNA arch created by hydrodynamic flow, is less than that without ATP. In support of this notion, a recent study has demonstrated that translocase activity of SMC5/6 requires ATP hydrolysis.¹⁹² In this fashion, SMC5/6 without ATP, perhaps using an alternative DNA binding site, may be able to more readily oligomerize and form clusters that facilitate DNA compaction.

Another explanation for this paradoxical observation accounts for the presence of hydrodynamic flow. In particular, SMC5/6 with ATP may bind more tightly to DNA and consequently may not respond to the flow. In contrast, SMC5/6 without ATP through its differential binding mode may slide on the DNA more readily in the same direction as the flow, resulting in more trapped SMC5/6 clusters at the tip of the arch. In this way, the flow may artificially increase the local concentration of SMC5/6 complexes that become trapped at the tip of the arch, particularly for those without ATP bound. Further experiments using lower flow rates and labeled SMC5/6 complexes will be useful to clarify these aforementioned mechanisms.

5.9 Cellular implications of SMC5/6 compaction modes

The observation of ATP-independent mode of compaction in SMC5/6 would hold several implications *in vivo*. First, local compaction can create a favorable environment to promote

biochemical reactions such as DNA repair and replication. For example, the increased local concentration DNA and proteins would increase the possibility of homologous recombination during DNA repair, thereby increasing its efficiency. Furthermore, increasing the local DNA concentration via compaction would promote DNA binding cooperativity. Second, local compaction by SMC5/6 can shield the desired DNA sites from undesirable non-specific enzymatic effects. In this way, the DNA intermediates formed during HR-repair may have less propensity to transition into more deleterious structures.¹⁹⁴

Given that ATP is present in the cellular environment, it can be speculated that SMC5/6 would switch between varied modes of compaction such as loop extrusion and DNA tethering depending on the relative abundance of ATP. It is entirely possible that *in vivo*, there would be other interacting factors that would govern the DNA-binding mode of SMC5/6. Elucidation of SMC5/6's DNA binding modes and their relation to SMC5/6's cellular function will remain an outstanding question awaiting further molecular insights.

5.10 Concluding remarks

While development of key technologies has enabled key advancements towards my understanding of eukaryotic genome organization, it raises additional outstanding questions concerning the underlying molecular mechanism. In this thesis, I describe my contributions towards understanding the regulation of eukaryotic genome organization by leveraging single-molecule approaches. I show a concentration-dependent regulatory interplay between Sox2 and H1 (Chapter 2). Further, I demonstrate that Sox2:DNA co-condensation can generate high forces that can be attenuated by the chromatin structure (Chapter 3). Lastly, I visualize in real time DNA compaction by the SMC5/6 complex. The results and methods outlined in this thesis will serve as a foundation through which researchers from diverse disciplines can built upon in their future work.

CHAPTER 6. Materials and methods

6.1 Protein purification and labeling

Sox2

Human Sox2 proteins were expressed and purified as previously described.⁸⁷ In brief, Sox2-FL and Sox2-HMGB constructs were cloned into the pET28B plasmid, expressed in Rosetta (DE3) plyS cells (Novagen #70956-3) in LB media at 37°C until reaching an OD₆₀₀ of ~0.6, and induced with 0.5 mM IPTG at 30 °C for 2 h. Cells were harvested, lysed, and purified using a Ni-NTA affinity column under denaturing conditions. Eluted Sox2 was refolded by changing to a zero-urea buffer using a desalting column (GE healthcare #17-1408-01). Further purification was performed by gel filtration on a Superdex 200 10/300 GL column (GE Healthcare). Fluorescence labeling was performed as previously described.⁸⁷ In brief, Cy5 or Cy3 maleimide (GE healthcare) was mixed with Sox2 at a molar ratio of ~2:1. For Sox2-FL, the dye was conjugated to the only native cysteine C265. For Sox2-HMGB, a K42C mutation was introduced by site-specific mutagenesis. Free dye was removed by gel filtration on a Superdex 200 10/300 GL column.

Histone octamer

Developed by Sai Li

Recombinant histone octamers from *Xenopus laevis* were purified and labeled as previously described.⁸⁷ In brief, each of the four core histones (H2A, H2B, H3 and H4) was individually expressed in BL21 (DE3) cells, extracted from inclusion bodies, and purified under denaturing conditions using Q and SP ion exchange columns (GE Healthcare). Octamers were refolded by dialysis and purified by gel filtration on a Superdex 200 10/300 GL column. To label the octamer, single-cysteine constructs H2B T49C and H2A K120C were generated by site-directed mutagenesis and incubated with Cy3 and A488 maleimide at 1:5 molar ratio, respectively.

Linker histone H1

Developed by Wola Osunsade of the David Lab

His-Sumo-H1.4^{A4C}-GyrA-His was expressed and purified as described previously²³¹ with minor adjustments. Briefly, the construct was expressed in Rosetta DE3 cells overnight at 16°C. Cells were lysed and lysate incubated with Ni-NTA beads (Bio-Rad). 1 mM DTT was added to the eluent, and it was incubated with Ulp-1 (1:100 v/v) for 1 h at room temperature. Following this, 500 mM β-mercaptoethanol was added. The mixture was run on a Hi-Trap SP column, and fractions containing full-length H1.4^{A4C} were pooled and injected on a semi-preparative HPLC C18 column. Pure fractions of H1.4^{A4C} were pooled and lyophilized. Lyophilized H1.4^{A4C} was resuspended in H1 labeling buffer (6 M Guanidine, 20 mM Tris pH 7.5, 0.2 mM TCEP). It was mixed with 3 molar equivalents of Cy3 maleimide for 1 h at room temperature, followed by quenching with 1 mM β-mercaptoethanol. This was injected on a semi-preparative HPLC

C18 column. Pure fractions of Cy3-H1.4 were pooled and lyophilized. Cy3-H1.4 was resuspended in buffer (20 mM Tris pH 7.5, 200 mM NaCl) before use.

6.2 DNA construct preparation

As a proof of concept, I reconstituted recombinant nucleosomes using a set of Cy3 labeled and biotinylated 601-Widom DNA containing 48 bp by 38 of linker DNA (i.e. total 233 bp) to allow for sufficient H1 binding, as described previously literature, and 601-Widom nucleosome positioning sequence (NPS) variants with the engineered canonical Sox2-Oct4 tandem motif (CATTGTTATGCTAAT) in the nucleosomal dyad, as described by a previous study⁸⁷ (i.e. referred to hereinafter as 601-dyad). Nucleosomes were further purified using glycerol gradient. DNA constructs were PCR-ed using pairs of internally labeled amine-modified and biotinylated primers and subsequently purified. Sox2 construct was purified and labeled with Cy5, as accordingly described previously.⁸⁷ Linker histones were obtained by courtesy of Adewola Osunsade from David Lab (MSKCC), either through commercial NEB (e.g. H1.0 variant as discussed here in this report) or purified in-house.

6.3 Single-molecule TIRFM experiments

Visualization of Sox2 and H1 biomolecular condensate formation

Single-molecule imaging was conducted on a total-internal-reflection fluorescence microscope (Olympus IX83 cellTIRF) with 100x magnification/1.49 numerical aperture (NA) objective (UApoN Olympus) and visualized using Metamorph v7.8 software. PEG slides were prepared as previously described.⁸⁷ The assembled flow chamber was infused with 20 μ L of 0.2 mg/mL streptavidin (Thermo Fisher Scientific), incubated for 5 min, and washed with 250 μ L of T150 buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.0075% Tween). Biotinylated λ DNA (LUMICKS) was immobilized by slowly injecting a diluted 10-20 pM solution at a volume of 40-80 μ L over the course of 2 min. Afterwards, 250 μ L of T150 buffer was flowed into the chamber to wash away molecules that were not immobilized. For T4 ligase treatment, 20 μ L of 1:20 diluted T4 ligase (NEB) in T4 ligase buffer was flowed into the chamber, incubated for 10 min, and washed away with 250 μ L of T150 buffer. 100 μ L of T150 buffer containing YOPRO1 (20 nM unless specified otherwise) and an oxygen scavenging system (4% w/v glucose, 1.5 mg/mL glucose oxidase, 0.072 mg/mL catalase, 2 mM Trolox) was then flowed in to visualize immobilized λ DNA.

In nucleosome experiments, I adopted a previously described protocol with minor modifications.²³² In brief, *in situ* nucleosome formation was achieved by flowing in 15 nM of Cy3-labeled histone octamer and 30 nM of Nap1 in T150 buffer into the chamber followed by a 5-min incubation. The chamber was then flushed with 250 μ L of T150 to wash away any free histone octamer and Nap1.

A solution containing a specified concentration of Cy5-labeled Sox2 and the above imaging buffer (i.e. T150, YOPRO1, and oxygen scavenging system) was prepared, 50 μ L of which was flowed into the microfluidic chamber, and movies/images were recorded. H1 imaging was similarly performed with a specified concentration of Cy3-H1. Movies were recorded at room

temperature with a frame rate of 300 ms. 488-nm at 20% power and incident angle of 31.51°, 532-nm at 20% power and incident angle of 33.86°, and 640-nm lasers at 15% power and incident angle of 34.61°, respectively, were used to excite YOPRO1, Cy3, and Cy5/TOTO3 dyes, respectively. Movies were subsequently displayed and analyzed using plugins in ImageJ/FIJI.

Visualization of SMC5/6 DNA compaction

The imaging was similarly performed but with the following modification. After immobilization of λ DNA, 10 nM of SMC5/6 holoenzyme construct in imaging buffer (50 mM Tris, pH 7.5, 50 mM Kglu, 2.5 mM MgCl₂, 1 mM DTT +/- 2 mM ATP, and 100 nM SxO) was flowed in at ~100 μ L/min at an orthogonal direction to that of the immobilized λ DNA for 3 mins. Movie condition was otherwise similarly performed as described above with exception to the 532 nm laser, which is set at 15% power and incident angle at 31.2°.

Colocalization single-molecule spectroscopy (CoSMoS) of Sox2 and H1 binding

Biotinylated constructs containing nucleosomes and naked DNA were immobilized on passivated PEG-slides, which were prepared as previously described. A washing buffer containing 50 mM Tris, 150 mM NaCl, pH 7.5 (i.e. T150 buffer) was flowed into the injection chamber to wash away molecules that were not immobilized on streptavidin surface. Solution containing 2nM of Cy5-labeled Sox2 with or without linker histone H1 was injected with imaging buffer containing 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, +/- 8% glycerol, 0.075% Tween-20 and oxygen scavenging system (2% (w/v) glucose, 1.5 mg/ml glucose oxidase, 0.072 mg/ml catalase, 2mM Trolox. The binding experiment of Sox2 were performed with H1 concentration titrated up in the same chamber, with T150 washing buffer flowed in before further injection of Sox2/H1 solution. For labeled Cy3-H1.4 experiments, 10 pM of Cy3-H1.4 was flowed into the chamber with identical buffer conditions as above.

Movies were recorded at room temperature with frame rate of 300 ms. The general imaging scheme is done as follows: 1) the positions of immobilized substrates (i.e. Cy3 labeled DNA/nucleosomes), were determined in the initial 100 frames by turning on the 532-nm laser (i.e. corresponding Cy3 channel), 2) 532-nm laser was turned off, 3) Cy5-labeled Sox2 binding dynamics were obtained by turning on the 640-nm laser (i.e. corresponding to Cy5 channel). The Sox2 binding events were obtained by colocalizing the 532 and 640 nM channel and single-molecule trajectories were extracted and analyzed using SPARTAN software suites. TF binding events were determined using automated cutoff signal to noise ratio in SPARTAN and fluorescence intensity threshold. The lifetime of Sox2 dwell time on substrate was determined through in-house MATLAB scripts.

6.3.1 TIRFM data analysis

Analysis of DNA envelope width and fluctuation radius

I followed an analysis pipeline as previously described.³⁰ In brief, time-averaged projections of DNA images were generated in conditions with/without proteins. Transverse line profile of the DNA intensity was generated by drawing a line perpendicular to the middle of the DNA,

which gives the maximum DNA width. Background was subtracted off these profiles, and a Gaussian curve was fitted to each line profile. The DNA envelope width and fluctuation radius were defined as two times the standard deviation of the fitted Gaussian curve.

Estimation of DNA content and Sox2 counts in a cluster

The YOPRO1 intensity profile was extracted and background subtracted. The estimated DNA content within each cluster was calculated as similarly described⁶⁵ and shown below:

$$DNA\ content\ (bp) = \frac{DNA\ intensity\ in\ cluster \times 48,502\ (\lambda DNA\ length)}{Total\ DNA\ intensity}$$

To estimate the number of Sox2 molecules in each cluster, I extracted the Cy5 intensity profile that colocalized with λ DNA after subtraction of background signals. I then extracted the intensity profiles of Cy5-Sox2 non-specifically adsorbed to surface in the same field of view, which I assumed as monomers. The average intensity of the monomer was calculated. The number of Sox2 molecules within each cluster on λ DNA was calculated by dividing the cluster intensity, which is defined as the local maximum of the fluorescence intensity profile, by the average monomer intensity.

Colocalization analysis

Time-averaged projection of the images in each fluorescence channel was generated, and background was subtracted. The regions of interest were segmented and extracted for further analysis. Pearson's correlation coefficients in each condition were calculated using JaCoP plugin in Fiji.²³³ Costes' randomized control, which describes the correlation between randomly shuffled pixels of two compared images,²³⁴ was also calculated using the JaCoP plugin.

Condensation time analysis

Each immobilized λ DNA molecule in a field of view was individually monitored, and the time when a molecule condensed was defined as the transition at which the molecule completely lost slack/fluctuations. I subsequently ranked the condensation times and recorded the 75th and 25th percentile values (T_{75} and T_{25} , respectively). The average condensation time ($T_{condense}$) was calculated as $T_{75}-T_{25}$.

Mobility analysis

Developed by John Watters

Kymographs were extracted from TIRF microscopy movies using the kymographBuilder plugin in Fiji. The Sox2 foci were then manually extracted using the Kymotracker 'greedy' tracking algorithm.^{235,236} Early and late Sox2 condensate events were recorded from movies taken ~ 5

min and ~15 min after Sox2 injection, respectively. One-dimensional mean squared displacement (MSD) was then applied using a maximum delay time of 4 sec (0.3-sec time steps) using a custom python script written based on the description and methods from the @msdanalyzer MATLAB per-value class.²³⁷ A Savitzky-Golay filter (third order polynomial with an eleven-frame window) was applied to smooth traces in preparation for MSD analysis. Diffusion coefficients were only calculated if the goodness of the linear fit was greater than 0.8. Approximately 67% of early Sox2 traces and 50% of late Sox2 traces met the required parameters for diffusion coefficient fitting.

Binding events ratio

The binding events ratio is calculated as follows:

$$\text{Binding events ratio} = \frac{\text{Number of colocalized, bound Sox2/H1 traces}}{\text{Number of immobilized DNA or nucleosome molecules}}$$

Relative DNA compaction

The relative DNA compaction, which measures the relative extent of DNA that gets compacted of each analyzed molecule, is calculated as follows:

$$\text{Relative DNA compaction} = 1 - \frac{\text{Measured DNA length at 1 min}}{\text{Measured DNA length at 0 min}}$$

Relative compaction rate

The relative compaction rate is derived from fitting a non-linear curve onto data points measuring the relative DNA length in each measured time point (Relative DNA length vs time, similar to **Figure 34D**). The relative DNA length is obtained by dividing the DNA length in each time point to the DNA length at time 0 min, which corresponds to the most elongated DNA segment. After fitting a non-linear curve onto each data points for each tether, I computed the highest slope on the fitted curve, which corresponds to the relative compaction rate.

6.4 Optical tweezers experiments

In collaboration with Jeremy Chang

Single-molecule optical tweezers experiments were performed at room temperature on a LUMICKS C-trap combining confocal fluorescence microscopy and dual-trap optical tweezers as previously described.⁹⁶ In brief, we trapped two streptavidin-coated polystyrene beads (Spherotech) with a 1064-nm trapping laser and moved these beads to a channel containing biotinylated λ DNA (LUMICKS). Single DNA tethers were selected based on the force-extension curve. The DNA tether was then moved into a channel containing Cy3-labeled Sox2 in T150 buffer. Cy3-Sox2 on DNA was visualized by confocal scanning with a 532-nm excitation laser. Correlative force and fluorescence measurements were made under different operation modes (force clamp mode, passive mode, or pulling mode)⁹² as specified in the figure legends. Nucleosomal DNA experiments were similarly performed. To assemble nucleosomes *in situ*, a single λ DNA tether was moved into a channel containing 12 nM of A488-H2A histone octamer

and 48 nM Nap1 in HR buffer (30 mM Tris-OAc pH 7.5, 20 mM Mg(OAc)₂, 50 mM KCl, 1mM DTT, 40 µg/mL BSA), and incubated at a fixed trap distance of 10 µm for 20 sec under flow and another 20 sec without flow. The tether was then moved into another channel containing 0.5 mg/mL salmon sperm DNA in HR buffer, in which a flow was applied for 30 sec to remove free histones and Nap1. Before moving to the Sox2 protein channel, the force was reset to zero to remove any influence of nucleosome wrapping on the force reading. Force and fluorescence data were generated via Bluelake software v2.1.5 (LUMICKS) and processed using a custom GUI Python script titled “C-Trap.h5 File Visualization GUI” (<https://harbor.lumicks.com/single-script/c5b103a4-0804-4b06-95d3-20a08d65768f>).

6.5 Electrophoretic mobility shift assay (EMSA)

In collaboration with Htet Ng.

DNA substrate was prepared via PCR and gel extraction of a 233-bp construct containing the Sox2 binding motif engineered into a 601 sequence (Supplementary Table 1) as previously described.⁸⁷ 10 nM of DNA substrate was incubated with Sox2 and HMGB constructs in T150 buffer at room temperature for 30 min. The reaction mixture was loaded onto a 5% non-denaturing polyacrylamide gel, which was run in 0.5× Tris-Borate-EDTA at 4°C at 100 V for 90 min, stained with SYBR Gold (Invitrogen), and visualized using a Typhoon FLA7000 gel imager (GE Healthcare).

6.6 Statistical analysis

Statistical tests and *P* values were reported in the figure legends (ns, not significant; * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001).

CHAPTER 7. References

1. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).
2. H., F. E. & Tom, M. Molecular basis and biological function of variability in spatial genome organization. *Science (80-.)*. **365**, eaaw9498 (2019).
3. Chen, P., Li, W. & Li, G. Structures and Functions of Chromatin Fibers. *Annu. Rev. Biophys.* **50**, 95–116 (2021).
4. Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* **19**, 192–206 (2018).
5. McGinty, R. K. & Tan, S. Nucleosome structure and function. *Chem. Rev.* **115**, 2255–2273 (2015).
6. Fierz, B. & Poirier, M. G. Biophysics of Chromatin Dynamics. *Annu. Rev. Biophys.* **48**, 321–345 (2019).
7. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665 (2018).
8. Inukai, S., Kock, K. H. & Bulyk, M. L. Transcription factor–DNA binding: beyond binding site motifs. *Curr. Opin. Genet. Dev.* **43**, 110–119 (2017).
9. Miroshnikova, Y. A., Nava, M. M. & Wickstro, S. A. Emerging roles of mechanical forces in chromatin regulation. *J. Cell Sci.* **130**, 2243–2250 (2017).
10. Miroshnikova, Y. A. & Wickström, S. A. Mechanical Forces in Nuclear Organization. *Cold Spring Harb. Perspect. Biol.* (2021). doi:10.1101/cshperspect.a039685
11. Wagh, K. *et al.* Mechanical Regulation of Transcription: Recent Advances. *Trends Cell Biol.* **31**, 457–472 (2021).
12. Dupont, S. & Wickström, S. A. Mechanical regulation of chromatin and transcription. *Nat. Rev. Genet.* **0123456789**, (2022).
13. Bouck, D. C. & Bloom, K. Pericentric Chromatin Is an Elastic Component of the Mitotic Spindle. *Curr. Biol.* **17**, 741–748 (2007).
14. Amar, K., Wei, F., Chen, J. & Wang, N. Effects of forces on chromatin. *APL Bioeng.* **5**, (2021).
15. Strickfaden, H. *et al.* Condensed Chromatin Behaves like a Solid on the Mesoscale In Vitro and in Living Cells. *Cell* **183**, 1772–1784.e13 (2020).
16. Le, T. T. *et al.* Synergistic Coordination of Chromatin Torsional Mechanics and Topoisomerase Activity. *Cell* **179**, 619–631 (2019).
17. Feric, M. & Misteli, T. Phase Separation in Genome Organization across Evolution. *Trends Cell Biol.* 1–15 (2021). doi:10.1016/j.tcb.2021.03.001
18. Narlikar, G. J. Perspectives Phase-separation in chromatin organization. *J. Biosci.* **89**, 1–5 (2020).

19. Razin, S. V & Ulianov, S. V. Divide and Rule: Phase Separation in Eukaryotic Genome Functioning. *Cells* **9**, (2020).
20. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
21. Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **30**, 39–58 (2014).
22. Alberti, S. & Hyman, A. A. Biomolecular condensates at the nexus of cellular stress , protein aggregation disease and ageing. *Nat. Rev. Mol. Cell Biol.* **22**, (2021).
23. Alberti, S., Gladfelter, A. & Mittag, T. Primer Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **176**, 419–434 (2019).
24. Louise, J. *et al.* Protein condensates as aging Maxwell fluids. *Science (80-.)*. **370**, 1317–1323 (2020).
25. Li, P. *et al.* Phase transitions in the assembly of multivalent signalling proteins. *Nature* **483**, 336–340 (2012).
26. Strulson, C. A., Molden, R. C., Keating, C. D. & Bevilacqua, P. C. RNA catalysis through compartmentalization. *Nat. Chem.* **4**, 941–946 (2012).
27. Miete, C. *et al.* Gai2-induced conductin/axin2 condensates inhibit Wnt/ β -catenin signaling and suppress cancer growth. *Nat. Commun.* **13**, 674 (2022).
28. Klosin, A. *et al.* Phase separation provides a mechanism to reduce noise in cells. **468**, 464–468 (2020).
29. Shin, Y. *et al.* Liquid Nuclear Condensates Mechanically Sense and Restructure the Genome. *Cell* **175**, 1481–1491 (2018).
30. Quail, T. *et al.* Force generation by protein – DNA co-condensation. *Nat. Phys.* **17**, 1007–1012 (2021).
31. Boija, A., Klein, I. A. & Young, R. A. Biomolecular Condensates and Cancer. *Cancer Cell* **39**, 174–192 (2021).
32. Cai, D., Liu, Z. & Lippincott-Schwartz, J. Biomolecular Condensates and Their Links to Cancer Progression. *Trends Biochem. Sci.* **46**, 535–549 (2021).
33. Zbinden, A., Pérez-Berlanga, M., De Rossi, P. & Polymenidou, M. Phase Separation and Neurodegenerative Diseases: A Disturbance in the Force. *Dev. Cell* **55**, 45–68 (2020).
34. Alberti, S. & Dormann, D. Liquid–Liquid Phase Separation in Disease. *Annu. Rev. Genet.* **53**, 171–194 (2019).
35. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).
36. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains. *Cell* **175**, 1842–1855 (2018).
37. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. **3958**, (2018).

38. Won-Ki, C. *et al.* Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science (80-.)*. **361**, 412–415 (2018).
39. Chong, S. *et al.* Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
40. Sharma, R. *et al.* Liquid condensation of reprogramming factor KLF4 with DNA provides a mechanism for chromatin organization. *Nat. Commun.* **12**, 5579 (2021).
41. Basu, S. *et al.* Unblending of Transcriptional Condensates in Human Repeat Expansion Disease. *Cell* **181**, 1062–1079 (2020).
42. Long, Q. *et al.* Phase separation drives the self-assembly of mitochondrial nucleoids for transcriptional modulation. *Nat. Struct. Mol. Biol.* **28**, 900–908 (2021).
43. Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225.e24 (2021).
44. Hansen, J. C., Maeshima, K. & Hendzel, M. J. The solid and liquid states of chromatin. *Epigenetics Chromatin* **14**, 50 (2021).
45. Rippe, K. Liquid–Liquid Phase Separation in Chromatin. *Cold Spring Harb. Perspect. Biol.* **14**, (2022).
46. Zhang, M. *et al.* Molecular Organization of the Early Stages of Nucleosome Phase Separation Visualized by Cryo-Electron Tomography. *bioRxiv* 2021.09.01.458650 (2021). doi:10.1101/2021.09.01.458650
47. Gibson, B. A. *et al.* Organization of Chromatin by Intrinsic and Regulated Article Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **179**, 470-484.e21 (2019).
48. Larson, A. G. *et al.* Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
49. Keenen, M. M. *et al.* HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. *Elife* **10**, e64563 (2021).
50. Shakya, A., Park, S., Rana, N. & King, J. T. Article Liquid-Liquid Phase Separation of Histone Proteins in Cells : Role in Chromatin Organization. *Biophysj* **118**, 753–764 (2020).
51. Leicher, R. *et al.* Single-stranded nucleic acid binding and coacervation by linker histone H1. *Nat. Struct. Mol. Biol.* **29**, 463–471 (2022).
52. Turner, A. L. *et al.* Highly disordered histone H1–DNA model complexes and their condensates. *Proc. Natl. Acad. Sci.* **115**, 11964–11969 (2018).
53. Li, B., Carey, M. & Workman, J. L. The Role of Chromatin during Transcription. *Cell* **128**, 707–719 (2007).
54. Bico, J., Reyssat, É. & Roman, B. Elastocapillarity: When Surface Tension Deforms Elastic Solids. *Annu. Rev. Fluid Mech.* **50**, 629–659 (2018).
55. Higashi, T. L. & Uhlmann, F. SMC complexes: Lifting the lid on loop extrusion. *Curr. Opin. Cell Biol.* **74**, 13–22 (2022).

56. Mirny, L. A. & Solovei, I. Keeping chromatin in the loop(s). *Nat. Rev. Mol. Cell Biol.* **22**, 439–440 (2021).
57. Vermunt, M. W., Zhang, D. & Blobel, G. A. The interdependence of gene-regulatory elements and the 3D genome. *J. Cell Biol.* **218**, 12–26 (2018).
58. Hassler, M., Shaltiel, I. A. & Haering, C. H. Towards a Unified Model of SMC Complex Function. *Curr. Biol.* **28**, R1266–R1281 (2018).
59. Uhlmann, F. SMC complexes: From DNA to chromosomes. *Nat. Rev. Mol. Cell Biol.* **17**, 399–412 (2016).
60. Mäkelä, J. & Sherratt, D. SMC complexes organize the bacterial chromosome by lengthwise compaction. *Curr. Genet.* **66**, 895–899 (2020).
61. Hirano, T. Condensins: universal organizers of chromosomes with diverse functions. *Genes Dev.* **26**, 1659–1678 (2012).
62. Perea-Resa, C., Wattendorf, L., Marzouk, S. & Blower, M. D. Cohesin: behind dynamic genome topology and gene expression reprogramming. *Trends Cell Biol.* **31**, 760–773 (2021).
63. Lau, A. C., Nabeshima, K. & Csankovszki, G. The *C. elegans* dosage compensation complex mediates interphase X chromosome compaction. *Epigenetics Chromatin* **7**, 31 (2014).
64. Aragón, L. The Smc5/6 complex: New and old functions of the enigmatic long-distance relative. *Annu. Rev. Genet.* **52**, 89–107 (2018).
65. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102–105 (2018).
66. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science (80-.)*. **366**, 1345–1349 (2019).
67. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science (80-.)*. **366**, 1338–1345 (2019).
68. Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. *Science (80-.)*. **358**, 672–676 (2017).
69. Gutierrez-Escribano, P. *et al.* A conserved ATP- And Scc2/4-dependent activity for cohesin in tethering DNA molecules. *Sci. Adv.* **5**, 1–16 (2019).
70. Kanno, T., Berta, D. G. & Sjögren, C. The Smc5/6 Complex Is an ATP-Dependent Intermolecular DNA Linker. *Cell Rep.* **12**, 1471–1482 (2015).
71. Kim, H. & Loparo, J. J. Multistep assembly of DNA condensation clusters by SMC. *Nat. Commun.* (2016). doi:10.1038/ncomms10200
72. Zernia, S., Kamp, D. & Stigler, J. The loader complex Scc2/4 forms co-condensates with DNA as loading sites for cohesin. *bioRxiv* 2021.12.14.472603 (2021).
73. Ryu, J. *et al.* Bridging-induced phase separation induced by cohesin SMC protein complexes. (2021).
74. Jerkovic', I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary

- methods. *Nat. Rev. Mol. Cell Biol.* **22**, 511–528 (2021).
75. James, F., Iain, W., A., B. W. & Josée, D. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **79**, 347–372 (2022).
 76. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
 77. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **160**, 1145–1158 (2015).
 78. Mcswiggen, D. T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells : diagnosis , caveats , and functional consequences. 1–16 (2019). doi:10.1101/gad.331520.119.
 79. Boopathi, R., Dimitrov, S., Hamiche, A., Petosa, C. & Bednar, J. Cryo-electron microscopy of the chromatin fiber. *Curr. Opin. Struct. Biol.* **64**, 97–103 (2020).
 80. Cuvier, O. & Fierz, B. Dynamic chromatin technologies: From individual molecules to epigenomic regulation in cells. *Nat. Rev. Genet.* **18**, 457–472 (2017).
 81. Chen, P., Li, W. & Li, G. Structures and Functions of Chromatin Fibers.
 82. Zhou, K., Gaullier, G. & Luger, K. Nucleosome structure and dynamics are coming of age. *Nat. Struct. Mol. Biol.* **26**, 3–13 (2019).
 83. Axelrod, D., Thompson, N. L. & Burghardt, T. P. Total internal reflection fluorescent microscopy. *J. Microsc.* **129**, 19–28 (1983).
 84. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5**, 507–516 (2008).
 85. Larson, J. D., Rodgers, M. L. & Hoskins, A. A. Visualizing cellular machines with colocalization single molecule microscopy. *Chem. Soc. Rev.* **43**, 1189–1200 (2014).
 86. Friedman, L. J. & Gelles, J. Mechanism of Transcription Initiation at an Activator-Dependent Promoter Defined by Single-Molecule Observation. *Cell* **148**, 679–689 (2012).
 87. Li, S., Zheng, E. B., Zhao, L. & Liu, S. Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *Cell Rep.* **28**, 2689–2703 (2019).
 88. Ticau, S., Friedman, L. J., Ivica, N. A., Gelles, J. & Bell, S. P. Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. *Cell* **161**, 513–525 (2015).
 89. Greene, E. C., Wind, S., Fazio, T., Gorman, J. & Visnapuu, M.-L. Chapter 14 - DNA Curtains for High-Throughput Single-Molecule Optical Imaging. in *Single Molecule Tools: Fluorescence Based Approaches, Part A* (ed. Walter, N. G. B. T.-M. in E.) **472**, 293–315 (Academic Press, 2010).
 90. Kim, E., Kerssemakers, J., Shaltiel, I. A., Haering, C. H. & Dekker, C. DNA-loop extruding condensin complexes can traverse one another. *Nature* **579**, (2020).

91. Ashkin, A. Acceleration and Trapping of Particles by Radiation Pressure. *Phys. Rev. Lett.* **24**, 156–159 (1970).
92. Bustamante, C. J., Chemla, Y. R., Liu, S. & Wang, M. D. Optical tweezers in single-molecule biophysics. *Nat. Rev. Methods Prim.* **1**, 25 (2021).
93. Roman, R. *et al.* Co-condensation of proteins with single- and double-stranded DNA. *Proc. Natl. Acad. Sci.* **119**, e2107871119 (2022).
94. Morin, J. A. *et al.* Sequence-dependent surface condensation of a pioneer transcription factor on DNA. *Nat. Phys.* **18**, 271–276 (2022).
95. Whitley, K. D., Comstock, M. J. & Chemla, Y. R. Chapter Six - High-Resolution Optical Tweezers Combined With Single-Molecule Confocal Microscopy. in *Single-Molecule Enzymology: Nanomechanical Manipulation and Hybrid Methods* (eds. Spies, M. & Chemla, Y. R. B. T.-M. in E.) **582**, 137–169 (Academic Press, 2017).
96. Wasserman, M. R. *et al.* Replication Fork Activation Is Enabled by a Single-Stranded DNA Gate in CMG Helicase. *Cell* **178**, 600–611 (2019).
97. Nguyen, T. *et al.* Chromatin sequesters pioneer transcription factor Sox2 from exerting force on DNA. *Biorxiv* (2022).
98. Mallik, R., Kundu, A. & Chaudhuri, S. High mobility group proteins: the multifaceted regulators of chromatin dynamics. *Nucl.* **61**, 213–226 (2018).
99. Malarkey, C. S. & Churchill, M. E. A. The high mobility group box: The ultimate utility player of a cell. *Trends Biochem. Sci.* **37**, 553–562 (2012).
100. Michal, Š. *Biochimica et Biophysica Acta* HMGB proteins : Interactions with DNA and chromatin. **1799**, 101–113 (2010).
101. Holmes, Z. E. *et al.* The Sox2 transcription factor binds RNA. *Nat. Commun.* **11**, 1805 (2020).
102. Hou, L., Srivastava, Y. & Jauch, R. Molecular basis for the genome engagement by Sox proteins. *Semin. Cell Dev. Biol.* **63**, 2–12 (2017).
103. Sánchez-Moreno, I., Coral-Vázquez, R., Méndez, J. P. & Canto, P. Full-length SRY protein is essential for DNA binding. *Mol. Hum. Reprod.* **14**, 325–330 (2008).
104. Travers, A. A. Priming the nucleosome: A role for HMGB proteins? *EMBO Rep.* **4**, 131–136 (2003).
105. Mandke, P. & Vasquez, K. M. Interactions of high mobility group box protein 1 (HMGB1) with nucleic acids: Implications in DNA repair and immune responses. *DNA Repair (Amst)*. **83**, 102701 (2019).
106. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
107. K., M. A. *et al.* Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science (80-)*. **368**, 1460–1465 (2020).
108. Cadigan, K. M. & Waterman, M. L. TCF/LEFs and Wnt Signaling in the Nucleus. *Cold*

- Spring Harb. Perspect. Biol.* **4**, (2012).
109. Sampson, E. M. *et al.* Negative regulation of the Wnt– β -catenin pathway by the transcriptional repressor HBP1. *EMBO J.* **20**, 4500–4511 (2001).
 110. Feric, M. *et al.* Self-assembly of multi-component mitochondrial nucleoids via phase separation. *EMBO J.* **40**, e107165 (2021).
 111. Krainer, G. *et al.* Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. *Nat. Commun.* **12**, 1085 (2021).
 112. Prendergast, L. & Reinberg, D. The missing linker: emerging trends for H1 variant-specific functions. *Genes Dev.* **35**, 40–58 (2021).
 113. Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* **16**, 1439–1453 (2015).
 114. Happel, N. & Doenecke, D. Histone H1 and its isoforms: Contribution to chromatin structure and function. *Gene* **431**, 1–12 (2009).
 115. Izzo, A. & Schneider, R. The role of linker histone H1 modifications in the regulation of gene expression and chromatin dynamics. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 486–495 (2016).
 116. Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014).
 117. Garcia-Saez, I. *et al.* Structure of an H1-Bound 6-Nucleosome Array Reveals an Untwisted Two-Start Chromatin Fiber Conformation. *Mol. Cell* **72**, 902–915.e7 (2018).
 118. Dombrowski, M., Engholm, M., Dienemann, C., Dodonova, S. & Cramer, P. Histone H1 binding to nucleosome arrays depends on linker DNA length and trajectory. *Nat. Struct. Mol. Biol.* **29**, 493–501 (2022).
 119. Flanagan, T. W. & Brown, D. T. Molecular dynamics of histone H1. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 468–475 (2016).
 120. Richardson, R. T. *et al.* Nuclear Autoantigenic Sperm Protein (NASP), a Linker Histone Chaperone That Is Required for Cell Proliferation*. *J. Biol. Chem.* **281**, 21526–21534 (2006).
 121. Kepert, J. F., Mazurkiewicz, J., Heuvelman, G. L., Tóth, K. F. & Rippe, K. NAP1 Modulates Binding of Linker Histone H1 to Chromatin and Induces an Extended Chromatin Fiber Conformation*. *J. Biol. Chem.* **280**, 34063–34072 (2005).
 122. George, E. M. & Brown, D. T. Prothymosin a is a component of a linker histone chaperone. *FEBS Lett.* **584**, 2833–2836 (2010).
 123. Karetsov, Z. *et al.* Prothymosin α modulates the interaction of histone H1 with chromatin. *Nucleic Acids Res.* **26**, 3111–3118 (1998).
 124. Heidarsson, P. O. *et al.* Release of linker histone from the nucleosome driven by polyelectrolyte competition with a disordered protein. *Nat. Chem.* (2022). doi:10.1038/s41557-021-00839-3
 125. Shen, X. & Gorovsky, M. A. Linker Histone H1 Regulates Specific Gene Expression but

- Not Global Transcription In Vivo. *Cell* **86**, 475–483 (1996).
126. Yuhong, F. *et al.* H1 Linker Histones Are Essential for Mouse Development and Affect Nucleosome Spacing In Vivo. *Mol. Cell. Biol.* **23**, 4559–4572 (2003).
 127. Fan, Y. *et al.* Histone H1 Depletion in Mammals Alters Global Chromatin Structure but Causes Specific Changes in Gene Regulation. *Cell* **123**, 1199–1212 (2005).
 128. Raouf, A. *et al.* Mammalian linker-histone subtypes differentially affect gene expression in vivo. *Proc. Natl. Acad. Sci.* **100**, 5920–5925 (2003).
 129. Kim, K. *et al.* Linker Histone H1.2 Cooperates with Cul4A and PAF1 to Drive H4K31 Ubiquitylation-Mediated Transactivation. *Cell Rep.* **5**, 1690–1703 (2013).
 130. Seung-Min, Y., Ju, K. B., Laura, N. T. & I., S. A. H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation. *Proc. Natl. Acad. Sci.* **110**, 1708–1713 (2013).
 131. Hansol, L., Raymond, H. & Cory, A.-S. Msx1 Cooperates with Histone H1b for Inhibition of Transcription and Myogenesis. *Science (80-.).* **304**, 1675–1678 (2004).
 132. Kim, K. *et al.* Isolation and Characterization of a Novel H1.2 Complex That Acts as a Repressor of p53-mediated Transcription *. *J. Biol. Chem.* **283**, 9113–9126 (2008).
 133. Iwafuchi-Doi, M. *et al.* The Pioneer Transcription Factor FoxA Maintains an Accessible Nucleosome Configuration at Enhancers for Tissue-Specific Gene Activation. *Mol. Cell* **62**, 79–91 (2016).
 134. Wen, Y. *et al.* Dense Chromatin Activates Polycomb Repressive Complex 2 to Regulate H3 Lysine 27 Methylation. *Science (80-.).* **337**, 971–975 (2012).
 135. An, W., van Holde, K. & Zlatanova, J. The Non-histone Chromatin Protein HMG1 Protects Linker DNA on the Side Opposite to That Protected by Linker Histones *. *J. Biol. Chem.* **273**, 26289–26291 (1998).
 136. Thomas, J. O. & Travers, A. A. HMG1 and 2, and related architectural DNA-binding proteins. *Trends Biochem. Sci.* **26**, 167–174 (2001).
 137. Bustin, M., Catez, F. & Lim, J. The Dynamics of Histone H1 Function in Chromatin. **17**, 617–620 (2005).
 138. Catez, F. *et al.* Network of Dynamic Interactions between Histone H1 and High-Mobility-Group Proteins in Chromatin. **24**, 4321–4328 (2004).
 139. Postnikov, Y. V. & Bustin, M. Functional interplay between histone H1 and HMG proteins in chromatin. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1859**, 462–467 (2016).
 140. Zlatanova, J. & van Holde, K. Binding to four-way junction DNA: a common property of architectural proteins? *FASEB J.* **12**, 421–431 (1998).
 141. Cato, L., Stott, K., Watson, M. & Thomas, J. O. The Interaction of HMGB1 and Linker Histones Occurs Through their Acidic and Basic Tails. *J. Mol. Biol.* **384**, 1262–1272 (2008).

142. Schultz, T. F., Spiker, S. & Quatrano, R. S. Histone H1 enhances the DNA binding activity of the transcription factor EmBP-1. *J Biol Chem* **271**, 25742–25745 (1996).
143. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
144. Michael, A. K. *et al.* article: Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science (80-.)*. **1465**, 1460–1465 (2020).
145. Tanaka, H. *et al.* GATA3 with nucleosomes. *Nat. Commun.* (2020). doi:10.1038/s41467-020-17959-y
146. Mivelaz, M. *et al.* Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor Article Chromatin Fiber Invasion and Nucleosome Displacement by the Rap1 Transcription Factor. *Mol. Cell* **77**, 488-500.e9 (2020).
147. Donaghey, J. *et al.* Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
148. Liu, Z. & Kraus, W. L. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Mol. Cell* **65**, 589-603.e9 (2017).
149. Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
150. Chronis, C. *et al.* Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell* **168**, 442-459.e20 (2017).
151. Zhu, F. *et al.* The interaction landscape between transcription factors and the nucleosome. *Nature* 76–81 (2018). doi:10.1101/240598
152. Takahashi, K. & Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat. Rev. Mol. Cell Biol.* **17**, 183–193 (2016).
153. Schaefer, T. & Lengerke, C. SOX2 protein biochemistry in stemness , reprogramming , and cancer : the PI3K / AKT / SOX2 axis and beyond. *Oncogene* **39**, 278–292 (2020).
154. Sarkar, A. & Hochedlinger, K. The Sox family of transcription factors: Versatile regulators of stem and progenitor cell fate. *Cell Stem Cell* **12**, 15–30 (2013).
155. Masui, S. *et al.* Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).
156. Li, M., Carlos, J. & Belmonte, I. Deconstructing the pluripotency gene regulatory network. *Nat Cell Biol* **20**, 382–392 (2018).
157. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
158. Chen, X. *et al.* Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133**, 1106–1117 (2008).
159. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
160. Hou, L. *et al.* Concurrent binding to DNA and RNA facilitates the pluripotency

- reprogramming activity of Sox2. **48**, 3869–3887 (2020).
161. Ferrari, S. *et al.* SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* **11**, 4497–4506 (1992).
 162. Reményi, A. *et al.* Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev.* **17**, 2048–2059 (2003).
 163. Scaffidi, P. & Bianchi, M. E. Spatially Precise DNA Bending Is an Essential Activity of the Sox2 Transcription Factor. *J. Biol. Chem.* **276**, 47296–47302 (2001).
 164. Yesudhas, D., Badoo, M., Anwar, M. A., Panneerselvam, S. & Choi, S. Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors. *Genes (Basel)*. **8**, (2017).
 165. Jagga, B. *et al.* Structural basis for nuclear import selectivity of pioneer transcription factor SOX2. *Nat. Commun.* **12**, 1–11 (2021).
 166. Weina, K. & Utikal, J. SOX2 and cancer: current research and its implications in the clinic. *Clin. Transl. Med.* **3**, e19 (2014).
 167. Li, S., Zheng, E. B., Zhao, L. & Liu, S. repeat: Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Article Nonreciprocal and Conditional Cooperativity Directs the Pioneer Activity of Pluripotency Transcription Factors. *CellReports* **28**, 2689-2703.e4 (2019).
 168. Sherwood, R. I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
 169. Shrinivas, K. *et al.* Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol. Cell* **75**, 549-561.e7 (2019).
 170. Zuo, L. *et al.* Loci-specific phase separation of FET fusion oncoproteins promotes gene transcription. *Nat. Commun.* **12**, 1491 (2021).
 171. Wei, M. *et al.* Nucleated transcriptional condensates amplify gene expression. *Nat. Cell Biol.* **22**, (2020).
 172. Wu, J. *et al.* Modulating gene regulation function by chemically controlled transcription factor clustering. *Nat. Commun.* **13**, 2663 (2022).
 173. Terlecki-Zaniewicz, S. *et al.* Biomolecular condensation of NUP98 fusion proteins drives leukemogenic gene expression. *Nat. Struct. Mol. Biol.* **28**, 190–201 (2021).
 174. Dolgin, E. Drug startups coalesce around condensates. *Nat. Biotechnol.* **39**, 123–125 (2021).
 175. A., K. I. *et al.* Partitioning of cancer therapeutics in nuclear condensates. *Science (80-.)*. **368**, 1386–1392 (2020).
 176. Biebricher, A. S. *et al.* The impact of DNA intercalators on DNA and DNA-processing enzymes elucidated through force-dependent binding kinetics. *Nat. Commun.* **6**, 7304 (2015).

177. Murade, C. U., Subramaniam, V., Otto, C. & Bennink, M. L. Interaction of oxazole yellow dyes with DNA studied with hybrid optical tweezers and fluorescence microscopy. *Biophys. J.* **97**, 835–843 (2009).
178. Xue, B., Oldfield, C. J., Van, Y., Keith, A. & Uversky, V. N. Protein intrinsic disorder and induced pluripotent stem cells. *Mol. Biosyst.* **8**, 134–150 (2012).
179. Bustamante, C., Chemla, Y. R., Forde, N. R. & Izhaky, D. Mechanical Processes in Biochemistry. *Annu. Rev. Biochem.* **73**, 705–748 (2004).
180. Ohta, S. *et al.* Proteomics Analysis with a Nano Random Forest Approach Reveals Novel Functional Interactions Regulated by SMC Complexes on Mitotic Chromosomes *. *Mol. Cell. Proteomics* **15**, 2802–2818 (2016).
181. Venegas, A. B., Natsume, T., Kanemaki, M. & Hickson, I. D. Inducible Degradation of the Human SMC5/6 Complex Reveals an Essential Role Only during Interphase. *Cell Rep.* **31**, (2020).
182. Decorsière, A. *et al.* Hepatitis B virus X protein identifies the Smc5/6 complex as a host restriction factor. *Nature* **531**, 386–389 (2016).
183. Diaz, M. & Pecinka, A. Scaffolding for repair: Understanding molecular functions of the SMC5/6 complex. *Genes (Basel)*. **9**, (2018).
184. Yu, Y. *et al.* Integrative analysis reveals unique structural and functional features of the Smc5/6 complex. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
185. Palecek, J. J. & Gruber, S. Kite Proteins: a Superfamily of SMC/Kleisin Partners Conserved Across Bacteria, Archaea, and Eukaryotes. *Structure* **23**, 2183–2190 (2015).
186. Kolesar, P., Stejskal, K., Potesil, D., Murray, J. M. & Palecek, J. J. Role of Nse1 Subunit of SMC5/6 Complex as a Ubiquitin Ligase. *Cells* **11**, (2022).
187. A., A. E. *et al.* Nse2, a Component of the Smc5-6 Complex, Is a SUMO Ligase Required for the Response to DNA Damage. *Mol. Cell. Biol.* **25**, 185–196 (2005).
188. Zabradý, K. *et al.* Chromatin association of the SMC5/6 complex is dependent on binding of its NSE3 subunit to DNA. *Nucleic Acids Res.* **44**, 1064–1079 (2016).
189. Vondrova, L. *et al.* A role of the Nse4 kleisin and Nse1/Nse3 KITE subunits in the ATPase cycle of SMC5/6. *Sci. Rep.* **10**, 9694 (2020).
190. Taschner, M. *et al.* Nse5/6 inhibits the Smc5/6 ATPase and modulates DNA substrate binding. *EMBO J.* **40**, 1–23 (2021).
191. Hallett, S. T. *et al.* Nse5/6 is a negative regulator of the ATPase activity of the Smc5/6 complex. *Nucleic Acids Res.* **49**, 4534–4549 (2021).
192. Pradhan, B. *et al.* The Smc5/6 complex is a DNA loop extruding motor. 1–19 (2022).
193. Yu, Y. *et al.* Cryo-EM structure of DNA-bound Smc5 / 6 reveals DNA clamping enabled by multi-subunit conformational changes. 1–10 (2022).
doi:10.1073/pnas.2202799119/-/DCSupplemental.Published
194. Palecek, J. J. SMC5/6: Multifunctional player in replication. *Genes (Basel)*. **10**, (2019).
195. Nasim, A. & Smith, B. P. GENETIC CONTROL OF RADIATION SENSITIVITY IN

- SCHIZOSACCHAROMYCES POMBE. *Genetics* **79**, 573–582 (1975).
196. De Piccoli, G. *et al.* Smc5–Smc6 mediate DNA double-strand-break repair by promoting sister-chromatid recombination. *Nat. Cell Biol.* **8**, 1032–1034 (2006).
 197. Potts, P. R., Porteus, M. H. & Yu, H. Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *EMBO J.* **25**, 3377–3388 (2006).
 198. Lena, S. *et al.* Postreplicative Formation of Cohesion Is Required for Repair and Induced by a Single DNA Break. *Science (80-.)*. **317**, 242–245 (2007).
 199. Kegel, A. & Sjögren, C. The Smc5/6 complex: More than repair? *Cold Spring Harb. Symp. Quant. Biol.* **75**, 179–187 (2010).
 200. Bermúdez-López, M. *et al.* ATPase-Dependent Control of the Mms21 SUMO Ligase during DNA Repair. *PLOS Biol.* **13**, e1002089 (2015).
 201. Potts, P. R. & Yu, H. The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nat. Struct. Mol. Biol.* **14**, 581–590 (2007).
 202. Torres-Rosell, J. *et al.* SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions. *Nat. Cell Biol.* **7**, 412–419 (2005).
 203. Betts Lindroos, H. *et al.* Chromosomal Association of the Smc5/6 Complex Reveals that It Functions in Differently Regulated Pathways. *Mol. Cell* **22**, 755–767 (2006).
 204. Bonner, J. N. *et al.* Smc5/6 Mediated Sumoylation of the Sgs1–Top3–Rmi1 Complex Promotes Removal of Recombination Intermediates. *Cell Rep.* **16**, 368–378 (2016).
 205. van der Crabben, S. N. *et al.* Destabilized SMC5/6 complex leads to chromosome breakage syndrome with severe lung disease. *J. Clin. Invest.* **126**, 2881–2892 (2016).
 206. Stephan, A. K., Kliszczak, M. & Morrison, C. G. The Nse2/Mms21 SUMO ligase of the Smc5/6 complex in the maintenance of genome stability. *FEBS Lett.* **585**, 2907–2913 (2011).
 207. Agashe, S. *et al.* Smc5/6 functions with Sgs1–Top3–Rmi1 to complete chromosome replication at natural pause sites. *Nat. Commun.* **12**, 2111 (2021).
 208. Jeppsson, K. *et al.* The Chromosomal Association of the Smc5/6 Complex Depends on Cohesion and Predicts the Level of Sister Chromatid Entanglement. *PLOS Genet.* **10**, e1004680 (2014).
 209. Laflamme, G. *et al.* Structural Maintenance of Chromosome (SMC) Proteins Link Microtubule Stability to Genome Integrity *. *J. Biol. Chem.* **289**, 27418–27431 (2014).
 210. Alt, A. *et al.* Specialized interfaces of Smc5/6 control hinge stability and DNA association. *Nat. Commun.* **8**, 14011 (2017).
 211. Serrano, D. *et al.* Article The Smc5 / 6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine II II Article The Smc5 / 6 Core Complex Is a Structure-Specific DNA Binding and Compacting Machine. *Mol. Cell* **80**, 1025–1038.e5 (2020).
 212. Gutierrez-Escribano, P. *et al.* Purified Smc5/6 Complex Exhibits DNA Substrate

- Recognition and Compaction. *Mol. Cell* **80**, 1039-1054.e6 (2020).
213. Etheridge, T. J. *et al.* Live-cell single-molecule tracking highlights requirements for stable Smc5/6 chromatin association in vivo. *Elife* **10**, e68579 (2021).
 214. W., B. N. A. *et al.* Single-molecule measurements reveal that PARP1 condenses DNA by loop stabilization. *Sci. Adv.* **7**, eabf3641 (2021).
 215. Banigan, E. J. & Mirny, L. A. Loop extrusion: theory meets single-molecule experiments. *Curr. Opin. Cell Biol.* **64**, 124–138 (2020).
 216. Vernerri, P. *et al.* Dynamical reorganization of the pluripotency transcription factors Oct4 and Sox2 during early differentiation of embryonic stem cells. *Sci. Rep.* **10**, 1–12 (2020).
 217. Courchaine, E. M. *et al.* DMA-tudor interaction modules control the specificity of in vivo condensates. *Cell* **184**, 3612–3625 (2021).
 218. Xia, P. *et al.* Sox2 functions as a sequence-specific DNA sensor in neutrophils to initiate innate immunity against microbial infection. *Nat. Immunol.* **16**, 366–378 (2015).
 219. Wei, F. *et al.* Stress fiber anisotropy contributes to force-mode dependent chromatin stretching and gene upregulation in living cells. *Nat. Commun.* **11**, 4902 (2020).
 220. Tajik, A. *et al.* Transcription upregulation via force-induced direct stretching of chromatin. *Nat. Mater.* **15**, 1287–1296 (2016).
 221. Keenen, M. M. *et al.* article: HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. *Elife* **10**, e64563 (2021).
 222. Williams, J. F. *et al.* Phase separation enables heterochromatin domains to do mechanical work. *bioRxiv* 2020.07.02.184127 (2020). doi:10.1101/2020.07.02.184127
 223. Patel, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **162**, 1066–1077 (2015).
 224. Flanagan, T. W. & Brown, D. T. Biochimica et Biophysica Acta Molecular dynamics of histone H1 ☆. *BBA - Gene Regul. Mech.* **1859**, 468–475 (2016).
 225. Yue, H., Fang, H., Wei, S., Hayes, J. J. & Lee, T. H. Single-Molecule Studies of the Linker Histone H1 Binding to DNA and the Nucleosome. *Biochemistry* **55**, 2069–2077 (2016).
 226. Borgia, A. *et al.* Extreme disorder in an ultrahigh-affinity protein complex. *Nat. Publ. Gr.* (2018). doi:10.1038/nature25762
 227. Cotton, R. W. & Barbara, A. Nucleosome dissociation at physiological ionic strengths. *Nucleic Acids Res.* **9**, 445–457 (1981).
 228. Hagerman, T. A. *et al.* Chromatin Stability at Low Concentration Depends on Histone Octamer Saturation Levels. *Biophys. J.* **96**, 1944–1951 (2009).
 229. Baumann, C. G. *et al.* Stretching of Single Collapsed DNA Molecules. *Biophys. J.* **78**, 1965–1978 (2000).
 230. Guilhas, B. *et al.* ATP-Driven Separation of Liquid Phase Condensates in Bacteria. *Mol. Cell* **79**, 293-303.e4 (2020).

231. Osunsade, A. *et al.* A Robust Method for the Purification and Characterization of Recombinant Human Histone H1 Variants. *Biochemistry* **58**, 171–176 (2018).
232. Crickard, J. B., Moevus, C. J., Kwon, Y., Sung, P. & Greene, E. C. Rad54 Drives ATP Hydrolysis-Dependent DNA Sequence Alignment during Homologous Recombination. *Cell* **181**, 1380–1394 (2020).
233. Bolte, S. & Cordelières, F. P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
234. Costes, S. V. *et al.* Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophys. J.* **86**, 3993–4003 (2004).
235. Sbalzarini, I. F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195 (2005).
236. Mangeol, P., Prevo, B. & Peterman, E. J. G. KymographClear and KymographDirect: two tools for the automated quantitative analysis of molecular and cellular dynamics using kymographs. *Mol. Biol. Cell* **27**, 1948–1957 (2016).
237. Tarantino, N. *et al.* TNF and IL-1 exhibit distinct ubiquitin requirements for inducing NEMO–IKK supramolecular structures. *J. Cell Biol.* **204**, 231–245 (2014).