MOLECULAR LEVEL INSIGHTS INTO THE FUNCTIONAL RELATIONSHIP BETWEEN CHROMATIN STRUCTURE AND THYMINE DNA GLYCOSYLASE

&

L-RNA INHIBITORS OF THE POLYCOMB REPRESSIVE COMPLEX 2

A Dissertation

by

CHARLES EDWARD DECKARD III

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jonathan Sczepanski
Committee Members,	Wenshe Liu
	Tadhg Begley
	Matthew Sachs
Head of Department,	Simon North

December 2020

Major Subject: Chemistry

ABSTRACT

Active DNA demethylation by Thymine DNA glycosylase (TDG) is essential for embryogenesis and genetic reprogramming, and like all transactions involving DNA, TDG-dependent demethylation must occur within the structural constraints of chromatin. Although a relationship between TDG and chromatin can sometimes be inferred, a molecular understanding is lacking. To address this knowledge gap we developed a method for introducing DNA modifications into chromatin with single nucleotide resolution and then applied this approach to engineering synthetic chromatin with a substrate for studying TDG catalysis, the demethylation intermediate 5-formylcytosine. Subsequent analysis of TDG reactivity within various chromatin ‘environments' provided the first direct evidence that chromatin structure regulates demethylation by TDG. This platform also enabled the first biochemical investigations into the influence of TDG on chromatin structure. While genomic architectures like chromatin loops and phase separated condensates have been observed at genes regulated by TDG, decoupling the direct contribution of TDG to these structures from the complex milieu of regulatory factors inside the cell remains a significant challenge. Utilizing synthetic chromatin, we discover that TDG destabilizes local fiber compaction while, simultaneously, promoting chromatin oligomerization (i.e. condensation) via bridging interactions mediated by the intrinsically disordered N-terminal domain. Overall, this work establishes a new paradigm for TDG as a dynamic regulator of chromatin structure and organization.

Finally, a novel therapeutic approach for targeting PRC2 based on mirror image RNAs is described. PRC2 is a long non-coding (lnc)RNA-associated chromatin modifying

enzyme that is frequently dysregulated in breast, prostate, ovarian, and esophageal cancers and, therefore, is a prime drug target. Inspired by evidence that PRC2 promiscuously (i.e. non-specifically) interacts with RNAs that contain guanine quadruplexes (G4s), we sought to develop inhibitors that mimicked these PRC2-interacting motifs. In this regard we exploited the synthetic enantiomer of naturally occurring D-RNA (L-RNA) and established that PRC2 binds RNA irrespective chemistry and that L-RNA G4s are potent inhibitors of PRC2-substrate interactions.

DEDICATION

This dissertation is dedicated to my parents, Jennifer and Charles Deckard Jr., and my sisters, Dakota, Amber, and Brooke. If not for their constant love and support throughout my life, whether it pertained to sports or graduate school, I would not be the person I am today, nor would I have been able to pursue a Ph.D. in chemistry.

ACKNOWLEDGEMENTS

I would like to express my great appreciation to Dr. Jonathan Sczepanski for giving me the opportunity to join his lab. His patience in the early days, as well as constant guidance, mentorship, and high standard of practicing science helped shape my development as both a mentor and researcher. I would also like to thank my committee members – Dr. Wenshe Liu, Dr. Tadgh Begley, and Dr. Matthew Sacchs – for their expertise and constructive criticism during classes and my preliminary examination. Their thoughtful advice and teachings were extremely helpful in developing my approach to potential research projects.

A special thank you also goes out to Dr. Deb Banerjee, Dr. Joanna San Pedro, and Dr. Brian Young for the countless hours they invested in teaching and training me when I first entered graduate school. In addition, I would like to thank Dr. Sougata Dey, Dr. Adam Kabza, Dr. Yu Zeng, and Nandini Kundu for always being willing to discuss experiments and give advice. Without the generous help and guidance offered by everyone in the Sczepanski lab my successes in research, and graduate school in general, would not have been possible. Lastly, I would like to thank my family, as well as the friends and colleagues I have had the opportunity to meet while at Texas A&M. All of these people contributed greatly to my positive experience of grad school and helped make these years in College Station an enjoyable experience.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Jonathan Sczepanski, Wenshe Liu, and Tadhg Begley of the Department of Chemistry and Professor Matthew Sachs of the Department of Biology.

The AFM data depicted in Chapter 2 was provided by Dr. Megan Elinski and Meridith Buzbee from the Batteas laboratory in the Department of chemistry at Texas A&M University. The UDG/APE1 kinetics discussed in Chapter 2 and Appendix A were conducted by fellow lab member Dr. Deb Banerjee. Dr. Banerjee also constructed the laboratory's first 12×601-containing plasmid, which is referenced throughout this work. The unpublished confocal microscopy experiments presented in Chapter 4 were conducted by Dr. Justin Smolen of the Department of Chemistry from Texas A&M University. The APE1 variants described in Figure C.14 were provided by Alexander Flemming from the Burrows laboratory, Department of Chemistry at the University of Utah.

All other work conducted for thise thesis was completed by the student independently.

Funding Sources

This work was supported by the Cancer Prevention and Research Institute of Texas under Grant Number [RR1500038] and The Welch Foundation under Grant Number [A1909].

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

The genetic blueprint for all life on Earth is written using the same four-letter language. These four letters (A, G, C and T) represent the four major bases of DNA and their exact sequence encodes the instructions for building the protein machinery required to operate a cell. The information stored by DNA is translated into proteins by a third biopolymer, RNA which, together constitute the central dogma of molecular biology and is the framework that governs all information flow in biology (Figure 1.1). While the four DNA bases have remained constant over the course of evolution, the size and complexity of what constitutes a genome has changed dramatically, as have the mechanisms in which it is stored within the cell. The genomes of complex eukaryotic organisms, from plants to humans, are encoded by billions of base pairs of DNA and, due to their large size, could not possibly fit inside a cell without some form of co-packaging partner. Ultimately, DNA-packaging is achieved through hierarchical organization into nucleoprotein complexes known as chromatin, the structure of which is expected to regulate how the underlying genetic information is processed, repaired, and read.

Due to its complex nature, chromatin has remained a challenge to study and thus a detailed understanding of its structure function relationships has remained elusive. In Chapter 1, I provide a literature review and relevant background on research related to the study of DNA and chromatin. Chapter 2 discusses the development of a methodology for introducing synthetic DNA modifications into chromatin *in vitro*. Chapters 3 and 4

describe research focused on understanding the relationship between chromatin structure and active DNA demethylation, with emphasis on the effects that chromatin has on catalysis by thymine DNA glycosylase (TDG) (Chapter 3) as well as how TDG, in turn, influences chromatin structure (Chapter 4). Finally, Chapter 5 details the discovery and investigation of interactions between L-nucleic acids and the promiscuous RNA-binding protein: Polycomb Repressive Complex 2 (PRC2).

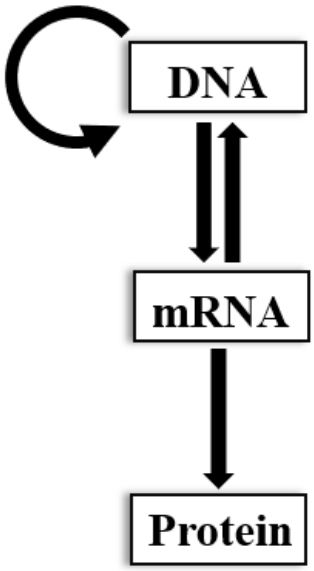


Figure 1.1: The central dogma of molecular biology. Arrows represent the directional flow of information as it occurs naturally.

1.1 DNA, Nucleosomes, and Chromatin

1.1.1 DNA and the genome

Deoxyribonucleic acid (DNA) is a relatively stable nucleotide biopolymer and its constituent bases are capable of forming sequence-specific interactions through a distinct pattern of hydrogen (H-) bonding known as Watson Crick (WC) base pairing: Adenine (A) pairs with Thymine (T) via two H-bonds and Guanine (G) pairs with Cytosine (C) via three H-bonds (Figure 1.2) [1]. WC base-pairing enables molecular recognition and is the basis for template-dependent synthesis which is important for ensuring accurate transmission of information between generations. In DNA, and ribonucleic acid (RNA), individual nucleotides are covalently linked to each other via phosphodiester linkages between the 3'-hydroxyl of one monomer and the 5'-phosphate of the next (5'→3' polarity) [1]. DNA naturally adopts a B-form helix due to the preferred C2'-endo sugar conformation of its sugar whereas, owing to the presence of the 2'-hydroxyl group, the ribose sugar is predominantly C3'-endo and the helical structure of RNA is mainly A-form [2], [3] (Figure 1.3). These subtle chemical differences render RNA flexible and capable of folding into myriad secondary structures (see Chapter 5) [4]–[7], making RNA well-suited for the wide range of functions it serves within the cell. However, a tendency to fold intra-molecularly, combined with increased sensitivity to alkaline conditions due to the 2'-hydroxyl, renders RNA less than ideal for storing the entire *genome* of complex life.

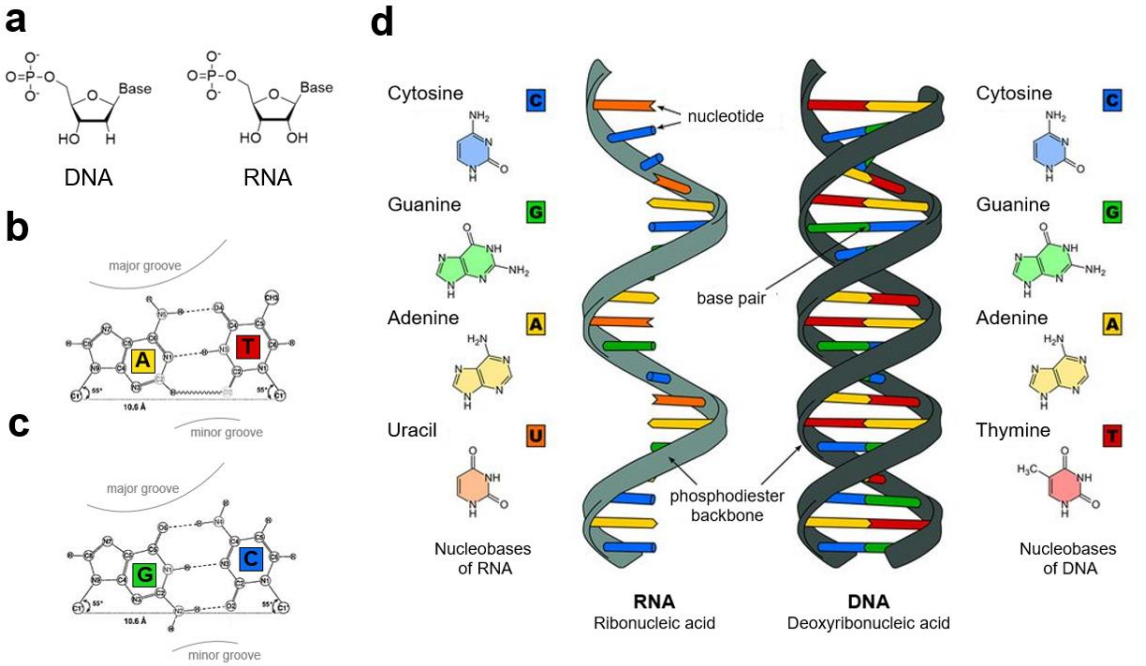


Figure 1.2: Structure of DNA, RNA, and their constituent nucleobases. (a) Chemical structure of DNA and RNA. (b) A-T and (c) G-C base pair with the Watson Crick (WC) interface, major-, and minor-grooves labelled. (d) Cartoon depiction of the structure of DNA and RNA. This figure was adapted from refs. [8] and [9].

The term "genome" refers to an organism's complete set of genetic information and can include anywhere from a few thousand, to tens of thousands, of individual *genes* [10]–[13]. A gene is the minimal coding DNA sequence that, when transcribed and/or translated, produces a functional RNA or protein product. There is not a clear correlation between organism complexity and number of genes [12], which is reflected by the fact that humans contain ~25,000 protein-coding genes compared to the >50,000 genes that have been detected in some protozoa [12], [14].

There is substantial evidence that non-coding (nc)RNAs serve functions just as essential as proteins, for there are many more non-protein coding than coding genes in the human genome (see Chapter 5). Nonetheless, evolving biological complexity requires a

significant number of DNA base pairs (e.g. ~3.5 billion bp for human haploid) and presents the cell with the challenge of fitting this information in a small nucleus. This storage problem is complicated by the rod-like structure of DNA, DNA has a tendency to undergo strand breaks from physical strain and damaging agents [15]–[17], all of which are life threatening to the cell (see section *1.2.1* and *1.2.2*). To facilitate the safe storage of DNA and to facilitate its condensation a class of packaging proteins known as histones has evolved in eukaryotes. Histones are highly conserved proteins and are essential components of chromatin and its fundamental subunit: the nucleosome.

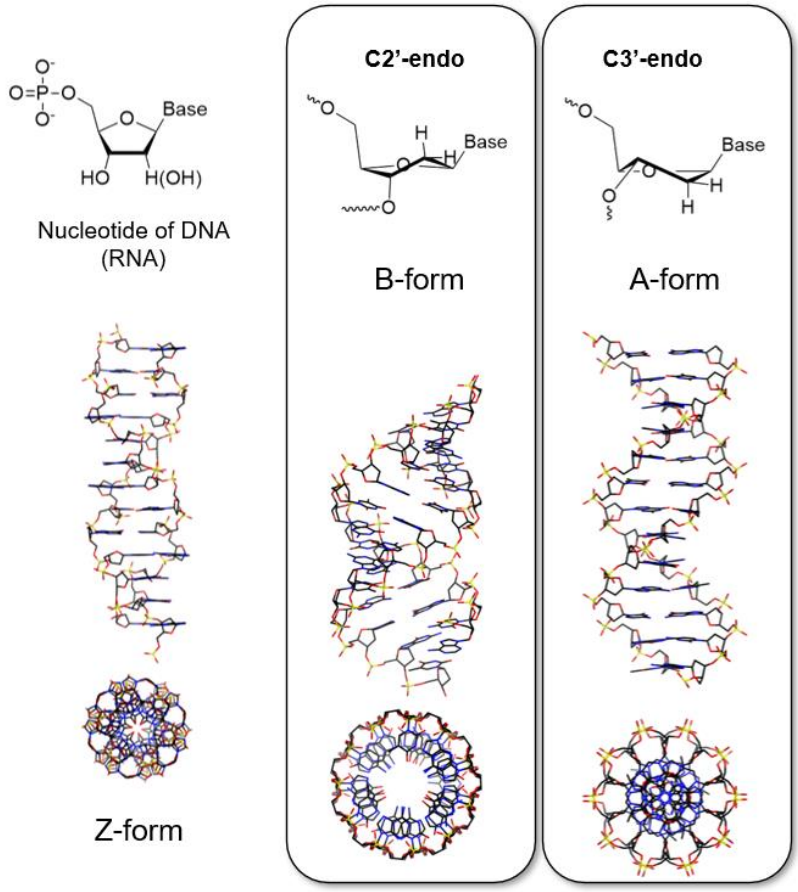


Figure 1.3: Chemical structures of DNA and RNA. The two most common sugar conformations and associated helical structures are depicted in boxes to the right and the lesser common Z-form DNA, which is characterized by a left handed helix resulting from a *syn* base-sugar conformation, is shown in the bottom left hand corner. This figure was adapted from ref. [18].

1.1.2 The Nucleosome Core Particle

It was extremely challenging to pin down either protein or nucleic as the true source of heredity. This was because nuclear extracts from eukaryotic cells contained extensive quantities of a substance having equal parts DNA and protein (i.e. histones). This mysterious substance was named chromatin because of its appearance when stained with dyes and it remained a challenge to study for decades following its identification [19]. Major questions surrounding DNA's role in genetics were answered through the 1940's and 50's when Avery and co-workers demonstrated virulent DNA transformed into bacteria was still lethal [20]. These initial findings were confirmed by Hershey and Chase, who showed the protein component of viruses remained bound to the surface of cells while DNA was the genetic material that was injected and replicated inside them, proving DNA could transmit genetic information [21]. These observations, combined with Watson and Crick's description of the double helix structure of DNA from crystals obtained by Rosalind Franklin, led to DNA's template-dependent mechanism of replication and synthesis to be realized [1].

While research involving DNA flourished, early efforts to characterize the abundant protein component of chromatin were complicated by proteases and solubility issues. Acid extraction of nuclei yielded stoichiometric quantities of four distinct histones, H2A, H2B, H3, and H4, and a fifth (H1) with approximately half abundance [22]–[25]. This revealed some insight into the potential modular nature of chromatin but, still, the tendency of histones to aggregate, both individually and in mixtures, rendered a molecular understanding of DNA-histone complexes out of reach. A substantial breakthrough came

when Roger Kornberg reported the X-ray diffraction pattern of native chromatin [22]. Kornberg's data demonstrated that chromatin consisted of a repeating unit, and this repeating unit could only form in the presence of H2A, H2B, H3, and H4 (the major histones). From this, Kornberg proposed that chromatin's structure was based on a repeating unit, the nucleosome, which consists of ~200 bp of DNA and 2 copies of each of the four major histones. Because H1 is present at half abundance and is not required for the repeating unit to form, he proposed it bound outside of the nucleosome. Decades of subsequent work in the chromatin field would support this early theory and shed more light on the intricate mechanisms of chromatin structure and organization [24]–[27] (Figure 1.4).

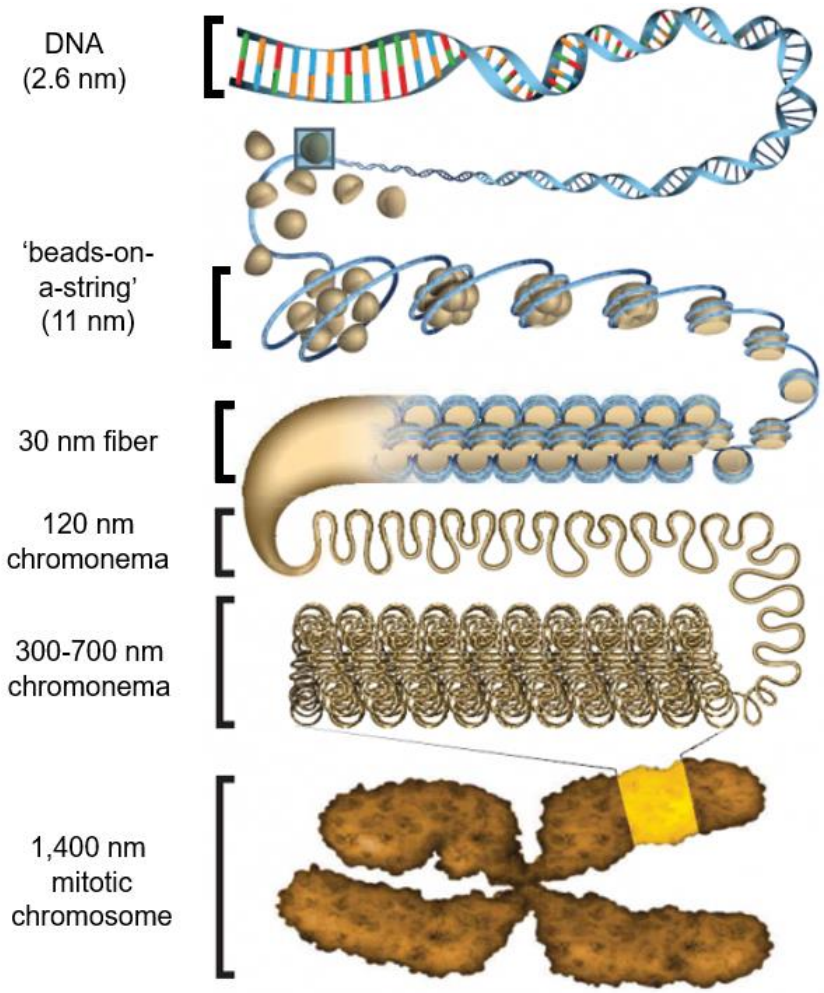


Figure 1.4: Hierarchal depiction of DNA condensation into nucleosomes and chromatin. This figure was adapted from ref. [27].

Early efforts to characterize nucleosomes relied heavily on strategic digestion of chromatin isolated from nuclei with proteases and nucleases. For example, partial digestion of intact chromosomes with micrococcal nuclease (MNase) from *Staphylococcus aureus* produced nucleosome bound DNA fragments ranging from ~160-240 bps in length indicating that nucleosomes, on average, were spaced between ~20-80 bps apart depending on the chromatin [28], [29]. Because MNase's activity towards DNA is dramatically reduced when a protein is present, only free DNA gets digested. In the context of chromatin, free DNA occurs between nucleosomes in the linker region. Interestingly, regardless of the chromatin source, prolonged digestion yielded a homogenous population of nucleosome core particles (NCPs) containing ~147 bp of DNA. These NCPs were homogenous enough to crystalize and provided a low resolution (20 Å) look at the cylindrical/disk-like structure of endogenous nucleosomes [30]. Additionally, this structure demonstrated NCPs were semi-symmetrical about a dyad axis and placed the histone core in the center, serving as a spool around which the DNA was wrapped [30].

As expected, further probing these NCPs with DNase, which unlike MNase is a nuclease capable of cleaving the phosphodiester backbone of nucleosomal DNA, revealed the DNA backbone was most accessible when it was facing outward, away from the histone octamer surface [29], [30]. Interestingly, the DNase-cleavage pattern indicated DNA was most reactive at intervals of 10*n* bp, suggesting the DNA helix repeat length is condensed in nucleosomes relative to natural B-form DNA. These findings were later corroborated by hydroxyl radical-foot printing experiments which offer an unbiased survey of the DNA backbone's solvent accessibility [31].

Researchers supplemented these nuclease analyses with proteolytic cleavage experiments using native and *in vitro* assembled nucleosomes. By comparing the histone degradation pattern of intact nucleosomes and nucleosomes having various combinations of histone proteins, the structural significance of the H3-H4 tetramer in forming the nucleosome core was established [29]. These studies demonstrated that H3-H4 tetramers bound ~80 bp of DNA near the dyad (40 bp on either side), are essential for nucleosome formation, and alone are sufficient to produce stable nucleo-histone complexes (i.e. tetrasomes) that somewhat resemble the NCP [29], [32]. By the 1990s a picture of the nucleosome was coming together, but it wasn't totally clear until Luger *et. al* reported a high-resolution structure of the nucleosome in 1997 [27] (Figure 1.5).

This ground-breaking structure revealed that all major histones (H2A, H2B, H3 and H4) share two distinct structural elements: a globular core and two highly unstructured tail domains at their N- and C- termini (Figure 1.5) [27]. The globular domain of each histone is constructed by a helix-loop-helix motif which contributes to the formation of a protein scaffold around which DNA can be directly wrapped (~1.7 times); the unstructured N- and C-terminal tails extend beyond the core nucleosome, making contacts with DNA and histones on proximal nucleosomes (Figure 1.5) [27], [33], [34]. The physical contacts between histones and DNA involve electrostatic and hydrogen bonding interactions with the phosphate backbone, as well as hydrophobic interactions with the sugar group, and importantly, do *not* involve the DNA major groove and are therefore not sequence-specific. With that being said, the histone octamer is not a perfect sphere and thus, wrapping of the DNA double helix around an irregular, but symmetric, histone core

requires heterogenous bending of the DNA molecule and presents a thermodynamic barrier to nucleosome formation that relies on DNA bending [27], [35], [36].

An ideal DNA packaging protein would exhibit *zero* sequence bias, however, inherent differences in local DNA flexibility and the irregular structure of the octamer core creates a thermodynamic landscape where local minima exist at DNA sequences most capable of bending/contorting at intervals conducive for nucleosome formation. Most notably, significant bending occurs at sites flanking the nucleosome dyad, around super-helical location (SHL) +/- 1.5, where the DNA double helix must accommodate the H3-H4 interface (Figure 1.5). There is not a set genetic code for nucleosome positioning, but the requirement for sharpest bending to occur with 10bp periodicity (when the major groove of the DNA faces inwards towards the nucleosome core and the minor groove faces outward) does bias sequences with periodical repeats of TA/AA/TT motifs recurring every ~10bp towards being stronger nucleosome binders [37]–[39]. This is supported by analysis of naturally occurring, and synthetic, DNA sequences which indicate the strongest nucleosome positioning sequences tend to have GC-rich sequences oscillating out of phase with these AT-rich motifs [37]–[39].

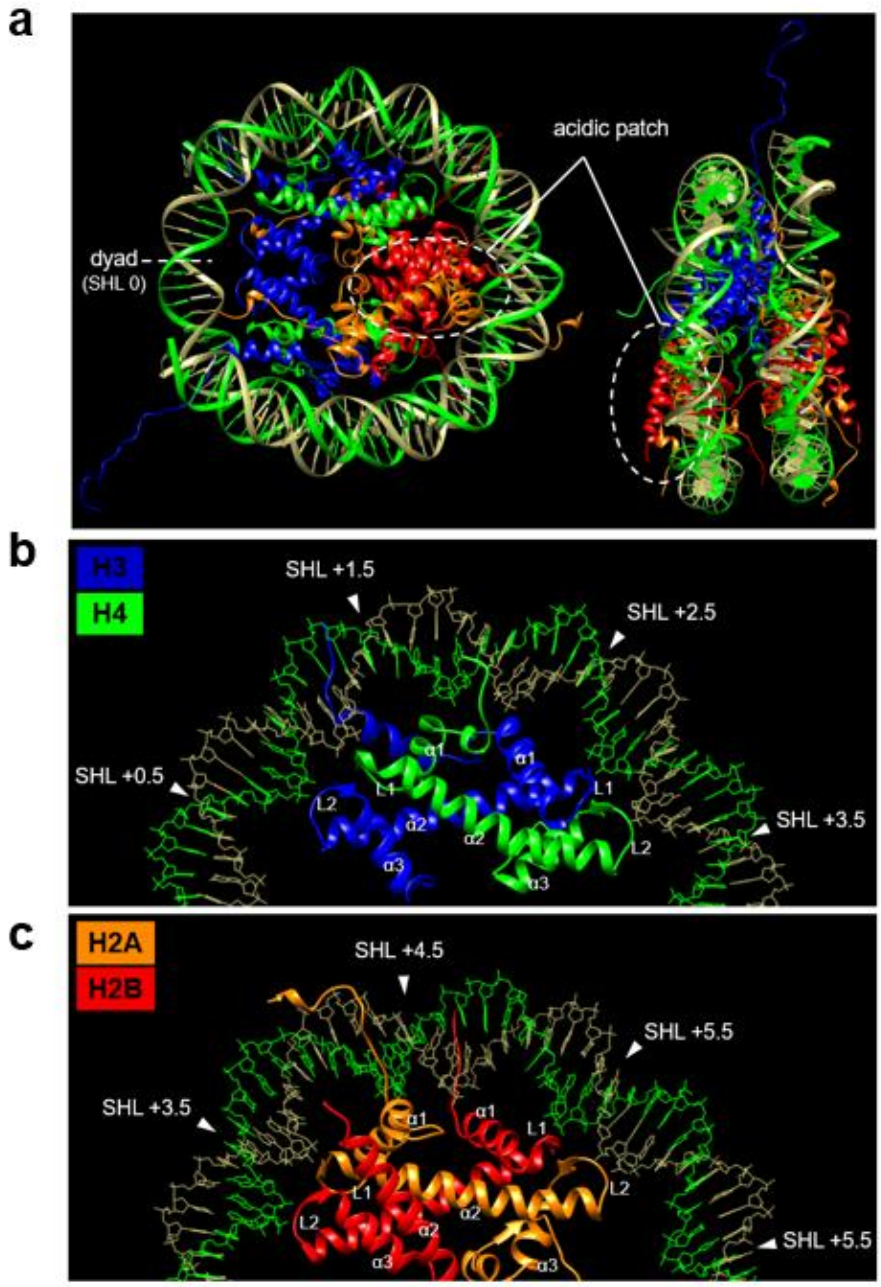


Figure 1.5: High resolution structure of the nucleosome core particle. (a) The Luger nucleosome with the dyad and H2A/H2B acidic patch indicated. (b) Zoomed in view of the H3-H4 and (c) H2A-H2B folded structures interfacing with DNA inside a nucleosome. α= alpha helix, L= loop. This figure was adapted from ref. [27] (pdb: 1aoi.).

1.1.3 Chromatin Structure

1.1.3.1 Primary and secondary structure of chromatin

It was apparent from early work that chromatin was dynamic. Electron micrographs revealed the primary structure of chromatin resembled beads-on-a-string, but was capable of undergoing reversible *intra*-molecular compaction in the presence of cations, including Na^{+}, Mg^{2+}, Ca^{2+}, and spermidine (Spd^{3+}), to produce condensed secondary structures known as chromatin fibers [22], [24], [40], [41]. Cations neutralize the DNA's negatively charged backbone and reduce electrostatic repulsion between DNA strands, enabling *inter*-nucleosomal (*intra*-fiber) interactions mediated by the histone tail domains to form [33], [34], [42]. Perhaps the best example of such an interaction is the H4-tail/H2A acidic patch, which involves K16, R19, K20 and R23 from the H4 N-termini, most notably H4K16, and a set of 8 spatially clustered Asp/Glu residues on the H2A/H2B dimer surface (Figure 1.5, H2A: E56, E61, E64, D90, E91, E92 and H2B: E102, E110) [34], [43]. Together, these histone domains constitute a docking mechanism that facilitates stacking of proximal nucleosomes on top of each other. Non-specific ionic interactions between the histones and nearby DNA further contribute to chromatin folding.

Histone tails mediate the folding of nucleosome arrays into the basic organization unit of chromatin: the 30 nm fiber (see Figures 1.4 and 1.6). The 30 nm fiber is expected to have implications on gene regulation, yet still, its predominant architecture has yet to be established. Two of the main competing models are the one-start (solenoid) and two-start (zig-zag) (Figure 1.6) [26], [44]–[46]. In the one-start model nucleosomes interact

with their neighbors (n + 1) and stack directly on top of each other to form a solenoid-like coil [26], [44]. In the two-start model the linker DNA is rigid and linear. This straightening of the linker DNA positions neighboring nucleosomes side-by-side in 3D space and creates a zig-zag pattern of folding in which every other nucleosome (n + *2*) stacks on top of each other [45], [46]. Some of the best evidence that supports of the zig-zag model comes from the crystal structure of a tetra-nucleosome unit of chromatin, and the cryo-electron microscopy (Cryo EM) structure of a chromatin fiber, which show a two-start chromatin helix [45], [46].

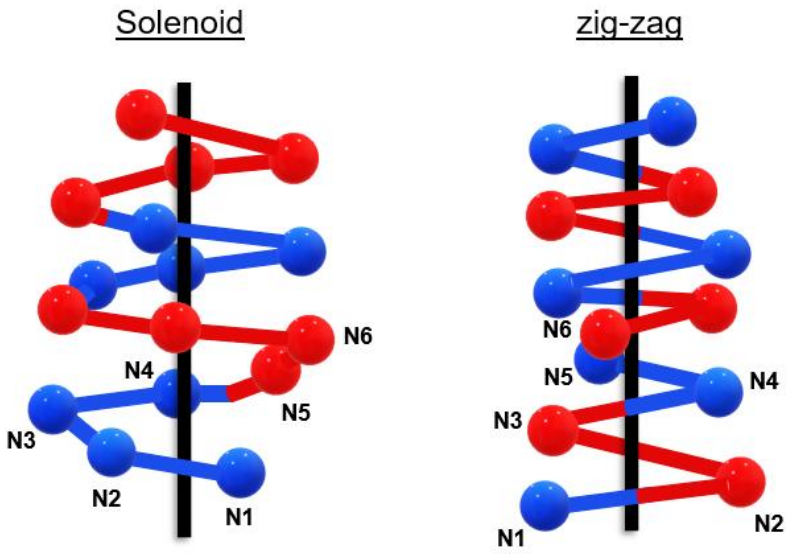


Figure 1.6: Ball and stick representation of the solenoid (one start) and zig-zag (two start) models of the 30 nm fiber. Red and blue spheres indicate nucleosomes which are connected by linker DNAs portrayed as sticks. This figure was adapted from ref. [47].

Structural data indicates shorter DNA linker lengths, which are inherently less flexible, tend to favor a zig-zag chromatin fiber fold, chromatin having more flexible, longer, linker DNAs tends to adopt a solenoid fiber coil, however these preferences can change when linker histones, or other DNA-binding proteins, are present [45], [48]–[50].

Interestingly, some chromatin structural studies suggest the ‘zig-zag’ and ‘solenoid’ models are not mutually exclusive and can co-exist in the same chromatin fiber [45], [48], [49], [51].

Due to non-uniform nucleosome spacing and an abundance of DNA-binding factors *in vivo* it is not likely that a single chromatin secondary structure dominates. An alternative theory to the one- and two-start models suggests that chromatin may be irregular and unstructured in the nucleus, forming an interdigitated framework that can be described by ‘polymer melt’. Somewhat in agreement with this, imaging of native chromatin while still inside the cell revealed that neither version of the archetypical fibers (‘zig-zag’ or ‘solenoid’) were detected in interphase human chromatin [51], [52]. Instead, chromatin visualized in the nucleus had diameters ranging from 5-25 nm and, rather than being regularly folded, local fiber architecture was disordered [51] and could be sub-divided into heterogenous segments with variable nucleosome density and topologies [52]. Ultimately, each of these varying models are supported in their by experimental evidence, therefore it is likely all of these chromatin architectures occur within the cell.

1.1.3.2 Tertiary and higher order structures

The contribution of order and disorder to chromatin organization increases as the structures progress up the hierarchy from nucleosome to chromosome. Apart from the mitotic chromosome, chromatin spends most of its time dispersed within functional genetic compartments in the nucleus. Such compartments partition the genome into transcriptionally active and silent domains, euchromatin and heterochromatin,

respectively [53]–[55]. On a macroscopic scale, this division of the genome is realized in the form of chromosome territories, which comprise millions of base pairs (Mbp) of DNA that, themselves, are further sub-divided into topographically associated domains (TADs) (Figure 1.7) [51], [53]. Architectural proteins, like the CCCTCF-binding factor (CCTCF) and cohesin, play an important role in establishing these compartments, the properties of which are influenced by tertiary chromatin interactions [56], [57].

In vitro, *c*hromatin fibers self-associate (i.e. oligomerize) in a histone-tail dependent manner via non-specific histone-DNA and H4-H2A interactions between separate fibers, or distal regions of the same fiber [22], [24], [40], [41]. *Inter*-fiber binding events are mediated primarily by H4, however, H3 contributes moderately to these interactions and H2A~H2B contribute minimally [41], [42], [58], [59]. *Inter*-molecular interactions are enhanced by cationic salts and some chromatin binding proteins [41], [42], [58], [59]. Chromatin oligomerization, also known as condensation, can produce multi-fiber complexes with properties ranging from densely packed, heterochromatic, and solid-like, to active, euchromatic, and liquid-like. Although chromatin condensation can be attributed to the histone tails and architectural proteins, local features like nucleosome density and linker DNA length produce fibers with differing higher order outcomes. For instance, one study found short, regularly spaced nucleosomes and a ‘zig-zag’ chromatin fiber were enriched in heterochromatin, whereas nucleosome depleted regions and a solenoid architecture were indicative of euchromatin [60]. Furthermore, chromatin looping, a tertiary structure that involves physical contacts between regions of chromatin on the same fiber that has implications for gene expression and domain organization

(Figure 1.7), is also expected to occur and to depend on nucleosome density and linker DNA length [61], [62].

Recent efforts to understand chromatin structure have benefited greatly from computational models that treat chromatin as a flexible polymer with self-interacting and self-avoiding domains [50], [60], [62], [63]. Molecular simulations using such platforms have provided additional insight into how structural variables (e.g. nucleosome density and spacing) contribute to higher order chromatin structure and have been validated against biological data. These relationships are supported by a multitude of genomic sequencing experiments which, have enabled researchers to map chromatin secondary and tertiary structure on a genomic scale. For instance, DNase- and MNase-seq, and more recently Assay for Transposase-Accessible Chromatin (ATAC-seq), have enabled quantification and mapping of nucleosome density *in vivo* by identifying accessible (i.e. euchromatic and active) regions of the genome [64]–[67]. Additionally, researchers have been able to probe the three-dimensional organization of the genome by enzymatically ligating proximal DNAs within fixed nuclei prior to downstream sequencing; this has been accomplished using chromatin conformation capture (3C), chromatin interaction analysis by paired-end tagging (ChIA-PET), and subtle derivatives thereof [53], [64], [68]–[70].

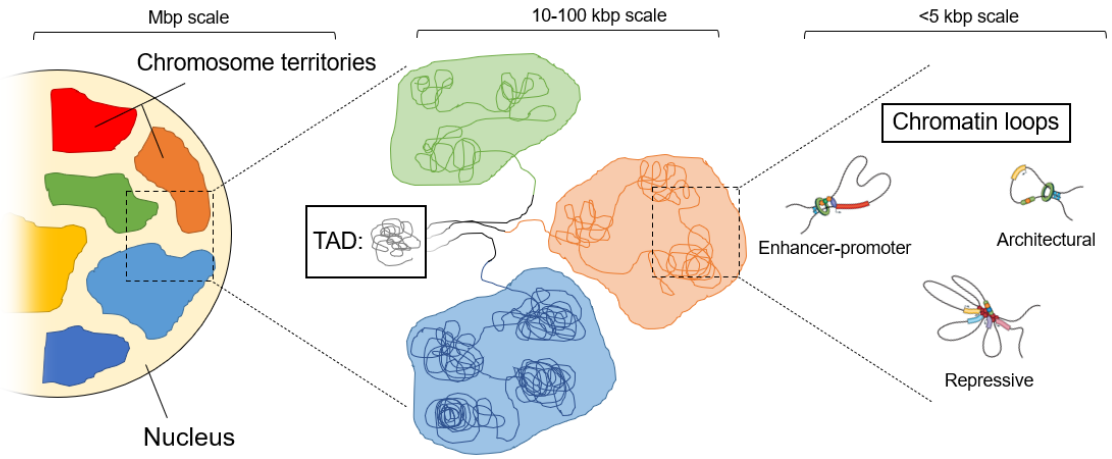


Figure 1.7: Schematic depicting the different scales of chromatin tertiary structure within the cell, from Mega base pair chromosome territories → kilo base pair topographically associated domains → local chromatin domains established via chromatin loops. This figure was adapted from ref. [71].

1.1.4 Chromatin epigenetics

As previously stated, the structure of native chromatin is highly dynamic and heterogenous *in vivo*. This is because a variety of mechanisms to establish local chromatin states that can change depending on the state of the cells. These mechanisms include ATP-dependent chromatin remodeling enzymes, like ISWI and SWI/SNF, which establish nucleosome domains with defined linker lengths [72]. Additionally, linker histones, heterochromatin associated proteins, and transcription factors can bind DNA and create both local and higher order chromatin architectures that are important for transcriptional control [59], [73]–[75]. Ultimately, however, primary sequence information alone is not sufficient to coordinate the elaborate cross-talk required to regulate chromatin state on a cell-by-cell basis. Instead, this is achieved by direct chemical modification to the underlying DNA and histones. The exact identity and distribution of these modifications

add an extra layer of information to the underlying genetic code and this additional information is generally known as epigenetic information.

1.1.4.1 DNA epigenetics

The major DNA epigenetic modification in humans is 5-methylcytosine (5mC). DNA methylation in the form of 5mC is established by DNA methyltransferase (DNMT) enzymes 1, 3a, and 3b. DNMT1 maintains symmetric methylation of CpG dinucleotides after DNA replication and DNMTs 3a/b catalyze the *de novo* methylation of cytosine. Adding a methyl group to C5 alters the chemistry of the DNA major groove and, due to Van der Waals repulsion between the hydrophobic methyl group and the negatively charged phosphate groups of adjacent nucleotides, reduces local DNA flexibility [76]. This methylation-induced rigidity of dsDNA has been shown to negatively influence nucleosome binding affinity while also enhancing intra-molecular compaction within methylated nucleosome arrays [36], [76]–[78]. Additionally, 5mC serves as an epitope for methyl-CpG binding domain (MBD) containing proteins, which have a high affinity for methylated CpG dinucleotides and are capable of established transcriptionally inactive chromatin domains [79].

While 5mC is the major modification in human DNA, representing ~1% of all cytosines, three additional modified cytosine forms have been detected in mammals: 5-hydroxymethyl-, 5-formyl-, and 5-carboxy-cytosine (5hmC, 5fC, and 5caC, respectively). These derivatives are produced enzymatically via oxidation of 5mC (see *1.2.3. Active DNA demethylation*) and have been implicated in a range of genomic processes.

Accumulation of these oxidized 5mC derivatives can counteract the negative effects of DNA methylation on transcription [80]; however, the outcomes associated with these modifications is highly context specific and can also lead to gene silencing [66], [67], [81]–[83]. In many cases, the gene-suppressing functions of oxidized cytosine derivatives, including 5fC and 5caC, are implicated in *epigenetic priming*, a process that involves maintaining a gene in a temporarily silent state until an appropriate stimuli/signal triggers gene expression (see *1.2.3. Active DNA demethylation).*

All oxidized cytosine derivatives have a hydrophilic substituent at the C5 position and, therefore, present a dramatically different major groove to chromatin-associated proteins and DNA-binding factors than 5mC. In further contrast to 5mC, the presence of even just one 5hmC, 5fC, or 5caC in a DNA molecule can enhance the DNA's flexibility, as well as its ability to bind nucleosomes [76], [81].

1.1.4.2 Histone Epigenetics

Ultimately, the methylation status of individual DNA sequences helps govern the establishment and distribution of histone post-translational modifications (PTMs) and is tightly coupled. Compared to DNA, histones undergo a wide range of enzymatically catalyzed covalent modifications. These range from small chemical additions like acetylation, methylation, and phosphorylation [84]–[87], to large modifications including the addition of ubiquitin, ubiquitin-like proteins, and ADP-ribose polymers [88]–[91]. Due to an abundance of nucleophilic lysine side chains, and multiple arginine, serine, and tyrosine residues, the exact location and number of PTMs that occur on an individual

histone, and ultimately a nucleosome, can vary substantially. In general, most histone PTMs occur at lysine residues and are established by histone writers and removed by erasers.

Even though the genetic outcomes associated with histone epigenetics is extremely context-dependent, some generalizing statements can be made. For instance, because lysine residues near the nucleosome core contribute to DNA-binding through electrostatic interactions, PTMs that alter lysine charge states can reduce nucleosome stability and increase DNA accessibility [92]. This effect has been reported for modifications that neutralize lysine's natural positive charge, like acetyl-, propionyl-, butyryl-, and crotonyl-lysine [84], [92]–[95], and is expected to be even more pronounced in nucleosomes harboring other acyl-lysine PTMs, like malonyl-, succinyl-, and glutaryl-lysine, which have charge inverting (+1 → -1) effects [92], [96]. In addition to mediating stability at the nucleosome level, positively charged lysine residues on the histone tail domains interact with DNA on neighboring nucleosomes, as well as the acidic patch on the H2A/H2B dimer surface, and facilitate intra-molecular compaction of nucleosome arrays. Similar to the disruptive effects exerted by acyl-lysine modifications at the nucleosome-level, charge-altering acyl-lysine modifications on the histone tails can also disrupt higher order chromatin folding by antagonizing these *inter*-nucleosomal contacts [85], [94]–[98].

Histone acetyltransferases (HATs) primarily catalyze the transfer of an acetyl group from an acetyl-CoA donor to the ε-amino group of histone lysine residues, but some catalyze the broad spectrum of acyl transfers outlined above. Other enzyme-catalyzed histone PTMs that alter the charge and structure of native lysine, such as phosphorylation

[99], [100], ubiquitination, and ADP-ribosylastion, enhance DNA accessibility at the nucleosome level and antagonize lysine-dependent chromatin folding pathways [86], [89], [91], [99], [101]. In the case of lysine methylation, even though all possible lysine methylation states (me^1-, me^2-, and me^3-) remain cationic, their associated structural outcomes with respect to chromatin and DNA accessibility is more complicated and can be activating (e.g. H3K4$me^{1/2/3}$) or repressing (e.g. H3K9$me^{1/2/3}$ and H3K27 $me^{1/2/3}$) depending on the interactions of the reader protein.

In summary, epigenetic modifications of DNA and histones coordinate specific chromatin environments and facilitate biological processes including gene expression, recombination, replication, and DNA repair. In efforts to gain in depth, molecular-level, insights into the structure-function relationships of chromatin and the myriad chemical modifications that it undergoes *in vivo* researchers have exploited a variety of chemical and molecular biology techniques to introduce histone and DNA modifications into chromatin with residue-level selectivity. Indeed, most of the structural outcomes described in this section can be recapitulated by *synthetic chromatin* strategies; common methods for preparing synthetic chromatin in the laboratory will be reviewed briefly in the section below.

1.1.5 Synthetic Chromatin

Many of the advancements made in our understanding of how DNA sequences govern nucleosome formation and stability came from work conducted in the laboratory of the late Jonathon Widom. One of the most significant contributions from his laboratory

was the discovery of the synthetically derived 601 sequence, which has remarkably strong nucleosome binding capabilities [35]. Although other native positioning elements, such as the alpha satellite and Sea Urchin 5S rRNA gene sequences can produce reliable nucleosomes [36], [37], since first being reported, the 601 sequence has become the gold standard for in vitro nucleosome reconstitution and has enabled unprecedented structural insights into nucleosome and chromatin structure [38]. The 601 sequence, and subtle derivatives thereof, are some of the early developments in the field of 'synthetic chromatin' (see 1.1.5) and have since been exploited by researchers to probe the reactivity of specific nucleotides within a unique nucleosomal (and chromatin) environments [39]–[41], as well as to investigate the relationship between histone post-translational modifications and chromatin structure [42]–[44].

In vitro-reconstituted chromatin fibers, typically assembled from recombinant histone proteins and DNA generated via PCR or plasmid digestion, are the basis for synthetic chromatin. As mentioned above, some of the early applications of synthetic chromatin were focused on determining the influence of a DNA sequence's position within the nucleosome, as well as the extent of local array folding, on DNA reactivity [102]–[104]. These studies primarily utilized generic and sequence-specific endonucleases which react with DNA by hydrolyzing the phosphodiester backbone, creating a free 3'-hydroxyl and 5'-phosphate, and creating double-stranded breaks [102], [104], [105]. Digestion patterns revealed that the accessibility of individual nucleotides to DNA-binding factors is highly dependent on their positions within chromatin (Figure 1.8) and suggested DNA-binding factors may rely on spontaneous fluctuations in DNA-

nucleosome conformations (i.e. DNA-breathing) to gain access to target sequences (Figure 1.8) [103], [106], [107]. The basic trends in DNA accessibility observed from these studies of synthetic chromatin reflect protein-DNA interactions in a very *general* sense, however, they appear to be representative of how most DNA-binding factors are regulated by chromatin structure.

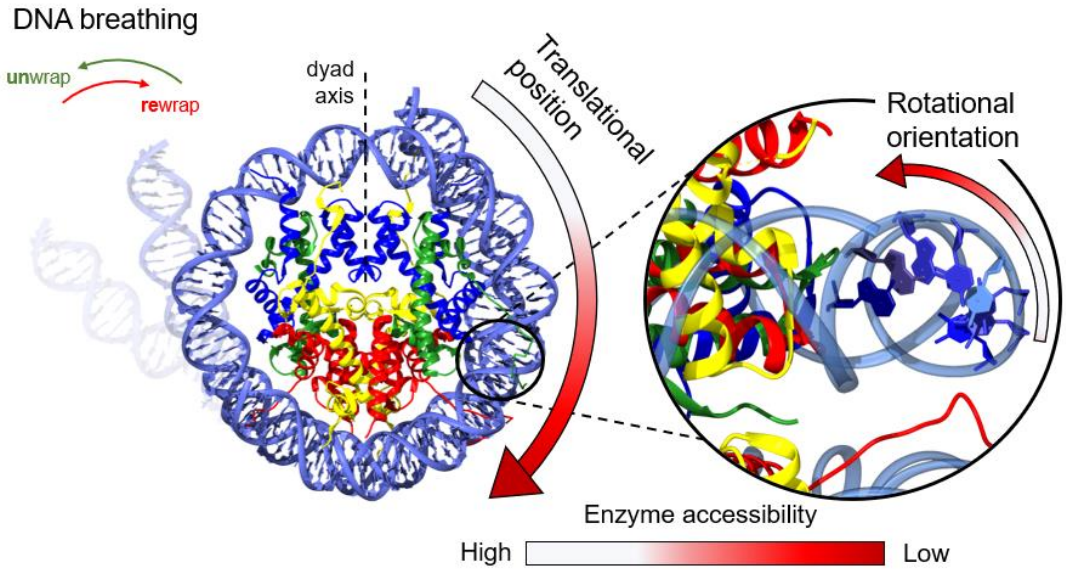


Figure 1.8: The ability of enzymes to access specific DNA sequences within nucleosomes depends on the translational position and rotational orientation of the target nucleotide (pdb: 1kx5). Accessibility is expected to be mediated, in part, by spontaneous unwrapping events known as 'DNA breathing'.

1.1.5.1 Protein Engineering Strategies

There are a number of protein engineering strategies at the disposal of researchers, each capable of producing milligram quantities of histones harboring site-specific post-translational modifications (PTMs) for incorporation into chromatin [88], [101], [108]–[110]. The most straight forward approach is to generate the PTM enzymatically; however, maintaining site-specific control over histone modifying enzymes is not always possible

when doing this. Native chemical ligation (NCL), expressed protein ligation (EPL) cysteine-selective chemistries, and non-canonical amino acid (ncaa) incorporation are all reliable chemical biology alternatives that are therefore used instead [88], [101], [108]–[110].

NCL and EPL employ similar mechanisms whereby trans-esterification covalently attaches a protein with an N-terminal cysteine to a peptide having a C-terminal thioester. Spontaneous S-N acyl shift then yields a native peptide bond between the two, once separate, protein fragments [111]. Apart from protein ligation, the thiol group of cysteine affords a number of other chemical approaches for installing PTMs site-specifically. For instance, disulfide bond formation has been exploited to reversibly attach large ubiquitin-like modifiers to histone proteins [110]; electrophilic mono-, di-, and tri-methylamines have been used to convert cysteines to methyl-lysine analogues [108], [112]; and thiol-reactive moieties, like maleimide, have been used to label histones with chemical probes for applications in studying chromatin structure [89], [101], [113]. Incorporation of ncaa into histones by using novel tRNA/tRNA synthetase pairs is yet another option for generating proteins with defined PTMs and has the added benefit of allowing the modified protein to be purified directly from recombinant sources [114]–[116].

1.1.5.2 DNA Engineering Strategies

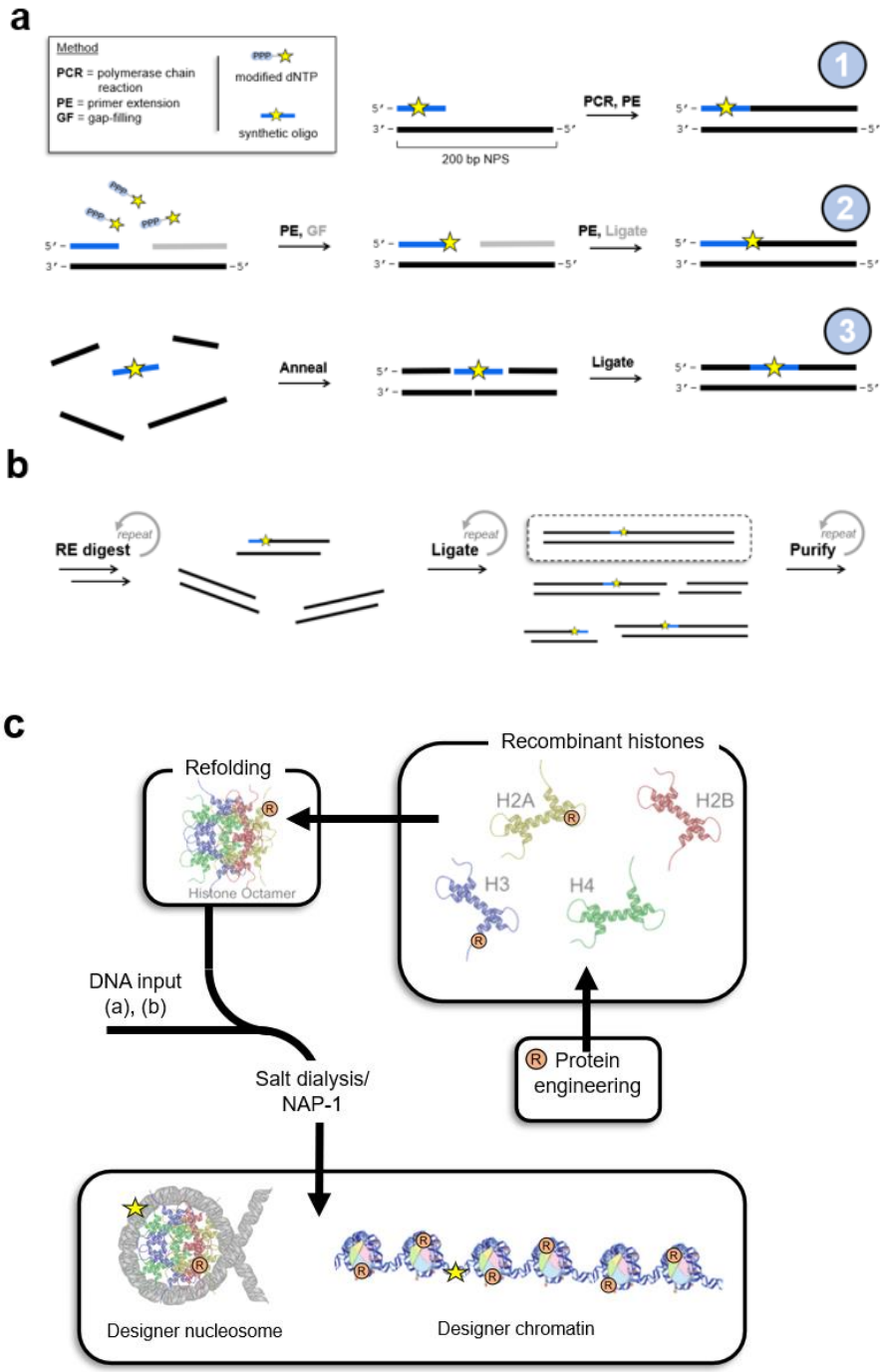
In contrast to histone PTMs, there are far fewer methods for incorporating DNA modifications into large DNA templates with the same precision and efficiency. Some of these challenges are discussed further in Chapter 2 and were the motivation for our

laboratory to pursue a new method for generating site-specifically modified DNA molecules suitable for reconstituting synthetic chromatin. In this section traditional approaches for assembling such templates will be discussed, of which the most common is piece-by-piece assembly through enzymatic ligation [117]–[119] (summarized in Figure 1.9a-b). Such bottom-up engineering is required because of the size limitations of chemical DNA synthesis, which precludes the efficient preparation of DNA oligomers >200 nucleotides [120]. As a result, mixtures of many small DNA duplexes, having complementary overhangs, are often ligated together, directionally, to create larger DNAs suitable for chromatin reconstitutions. When the modification of interest is located internally within the DNA molecule, such piece-by-piece ligation approaches are especially required [117]–[119], [121], [122]. On the other hand, when the modification is near the 5'- or 3'- termini or does not require precise positioning, the modification can be introduced into the underlying sequence via primer extension or PCR using modified primers or dNTPs [81], [117], [123], [124] (Figure 1.9a-*Method 1*). However, the inability of some modifications to be recognized by DNA polymerases could hamper template synthesis via standard PCR and, to overcome some of the limitations of polymerase-mediated synthesis, ligation-based methods are frequently used in conjunction with synthesis methods.

Even though DNA polymerases are incapable of selectively incorporating modified nucleotide triphosphates during templated synthesis, in some cases, polymerases can be coaxed into adding a modified nucleotide triphosphate to the 3'-end of a 5'-overhang, and then the rest of the template can be assembled via ligation (Figure 1.9a-

Method 2). Besides enzymatic polymerization and ligation, which generate DNA molecules with phosphodiester linkages identical to those present in endogenous DNA, chemical alternatives to the traditional phosphodiester bond have been pursued. Perhaps the most notable alternative is the use of triazole linkages generated through Cu^{2+}-catalyzed click reactions between DNAs having 3'- alkynes and 5'- azido groups [125], [126]. By appending these functionalities to duplexes having complementary overhangs, just as above, large DNA molecules can be assembled in a single mixture. Even though unnatural triazole bonds moderately alter local DNA structure and thermodynamic properties (i.e. melting temperature), a synthetic gene assembled via these synthetic linkages was a viable template for PCR and was capable of being transcribed in bacteria cells [125]. Nevertheless, the DNA templates best suited for quantitative biochemical analyses are those which most closely resemble that of native DNA. Once a modified DNA template suitable reconstituting chromatin has been prepared, *synthetic* nucleosomes and chromatin can be reconstituted using salt-dialysis or enzymatic methods (Figure 1.9c)

Figure 1.9: DNA engineering strategies for assembling synthetic chromatin fibers having precisely positioned DNA modifications. (a) Overview of "3" common strategies for achieving site-selective incorporation of modified nucleotides into ~200 bp DNA templates. Method 1: DNA modification is installed within primer via chemical synthesis and then incorporated into the template by PCR/PE. Method 2: Polymerase-dependent extension (or gap-filling) of free 3'-ends in the presence of the desired modified triphosphate. This method requires an additional gap-filling/ligation, or primer extension step. Method 3: Ligation-based assembly of a DNA template by annealing a mixture of complimentary short, synthetic oligonucleotides. Strategic overlay of the short DNAs allows assembly of a full-length duplex and subsequent ligation results in an intact template. (b) Assembly of chromatin-sized DNA arrays from site-specifically modified DNA and nucleosome positioning repeat sequences generated via some combination of the approaches outlined in (a). Restriction enzyme digestion produces complimentary sticky ends that can be ligated enzymatically to generate larger DNA oligomers. The desired, full-length, DNA must be purified by gel electrophoresis, HPLC, or size-selective precipitation; depending on the size and complexity of the template DNA molecule, multiple iterations of digestion, ligation, and purification may be required. (c) Methods for reconstituting synthetic chromatin (NAP-1: nucleosome assembly protein 1).


a
Method
PCR = polymerase chain reaction
PE = primer extension
GF = gap-filling
modified dNTP
synthetic oligo
PCR, PE
200 bp NPS
PE, GF
PE, Ligate
Anneal
Ligate
b
RE digest
repeat
Ligate
repeat
Purify
repeat
c
Recombinant histones
Refolding
H2A
H2B
H3
H4
Histone Octamer
DNA input
(a), (b)
Salt dialysis/
NAP-1
Protein
engineering
Designer nucleosome
Designer chromatin

1.2 DNA Repair

1.2.1 Overview of repair pathways

Genome integrity is maintained through the coordinated efforts of multiple DNA repair pathways. Each pathway is responsible for repairing specific types of DNA damage and are active at different times during a cell's life. This section will briefly overview the major DNA repair pathways, with an emphasis on Base Excision Repair (see **1.2.1.4**).

1.2.1.1 Mismatch repair

DNA mismatch repair (MMR), also referred to as methyl-directed repair (MDR), is responsible for detecting and repairing errors made by DNA polymerases during replication and recombination. This pathway is most active during S phase and is specific to the newly synthesized daughter strand. MMR is initiated when a Mutator S (MutS) homodimer identifies and binds a DNA mismatch. This initial recognition signals a MutL protein to bind the DNA downstream on the template strand and then, through MutS-MutL binding, create a DNA loop [127], [128]. The resulting loop is processed by MMR-related nucleases and is effectively converted to a large gap via removal of the damaged portion of the newly synthesized daughter strand; this gap is then re-filled and sealed by DNA polymerase and DNA ligase [127], [128].

1.2.1.2 Double-strand break repair

dsDNA breaks, which can arise from factors including ionizing radiation and physical strain, pose a serious threat to the cell. dsDNA breaks are repaired via two pathways: Non-homologous end-joining (NHEJ) and Homologous repair (HR) [16], [17], [129], [130]. NHEJ is the most direct approach to repairing dsDNA breaks and involves recognition of newly formed DNA ends by X-ray Cross Complementation (XRCC) protein isoforms [16], [17]. These XRCC proteins interact with other structural proteins to bridge the two broken DNA ends and facilitate their re-alignment. XRCC dimers also act as scaffolds for nucleases, polymerases, and ligases, which process the DNA termini to ensure they are compatible, and then re-attach them via ligation [16], [17]. NHEJ offers a rapid response to repairing double-stranded breaks; however, because it occurs independent of a template and does not require sequence homology, deletions and other alterations can sometimes be introduced when this pathway is used.

In contrast to NHEJ, which can repair dsDNA breaks at any point in the cell cycle and is highly error-prone, HR is template-dependent and only available during S/G2 phase when sister chromatids are accessible [16], [17], [129], [130]. During HR, the dsDNA breaks are processed by helicases and exonucleases to generate large (>100 nt) single stranded regions on their 3'-ends [102], [103]. Then, through interactions with proteins involved in recombination, the free 3'-ends of these broken DNA fragments invade the homologous sequence on sister chromosomes and are used in what is essentially a primer-extension reaction, in which the intact chromosome serves as the template for re-

synthesizing the DNA of the damaged strand. Subsequent end-filling and ligation regenerates an intact DNA molecule on the damaged chromosome.

1.2.1.3 Nucleotide Excision Repair

Nucleotide Excision Repair (NER) is the primary response to bulky DNA lesions that significantly perturb the DNA helix [131]–[133]. These lesions include DNA photo-adducts and covalent DNA-protein cross-links. NER employs a cut-and-fill mechanism whereby the DNA is cleaved at sites adjacent to the lesion and then replaced with a freshly synthesized strand. The NER pathway is highly conserved. In prokaryotes UV-induced repair proteins (Uvr) first locate pyrimidine-pyrimidine dimers (UvrA-UvrB), then hydrolyze the DNA backbone downstream (UvrB) and upstream (UvrC) of the DNA damage site [131], [133]. Helicase activity from UvrD then removes the damaged single strand and forms a gap that is subsequently filled by a DNA polymerase and sealed by a DNA ligase[131], [133]. In eukaryotes the mechanism is more complex and is facilitated by additional factors, but proceeds through a similar mechanism [132].

1.2.2 Base Excision Repair

The Base Excision Repair (BER) pathway, summarized in Figure 1.10, is responsible for repairing DNA lesions that occur directly on the bases themselves [134]–[136]. Due to their electron-rich aromatic rings and nucleophilic heteroatoms, purine and pyrimidine bases are the most reactive components of DNA and frequently undergo alkylation, oxidation and hydrolysis *in vivo* and *in vitro*. In most cases, the resulting modification has negative effects on the DNA's stability and can alter the integrity of its sequence-encoded information. For instance, 3-methyladenine cannot be replicated by DNA polymerases, and 7-methyl-guanine (and -adenine) are prone to spontaneous depurination, which leads to cytotoxic apurinc/apyrimidinic (AP) sites [134]–[136]. Additionally, other DNA bases such as 8-oxoguanine and uracil (which results from hydrolytic deamination of cytosine) exhibit altered WC base-pairing patterns and can be mistakenly read by polymerases during replication, resulting in mutagenic G:C→T:A and G:C→A:U transitions if not repaired [119], [137]–[139].

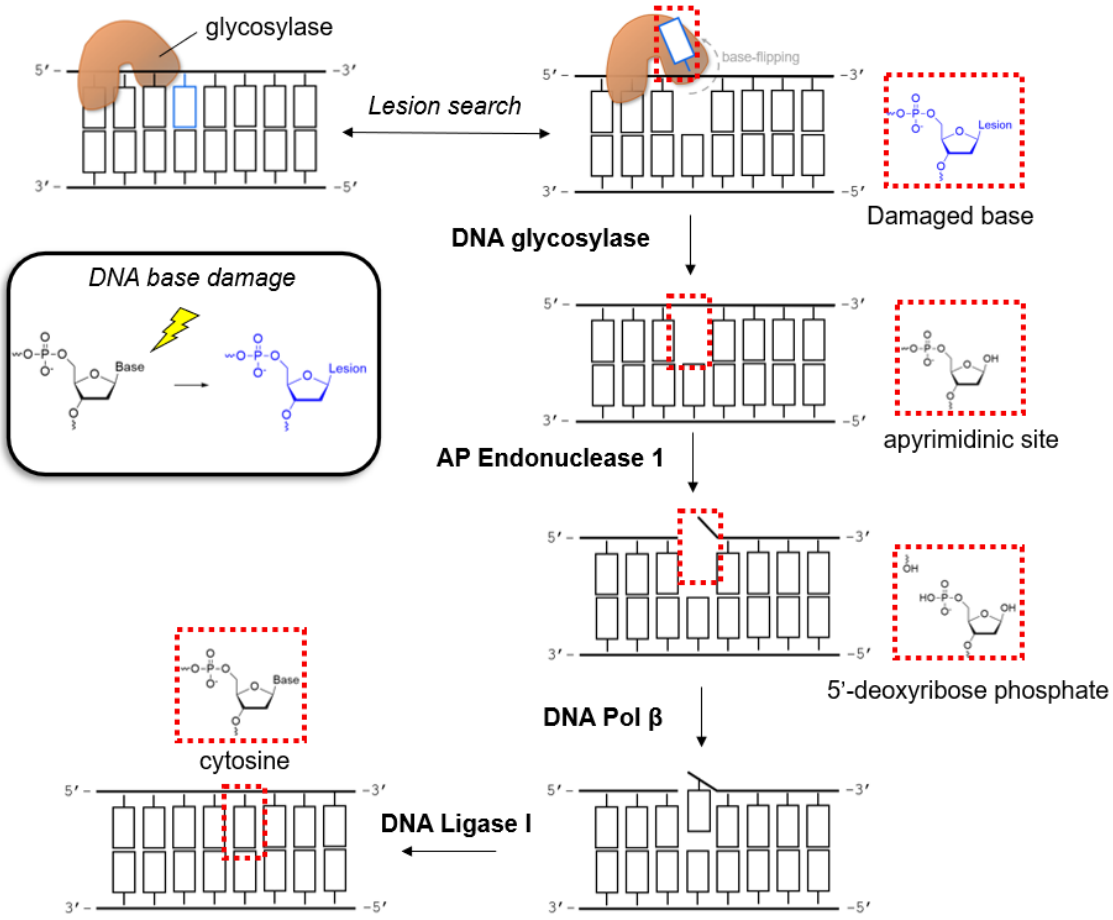


Figure 1.10: BER pathway overview.

BER is the frontline defense against lesions affecting the bases and, as its name implies, is initiated by the removal (i.e. excision) of a damaged base by a specific DNA glycosylase. There have been 11 distinct DNA glycosylases reported to date [135], [136], each of which is capable of recognizing and cleaving the *N*-glycosidic bond of their cognate base to generate an AP site that is subsequently processed by AP endonuclease 1 (APE1) or an AP lyase modality within the DNA glycosylase itself (see **1.2.2.1**) [135], [140]. Next, AP-processing by APE1 results in phosphodiester bond cleavage 5'- to the damaged nucleotide and creates a 5'-deoxyribose phosphate which is ultimately removed by DNA polymerase β (Pol β) and then replaced using a deoxynucleotide triphosphate (dNTP) complementary to the DNA base paired opposite the damage site (Figure 1.10) [135], [136], [141]. The DNA nick produced by Pol β is then sealed by DNA Ligase III, restoring the intact duplex structure [135], [136], [142].

1.2.2.1 DNA glycosylases

DNA glycosylases fall into two general classes: mono-functional and bi-functional. The major difference is that, in addition to catalyzing *N*-glycosidic bond hydrolysis, bifunctional glycosylases also maintain an active site lysine residue capable of forming a Schiff base with the C1'-aldehyde of AP sites, facilitating elimination across the C2'-C3' bond and cleavage of the phosphodiester linkage 3'- to the AP site [135], [140]. Whether monofunctional or bifunctional, DNA glycosylases share a similar strategy for identifying and reacting with damaged bases.

The first challenge glycosylases must overcome is *locating* their cognate DNA substrates within a human genome of ~7×10^9 total base pairs. The exact details regarding the mechanism of lesion search and recognition by DNA glycosylases varies from enzyme to enzyme but, in general, appears to rely on a combination of diffusion and non-specific ionic interactions. It is statistically improbable for a glycosylase, floating around the nucleus in 3-dimensional (3D) space, to rely on random collisions with a damaged DNA molecule to produce enzyme-substrate complexes conducive for chemistry. Instead, it is expected that positively charged residues dispersed along the glycosylase surface helps to maintain the enzyme in close proximity to the negatively charged DNA backbone [143]–[146]. This prolonged proximity mediated by electrostatic attraction, coupled to random diffusion of the enzyme under physiological conditions, results in the effective 1D translocation of the glycosylase along the DNA backbone [144]–[146]. There are several competing theories to explain this translocation phenomena, but they all seem to have some elements of the 'hop-and-slide' mechanism whereby the DNA glycosylase will randomly translocate, or 'tumble', along a DNA molecule for some length in search for lesions before dissociating, briefly, and then re-binding to either the same, or a different, DNA molecule to continue the search process.

DNA scanning by glycosylase enzymes involves constant sampling of nucleotides within the DNA duplex. While some DNA damage imparts local changes to the DNA helix that could readily be detected by surface level interactions, base modifications often offer little structural changes to the DNA helix and require a more invasive mode of searching. In this regard, all known glycosylases employ a base-flipping mechanism that

involves an aliphatic or cationic side chain which intercalates into the DNA duplex and flips the base in question into the enzyme's active site [140], [143], [147], [148]. To overcome the energetic penalty associated with breaking base pairing and base stacking interactions glycosylase enzymes rely on the free energy associated with binding, as well as favorable interactions mediated by the intercalator residue (e.g. an arginine residue that facilitates base-flipping).

DNA glycosylases must perform base-flipping even on undamaged bases, but their active sites are so finely tuned for recognizing specific modifications that the half-life of non-specific DNA-complexes for most glycosylases is short-lived, where a 'specific' complex refers to one that is conducive with product formation and a 'non-specific' complex refers to a complex that cannot yield a product. Once a suitable DNA base has been flipped into the active site an oxocarbenium transition state begins to form as the *N*-glycosidic bond lengthens; then, an activated water molecule strategically positioned within the active site attacks the electrophilic C1' and completes a catalytic event for the glycosylase [147], [149], [150]. Some glycosylases rapidly dissociate from their AP products and move on to continue their search for DNA damage, whereas others remained bound for prolonged periods until displaced by APE1 or other repair factors [135], [140], [151], [152].

1.2.2.2 BER within nucleosomes

As discussed in section *1.1.5* (*Synthetic chromatin*), the nucleosome introduces physical barriers to DNA-binding that have been shown to inhibit repair by BER enzymes.

The two main factors which contribute to nucleosomal DNA reactivity towards nucleases (see Figure 1.8): translational positioning relative to the dyad and rotational orientation relative to the histone octamer surface, have been shown to modulate BER enzyme catalysis with the same general position-dependent effects. In summary, (*1*) BER is *least efficient* near the dyad and on lesions rotated inward towards the histone octamer, (*2*) BER is *most efficient* on lesions facing outward from the histone octamer and located near the nucleosome entry/exit site [136], [153]–[157], and (*3*) BER at linker DNA occurs with similar efficiency as with free DNA [119], [124], [158] in the absence of higher order chromatin structure and linker histones [119], [124], [158]. These general trends in reactivity have been determined from *synthetic chromatin* systems in which nucleosomes and chromatin with site-specifically positioned DNA lesions including 2'-deoxyuridine [154], [157]–[159], 8-oxoguanine [119], [157], and 5-formylcytosine were used to studying BER initiation by Uracil DNA glycosylase (UDG), 8-oxoguanine glycosylase (OGG1), and Thymine DNA glycosylase (TDG) [124], respectively. Similar reconstituted systems have enabled investigation of the effects of lesion positioning within a nucleosome on downstream BER proteins, like APE1 [156], DNA pol β [160], and DNA Ligase I [142]. Lastly, even though all BER enzymes are generally inhibited by nucleosomes, some enzymes in particular are much more negatively affected than others. TDG and OGG1 are two examples of enzymes that are strongly inhibited by nucleosomes and this could reflect the need to tightly regulate their transcription-related activities [137], [161], [162],[124], [153], [157].

1.2.3 Active DNA demethylation

As discussed above, DNA methylation plays a central role in regulating chromatin structure and gene expression. Due to its stability, 5mC was long thought to be a static modification that was only removed through inhibition of DNMT1-mediated maintenance methylation. Eventually, however, it became clear that a *passive*, replication-dependent, mechanism could not account for the rates of demethylation observed on a genomic-scale in early embryos [163], [164], or at hormone-responsive genes in somatic tissue [64], [68], [75], [165], [166]. Together, these observations contributed to the discovery of an *active* mechanism for removing 5mC from the genome, known as active DNA demethylation (Figure 1.11).

Active DNA demethylation proceeds through BER enzymes. In plants, the DNA glycosylases ROS1 (DML1), DML2, DML3, and DME initiates BER at 5mC sites, directly, by cleaving the *N*-glycosidic bond and generating an AP site [167]. Interestingly, there have been no mammalian glycosylases discovered to date with this capability. Instead, there are two main pathways for demethylation in mammals which, like plants, also requires the cell's BER machinery [168]–[171]. The major difference is that before BER is initiated by a DNA glycosylase the methylated base is first converted to a demethylation intermediate by activation induced-cytosine deaminases (AID/APOBEC) or enzymes from the ten-eleven translocation (TET1-3) family of iron oxidases (Figure 1.11). AID/APOBEC enzymes deaminate 5mC to generate G:T mismatches that are substrates for removal by the G:T-specific mismatch glycosylase TDG, or the methyl-domain binding 4 glycosylase MBD4.

Although AID/APOBEC are active in the cell, the primary pathway for initiation of 5mC demethylation is TET-mediated oxidation [169], [172]. TET enzymes catalyze the iterative oxidation of 5mC to 5hmC, 5fC, and 5caC. The two most oxidized 5mC derivatives, 5fC and 5caC, are substrates for Thymine DNA glycosylase (TDG) which converts these modifications to AP sites before handing them off to downstream BER factors [152], [168].

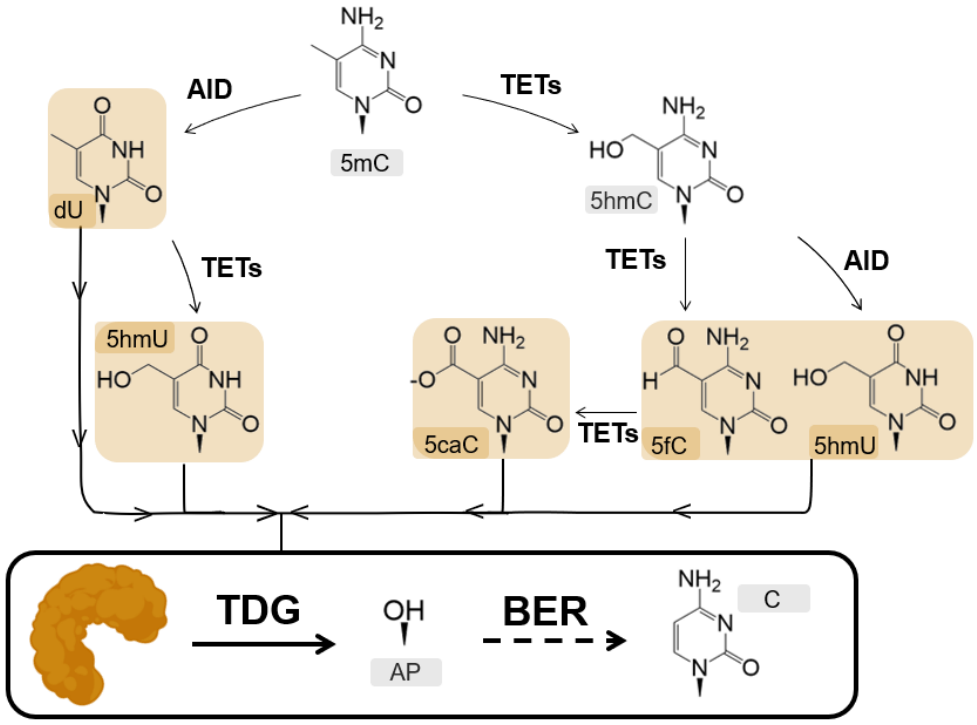


Figure 1.11: TDG-dependent BER is a pipeline for active DNA demethylation.

1.2.3.1 Thymine DNA glycosylase

TDG is capable of cleaving the *N*-glycosidic bond of many natural and unnatural modified-cytosine and uracil nucleobases, with key determinants for activity being the presence of an electron withdrawing substituent at the C5 position and a guanine base

located opposite the damaged base [161], [173]–[175]. Substrates for TDG base excision include thymine glycol, C5-halogenated pyrimidines, and 5-hydroxy- or 5-hydroxymethyl-uracil, among others [161], [174], [175]. Even though TDG exhibits remarkable substrate promiscuity, which is extremely uncommon for a DNA glycosylase, there is little evidence to suggest TDG's biologically relevant targets include all of the substrates it has been reported to cleave *in vitro*. Mainly, endogenous TDG initiates BER at G:T and G:U mismatches. Additionally, as the only glycosylase (known to be) capable of removing 5fC and 5caC from DNA in mammals, TDG also plays an essential role in regulating DNA methylation dynamics throughout the genome [169], [170]. Interestingly, there is now convincing evidence to suggest that 5fC and 5caC are stable modifications capable of regulating gene expression, not just simply demethylation intermediates, which places TDG in the pivotal role of determining whether 5fC and 5caC sites are retained as epigenetic marks or removed via BER [76], [81], [82], [176], [177].

Like all known glycosylases, TDG scans DNA for potential substrates by actively 'flipping' potential target bases out of the DNA duplex into its active site. While most glycosylases employ aliphatic side chains to perform this base-flipping function, TDG uses a critical arginine residue (R275) that, in addition to base-flipping, forms stabilizing π-cation interactions with the DNA in place of the protruded base [178]. Arginine 275 is juxtaposed within an intercalation loop near the active site of TDG that forms sequence-specific hydrogen bonding interactions with the CpG dinucleotide on the adjacent DNA strand, and the neighboring G on the same strand as the modified cytosine (Figure 1.12a-b) [147], [173]. These hydrogen bonding contacts, plus considerable ionic interactions

with the phosphate backbone involving residues from the catalytic domain and the highly basic N-terminal domain contribute to a large footprint (~20 bp) and high stability for TDG-DNA complexes (Figure 1.12c).

The dissociation constants of TDG and DNA substrates harboring a G:T/U mismatch or oxidized cytosine derivatives are in the low nanomolar range (50-150 nM) [178], [179]. Following substrate recognition, TDG catalysis proceeds through an S_N1-like mechanism (Figure 1.12d) which involves stabilizing interactions with the modified base to promote an oxocarbenium intermediate. This intermediate is fully hydrolyzed to product by an active site water molecule, which is positioned for nucleophilic attack via Asn_{140}. After product formation, Asn_{140} maintains hydrogen-bond contacts with the C1'-OH and the abasic sugar remains stably bound in a hydrophobic pocket. These AP-interactions, in addition to the already existing CpG- and phosphate- contacts, result in an extremely high affinity for the product of its own reaction, with TDG-AP complexes having K_d values < 1 nM [178]. Because of this, TDG catalytic models are often generated using single turnover conditions (see Chapter 3) [173], [178], [180].

To summarize, *in vitro* binding studies have demonstrated that TDG's ability to form stable complexes with DNA depends minimally on the identity of the (un)modified cytosine but, instead, is highly dependent on the presence and density of CpG dinucleotides [124], [179]. Although one might expect that tight, non-specific, DNA-binding could be a fatal flaw for a glycosylase, this unique ability of TDG to remain stably bound at CpG dinucleotides allows TDG to serve as a scaffold for nucleating transcription factor complexes [75], [181] and proteins involved in BER [152], [170]. In fact, there is

accumulating evidence to suggest the BER-associated catalytic functions of TDG may be generally less important than its role as a scaffold and transcriptional co-activator (see Chapter 4) [64], [68], [162].

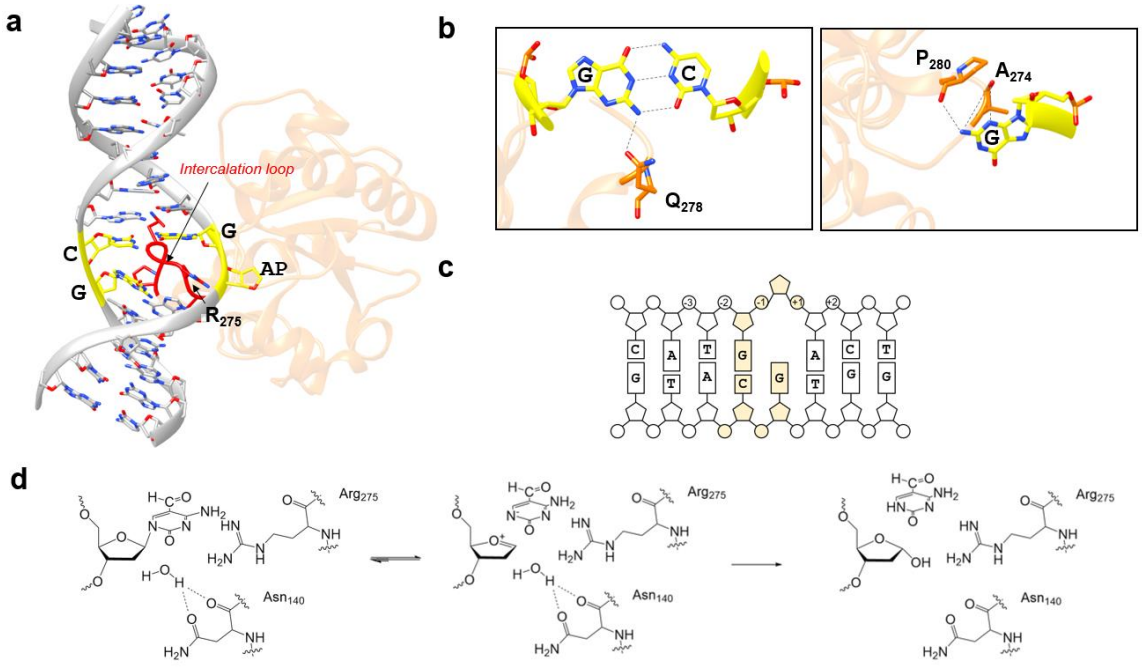


Figure 1.12: TDG structure and mechanism. (a) Structure of TDG-product complex within a CpG dinucleotide; CpG-recognition and base-flipping involves an intercalation loop (red). (b) CpG-specific interactions mediated by the intercalation loop involve Glu_{278} (top), Ala_{274}, and Pro_{280} (bot). (c) Schematic depicting the direct physical contacts between TDG's catalytic domain and DNA. (d) TDG catalytic mechanism. This figure was adapted from refs. [147], [173].

The biological importance of TDG is exemplified by the embryonic lethal phenotype of *TDG knockout* in mice [182], as well as the dramatically altered epigenome and transcriptome profiles that result from TDG deletion in embryonic stem cells (ESCs) [66], [182], [183]. Depleting TDG leads to accumulation of 5fC and 5caC at genomic features associated with important gene-regulatory functions, such as promoters and enhancers [66], [67], [183], and results in a global increase in DNA methylation [182],

[184]. In addition, loss of TDG alters the levels and distribution of histone PTMs associated with poised and transcriptionally active promoters and enhancers, including histone acetylation and H3K4 methylation [66], [182], [183]. In addition to roles in development, TDG-mediated DNA demethylation also plays a key role in transcriptional activation at a subset of hormone-responsive genes, where DNA methylation/demethylation occurs rapidly and in a cyclic, stimuli-dependent, fashion [64], [68], [165], [166].

1.2.3.2 Chromatin landscapes associated with DNA demethylation

As mentioned previously, 5mC accounts for nearly 1% of all cytosines, where 5hmC, 5fC, and 5caC only account for 0.1 and 0.01-0.001%, respectively [183], [185], [186]. Although these fractions might appear insignificant the establishment of these modifications by TET enzymes is tightly regulated and occurs almost exclusively within CpG dinucleotides of gene regulatory elements (GREs), such as enhancers and promoters, where they encode specific epigenetic instructions related to demethylation and transcription [66], [183], [185]–[187]. GREs do not encode protein products, but instead regulate the expression of nearby (*cis*) or distal (*trans*) genes by producing ncRNA transcripts, known as enhancer (e)RNAs, that enhance the recruitment of transcriptional machinery to target genes. These eRNAs and chromatin architectural proteins also contribute to 3D genome architectures, like enhancer promoter loops and other transcriptionally competent higher order structures.

TDG-dependent demethylation at enhancers and promoters coincides with the placement of histone variants H2A.Z and H3.3 near these sites [67], [165]. Both variants are structurally similar to their non-variant counterparts, but subtle residue changes leads to reduced stability and enhanced DNA accessibility at the nucleosome level when they are present [67], [188], [189]. The main source of instability in these nucleosomes is H3.3 [190]. The subtle differences between H3.3 and H3.1 reside on the globular domain at the H2A/H2B dimer interface [188], [191]. The position of these altered side chains decreases octamer stability and favors dimer eviction during transcription [186], [188], [190], [191]. Enhanced DNA accessibility at the nucleosome level is also associated with alternative chromatin architectures mediated by the H2A.Z acidic patch [191], [193]–[195], which is enlarged relative to the corresponding H2A acidic patch and has been shown to both *promote* intra-fiber compaction and *inhibit* inter-fiber oligomerization (i.e. condensation) [195]. Nucleosomes containing the histone variants H2A.Z and H3.3 are inherently unstable and primed for demethylation and transcription [193].

The de-stabilizing effects mediated by variant histones is enhanced by acetylation, which occurs on most histones at active enhancers (see 1.1.4.2) [191]. Acetylation is a hallmark feature of transcriptionally active chromatin and is deposited on H2A.Z, H3.1, H3.3, and H4, with notable sites including: H3K18ac, H3K27ac, H3K122ac, H4K16ac, and H2A.Zac (K4, K7, K11) [64], [84], [93], [196], [197]. These acetylation sites are enriched in poised and active promoters, and enhancers, and overlap considerably with 5fC and 5caC.

One contributing factor to the observed distribution of histone acetylation and oxidized 5mC derivatives throughout the genome is the direct, physical, interaction of TDG with HATs like CBP/p300, which directly links TDG-dependent demethylation to transcription [162]. TDG stimulates acetylation of H3 by p300 *in vitro* and *in vitro* [198] and TDG/p300 form ternary complexes with the transcriptional regulators estrogen receptor alpha (ERα), retinoic acid receptor (RAR), growth arrest and DNA damage inducible alpha (GADD45α) [64], [75], [181], and these complexes have direct effects on chromatin structure and gene expression. Currently, models in which TDG recruits p300 to bound genomic sites, or where TDG-p300 complexes are co-recruited via interactions with other regulatory factors, are favored. In addition to directing CBP/p300 acetylation to the underlying chromatin, TDG itself is also acetylated at lysine residues within its N-terminal domain (K70, K94, K95 and K98) and these PTMs of TDG regulate its DNA-binding and catalytic activity, as well as interactions with other proteins, including APE1 [162], [199].

In addition to the known transcriptional coactivators CBP/p300, and the various nuclear receptors mentioned before, TDG has also been indirectly linked to numerous other TFs like CTCF, FOXA1, GATA3, MED3, and BRD4 [64], [200]–[203]. Interestingly, TFs like GATA3 and BRD4, among others, co-localize with steroid receptors known to interact with TDG, and also bind to GREs implicated in DNA demethylation. Intriguingly, recent biochemical experiments have demonstrated that many transcriptionally competent higher order chromatin structures require factors that interact with TDG. These structures include enhancer-promoter loops [64], [68], [75] and,

recently, a form of chromatin condensation which equates to liquid-liquid phase separation. In light of recent evidence it appears TDG could be directly involved in these structural processes that coordinate 3D genomic architecture at sites marked for demethylation and transcriptional activation [64]. The higher order chromatin structures associated with active DNA demethylation, and TDG's direct role in this process, will be explored in Chapter 4.

1.3 Thesis Aims

Despite evidence that chromatin organization and TDG-mediated DNA demethylation are tightly coordinated *in vivo*, a molecular-level understanding of the structure-function relationships of TDG and chromatin has remained elusive. Overcoming this knowledge gap requires moving beyond *in vivo* studies and necessitates quantitative *in vitro* experiments using chemically defined systems. Unfortunately, due to the heterogeneity and combinatorial complexity of native chromatin, such experiments cannot be performed on chromatin isolated from cells. In the absence of a straight-forward approach for assembling chemically defined nucleosome arrays (i.e. synthetic chromatin) suitable for studying TDG-dependent demethylation, an initial focus of this research was the "Development of a 'plug-and-play' approach for assembling chromatin with site-specifically positioned DNA modifications" (Chapter 2). This methodology enables single-base resolution control over any synthetically accessible DNA modification and its position within a defined array of 12 nucleosomes.

This ‘plug-and-play’ approach was used to engineer chromatin fibers with DNA demethylation intermediates (5fC) and fluorescent probes for measuring inter-nucleosomal distances. Work with this system revealed that “Chromatin structure and the pioneer transcription factor FOXA1 regulate TDG-mediated removal of 5fC from DNA” (Chapter 3) and offered the first ever quantitative insight into chromatin structure’s role in regulating demethylation by TDG. Lastly, this body of work established a novel function for TDG as a chromatin architectural protein. By broadening the scope of our investigations to consider the implications of non-specific DNA-binding by TDG on local and higher order chromatin structure, it was discovered that “Reversible chromatin condensation by thymine DNA glycosylase” (Chapter 4) could be a mechanism for coupling chromatin condensation (i.e. phase separation) to genomic sites bound by TDG.

Building off the general chemical biology techniques applied to studying chromatin structure and TDG, a second research focus was the discovery and characterization of a novel class of PRC2 inhibitors based on mirror image L-RNA G-quadruplexes (Chapter 5). Not only does this work represent a new modality for targeting nucleic acid-binding proteins in disease, but it represents the first known example of an endogenous protein interacting with a nucleic acid sequence independent of chirality.

CHAPTER 2

DEVELOPMENT OF A 'PLUG-AND-PLAY' APPROACH FOR ASSEMBLING CHROMATIN WITH SITE-SPECIFICALLY POSITIONED DNA MODIFICATIONS

Chemically defined nucleosome arrays (i.e. synthetic chromatin) have become a common tool for elucidating fundamental molecular mechanisms of chromatin regulation by enabling researchers to carry out quantitative measurements under precisely defined experimental conditions [99], [115], [116]. In particular, the availability of multiple protein engineering strategies for introducing site-specific post-translational modifications (PTMs) has enabled assembly of synthetic chromatin fibers harboring defined arrangements of modified histones (see 1.1.5.1) [88], [90], [96], [204]. These chromatin substrates have been instrumental in revealing how specific histone PTMs alter DNA accessibility and local chromatin structure, in many cases providing molecular details that could not be obtained using simpler systems (i.e., mono- or di-nucleosomes) [89], [117], [118], [205].

Unfortunately, even though synthetic chromatin approaches offer a platform for investigating the functional relationship between chromatin structure and DNA modifications, the technical challenge of generating site-specifically modified DNA templates (>2000 bp) suitable for reconstituting nucleosome arrays is a major challenge

*Parts of this chapter have been reprinted with permission from Banerjee, D. R.[†]; Deckard III, C.E. [†]; Elinski, M. B.; Buzbee, M. L.; Wang, W. W.; Batteas, J. D.; Sczepanski, J. T., Plug-and-Play Approach for Preparing Chromatin Containing Site-Specific DNA Modifications: The Influence of Chromatin Structure on Base Excision Repair, *Journal of the American Chemical Society*, **2018**, *140* (26), 8260–8267.

for at least two reasons: first, most nucleotide modifications are incompatible with enzyme dependent replication (i.e. PCR); and second, DNA polymerases are unable to install modified nucleoside triphosphates in a site-specific manner. In the past, multistep ligation-based strategies, in which many short oligonucleotides are hybridized and subsequently ligated to generate the desired modified DNA duplex, have been sufficient for producing chromatin templates (see Figure 1.7) [97], [118], [204]. However, ligation efficiencies drop precipitously with increasing number of individual DNA fragments, thus making the isolation of useable quantities of pure, full-length, template time-consuming and difficult.

In the absence of a straightforward approach to prepare synthetic chromatin containing precisely positioned DNA modifications, we developed a methodology, termed 'plug-and-play', that enables the efficient and stream-lined incorporation of virtually any synthetically accessible DNA modification into chromatin fibers, *in vitro*, with single-nucleotide resolution. The end goal of this project was to investigate the structure function-relationship between chromatin structure and TDG, an area that is poorly understood, therefore we first sought to validate our system by applying it to UDG and APE1 [136], [144], [154], [155], [158]. We initially chose this system because UDG and APE1 are well characterized in DNA- and mononucleosome-only systems. Additionally, dU lesions resulting from hydrolytic deamination of cytosine and enzymatic deamination by AID/APOBEC [138], [184], [206] frequently occur within a CpG context, which are extremely relevant to studying TDG. This chapter describes the rationale design and development of 'plug-and-play' chromatin templates, as well as the experimental

procedures required to introduce both natural and synthetic DNA modifications into a requisite chromatin template.

2.1 Results

2.1.1 Rational Design of Modular Chromatin Templates

In order to construct nucleosome arrays with site-specifically modified DNA, we first focused on developing a straightforward approach for inserting modified nucleotides into the requisite 12×601 DNA template (*see 2.3.2* and Figure A.1-2). In particular, we sought to avoid the inefficient joining together of multiple DNA fragments using intermolecular ligation. Instead, we chose to employ a strategy whereby sequence-specific nicking endonucleases (or "nickases") are used to nick the DNA at two proximal sites resulting in formation of a short gap upon melting [207], [208]. This gap is then filled with a DNA oligonucleotide carrying the desired modification and the nicks subsequently resealed to generate an intact DNA template (Figure 2.1a). Because the sealing of DNA nicks by T4 DNA ligase is extremely efficient, this approach streamlines the purification process. Overall, this strategy reduces the time and increases the yield of modified DNA compared to previously employed ligation-based methods.

The strand-exchange reaction outlined above offers a general approach for modifying plasmid DNA [207], [208] but, within the context of 12×601 DNA, this method enables precise control over the translational and rotational setting of the corresponding DNA modifications and its environment within a defined chromatin fiber (Figure 2.1). There are several important considerations to be made when designing these 'nickable'

templates: where the DNA modification(s) of interest will be positioned within the 12×601; which base pairs from the native 601s can be altered without compromising nucleosome binding; and what is the melting temperature and relevant secondary structure of the modified ssDNA fragment.

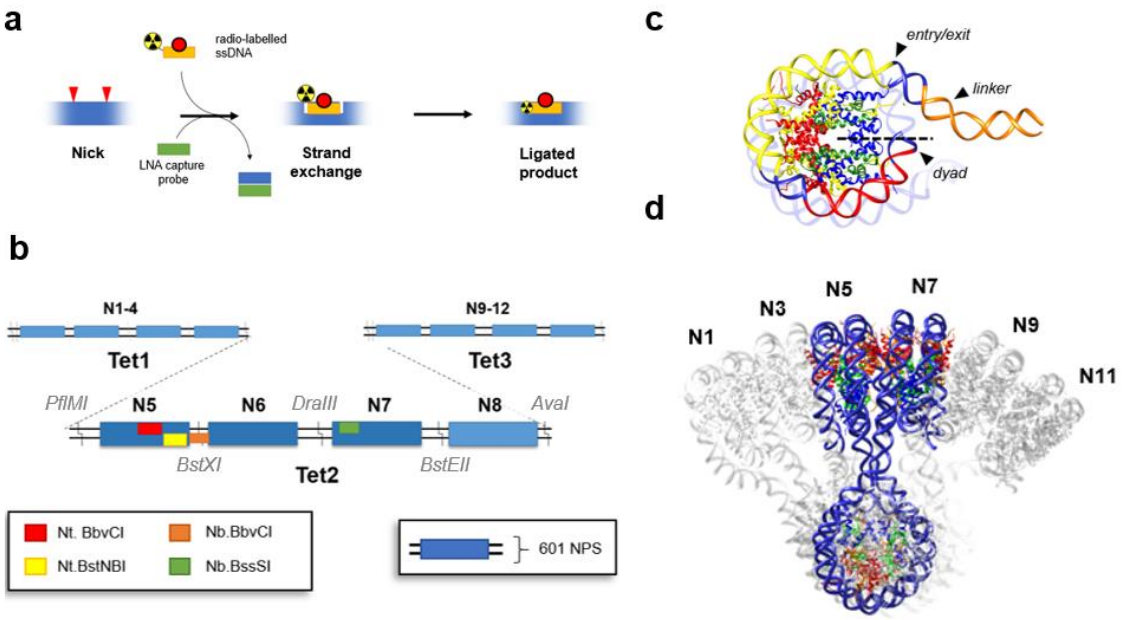


Figure 2.1: Plug-and-play chromatin. (a) Schematic depiction of the plug-and-play method and 12×601 DNA array (b) used in this work. Arrays consisted of 12 repeats of the Widom nucleosome positioning sequence (NPS). Colored rectangles on the 12×601 template indicate excisable ssDNA fragments enabled via digestion with the nicking endonucleases indicated in the bottom left. (c) Plug-and-play constructs offer complete coverage over all three major translational regions of interest within the nucleosome (orange= linker DNA, yellow= nucleosome entry/exit, red= dyad). (d) Structural representation of the 12×601 nucleosome arrays employed in this work. The chromatin image shown here was constructed by fitting the crystal structure of the tetranucleosome (PDB: 1zbb) into the published electron density map (pdb: 1zbb, emd: 2600) using Chimera.

Due to the symmetrical nature of the nucleosome about the dyad (Figure 2.1c), each 177 bp nucleosome within the 12×601 array contains ~88 unique translational environments. To allow for complete coverage over all sites within this region it was

partitioned into three distinct segments (Figure 2.1b-c). Each segment was flanked by a 'nickase' recognition sequence, which were introduced within the 601 at sites estimated to have a low probability of impacting the DNA's ability to form a stable nucleosome (Figure A.2). Mutations to TA/AT/TT/AA repeat sequences, especially near SHL (+/-) ~1.5, and GC/CG/CC/GG rich sequences near the dyad were generally avoided, with less care paid to sequence changes near the entry/exit site relative to sequences located internal to the nucleosome [15]–[18]. Indeed, DNase footprinting analysis of a 601-mutant mononucleosome containing the Nt.BstNBI nickase recognition site, as is it occurs within nucleosome N5 of **12-601-Nt** and **12-601-Nt/SI**, confirmed that these minor sequences variations did not significantly alter the rotational positioning of the DNA relative to the native 601 sequence (Figure A.5). Additionally, the target sequences' melting temperature (T_m) and tendency to dimerize and form secondary structures should be taken into consideration. For these reasons, the excisable 601 fragments were designed to be 14-33 nucleotides (nts) in length with melting temperatures (T_m) in the range of 45-65 ˚C, if possible; the DNA modification, was placed ≥3 nts away from the ligation junction; and, in cases near the dyad, additional consideration is required for Guanine-, and GC-rich sequences that have a tendency to form secondary structures like G-quadruplexes, hairpins, and homo-dimers that could potentially compete with hybridization of the ssDNA insert to the 601 template [209], [210]. For example, when targeting sequences having multiple consecutive guanines it is best to perform subsequent exchanges in buffers lacking monovalent cations known to stabilize these structures, such as K^+. An important benefit of the current platform is that the chosen nicking endonuclease recognition

sequences allow modification of either strand of DNA by using the orthogonal nickase (e.g. Nb.BbvCI and Nt.BbvCI, which cut bottom and top, respectively).

2.1.2 Reconstitution and Characterization of 'Plug-and-Play' Chromatin

To demonstrate the feasibility and versatility of this approach, the 'plug-and-play' arrays outlined in Figure 2.1 were used to generate DNA templates harboring a variety of lesions and fluorescent probes with applications in studying the relationship between chromatin structure and BER. Specifically, **12-601-Nt** and **12-601-Nb** were used to generate templates harboring a 2'-deoxyuridine (dU) or an *O*-nitrobenzyl (ONV) protected abasic site at the +49 and +88 translational positions of N5, **12-601-dU49**(**dU88**) or **12-601-ONV49**(**ONV88**) (Figure 2.2, Figure A.6), respectively; and **12-601-Nt/Nb** was used to generate a dually-modified template that contained Forster Resonance Energy Transfer (FRET) pairs, Cyanine 3 and 5, on adjacent 601s (Figure 2.3) [117]. Once prepared, the intact DNAs were reconstituted into 12mer nucleosome arrays using well-established protocols (see methods *2.3.7* and *2.3.8*). Prior to reconstitution, each 12×601 DNA template was released from its corresponding plasmid backbone via digestion with EcoRV and the remaining plasmid DNA was digested into shorter DNA fragments with DraI and HaeII. The presence of plasmid DNA contaminants did not affect the reconstitution efficiency and enabled us to eliminate a purification step in the process.

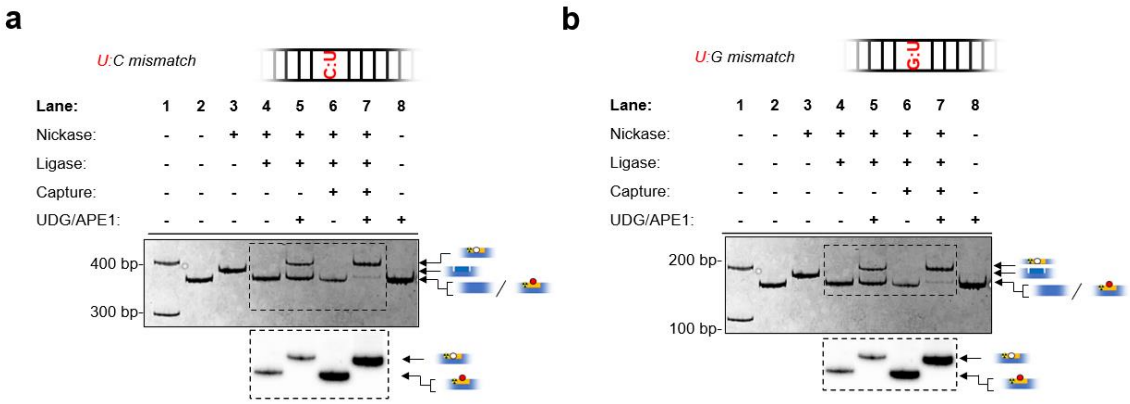


Figure 2.2: Site-specific incorporation of dU into 12×601 DNA templates. (a) Representative electrophoretic mobility shift assay demonstrating the insertion of a dU-containing oligonucleotide (**N5_dU+49**) into template **12-601-Nt**. All reactions were carried out on **pUC-12-601-Nt** and the 601 DNA fragment (N5) containing the modified site was excised via PflMI and BstXI restriction digestion (Figure 2.1b, A.1) and analyzed by native PAGE. (b) The same analysis described in part (a) except for the insertion of **N5_dU+88** into **pUC-12-601-Nb**. The 601 fragments (N5-N6) depicted were excised with PflMI and DraIII restriction digestion (Figure 2.1b, A.1).

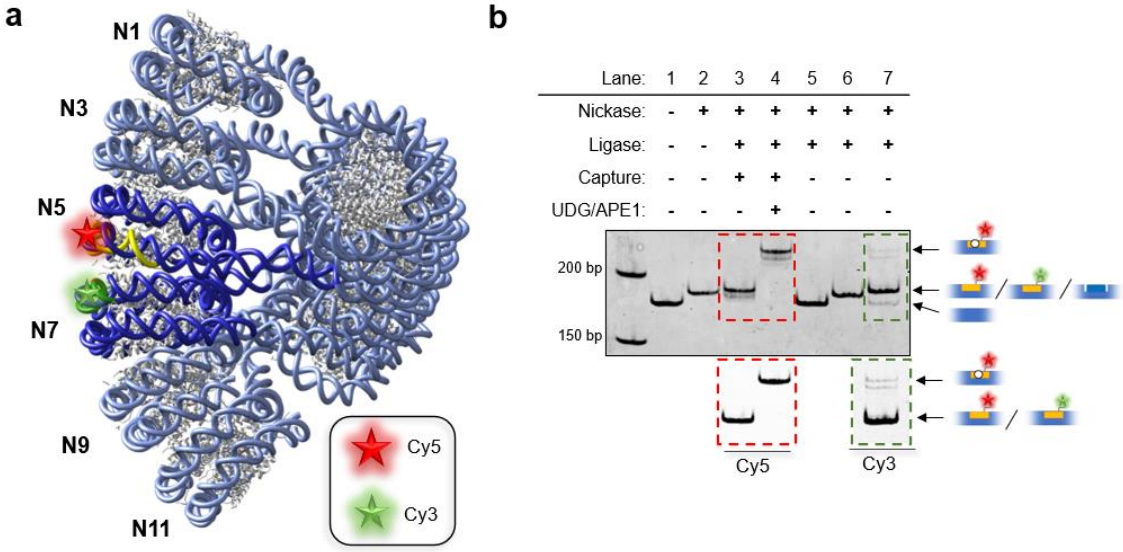


Figure 2.3: Dual insertion of fluorescent modifications within the **12-601-Nt/SI** DNA array. (a) Cryo-electron microscopy structural model of the 30 nm chromatin fiber (12×601, pdb: 1zbb, emd: 2600) showing the locations of the Cy3 and Cy5 modifications, as well as the corresponding Nt.BstNBI (yellow) and Nb.BssSI (green) excisable DNA regions (see Figure 2.1). (b) Electrophoretic mobility shift assays demonstrating simultaneous insertion of Cy5-labelled **N5_dU+49_Am+39** and Cy3-labelled **N7_Am.T-39** into **12-601-Nt/SI**. The gel was imaged first with the indicated fluorescent channel (see below) and then stained and visualized with EtBr.

Optimal reconstitution conditions were determined by histone octamer titration (Figure 2.4 and A.6), and fully saturated nucleosome arrays were separated from the digested plasmid DNA via selective magnesium-induced precipitation (Figure 2.4) [58], [211]. Native gel electrophoresis of the reconstituted products revealed a single discrete band consistent with a homogeneous population of nucleosome arrays (Figure 2.4), restriction enzyme digests of the reconstituted arrays demonstrated full nucleosome occupancy (Figure Appendix A.7a), and partial digestion with micrococcal nuclease (MNase) confirmed the presence of 12 positioned nucleosomes (Figure A.7b). These data suggest that incorporation of a single dU residue does not significantly influence the overall structure of the nucleosome array and is consistent with several previous studies showing that the presence of a single, or even multiple, dU residues does not adversely affect mononucleosome and oligonucleosome array formation [122], [141], [212]. The translational setting of dU49 in relation to the underlying nucleosome was also verified through a series of nuclease digestions (Figure A.8).

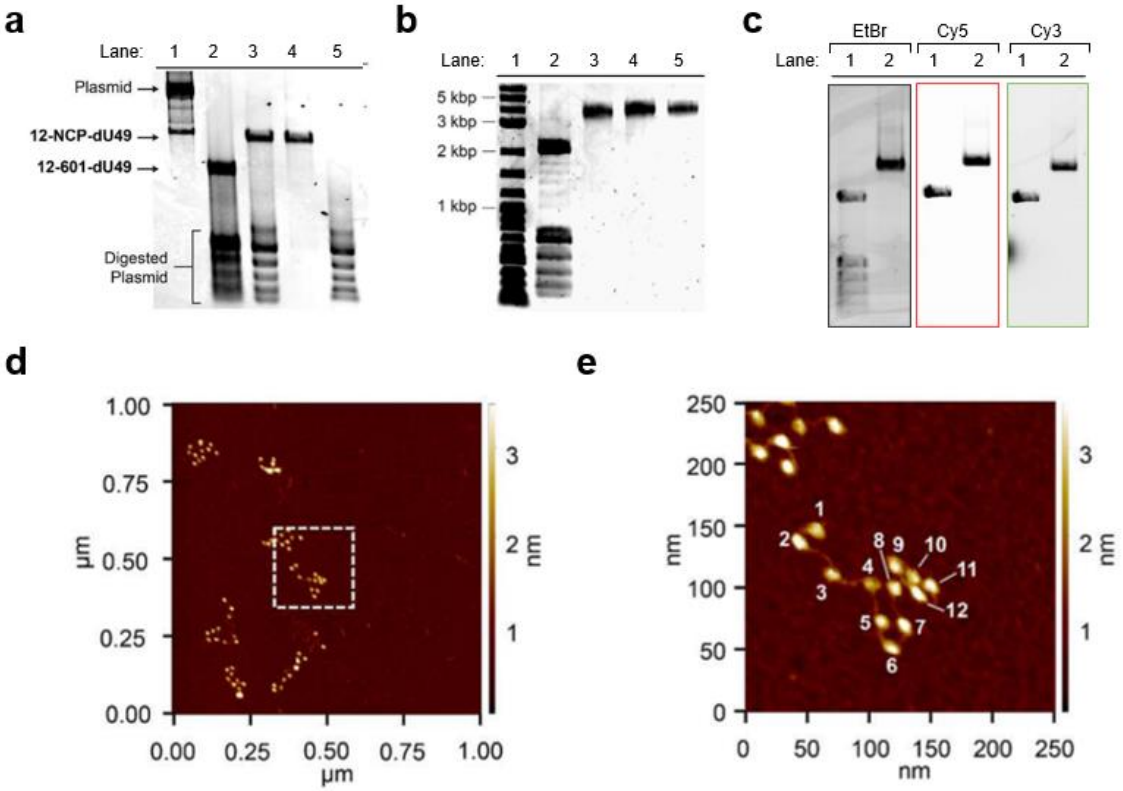


Figure 2.4: Characterization of modified 12-mer nucleosome arrays. (a) Representative agarose gel (0.7%, 0.2× TBE) demonstrating selective Mg^{2+}-precipitation of **12-NCP-dU49**. Lane 1, undigested pUC19 plasmid containing the **dU49** insert; lane 2, EcoRV-digested plasmid from lane 1 showing the release of **12-601-dU49**; lane 3, reconstituted **12-NCP-dU49** prior to selective Mg^{2+} precipitation; lane 4, pure **12-NCP-dU49** obtained from the pellet following selective Mg^{2+} precipitation of the mixture in lane 3; lane 5, remaining supernatant following selective Mg^{2+} precipitation of the mixture in lane 3. (b) Purified 12-mer nucleosome arrays. Lane 1, **12-601-dU49** prior to reconstitution (for reference); lane 2, wild-type 12×601 nucleosome array; lane 3, **12-NCP-dU49**; lane 4, **12-NCP-dU88**. (c) Agarose gel (0.6%, 0.2x TBE) depicting crude, dually labelled **12-601-Nt/SI** DNA, before and after reconstitution and Mg^{2+}-induced precipitation to remove DNA impurities (Lanes 1 and 2, respectively). Lanes 1 and 2 were visualized with the three indicated fluorescent channels to confirm the presence of Cy3 and Cy5. (d,e) AFM tapping mode topography images of **12-NCP-dU49** in extended form. The white box in panel (d) highlights the image under high-resolution in panel (e).

In collaboration with the Batteas laboratory, we further validated the biochemical integrity of our modified arrays using atomic force microscopy (AFM) (Figure 2.4 and A.9). At low salt concentrations, the arrays adsorb onto the surface in an extended form, enabling individual nucleosomes on the array to be resolved and counted. At least two-dozen well-resolved 12-mer arrays were analyzed for each sample and the overall results demonstrate greater than 98% of the nucleosome positioning sequences in both 12mer substrates were incorporated into a nucleosome. In addition, we successfully obtained AFM images of APE1 bound to the nucleosome array (Figure A.7). This complex was prepared by first treating **12-NCP-dU49** with UDG to convert dU49 into an abasic site. After removing UDG via precipitation, excess APE1 was added to the sample along with glutaraldehyde (0.1%) to trap and stabilize the interaction for AFM imaging. As expected, APE1 bound specifically to the fifth nucleosome (N5, Figure 2.1) originally containing the dU lesion (dU49). In contrast, no complexes were observed between APE1 and undamaged chromatin (Figure A.6). Together, these unprecedented images further validate the authenticity of our substrates and demonstrate the potential utility of our system for single molecule characterization of BER component enzymes on chromatin.

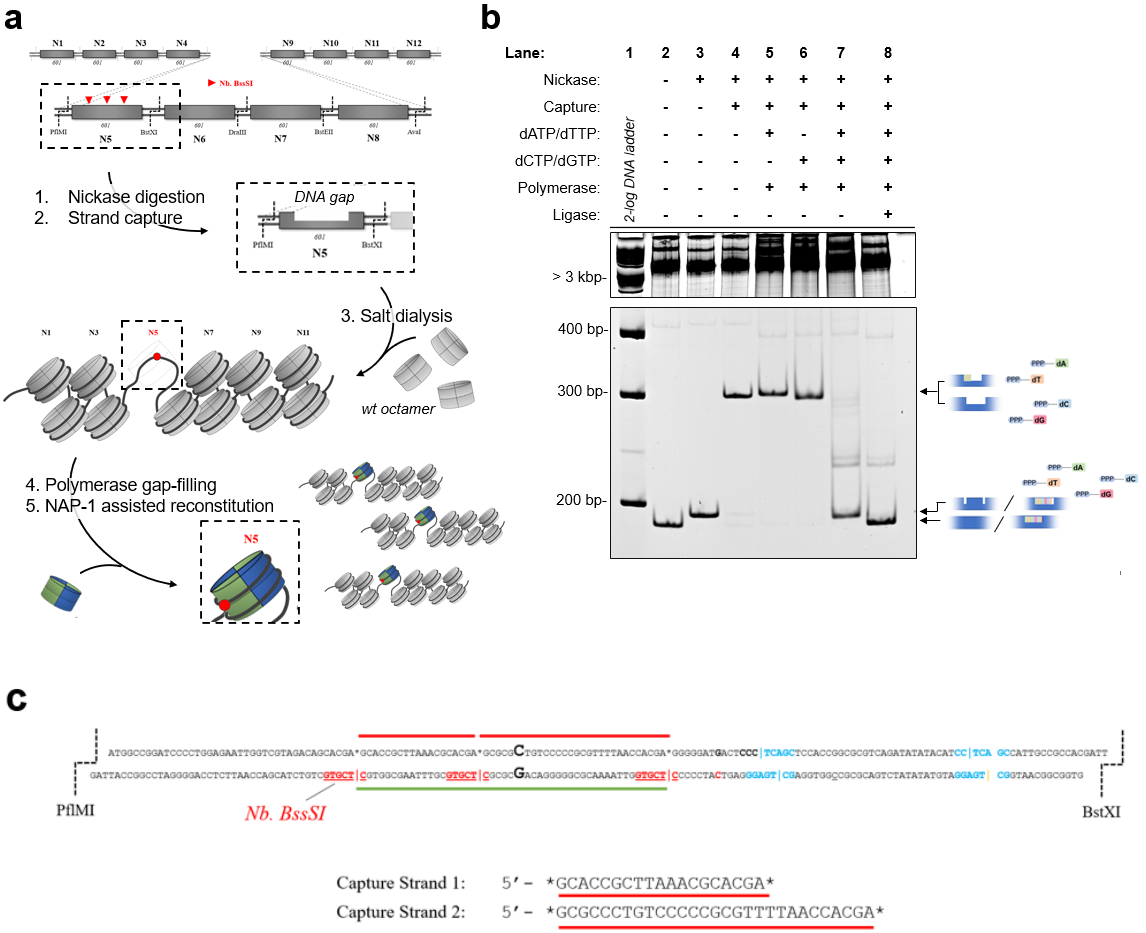


Figure 2.5: Applications of 'gapped' chromatin templates. (a) Proposed method for heterogenous chromatin assembly. A DNA gap is generated via capture probes and nucleosome formation is subsequently inhibited, selectively, at N5. This is achieved via a gap spanning the dyad (~24 nt on each side) which cannot bind a nucleosome during salt dialysis reconstitutions. The DNA gap is filled with a polymerase and then a modified histone octamer is installed at the open site via the nucleosome assembly protein (NAP-1). I have performed steps 1-3 and step 4-5 are currently being optimized by our laboratory. The plasmid for preparing this gapped chromatin template (**pUC-12-601-Gap**) and the requisite **12-601-Gap** DNA was been prepared and confirmed by sequencing (Table A.2). (b) Gap-filling reaction carried out on the N5 gapped **12-601-Gap** template after gap formation using T4 DNA polymerase (0.1 U/µL, 1.5 U total) and 0.2 mM dNTPs. 1 U Nickase and 1 U Ligase was used per 1 µg plasmid DNA. (c) Sequence details of the template and capture strands used to generate the **12-601-Gap**.

2.2 Conclusions and Broader Impact

In summary, a straightforward and efficient method for generating oligonucleosome arrays containing site-specific DNA modifications based on nicking endonucleases was developed. Using this approach, we assembled 12-mer nucleosome arrays containing site-specifically positioned dU residues and studied their repair in chromatin by UDG/APE1. This work led to the first experimental evidence that inter-nucleosome interactions mediated by histone tail domains (i.e., histone H4) play an important role in regulating BER within higher order chromatin structures (Appendix A). In addition to enabling specific internal DNA sites to be modified, the 'plug-and-play' method also affords regio-selective, dual-strand, and 3'- and 5'- termini modification. These applications of our 'plug-and-play' methodology weren't utilized for a complete project, however, exploratory studies yielded promising results which have motivated current and future projects. For instance, using 'gapped' 12×601 DNA templates we demonstrated nucleosomes could be prevented from forming, providing the basis for a heterogenous chromatin assembly project (see Figure 2.5a). One could also use a DNA polymerase lacking 5'-exonuclease activity to fill in this gap with the desired modified triphosphates (Figure 2.5b). We also generated chromatin templates having modified 3'- and 5'- ends for selective labelling and pull-down experiments (Appendix F).

Overall, this approach offers a modular platform for investigating a range of chromatin-related chemistries. Further broadening the applications of our 'plug-and-play' chromatin, it has recently enabled the *in vitro* characterization of two potential therapeutics: the first being L-RNA inhibitors of PRC2 (see Chapter 5) and the second

being a collaborative study with researchers from the Graduate School of Pharmaceutical Sciences at the University of Tokyo, where they are using the 'plug-and-play' strategy to investigate the selectivity of an acetylation drug catalyst with epigenetic reprogramming potential on chromatin structure and DNA repair.

2.3 Materials and Methods

2.3.1 General

Restriction enzymes (EcoRV, EcoRI, XbaI, TspRI, PflMI, BstXI, BstEII, DraIII-HF, AvaI, AluI, Nb.BbvCI, Nb.BssSI, HaeII, DraI, Nt.BstNBI), T4 DNA Ligase, UDG, APE1, T4 polynucleotide kinase, micrococcal nuclease (MNase), and shrimp alkaline phosphatase were purchased from New England Biolabs (Ipswich, MA). [γ^{32}P]-ATP was purchased from Perkin Elmer (Waltham, MA). N-Hydroxysuccinimide (NHS) esters of Cy3 or Cy5 dyes used in the labeling of **T2.N5_dU+49_Am+39** and **T2.N7_Am.T-39** (Table A.3) were acquired from Lumiprobe Life Science Solutions (Hallendale Beach, FL). Synthetic oligonucleotides (Table A.1- A.3) were either purchased from Integrated DNA Technologies (Coralville, IA) or prepared by solid-phase synthesis on an Expedite 8909 DNA/RNA synthesizer. DNA synthesis reagents and nucleoside phosphoramidites were purchased from Glen Research (Sterling, VA), and lock nucleic acid (LNA) phosphoramidites were obtained from Exiqon (Denmark, Kingdom of Denmark).

2.3.2 Construction of 12×601 DNA templates.

We began by assembling a "general" DNA template (**SI-12-601-WT**) composed of 12 repeating units of the 147-base pair (bp) "601" nucleosome positioning sequence with 30 bp of linker DNA separating each 601 unit (Figure A.1). This template was designed to be modular such that the four central 601 units (N5–8) could be easily replaced using a unique set of restriction enzymes. To facilitate assembly of full-length **SI-12- 601-WT** DNA (2094 bp), the 12×601 sequence was divided into three individual 4×601 "tetramers" (Tet1–3), each of which was prepared separately (Figure A.1). Individual 601 units within each tetramer were generated by PCR from the native 601 DNA sequence using primers designed to introduced unique restriction sites at both ends of the 601 unit (Figure A.1 and Table A.1). Following PCR, each 601 unit was treated with its corresponding restriction enzyme(s), pooled into separate reaction mixtures (N1–N4 for Tet1, N5–N8 for Tet2, and N9– N12 for Tet3), and ligated together to generate their corresponding tetramer. Purified tetramers were cloned into separate pUC19 vectors between the EcoRI and XbaI restriction sites and correct assembly was confirmed by sequencing (Eton Bioscience, Inc., San Diego, CA). The tetramer DNA (Tet1–3) were then released from their plasmid backbones using the corresponding set of restriction enzymes (EcoRI/PflMI for Tet1, PflMI/AvaI for Tet2, and AvaI/XbaI for Tet3), purified by agarose gel electrophoresis, and ligated together in a single reaction to generate the full-length **SI-12-601-WT** product (Figure A.1). The **SI-12-601-WT** product was subsequently cloned into the pUC19 vector between the EcoRI and XbaI sites, S3 generating plasmid **pUC-SI-12-601-WT** (Figure A.1c). Correct assembly of this plasmid

was verified via sequencing from both ends and by selective restriction digestion analysis (Figure A.1d,e).

Plasmid **pUC-SI-12-601-WT** was then used to prepare the 12-mer DNA templates depicted in Figure 2.1 (**12-601-Nt**, **12-601-Nb**, and **12-601-Nt/SI**). Individual 601 units (e.g. N5) within pUC-SI-12-601-WT were excised using the appropriate restriction enzymes and replaced with new 601 units containing the corresponding nickase recognition sites (Figure A.2). These new 601 units were generated by PCR as described above using primers listed in Table A.2. Plasmids containing **12-601-Nt**, **12-601-Nb**, and **12-601-Nt/SI** are referred to as **pUC-12-601-Nt**, **pUC-12-601-Nb**, and **pUC-12-601-Nt/SI**, respectively (Figure A.2). Each plasmid was verified by sequencing.

2.3.3 Preparation of dU- and ONV-containing 12-601 DNA templates.

Plasmids **pUC-12-601-Nt** and **pUC-12-601-Nb** (500 µg, 160 pmol) were digested for 1 hour with 500 units of Nt.BstNBI or Nb.BbvCI, respectively, in a 700 µL reaction mixture according to the manufacturer's recommended protocol using NEB buffer 3.1 (100 mM NaCl, 50 mM Tris-HCl, 10 mM MgCl2, 100 µg/mL BSA, pH 7.9) or "Cutsmart" buffer (50 mM KOAc, 20 mM Tris-acetate, 10 mM MgOAc, 100 µg/mL BSA, pH 7.9). To the digested plasmid was added 3.2 nmol of the 5′-[^{32}P]- labeled synthetic oligonucleotide insert- **N5_dU+49** or **N5_dU+88** for **12-601-dU49** and **12-601-dU88**, respectively; and **N5_ONV+49** or **N5_ONV+88** for **12-601-ONV49** and **12-601-ONV88**, respectively (Table A.3)- and 800 pmol of the corresponding lock nucleic acid (LNA) capture probe (**LNA.Capt_NCP** or **LNA.Capt_Link**, respectively; Table A.3). The

reaction mixture was then heated at 80°C for 20 minutes before being cooled to room temperature at -1 °C/min. Following the annealing step (~1 hour), 400 units of T4 DNA ligase and ATP (2 mM final concentration) were added to the mixture. The ligation reaction was allowed to proceed for 2 hours at room temperature and inactivated at 70 °C for 20 minutes. The efficiency of each step of the exchange process (nicking, insertion, and ligation) was carefully monitored in order to ensure complete insertion of the modified oligonucleotide (see Figure 2.2, 2.3 and Figure A.3). For this, ~1 pmol aliquots were taken and digested with either PflMI/BstXI (**pUC-12-601-Nt**) or PflMI/DraIII (**pUC-12-601-Nb**) to release the 601 unit(s) harboring the nickase sites (Figure 2.1 and A.1). Aliquots isolated following the ligation step from dU-containing samples were treated with UDG/APE1 (100 nM each) as described below, and ONV-containing samples were exposed to 365 nm lamp at 4˚C for 15 minutes prior to treatment with APE1. These aliquots were subsequently analyzed by 10% native PAGE (19:1 acrylamide:bisacrylamide). After the integrity of the modified DNA was confirmed, the reaction mixture was desalted by ethanol precipitation, and the modified 12×601 DNA template was removed from the corresponding plasmid backbone via digestion with EcoRV (Figure 2.4). The reaction mixture was also digested with DraI and HaeII (600 units each) in order to degrade the plasmid DNA to fragments ≤ 700 bp. The digest was monitored using 0.8% Agarose (1× sodium borate buffer). Following digestion, the reaction mixture was passed through a QIAquick spin column (Qiagen, Hilden, Germany) and the eluted DNA was used directly to reconstitute oligonucleosome arrays.

2.3.4 Dual modification of 12-601-Nt/SI.

Starting from plasmid **pUC12-601-Nt/SI**, Cy3 and Cy5 were installed using the oligonucleotide exchange process described above. Briefly**, pUC-12-601-Nt/SI** was digested sequentially using Nt.BstNBI and Nb.BssSI and labeled in a single exchange step using the corresponding fluorescently labelled oligonucleotide inserts (**N5_dU+49_Am+39** and **N7_Am.T-39**, respectively; Table A.3). As before, the efficiency of the exchange process was monitored by native PAGE (Figure 2.4).

2.3.5 Regio-selective modification via polymerase gap-filling.

Our strand exchange methodology enables modifications to be installed in chromatin with single nucleotide resolution, but it also allows for specific regions of a nucleosome to be modified via polymerase-dependent gap-filling. As a proof of concept we generated a gap spanning the dyad by performing the strand-exchange reaction using only **pUC-12-601-Gap** and the corresponding DNA capture strands in 25-fold excess (Figure 2.5, Table A.3). The resulting gap was then able to be filled using T4 DNA polymerase (0.1 Unit/μg DNA), 200 μM dNTP's in nucleosome buffer supplemented with 1 mM $MgCl_2$ and subsequently ligated with T4 DNA ligase. The gap was only completely filled when all four dNTP's were included in the reaction and, although no modified bases were included here, we expect this methodology would allow any polymerase-compatible DNA modification to be introduced region-specifically within chromatin.

2.3.6 Preparation of 1-601-dU49 and 1-601-dU88 DNA templates

The oligonucleotide exchange procedure described above was also used to prepare mononucleosome substrates **1-601-dU49** and **1-601-dU88** (Figure A.3) containing dU modifications at the same translational positions relative to nucleosome 5 (N5) in the corresponding arrays (**12-NCP-dU49** and **12-NCP-dU88**, respectively). For clarity, these DNAs are used (below) to generate mononucleosome substrates **1-NCP-dU49** and **1-NCP-dU88**, respectively. The unmodified 1×601 DNA template (**1-601-Nt**) corresponding to **1-601-dU49** was prepared by PCR from the native 601 DNA sequence using primers **N5_FWD.EcoRI/PflMI** and **N5_REV.Nt.BstNBI** (Table A.1, A.2 and Figure A.1. The unmodified 1×601 DNA template (**1-601-Nb**) corresponding to **1-601-dU88** was prepared similarly using primers **N5_FWD.EcoRI/PflMI** and **Nick.NCP_REV.Nb.BbvCI** (Table A.1, A.2; Figure A.3). The dU residues (dU49 and dU88) were then installed into **1-601-Nt** and **1-601-Nb** using the same oligonucleotide exchange process described above, employing **N5_dU+49** and **N5_dU+88** (Table A.3), respectively. Complete incorporation of the modified oligonucleotide was confirmed by native PAGE (Figure A.3).

2.3.7 Histone preparation and octamer refolding

Recombinant human histones (H2A.1, H2B.1, H3, H4, H4Δ-tail) were expressed and purified using established protocols [109], [121]. Histone octamers were refolded using standard protocol [213], and purified by size exclusion chromatography using a S5 SuperDex 200 10/300 GL column (GE Lifesciences, Boston, MA). Purified histone

octamer was stored in Octamer Buffer (2 M NaCl, 5 mM BME, 0.2 mM PMSF, 1 mM EDTA, 10 mM HEPES, pH 7.8).

2.3.8 Reconstitution of mononucleosomes and oligonucleosome arrays

The reconstitution of nucleosome arrays generally followed the protocol described in ref. 3, with some modifications. Crude plasmid digests (above) containing the modified 12×601 DNA templates (**12-601-dU49** or **12-601-dU88**) were mixed with varying molar ratios of refolded histone octamers in a 25 µL reaction mixture containing 2 M NaCl at 4 °C. Typical reconstitutions reactions contained 1.2-1.7 equivalent histone octamer relative to the DNA. The final concentration of 12×601 DNA in each sample was 0.1 µg/µL. The reaction mixture was then transferred to a 3.5k MWCO Slide-A-Lyzer MINI dialysis unit (Thermo Fisher Scientific, Waltham, MA) and floated over 500 mL of pre-cooled nucleosome high salt (HS) buffer (2 M NaCl, 0.1 mM PMSF, 10 mM HEPES, pH 7.8). Over a period of ~24 hours at 4 °C, the HS buffer was exchanged with 1.6 L low salt (LS) buffer (25 mM NaCl, 0.1 mM PMSF, 10 mM HEPES, pH 7.8) using peristaltic pump. When the buffer exchange was complete, reconstitution mixtures were spun at 13,000 ×g for 20 minutes to remove any aggregates and the arrays were purified from the plasmid DNA using selective Mg^{2+} precipitation (6 mM $MgCl_2$). After addition of $MgCl_2$, the samples were incubated on ice for 30 minutes, spun at 13,000 ×g for 20 minutes at 4 °C, and the pellet and supernatant were separated. The pellet containing the pure arrays was then dissolved in 25 µL of LS buffer. The reconstituted arrays (**12-NCP-dU49** and **12-NCP-dU88**) were analyzed by 0.6% agarose gel (0.2×TBE buffer, 12V/cm, 90 minutes)

and stained with ethidium bromide (Figure 2.4 and Figure A.4). For arrays lacking the H4 tail domain, which precludes selective Mg^{2+}-induced precipitation, the 12×601 DNA was purified by preparative agarose gel electrophoresis prior to reconstitution. For this purpose, the **12-601-dU49** containing plasmid was digested with EcoRV/DraI/HaeII as described before, then digests were separated in 1% Agarose gel (1× sodium borate buffer), and the **12-601-dU49** band was excised and eluted using QIAquick Gel Extraction kit (Qiagen). The purified DNA was then used in reconstitution with octamer_ΔH4 using the same protocol described above. Mononucleosomes were reconstituted using the identical salt dialysis method described above. Reconstituted mononucleosomes were analyzed by 5% native PAGE (59:1 acrylamide:bisacrylamide) and only those sample samples containing <5% free DNA were used further (Figure A.4). Mononucleosome and nucleosome array concentrations were determined by measuring the A_{260} in 0.1 M NaOH as reported previously.

2.3.9 DNase Footprinting

Reconstituted nucleosomes were treated with 2 unit DNase I in a 10 µL reaction mixture containing 1× DNase buffer (10 mM Tris, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$, pH 7.5) for 30s (free DNA) or 45s (nucleosomes). The reaction was stopped by adding 40 µL quenching buffer containing 200 mM EDTA, 1% SDS and 1 unit proteinase K. The proteins were then removed by extracting with phenol:chloroform:isoamyl alcohol (25:24:1, v/v, ThermoFisher Scientific) and the DNA desalted by ethanol precipitation. The resulting DNA pellet was resuspended in formamide loading buffer (90% formamide,

10 mM EDTA) and analyzed by 10% denaturing PAGE (19:1 acrylamide:bisacrylamide; Figure A.5). The alignment of the banding pattern for the native and Nt.BstNBI-modified mononucleosomes indicates that the rotational positioning is consistent across both DNA substrates.

2.3.10 Restriction digestion of oligonucleosome arrays

Nucleosome saturation was confirmed by digestion of 0.2 pmol oligonucleosome array with 10 units of either PflMI/BstXI (for **12-NCP-Nt** and **12-NCP-dU49**) or PflMI/DraIII (for **12-NCP-dU88**) in LS buffer supplemented with 2 mM $MgCl_2$. The respective naked DNAs (**12-601-Nt**, **12-601-dU49**, and **12-601-dU88**) were also digested under the same conditions, and both sets of samples (naked and arrays) were analyzed side-by-side on a 5% native PAGE gel (59:1 acrylamide:bisacrylamide; Figure A.7a). The presence of a nucleosome band as well as the absence of significant free DNA (<3%) demonstrates full nucleosome occupancies in these reconstituted arrays.

2.3.11 Partial micrococcal nuclease digestion of reconstituted arrays

The presence of 12 nucleosomes per array was further confirmed through partial micrococcal nuclease (MNase) digestion (Figure A.7b). Arrays (0.4 pmol) were digested with 1000 units of MNase for 60 sec on ice in presence of 5 mM $CaCl_2$. Reactions were stopped with the addition of 0.2% (v/v) SDS and 20 mM EDTA and filtered through a QIAquick spin column. The eluted DNA was analyzed on a 0.5% agarose gel (1× sodium

borate buffer) and visualized with ethidium bromide (Figure A.7. The presence of twelve bands confirmed 12 positioned nucleosomes on the array.

2.3.12 Analysis of the translational setting of dU49 and the integrity of nucleosome N5

Nucleosome array **12-NCP-dU49** (0.2 pmol) was digested with 10 units MNase at 37 ºC for 10 min in presence of 1 mM $CaCl_2$. The reaction was stopped with the addition of 0.2% (v/v) SDS and 20 mM EDTA and filtered through a QIAquick spin column. The eluted DNA was then separated by 10% native PAGE (19:1 acrylamide:bisacrylamide; Figure A.8), and the band corresponding to the fully digested array (~150bp) was excised and eluted overnight in buffer EB (200 mM NaCl, 10 mM EDTA, 10 mM Tris, pH 7.6). The eluted DNA was filtered through a 0.22 micron syringe filter (Millipore, Burlington, MA) and desalted using ethanol precipitation. The DNA pellet was resuspended in 20 µL Tris buffer (10 mM). The sample was then treated with 5 unites AluI for 60 minutes according to the manufacturer's recommended protocol using Cutsmart buffer (50 mM KOAc, 20 mM Tris-acetate, 10 mM MgOAc, 100 µg/mL BSA, pH 7.9). The products of the digestion were separated by 10% native PAGE (19:1 acrylamide:bisacrylamide) and visualized with ethidium bromide staining and autoradiography (Figure A.8).

2.3.13 Mg^{2+} precipitation analysis of reconstituted arrays

To analyze the aggregation behavior, an unmodified 12-mer array comprised of **12-NCP-Nt** was incubated in separate reaction mixtures each containing 0.2 pmol array, 25 mM NaCl, 0.1 mM PMSF, and 10 mM HEPES (pH 7.8) and either 0, 0.2, 2 or 5 mM

$MgCl_2$ in a total volume 20 µL at 37 ºC for 60 minutes. Aliquots were taken at 0, 10, 30 and 60 min and spun at 13000 × g for 10 minutes at 4 ºC. The supernatant was then analyzed by 0.6% agarose (0.2× TBE, 12 V/cm at 4 ºC; Figure A.9) and the extent of aggregation was estimated through measuring the percentage of un-precipitated DNA present in supernatant.

CHAPTER 3

CHROMATIN STRUCTURE AND THE PIONEERING TF FOXA1 REGULATE TDG-MEDIATED REMOVAL OF 5-FORMYLCYTOSINE FROM DNA

In addition to its well-known role initiating BER at mismatched pyrimidine bases (T/U) within the context of CpG dinucleotides [139], [214], Thymine DNA glycosylase (TDG) has recently been implicated in epigenetic regulation through its involvement in the removal of 5-methylcytosine (5mC) from genomic DNA [161], [215]. In mammals, enzymatic reversal of 5mC to cytosine occurs through TET/TDG-mediated BER (see *1.2.3 Active DNA demethylation*) [169],[169],[177],[168]. Although 5fC and 5caC are "committed" to removal by TDG, emerging evidence suggests that they are not simply intermediates in the demethylation pathway but may also possess unique regulatory functions [76], [81], [82], [123], [176]. This places TDG in the pivotal role of determining whether 5fC and 5caC are retained as potential epigenetic marks or removed via BER.

Despite the importance of TDG in the active DNA demethylation pathway, it is unclear how TDG catalysis is regulated *in vivo*. Recent evidence suggests chromatin structure is likely to play an important role. For example, wrapping DNA into mononucleosomes inhibits TDG-mediated BER of G·T mismatches *in vitro* [153]. Furthermore, TDG knockdown leads to 5fC and 5caC accumulation at genomic regions

*Parts of this chapter have been reprinted with permission from Deckard III, C. E.; Banerjee, D. R.; Sczepanski, J. T., Chromatin Structure and the Pioneering Transcription Factor FOXA1 Regulate TDG-Mediated Removal of 5-Formylcytosine from DNA, *Journal of the American Chemical Society*, **2019**, *141* (36), 14110-14114.

where chromatin is generally more accessible, implying that TDG-mediated demethylation occurs at these sites [183]. However, the functional relationship between chromatin structure and TDG-dependent DNA demethylation has not been directly addressed. Herein, we assembled synthetic chromatin harboring site-specifically positioned 5fC residues and characterized, for the first time, the impact of general and tissue-specific chromatin environments on TDG-mediated removal of 5fC. For the 'tissue-specific' chromatin studies we were motivated to examine FOXA1 and FOXA1-dependent enhancers because, even though substantial biological data points to a direct role for FOXA1 in regulating TDG-dependent demethylation (see *3.0.1*), quantitative biochemical data to support this is lacking.

3.0.1 Forkhead BoxA1 (FOXA1) Overview

FOXA1 is a pioneer transcription factor (TF) involved in liver development and gene regulation at hormone-responsive genes in breast and prostate tissue [216]. Unlike general transcription factors, which are typically inhibited by nucleosomes and chromatin, FOXA1 belongs to a unique class of TFs that have pioneering capabilities. As a pioneer factor FOXA1 invades compact chromatin to locate target sequences. This is possible due to a combination of specific and non-specific DNA-interactions that contribute to long-lived ternary complexes which equates to long search times. Although FOXA1 binds the consensus sequence: 5'-A(A/T)TRTT(G/T)RYTY [217], [218], FOXA1 also binds nucleosomes non-specifically with high affinity.

DNA-binding is facilitated by a winged-helix domain that resembles the helix-loop-helix structure of linker histone H1.1, except in that it has more pronounced loop (i.e.'wing') structures. Unlike H1.1, FOXA1 has been shown to open compacted chromatin and contribute to the formation of chromatin loops and transcriptionally activating higher order chromatin structures [74], [203], [218]. The gene-regulating functions of FOXA1 are tightly coupled to chromatin epigenetics and, importantly, DNA methylation states [165], [202], [218]. FOXA1 interacts with Pol β, DNA ligase, XRCC1, and PARP-1 and is expected to promote DNA demethylation by nucleating BER complexes at oxidized 5mC sites [165], [202]. Interestingly, the genomic distribution of FOXA proteins highly overlaps with the histone variant H2A.Z and, together, are two essential features of active enhancers [203], [216], [219].

3.1 Results

3.1.1 Nucleosomes and Chromatin Structure Inhibit TDG

We prepared nucleosome arrays consisting of 12 copies of the "Widom 601" nucleosome position sequence [35], each of which was separated by 30 bp of linker DNA (Figures 3.1 and B.2) [158]. Each substrate contained a single, site-specifically positioned 5fC residue, which was incorporated using the 'plug-and-play' approach described in Chapter 2 [158]. Two different positions were examined: one within the nucleosome-bound region of the 601 DNA (49 nucleotides from the N5 dyad; 5fC49) and one centered within the adjacent linker DNA (88 nucleotides from the N5 dyad; 5fC88) (Figures 3.1a

and B.2). Incorporation of 5fC at these positions was expected to provide insight into the role of nucleosome positioning on the removal of 5fC. Although nucleosomes are known to inhibit DNA glycosylases relative to linker DNA [119], [158], 5fC49 is oriented outward relative to the histone octamer surface where it is presumably more accessible to TDG. We note, however, that we cannot rule out that positioning 5fC elsewhere may lead to different outcomes. The arrays described above were reconstituted using histone octamers containing either canonical histone H4 or histone H4 lacking the N-terminal tail domain (residues 1−19, Δ4) (Figure B.3). Under the conditions used in our glycosylase assays (2 mM Mg^{2+}), canonical arrays fold into compact fibers similar to condensed chromatin in vivo (i.e., the 30 nm chromatin fiber), whereas H4 tailless arrays remain in a mostly extended "beads-on-a-string" conformation [33], [211]. This behavior was confirmed using a Förster resonance energy transfer (FRET)-based assay that measures the distance between nucleosome 5 (N5) and nucleosome 7 (N7) (Figures 3.1b-c and B.5) [117], [118]. Unlike canonical arrays (**12-NCP**), the FRET signal for H4 tailless arrays (**12-NCPΔ4**) was unchanged upon the addition of 2 mM Mg^{2+}, indicating that they remain in an extended conformation (Figure 3.1c). By using both canonical and H4 tailless arrays, we are able to probe TDG activity within two distinct structural states of chromatin under identical ionic conditions (i.e., 2 mM Mg^{2+}). The glycosylase activity of TDG was determined by measuring cleavage of the 5fC-containing strand following sodium hydroxide treatment of the abasic site product (Figure B.6). Due to significant product inhibition of TDG [152], all experiments were performed under single-turnover conditions using an excess of TDG (200 nM) over substrate (10 nM).

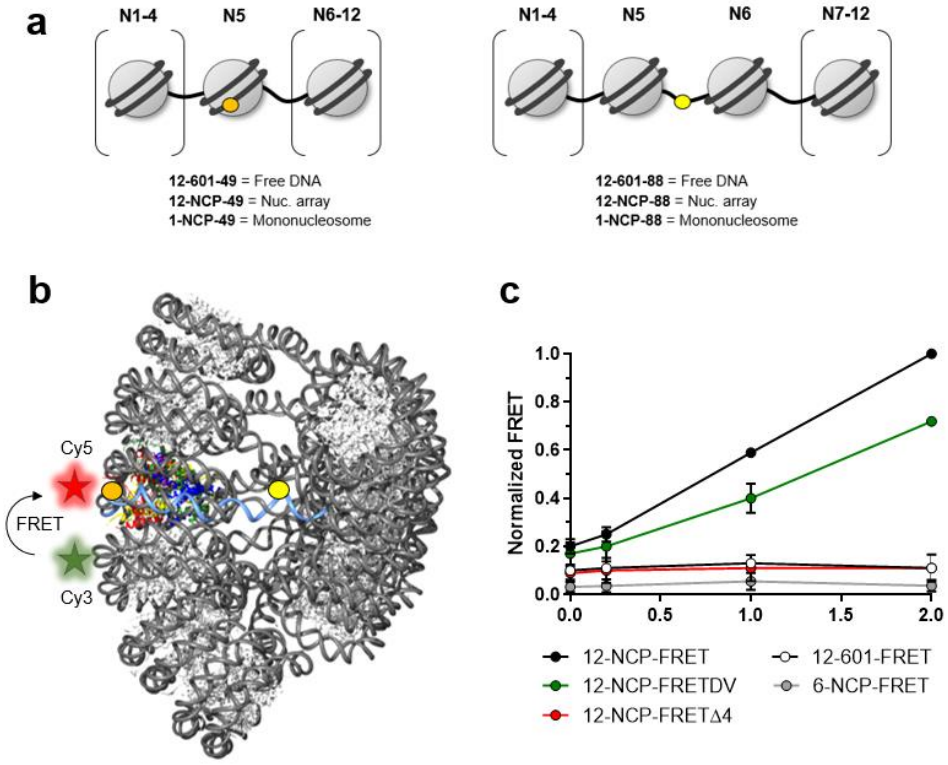


Figure 3.1: Chromatin substrates containing site-specifically incorporated 5fC residues. (a) Nucleosome arrays prepared in this study. (b) Cryo-electron microscopy structural model of the 30 nm chromatin fiber (12×601, EMD-2600) used in this study showing the locations of 5fC49 and 5fC88. The nucleosome surface (PDB ID: 1ZBB) was fitted to the electron density map using Chimera.20 Locations of the donor (Cy3) and acceptor (Cy5) for the FRET assay are indicated. (c) FRET analysis of nucleosome array compaction. Arrays were reconstituted using either canonical histone octamers (**12-NCP**), H4 tail deleted octamers (**12-NCPΔ4**), or **H2A.Z/H3.3** dual-variant octamers (**12-NCPDV**). All fluorescent intensities were normalized to canonical arrays (**12-NCP**) at 2.0 mM $Mg^{2.}$ As a control, we cut **12-NCP-FRET** in half at the DraIII restriction site such that the FRET dyes are on separate fibers (**6-NCP-FRET**).

We first examined the ability of TDG to excise 5fC49 positioned within the nucleosome core (Figure 3.2a). Compared to naked DNA (**12-601-49**), excision of 5fC49 was dramatically reduced on folded chromatin (**12-NCP-49**), resulting in only 8.5 ± 4% product formation after 30 min. Deletion of the histone H4 tail domain (**12-NCP-49Δ4**) only modestly increased the amount of 5fC49 that was excised by TDG (18.7 ± 0.9%).

Similar results were obtained for mononucleosomes containing an identically positioned 5fC site (**1-NCP-49**), which are unable to fold into higher order chromatin structures. Together, these results reveal that the majority of 5fC49 is inaccessible to glycosidic bond cleavage by TDG regardless of chromatin folding and that the histone octamer represents a major barrier for the removal of 5fC from chromatin.

Next, we examined the ability of TDG to excise 5fC88 from the linker DNA (Figure 3.2b). In contrast to 5fC49, TDG readily excised 5fC88 from both extended arrays (**12-NCP88Δ4**) and mononucleosomes (**1-NCP-88**), reaching ~80% product formation after 30 min. On compact arrays (**12-NCP-88**), however, TDG activity was significantly reduced (max extent = 33.8 ± 8.3%), which may reflect the introduction of steric clash between TDG and histone proteins upon chromatin folding. Alternatively, the limited flexibility of linker DNA within compact 30 nm fibers may prevent DNA bending by TDG [50], [220], which is required for target identification and glycosidic bond cleavage [143]. Irrespective of the structural basis, these results reveal that 5fC residues positioned between nucleosomes are susceptible to removal by TDG, although the level of chromatin folding impacts the process.

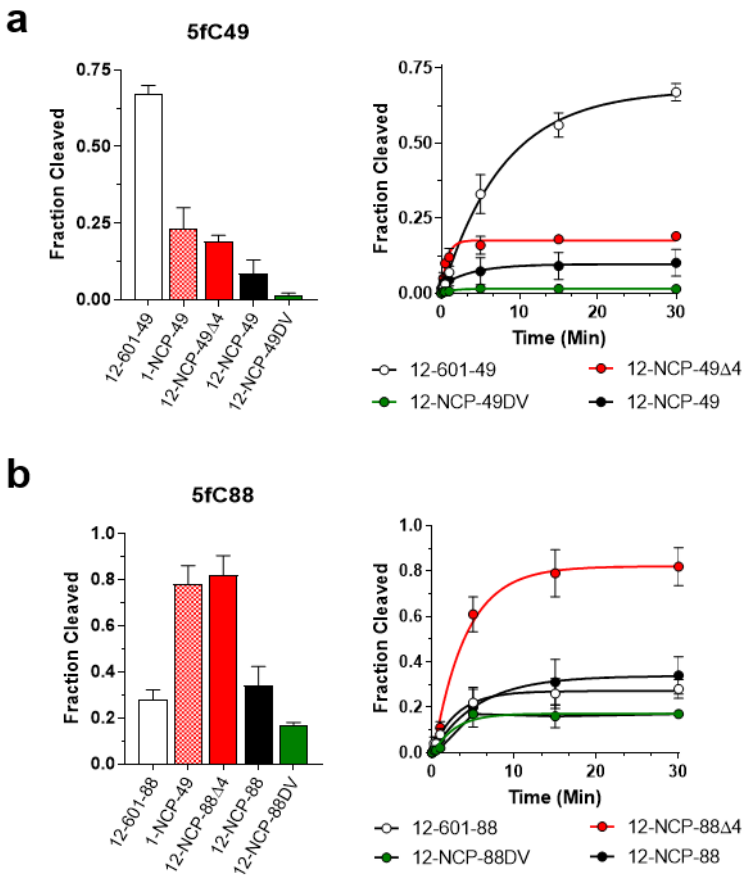


Figure 3.2: TDG-mediated removal of 5fC49 (a) and 5fC88 (b) from different chromatin environments. Bar graphs show the fraction of 5fC cleaved after 30 min. Error bars represent standard deviation from at least three independent experiments.

We observed that excision of 5fC88 on naked DNA (**12-601-88**) was reduced compared to mononucleosomes and extended arrays (Figure 3.2b), which is similar to what we previously observed for uracil DNA glycosylase (UDG) [158]. One possible explanation is that wrapping DNA into a nucleosome reduces the number of nonspecific TDG binding sites [179], which can compete with 5fC88 for binding to TDG [143]. Consistent with this hypothesis (and our observations for 5fC49), we found that TDG binds naked 601 DNA ~7-fold more tightly than 601 mononucleosomes (Figure 3.3a).

We also observed that excision of 5fC from **12-601-88** was less than **12-601-49** (Figure 3.2). This may reflect the high CpG content (i.e., TDG binding sites) near 5fC49, which could further facilitate its removal relative to 5fC88 (Figure B.1) [179].

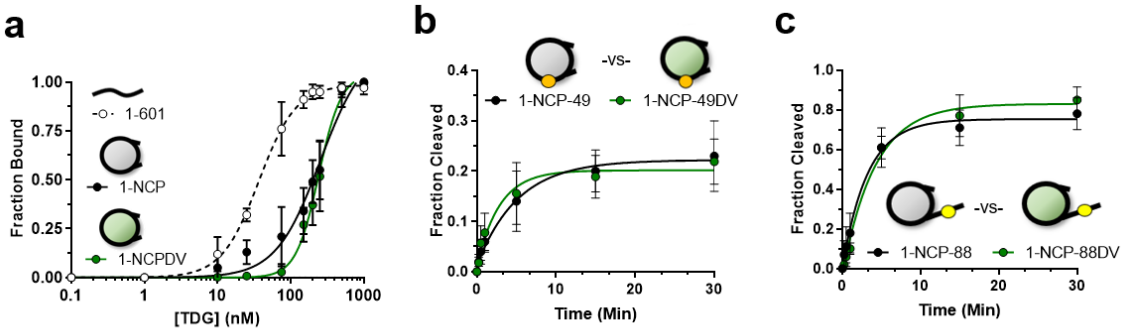
Collectively, the above results reveal that both chromatin folding and nucleosome positioning are capable of regulating TDG activity on DNA and, thus, are likely key contributors to the observed distribution of 5fC (and 5caC) throughout the genome. In particular, nucleosomes severely impede TDG activity, indicating that nucleosome-bound 5fC residues will be long lived in vivo. This is important because 5fC has recently been shown to form DNA−protein cross-links with histones [81], [123]. In addition, our results are consistent with the observation that 5fC sites overlap with nucleosomes at tissue-specific enhancers *in vivo* and provide a mechanistic explanation for the retention of 5fC at these sites [81].

3.1.2 H2A.Z/H3.3 and FOXA1 differentially regulate TDG

Finally, we extended these studies to a specific chromatin environment known to undergo active DNA demethylation in cells, namely forkhead box A1 (FOXA1) dependent enhancers [165]. In mammals, FOXA1 directs liver development and collaborates with hormone receptors to regulate prostate and breast-specific gene expression [216]. Importantly, recruitment of FOXA1 to enhancers leads to an “epigenetic switch” that induces local DNA demethylation through a mechanism that has been linked to BER [165], [202]. Whether FOXA1 activities contribute to the regulation of TDG, and thus 5fC metabolism, has yet to be delineated. Additionally, enhancers tend to be enriched with the

histone variants H2A.Z and H3.3 (including H2A.Z/H3.3 dual-variant nucleosomes), which help establish an environment permissive to TF-binding, including FOXA1 [194], [221]. Furthermore, 5fC colocalizes with both histone variants *in vivo* [183], suggesting a potential role for histones H2A.Z and H3.3 during DNA demethylation.

In order to replicate this environment in vitro, we assembled nucleosome arrays using histone octamers containing both H2A.Z and H3.3 (Figure B.3). Consistent with previous observations [190], H2A.Z/H3.3-containing arrays (**12-NCPDV**) were less compact than canonical arrays (**12-NCP**) as determined by FRET (Figures 3.1c). Nevertheless, the linker DNA within dual-variant arrays was found to be less accessible to restriction enzyme digestion than canonical arrays (Figure 3.4), suggesting that dual-variant arrays may actually adopt a more restrictive structure toward DNA-binding proteins. Accordingly, we found that excision of both 5fC49 and 5fC88 by TDG was further reduced on H2A.Z/H3.3-containing arrays (**12-NCP-49DV** and **12-NCP88DV**, respectively) relative to what we observed for canonical arrays (Figure 3.2). Most notably, TDG activity at 5fC49 was nearly abolished on H2A.Z/H3.3 arrays, with <2% of 5fC49 being converted to product. Interestingly, TDG activity on mononucleosomes was unaffected by the incorporation of the H2A.Z/H3.3 dual-variant octamer (Figure 3.3), indicating that the reduced glycosylase activity on dual-variant arrays was likely due to altered inter-nucleosome interactions. These data reveal that incorporation of H2A.Z/H3.3 into chromatin further inhibits TDG and, thus, may function to preserve the local epigenetic status (e.g., 5fC/5caC) of DNA.



Substrate	K_d
1-601	37 ± 2 nM
1-NCP	255 ± 49 nM
1-NCPDV	241 ± 12 nM

Substrate	K_{obs}
1-NCP-49	0.04 ± 0.01 min^{-1}
1-NCP-49DV	0.07 ± 0.02 min^{-1}

Substrate	K_{obs}
1-NCP-88	0.23 ± 0.03 min^{-1}
1-NCP-88DV	0.2 ± 0.02 min^{-1}

Figure 3.3: TDG activity is hampered by nucleosome structure independent of octamer identity. (a) Saturation plots for binding of TDG to naked 601 DNA (**1-601**) or 601-derived NCPs as determined by EMSA. Data were fit to a one site specific-binding equation with GraphPad Prism (v7.03) (b) Digestion of canonical and H2A.Z/H3.3 NCPs containing 5fC 49 or (c) 5fC88 by TDG. Error bars represent standard deviation from at least three independent experiments. Rates (k_{obs} min^{-1}) were calculated by fitting the digestion data to a single exponential equation and multiplying each rate constant by the Y_{max} as previously described [111], [122], [212].

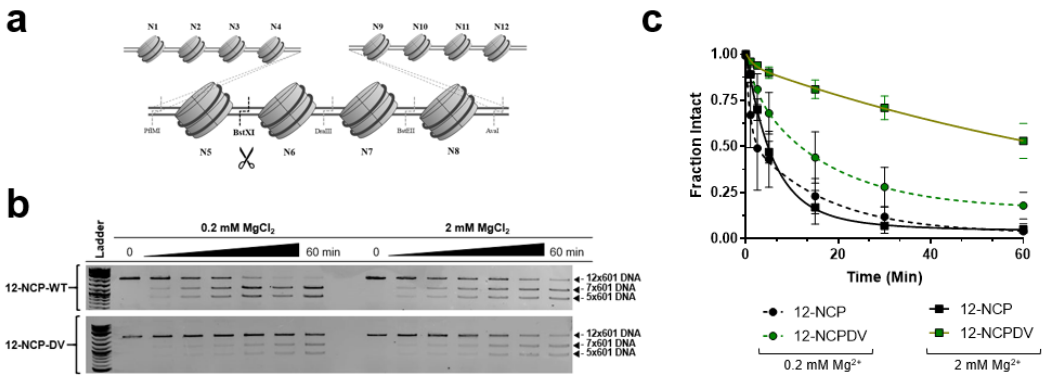


Figure 3.4: Restriction enzyme accessibility within canonical and H2A.Z/H3.3 (DV) arrays. (a) Schematic depicting the linker site probed by BstXI in accessibility assays. (b) Representative gel images of chromatin digestion reactions in the presence of 0.2 or 2 mM $MgCl_2$. (c) Plot depicting the time-dependent digestion of the indicated chromatin at both low and high Magnesium concentrations. Error bars represent standard deviation from at least three independent experiments.

In contrast, incubation of H2A.Z/H3.3-containing arrays with FOXA1 resulted in increased TDG activity at both 5fC sites (Figure 3.5). In fact, product formation at 5fC88 on compacted dual-variant arrays (**12-NCP-88DV**) and canonical arrays (**12-NCP-88**) in the presence of FOXA1 was nearly equivalent to what was observed for untreated extended arrays (**12-NCP-88Δ4**). Interestingly, FOXA1 did not stimulate TDG-mediated removal of 5fC49 from canonical arrays, indicating that this effect may be sensitive to the composition of the underlying histone octamer. Consistent with FOXA1's ability to open compact chromatin [74], we detected an increase in the inter-nucleosomal distance between N5 and N7 in the presence of FOXA1 for both canonical and dual-variant arrays (Figure 3.5e). In contrast, addition of either histone H1.1 or BSA resulted in further compaction (Figure 3.5f). These observations suggest a mechanism wherein FOXA1 stimulates TDG by altering the chromatin structure (presumably by increasing DNA accessibility).

Surprisingly, however, we found that FOXA1 also stimulated TDG on naked DNA (Figure 3.6). While future investigations are clearly needed to elucidate the exact mechanisms by which FOXA1 stimulates TDG activity on chromatin, these results demonstrate that, in addition to its chromatin remodeling activities [74], FOXA1 may promote DNA demethylation by stimulating TDG. It is worth noting that even in the presence of FOXA1, efficient removal nucleosome-bound 5fC residues will likely require additional factors, such as ATP-dependent chromatin remodelers [222].

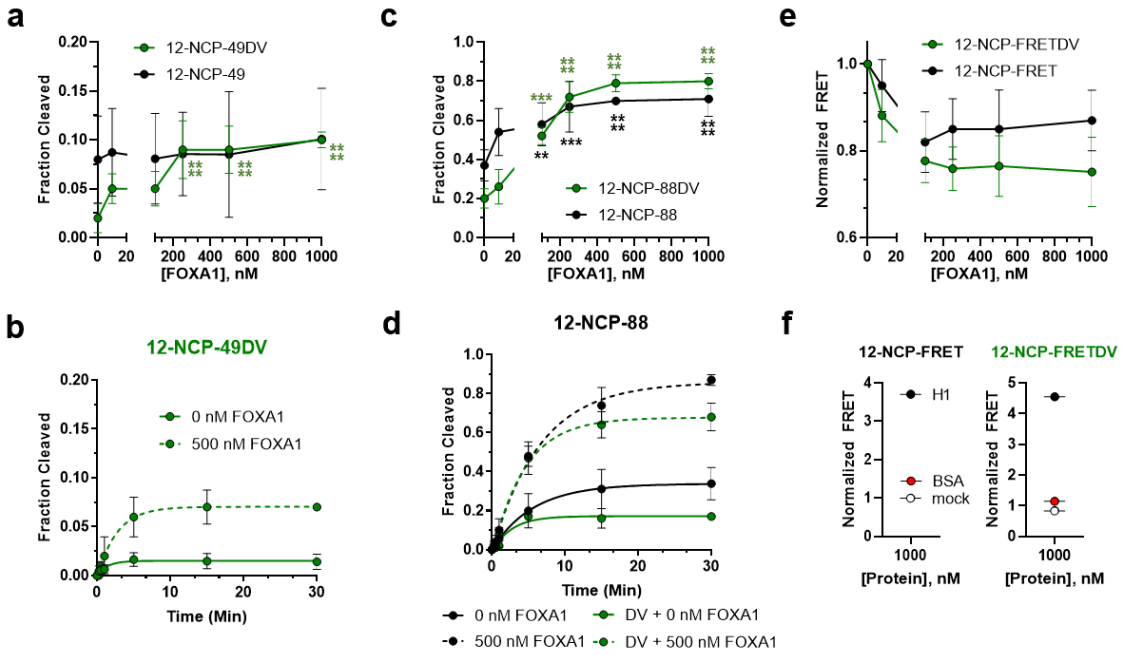


Figure 3.5: FOXA1 stimulates TDG activity on chromatin via fiber decompaction. (a) Fraction of 5fC49 or (c) 5fC88 cleaved following treatment of WT and DV arrays with TDG in the presence of FOXA1. (b) Time-dependent cleavage of the indicated 5fC49 or (d) 5fC88 array by TDG in the presence or absence of 500 nM FOXA1. Curves are fits to a single-exponential model. (e) FRET analysis of canonical (**12-NCP-FRET**) and dual-variant (**12-NCP-FRETDV**) arrays in the presence of FOXA1 (0–1000 nM) or (f) H1 or BSA. All FRET intensities were normalized to the same array in the absence of FOXA1. A significant difference ($p < 0.01$) in array compaction was detected for all FOXA1 concentrations except 10 nM. The concentration of chromatin (10 nM) and TDG (200 nM) were constant in all experiments, and all error bars represent standard deviation from at least three independent experiments.

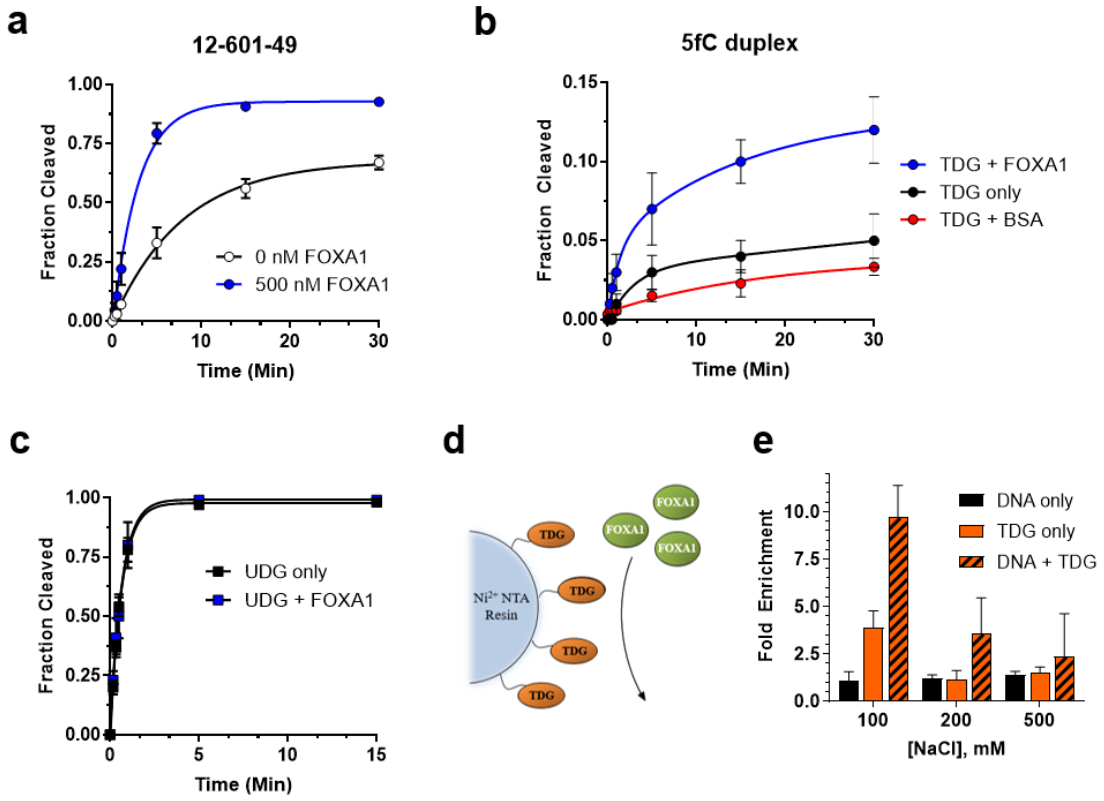


Figure 3.6: FOXA1 physically interacts with TDG and stimulates 5fC removal from free DNA. (a) Time-dependent cleavage of a **12-601-49** (10 nM) by TDG (200 nM) in the presence or absence of FOXA1. (b) Time-dependent cleavage of a 5fC-containing 33-mer (50 nM) by TDG (200 nM) in the presence or absence of FOXA1 or BSA (500 nM). (c) The same reaction depicted in (b) except with UDG (1 nM). Error bars represent standard deviation from at least three independent experiments. Curves are fits to a single-exponential model. (d) Schematic depicting FOXA1 pull-down using immobilized TDG. (e) Fold enrichment of FOXA1 eluted from Ni-NTA resin treated with DNA, (His-tagged) TDG, or DNA+TDG with increasing NaCl (see *3.3.15*).

3.2 Conclusions

In summary, we utilized novel nucleosome arrays containing positioned 5fC residues and H2A.Z/H3.3 to provide the first direct evidence that local chromatin states influence active DNA demethylation by regulating TDG. To the best of our knowledge, this was the first time active DNA demethylation has been studied using precisely modified nucleosome arrays. Our data provide fundamental mechanistic insights into the

observed genomic distribution of 5fC (and 5caC) and brings us closer to understanding how the DNA demethylation machinery is targeted throughout the genome. Our findings also showed that H2A.Z/H3.3-containing nucleosomes and FOXA1 may have opposing roles in controlling 5fC turnover. In particular, stimulation of TDG's glycosylase activity represents a novel role for FOXA1 in the DNA demethylation pathway and warrants further investigation.

3.3 Materials and Methods

3.3.1 General

Restriction enzymes (EcoRV, PflMI, BstXI, DraIII-HF, Nb.BbvCI, Nb.BssSI, HaeII, DraI), T4 DNA Ligase, APE1, T4 polynucleotide kinase, micrococcal nuclease (MNase), and shrimp alkaline phosphatase were purchased from New England Biolabs (Ipswich, MA). [γ^{32}P]-ATP was purchased from Perkin Elmer. N-Hydoxysuccinimide (NHS) esters of Cy3 or Cy5 dyes used in the labeling of **-53-Cy3** insert and **+51-Cy5** insert (Table B1) were acquired from Lumiprobe Life Science Solutions. Synthetic oligonucleotides (Table B1) were either purchased from Integrated DNA Technologies (IDT) or prepared by solid-phase synthesis on an Expedite 8909 DNA/RNA synthesizer using standard methods. DNA synthesis reagents and nucleoside phosphoramidites were purchased from Glen Research. Locked nucleic acid (LNA) phosphoramidites were obtained from Exiqon. The integrity of DNAs containing 5-formylcytosine was confirmed by mass spectrometry prior to use (Figure B.1).

3.3.2 Histone preparation and octamer refolding

Recombinant human histones (H2A.1, H2A.Z, H2B.1, H3, H3.3, H4, H4Δ-tail) were expressed and purified as described in Chapter 2 (see *3.3.7*) Histone octamers were refolded as previously described [121], [158], [223], and purified by size exclusion chromatography using a SuperDex 200 10/300 GL column (GE Healthcare Lifesciences). Purified histone octamers were stored at 4 °C in Octamer Buffer (2 M NaCl, 5 mM BME, 0.2 mM PMSF, 1 mM EDTA, 10 mM HEPES, pH 7.8). Histone H1.1 was purchased from Abcam (ab198676).

3.3.3 TDG expression and purification

The codon optimized gene for human thymine DNA glycosylase (UniProt identifier: Q13569-1) was purchased as a gBlock Gene Fragment from IDT and assembled by PCR as recommended by the manufacturer. The assembled DNA was cloned into a pET28a expression vector (Novagen) between the NdeI and HindIII restriction sites, generating plasmid **pET28a-hTDG**. Correct assembly of this plasmid was verified by DNA sequencing (Eton Bio).

TDG was expressed and purified using a modified version of the procedure originally reported by Kunz *et. al* [174]. The TDG plasmid assembled above (**pET28a-hTDG**; 10 ng) was transformed into BL21 (DE3) cells and the outgrowth (1 mL) was used to seed 4 × 25 mL cultures of Terrific Broth (TB) supplemented with 50 μg/mL Kanamycin. After shaking overnight at 37 °C, each 25 mL growth was used to innoculate 750 mL of TB supplemented with 2% glucose and 50 ug/mL Kanamycin. The cells were

grown to an OD_{600} ~0.800 at 30 °C with vigorous shaking and expression was induced with 0.5 mM IPTG at room temperature for 5 hours. Next, the cells were pelleted by centrifugation at 3900 rpm for 60 minutes using a swinging bucket rotor (all centrifuging steps were carried out at 4 °C) and the resulting cell pellet was resuspended in Buffer H0 (50 mM HEPES, pH 8.0, 300 mM NaCl, 10% glycerol, 1 mM PMSF); 40 mL of Buffer H0 was used per 1 L of cell culture. To begin lysis, the resuspended cell pellet was frozen at -80 °C (until solid) and then thawed on water at 4 °C. The cells were further lysed by sonicating on ice for 5 minutes, and lysates were cleared by centrifugation (10,000 × g, 60 minutes) and subsequently filtered with a 0.2 um syringe tip filter. The filtered lysate was passed over a 1 mL pre-packed TALON column (GE Healthcare Lifesciences) equilibrated with 5 column volumes (CV) of Buffer H0. The TDG-bound resin was then washed with 25 CV Buffer H1 (1 mM immidazole in Buffer H0), followed by 25 CV of Buffer H5 (5 mM immidazole in Buffer H0). Bound TDG was then eluted from the resin with 5 CV Buffer H500 (500 mM immidazole in Buffer H0). The TDG-enriched sample from the TALON filtration step was exchanged into Buffer HP50 (50 mM HEPES, pH 8, 50 mM NaCl 10% glycerol, 10 mM BME, 1 mM PMSF) using a 5mL HiTrap Desalting column (GE Healthcare Lifesciences). The protein sample was loaded onto a 5 mL Heparin column (GE Healthcare Lifesciecnes) that had been preequilibrated with 5 CV Buffer HP50. Bound proteins were eluted using a linear gradient (0% → 100%) of Buffer HP1000 (50 mM HEPES, pH 8, 1 M NaCl 10% glycerol, 10 mM BME, 1 mM PMSF) over 10 CV. Fractions containing TDG were pooled and directly loaded onto a Superdex

200 Increase 10/300 GL column. Pure TDG was isolated following isocratic elution with buffer HP50.

3.3.4 FOXA1 expression and purification

The codon optimized gene for full-length human FOXA1 (UniProt identifier: P55317-1) was purchased as a gBlock Gene Fragment from IDT and assembled by PCR as recommended by the manufacturer. The assembled DNA was cloned into the pET28a expression vector (Novagen) between the NcoI and XhoI restriction sites, generating plasmid **pET28a-FOXA1**). Correct assembly of this plasmid was verified by DNA sequencing (Eton Bio, San Diego, CA). FOXA1 was purified from E. coli as previously described [224].

3.3.5 Preparation of DNA substrates containing 5fC

Assembly of plasmids **pUC-12-601-49** and **pUC-12-601-88** containing unmodified 12×601 templates used to prepare **12-601-49** and **12-601-88**, respectively, was described previously [158]. A short oligonucleotide containing a single 5fC residue (Table B.1) was incorporated into these plasmids using the strand-exchange protocol described in Chapter 2 [158]. Briefly, plasmids **pUC-12-601-49** and **pUC-12-601-88** (500 μg, 160 pmol) were digested with 500 units of Nb.BbvCI nicking endonuclease for 1 hour in Cutsmart Buffer (50 mM KOAc, 20 mM Tris-acetate, 10 mM MgOAc, 100 μg/mL BSA, pH 7.9). To the digested plasmids was added 800 pmol of an LNA capture probe (**49-capture** for **pUC-12-601-49**; **88-capture** for **pUC-12-601-88**) (Table B1) and the reaction

mixture was heated at 80°C for 20 minutes before being slowly cooled to room temperature at 1 °C/minute. The resulting gap was then filled by the addition of 3.2 nmol of the appropriate 5'-[^{32}P]-labeled synthetic oligonucleotide insert containing either 5fC49 or 5fC88 (**N5_5fC49** or **N5_5fC88**, respectively) and subsequently the sample was heated at 50 °C for 2 minutes before being slowly cooled to room temperature at -1 °C/minute. After the gap-filling step, 400 units of T4 DNA ligase and ATP (2 mM final concentration) were added to the mixture. The ligation reaction was allowed to proceed for 4 hours at room temperature before inactivation of the ligase at 70 °C for 20 minutes. The efficiency of each step of the exchange process (nicking, insertion, and ligation) was carefully monitored in order to ensure complete (>95%) insertion of the modified oligonucleotide (Figure B.2). For this purpose, ~1 pmol aliquots were taken and digested with either PflMI/BstXI (**pUC-12-601-49**) or PflMI/DraIII (**pUC-12-601-88**) to release the 601 unit(s) harboring the 5fC-modified site. Aliquots were then treated with both TDG (0.5 μM) and APE1 (35 nM) in a reaction buffer containing 2 mM $MgCl_2$ at 37 °C for 30 minutes. The reactions were then analyzed by 10% native PAGE (29:1 acrylamide:bisacrylamide) (Figure B.2). After the integrity of the 5fC-containing DNA was confirmed, the modified 12×601 DNA template (**12-601-49** or **12-601-88**) was removed from the corresponding plasmid backbone via digestion with EcoRV. The plasmid was further digested with DraI and HaeII (600 units each) in order to degrade the plasmid DNA into fragments ≤ 700 base pairs (bp) in length. This digestion was monitored using 1% Agarose (1× sodium borate buffer). Following digestion, the modified 12×601 was purified via preparative agarose gel electrophoresis and extracted using a QIAquick

Gel Extraction Kit (Qiagen). The purified DNA was used directly to reconstitute oligonucleosome arrays (below).

DNA templates **1-601-49** and **1-601-88** used to assemble mononucleosomes containing 5fC (**1-NCP-49** and **1-NCP-88**, respectively) were prepared by PCR from the native 601 DNA sequence using either **N5_5fC49** (for **1-NCP-49**) or **N5_5fC88** (for **1-NCP-88**) and **CpG-601-FWD** as primers (Table B.1).

3.3.6 Reconstitution of mononucleosomes and nucleosome arrays

Reconstitution of both mononucleosomes and nucleosome arrays was carried out according to procedures described in Chapter 2. Immediately following the reconstitution step, samples were centrifuged at 13,000 × g for 20 min and the resulting pellets were discarded. Chromatin substrates were stored at 4 °C in nucleosome buffer (25 mM NaCl, 0.1 mM PMSF, 10 mM HEPES, pH 7.8) until use. Reconstituted arrays were analyzed by 0.6% agarose gel electrophoresis (Figure B.3) and reconstituted mononucleosomes were analyzed by 5% native PAGE (59:1 acrylamide:bisacrylamide) (Figure B.4).

3.3.7 Nucleosome saturation assay

Nucleosome saturation was confirmed by digestion of ~100 ng (~ 80 fmol) nucleosome arrays with 5 units of either PflMI/BstXI (**12-NCP-49**) or PflMI/DraIII (**12-NCP-88**) in nucleosome buffer supplemented with 2 mM $MgCl_2$. The respective naked DNAs were also digested under the same conditions, and both sets of samples (naked DNA and arrays) were analyzed side-by-side by 5% native PAGE (59:1, acrylamide:

bisacrylamide) (Figure B.3b,e). The presence of a nucleosome band as well as the absence of significant free DNA (<3%) demonstrates full nucleosome occupancies in these reconstituted arrays.

3.3.8 Partial micrococcal nuclease digestion of reconstituted arrays

The presence of 12 nucleosomes per array was confirmed through partial micrococcal nuclease (MNase) digestion. Arrays (200 ng, 160 fmol) were digested with 2 units of MNase for 1 minute at room temperature in nucleosome buffer supplemented with 1 mM $CaCl_2$. Reactions were stopped with the addition of 0.5% (v/v) SDS and 10 mM EDTA. The partially digested DNA was analyzed on a 1% agarose gel and visualized with ethidium bromide (Figure B.3c,f).

3.3.9 Analysis of chromatin compaction by FRET

Nucleosome arrays were fluorescently labeled with both Cy3 (Donor) and Cy5 (Acceptor) as previously reported [158], placing the dyes in the identical positions reported by Poirier et. al (Figure B.5) [117]. We refer to these substrates as **12-601-FRET** and **12-NCP-FRET** for naked DNA and arrays, respectively. This FRET system has been shown to measure Mg^{2+}-induced array compaction, reaching a maximally compact state concentrations exceeding 1 mM, as well as structural changes in the array due to protein binding. Array compaction was monitored via FRET using the methodology described by Hieb *et. al* [225]. Nucleosome arrays (10 nM) were equilibrated in nucleosome buffer supplemented with the indicated concentration of $MgCl_2$ at 37 ˚C for 5 minutes before

being transferred to a Nunc 384-Well Optical (glass) Bottom Plate (Thermofisher) (Figure B.5). When measuring the impact of either FOXA1, H1.1, or BSA on array compaction, the indicated amount of protein was added to the reaction mixture and incubated at 37 °C for 20 minutes prior to being transferred to the plate. The plate was imaged using a Typhoon FLA 9500 multimode imager (GE Healthcare Lifesciences) at 100 μm resolution using the laser settings and emission filters indicated below.6 Raw sample emission intensities for FRET (F_{raw}), Acceptor (A), and Donor (D) channels were quantified for the same area within each sample well using ImageQuant TL software (GE Healthcare Lifesciences).

F_{raw}: 532 nm ex.; 665 nm em.; PMT voltage: 675

D: 635 nm ex.; 665 nm em.; PMT voltage: 675

A: 532 nm ex.; 575 nm em.; PMT voltage: 625

All FRET intensities reported herein were corrected for spectral overlap as previously described [225]. Briefly, FRET (F) intensities were determined using Eq. 1, which corrects the raw FRET data (F_{raw}) for background resulting from donor bleed-through into the acceptor emission (X_d, Eq. 2), as well as background from direct acceptor excitation at the donor's excitation wavelength (Xa, Eq. 3) [225].

$$F_{corr} = F - (D \bullet X_d) - (A \bullet X_a)$$

X_d is obtained from the ratio of Fraw and D with the donor only sample (2).

$$X_d = F/D$$

X_a is obtained from the ratio of Fraw and A with the donor only sample (3).

$$X_a = F/D$$

Values of X_d and X_a were determined separately for each sample by measuring the donor- and acceptor- only samples under identical reaction conditions.

3.3.10 TDG digestion reactions

Reaction mixtures contained 10 nM of the indicated substrate in nucleosome buffer supplemented with 2 mM $MgCl_2$. Reactions were initiated by adding TDG (200 nM) and incubating at 37 ˚C for the indicated times. When FOXA1 was present, the indicated amount of the protein was added to the reaction mixture prior to TDG (Figure 3.5). For mononucleosomes, reactions were supplemented with 110 nM (11 equivalents) of unmodified mononucleosomes such that the ratio of modified to unmodified mononucleosomes, as well as total number of 601 units, was identical to the 12×601 arrays. Aliquots were removed at different times and quenched by the addition of 1% SDS (v/v) and Proteinase K (2 units/μL; New England Biolabs). After incubating at 37 ˚C for 15 minutes, the mixture was diluted to 125 μL with water and extracted with an equal volume of a solution containing phenol/chloroform/isoamyl alcohol (25:24:1). The DNA was then desalted by ethanol precipitation, and the resulting pellet was resuspended in CutSmart Buffer containing 5 units Nb. BbvCI in order to excise the 5'-[^{32}P]-labelled oligonucleotide insert from the much larger 12×601 DNA. The digestion was carried out for 45 minutes at 37 ˚C, at which point 0.1 M NaOH (final) was added and the sample was heated at 70 ˚C for 5 minutes to induce cleavage of any remaining abasic sites. The digested DNA was then resolved by 10% denaturing PAGE (19:1 acrylamide:bisacrylamide) and visualized using a Typhoon FLA 9500 gel imager. (Figure

B.6). The fraction cleaved was quantified using ImageQuant TL software (GE Healthcare, v8.1). At least three replicates were carried out for each substrate.

3.3.11 Electrophoretic mobility shift assays

The binding affinity of TDG to naked DNA and mononucleosomes was determined by EMSA. 5'-[^{32}P]-labelled 601 DNA or mononucleosomes (5 nM) were incubated with the indicated concentrations of TDG (0–1 μM) in nucleosome buffer supplemented with 0.2 mM $MgCl_2$ and 5% glycerol. The binding reactions were carried out at 37 ˚C for 20 minutes and were resolved by 5% native PAGE (59:1 acrylamide:bisacrylamide), which were run at 160 V for 30 minutes at 4 ˚C. The gel was visualized using a GE Typhoon gel imager and quantified using ImageQuant TL software (Figure B.7).

3.3.12 Restriction enzyme accessibility assay

The indicated nucleosome arrays (5 nM) were digested with BstXI (0.1 units/μL) in nucleosome buffer supplemented with either 0.2 or 2 mM $MgCl_2$ at 37 ˚C. Aliquots were taken at the indicated times and quenched with three volume equivalence of nucleosome buffer supplemented with 0.1% SDS, 2 units/μL Proteinase K, and 8% glycerol and then were heated at 60 ˚C for 10 minutes. Digested DNA was resolved by 1% agarose gel electrophoresis and visualized/quantified as described above (Figure 3.4).

3.3.13 TDG and UDG digestion of a short duplex DNA

A 5'-[^{32}P]-labeled 33 bp duplex was generated by annealing **N5_5fC49** with 2 equivalent of its compliment (**N5_5fC49-comp**; Table B.1) and was purified by 20% native PAGE (29:1 acrylamide:bisacrylamide). Reaction mixtures contained the above DNA duplex (50 nM) and the indicated protein effector (FOXA1, H1.1, or BSA) at 500 nM in nucleosome buffer supplemented with 2 mM $MgCl_2$. Reactions were initiated by the addition of TDG (200 nM) and were incubated at 37 ˚C for the indicated times (Figure 3.6). Aliquots were treated as described above and the digested DNA was resolved by 10% denaturing PAGE (19:1 acrylamide:bisacrylamide). The gel as visualized using a GE Typhoon gel imager and quantified using ImageQuant TL software. Digestion reactions with UDG (1 nM) were carried out identically using a substrate containing deoxyuridine (dU) in place of 5fC (N5_dU49; Table B.1).

3.3.14 Statistical analysis

All FRET data were presented as means and standard deviations. Statistical analysis was conducted using GraphPad Prism (v7.03). All data sets were first compared by unpaired one-way analysis of variance (ANOVA). For the FOXA1-FRET data, significant differences were determined between each concentration of FOXA1 tested and a control lacking any additional protein using a Dunnett's multiple comparisons test. In both cases, statistical significance was considered as P-values less than 0.05.

3.3.15 TDG-FOXA1 pulldown

For pull-down experiments the N-terminal His-tag of FOXA1 was removed via TEV digestion and >98% cleavage was confirmed via SDS PAGE (Figure B.8). TEV-treated FoxA1 (500 nM) was then equilibrated at 37 °C in 280 µL nucleosome buffer (0.2 mM $MgCl_2$) supplemented with either TDG only (200 nM), 1-601 DNA only (10 nM), or both TDG and 1-601 (200 nM & 10 nM, respectively). After a 30 minute incubation period the entire reaction mixture was passed through a poly prep column packed with 100 µL TALON resin (GE Lifesciences, Marlborough, MA), five times at room temperature – The TALON resin was blocked with 1 mg/mL BSA and washed with 10 mL nucleosome buffer prior to sample application.-- The protein-bound resin was then washed with 5 mL nucleosome buffer, followed by multiple 2.5 mL washes using buffers with increasing NaCl concentration (100, 200, and 500 mM). Finally, all proteins were eluted from the resin with 2.5 mL of elution buffer (500 mM imidazole).

Then, all 2.5 mL fractions exchanged into milli-Q H2O with a 10 kDa MWCO Ultra 4 Amicon, the volume was then measured by pipetting, and exactly half of each sample was lyophilized. The resulting pellet was re-suspended in 4 µL H_2O and spotted on a 0.2 µm nitrocellulose membrane (Bio-rad #1620112). The membrane was then blocked with 5% (w/v) fat-free milk and probed with a FOXA1-specific antibody (Abcam, #ab55178) at 1:275 dilution. The membranes were then treated with an Alexa647 anti-mouse antibody at 1:500 dilutions (Invitrogen- #A32728), imaged using a Typhoon FLA 9500 gel imager (PMT: 700 V, channel: Cy5), and quantified with Image Quant software. After subtracting background, which was determined by measuring an equivalent area of

a non-protein-bound portion of the membrane, all dot intensities were converted into %-retained values based on the average of 4 blots containing 2.5 % of the overall FOXA1 input. Because the TEV digestion did not go to 100% completion, a small fraction of FOXA1 in our pull-down assay was capable of binding the resin independent of TDG or DNA. We expected this background binding would be consistent across experiments and could therefore be used to control (normalize data) for equal sample loading. Although not ideal conditions, we believe this is a conservative approach towards identifying a potential interaction between TDG and FOXA1 (because FOXA1 molecules that remain bound to either DNA or resin-bound TDG until the imidazole wash would contribute to the background binding and ultimately deflate the normalized value for % retained).

CHAPTER 4

REVERSIBLE CHROMATIN CONDENSATION BY THYMINE DNA GLYCOSYLASE

Nucleosomes, the basic unit of chromatin, interact with each other over short distances to form locally compact structures (e.g. 30 nm diameter fibers) that modulate DNA accessibility at the level of single genes [211]. On a larger scale, long-range chromatin fiber contacts within, and between, chromosomes drive the condensation of chromatin into distinct structural domains that are key to genome organization and function [226]. Unraveling the molecular mechanisms underlying the formation and regulation of these locally and globally condensed chromatin structures, and identifying the protein factors involved, is essential for understanding the fundamental genomic processes of the cell.

Having established the relationship between chromatin structure and TDG catalysis (Chapter 3), we next wanted to examine the inverse relationship: the effect of TDG on chromatin structure. Beyond its catalytic roles, TDG has been shown to function as a transcriptional co-activator through its association with various transcription factors and activating histone modifiers, such as the histone acetyltransferases CBP and p300 [162], thereby coordinating the formation of a transcriptionally permissive chromatin state [161], [227]. TDG also mediates long-range physical contacts between promoters and enhancers at a subset of hormone responsive genes [64], [75]. While TDG's role in chromatin organization is generally viewed as indirect, we recognized that TDG possesses

several features that suggest it may have the capacity to alter chromatin structure directly: TDG binds non-specifically to DNA [26] and nucleosomes [124] *in vitro* and contains an intrinsically disordered lysine-rich regulatory domain that closely resembles the C-terminus of linker histone H1, a basic peptide known to promote chromatin fiber folding and condensation [113], [228], [229]. Herein, we now provide the first experimental evidence that TDG can directly alter chromatin structure through its physical interactions with DNA. Importantly, we show that TDG promotes condensation of chromatin fibers into higher-order oligomeric structures, thereby linking TDG-dependent pathways to long-range chromatin organization.

4.1 Results

4.1.1 TDG condenses chromatin through interactions with linker DNA

We first considered the ability of TDG to alter chromatin structure at the single-fiber level (we use "TDG" throughout to refer to the full-length human protein). In particular, we focused on TDG's ability to bind to and alter the structure of nucleosome arrays compacted into "30 nm" chromatin fibers. Previous biophysical studies have shown that nonspecific binding of transcription factors (TFs) to nucleosome arrays, and specifically to extra-nucleosomal (or "linker") DNA, causes array decompaction [117], [124]. Given TDG's high affinity for DNA, even in the absence of a target nucleobase [179], we reasoned that TDG may also drive chromatin decompaction through similar interactions. To test this, we assembled 12-mer nucleosome arrays containing fluorescent

donor and acceptor dyes that were placed at locations that allow nucleosome stacking interactions, and thus the overall compaction of the array, to be monitored by FRET (Förster resonance energy transfer) (Figure 4.1 and C.1) [117], [124]. Under the conditions used in our assay (2 mM Mg^{2+}), nucleosome arrays fold into maximally compact 30 nm fibers [211], which is accompanied by a characteristic increase in FRET (Figure 4.1a) [117], [124]. In the presence of 200 nM TDG, this FRET signal was reduced by ~40%, indicating that TDG induced decompaction of the arrays. In fact, this effect was more pronounced for TDG than with the pioneering TF FOXA1, which is known to actively initiate chromatin decompaction and promote DNA accessibility [74], [124]. We presumed that the relevant TDG binding occurred with the linker DNA because, compared to nucleosomal DNA, linker DNA more closely resembles typical B-form DNA [46], [220], is generally more accessible to DNA binding factors [102], and is a better substrate for TDG's glycosylase activity [124], [153]. Consistently, we found that TDG binds mononucleosomes containing linker DNA (30 bps) more strongly than those without (Kd = 59 ± 7 nM and 255 ± 49 nM, respectively) (Figure 4.1b), and protects linker DNA within nucleosome arrays from micrococcal nuclease (MNase) digestion (Figure 4.1d). Overall, these results strongly suggest that TDG drives chromatin fiber decompaction through nonspecific binding to linker DNA.

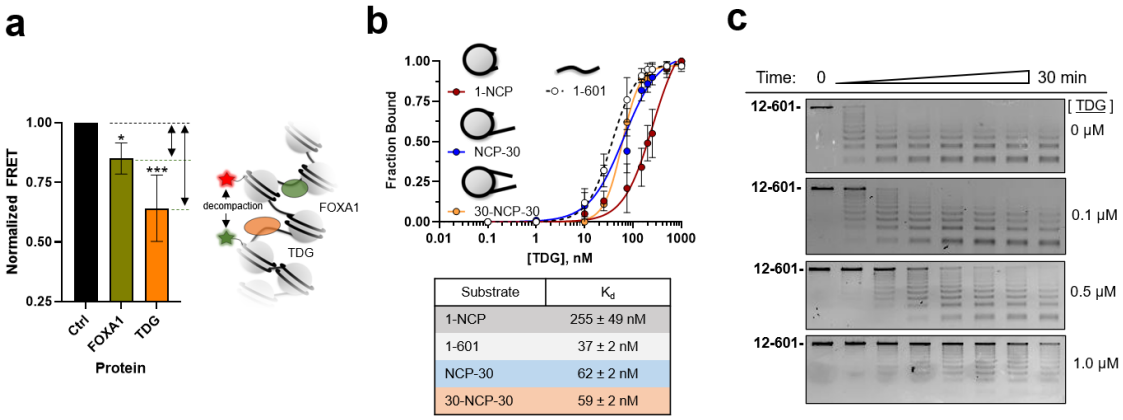


Figure 4.1: TDG locally de-compacts chromatin fibers via interactions with linker DNA. (a) FRET analysis of compact 12-mer arrays (2 mM Mg2+) in the presence of TDG (200 nM) or FOXA1 (1 μM). All fluorescent intensities were normalized to the 12-mer array in the absence of protein additives (Array). ***P < 0.001. (b) Saturation plots for binding of TDG to naked 601 DNA or mononucleosomes having different arrangements of linker DNA. The Kd is listed below each substrate. (c) MNase digestion of nucleosome arrays in the presence of TDG. The concentration of TDG (nM) used in each experiment is listed to the right. Error bars represent standard deviation from at least three independent experiments.

We next asked whether TDG's association with chromatin fibers influences their ability to undergo oligomerization (also referred to as "condensation"). *In vitro*, individual chromatin fibers undergo self-association into higher-order oligomeric structures at Mg^{2+} concentrations greater than 3-4 mM, a process that mimics the formation of long-range intra- and inter-fiber interactions observed in native chromatin [34], [211]. We incubated 12-mer nucleosome arrays with increasing concentrations of Mg^{2+}, removed the precipitated oligomers by centrifugation and quantified the unassociated fibers in the supernatant. Compared to Mg^{2+} alone, the presence of TDG resulted in a profound increase in array oligomerization (Figure 4.2a). A similar affect was observed for the monovalent cation K^+ (Figure 4.2b). Notably, significant precipitation of the arrays was observed even in the absence of these added salts, suggesting that TDG alone is capable of inducing chromatin condensation. Therefore, we excluded Mg^{2+} and K^+ from the following

experiments to ensure that the observed chromatin condensation could be attributed solely to TDG. Indeed, titration of 12-mer arrays with only TDG led to a concentration-dependent increase in precipitated material (Figures 4.2c and C.3), with the midpoint for array oligomerization occurring at ~200 nM TDG (~3:1 molar ratio of TDG to mononucleosome). For comparison, the midpoint for histone H1.1-induced oligomerization occurred at ~50 nM (~1:1 molar ratio of H1.1 to mononucleosome). The fact that TDG induced both nucleosome array decompaction, as well as inter-fiber oligomerization, at similar concentrations suggests that these two processes are coupled. We chose not to incubate TDG with arrays that had been pre-bound by H1.1 because these proteins are localized to different chromatin domains *in vivo*. For instance, TDG and its substrates, 5fC/5caC, are primarily localized to *active* promoter and enhancers [64], [67], [68], whereas linker histone H1.1 is depleted at these sites [230], [231]. This indicates that TDG will most often be bound to chromatin lacking H1.1.

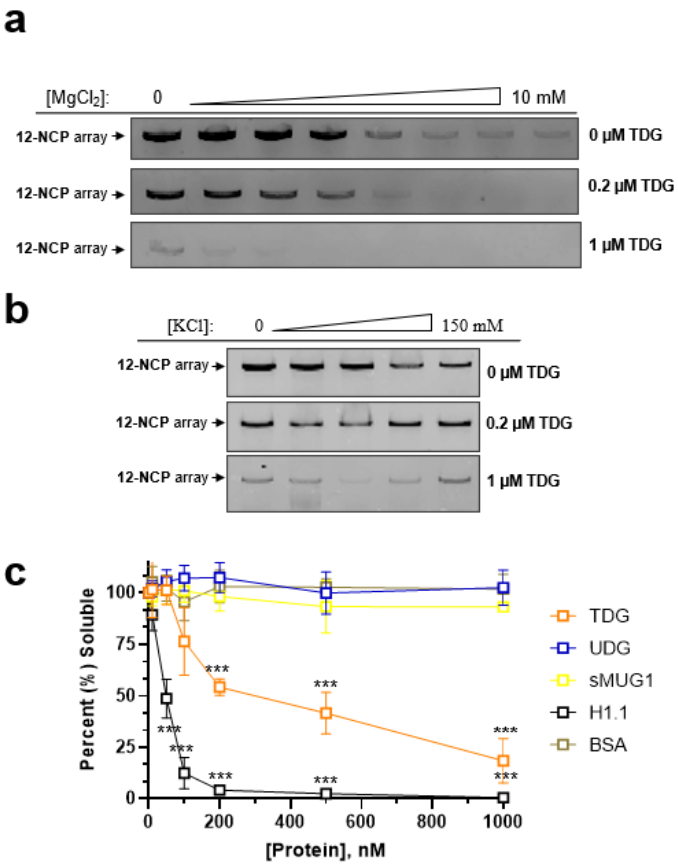


Figure 4.2: Analysis of TDG-chromatin condensates via precipitation. (a) Representative gel images showing the soluble fraction from Mg^{2+} dependent (0 – 10 mM) oligomerization experiments carried out in either the presence or absence of TDG. (b) The same experiment depicted in part (a) except with 0, 25, 50, 100, or 150 mM KCl. (c) Percent of chromatin remaining in solution (un-condensed) following incubation with the indicated protein.

In contrast to TDG, two related DNA glycosylases, uracil DNA glycosylase (UDG) and single-stranded mono-functional uracil glycosylase (SMUG1), as well as BSA, had no effect on chromatin solubility (Figures 4.2c and C.3). TDG-mediated array oligomerization still occurred in the absence of histone N-terminal tail domains (Figure 4.3a and C.3). Thus, histones tails are not essential for TDG-mediated chromatin condensation. This is in contrast to linker histones, which have been shown to require the histone tail domains to induce oligomerization [51]. This suggests that histone H1 and

TDG promote chromatin condensation through distinct mechanisms. Although the histone tail domains were mostly dispensable for array oligomerization by TDG, nucleosome cores are essential, as TDG failed to precipitate free 12-mer DNA (Figure C.3). Importantly, we found that oligomerization is not coupled to DNA sequence, as TDG precipitated chromatin reconstituted from human genomic DNA (Figure 4.3d and C.4) and, finally, it is worth emphasizing that the nucleosome arrays employed in this study lack substrates for TDG base excision, indicating that catalysis is not a requirement for chromatin condensation.

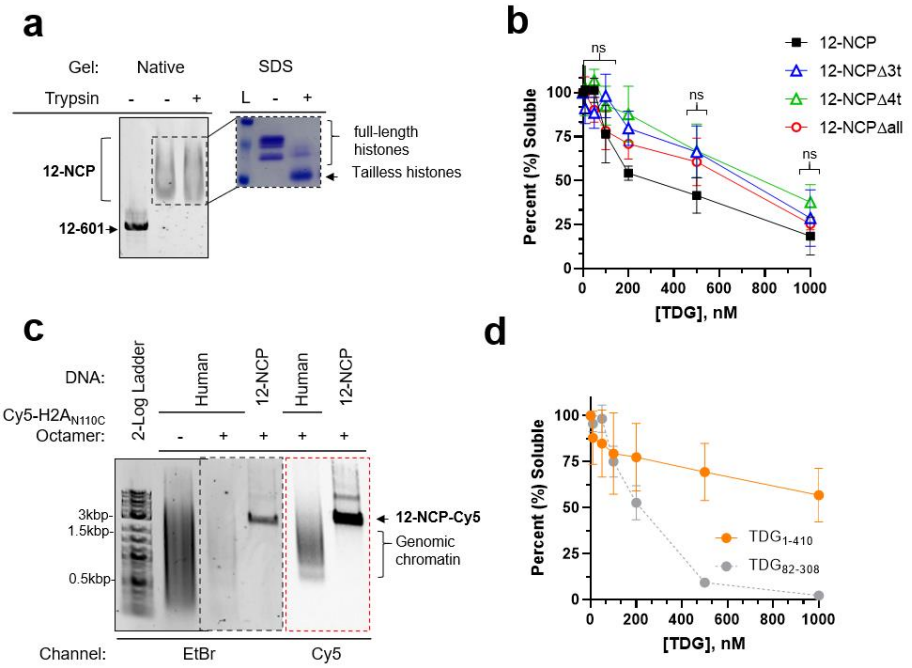


Figure 4.3: : TDG induces chromatin oligomerization independent of the histone tail domains and DNA sequence. (a) Agarose gel (0.6%) (left) and SDS PAGE (15%) (right) analysis of **12-NCP** arrays following digestion with Trypsin (2.5 Units/µL) to remove histone tail domains. (b) Representative gel images showing the soluble fraction of various tailless **12-NCP** arrays (or free **12-601** DNA) following treatment with increasing concentrations of TDG. Histone octamer composition is listed top the right of each gel image. (c) Solubility plot for the experiments shown in part (b). Error bars represent standard deviation from at least three independent experiments.

We further confirmed that the insoluble TDG-complexes were in fact oligomers, comprising multiple chromatin fibers, using an inter-fiber FRET-based assay (Figure 4.4 and C.5). Nucleosome arrays were labelled separately with either Cy3 (donor) or Cy5 (acceptor) dyes via maleimide conjugation to histone H2A bearing a N110C mutation, and the labelled arrays were mixed in a 1:1 ratio. Upon fiber oligomerization, which has been proposed to involve interdigitation of nucleosomes between different fibers [232], the dyes become close enough in space to allow for efficient FRET [90]. We first validated the method using Mg^{2+}, which is well known to induce chromatin fiber oligomerization. Consistently, titration of the donor/acceptor array mixture with increasing concentrations of Mg^{2+} resulted in a concentration dependent increase in FRET. Moreover, treatment of condensed arrays (5 mM Mg^{2+}) with MNase significantly reduced the FRET signal (Figure 4.4c), which is consistent with the inability of mononucleosomes to undergo Mg^{2+}-induced oligomerization [233]. These observations strongly suggest that the FRET system properly monitors inter-fiber oligomerization. We then applied the assay to TDG. Treatment of the donor/acceptor array mixture with 1 μM TDG, which induces nearly complete array precipitation (Figure 4.2.b), resulted in a pronounced increase in inter-fiber FRET relative to untreated arrays (Figure 4.4c). This FRET signal was similar to arrays treated with 5 mM Mg^{2+}, which also induces complete array precipitation. Collectively, these data indicate that the observed precipitation of nucleosome arrays by TDG is indeed due to inter-fiber oligomerization. Interestingly, incubation of MNase-treated arrays (i.e. mononucleosomes) with TDG resulted in a modest reduction in the inter-molecular FRET signal compared to what was observed for Mg^{2+} (5 mM) (Figure 4.3c). This difference

suggests that oligomerization by TDG involves more specific bridging interactions between nucleosomes and TDG, and is consistent with our observation that, despite binding tightly to DNA, TDG requires nucleosome cores to induce array oligomerization.

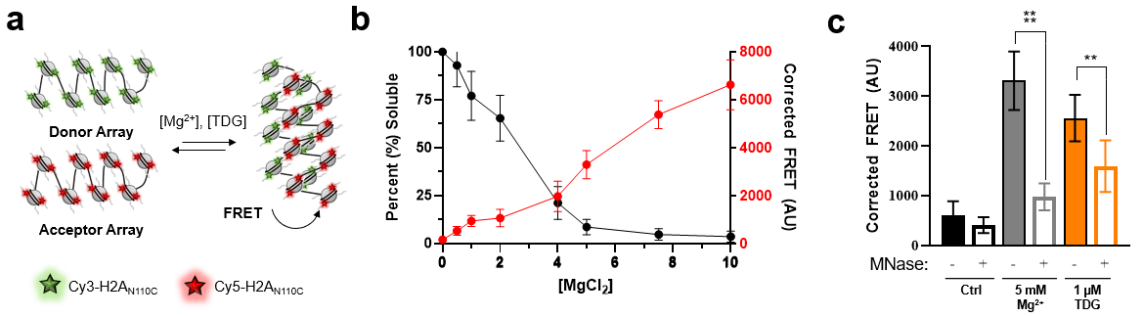


Figure 4.4: Analysis of chromatin condensation via FRET. (a) Schematic of inter-fiber FRET approach. (b) Mg^{2+}-induced oligomerization of nucleosome arrays. Precipitation data (black) is shown on the left Y-axis, and inter-fiber FRET data (red) is shown on the right Y-axis. (c) Comparison of the inter-fiber FRET signal for arrays treated with Mg^{2+} or TDG. Error bars represent standard deviation from at least three independent experiments.

4.1.2 The N- and C-terminal domains of TDG have opposing roles during chromatin condensation

The linker histone H1.1 contains a disordered positively charged C-terminal domain (CTD) that is responsible for stabilizing secondary chromatin structures and promoting chromatin condensation [113], [228], [229]. The N-terminal domain (NTD) of TDG (residues 1-110; Figure 4.5), which confers enhanced DNA binding and has other important regulatory functions, shares a number of similarities with the histone H1.1 CTD: they are both highly basic, mostly disordered, and have low sequence complexity (Figure 4.5b) [161], [215], [227]. This observation prompted us to ask whether TDG's NTD is responsible for mediating chromatin condensation. Indeed, we found that deletion of residues 1–110 ($TDG_{111-410}$) completely abolished TDG's ability to condense nucleosome

arrays (Figures 4.5 and C.6). Interestingly, deletion of TDG's CTD (residues 309–410; $TDG_{1\text{-}308}$ and $TDG_{82\text{-}308}$) had the opposite effect, instead promoting array oligomerization in the presence of the NTD. Notably, $TDG_{82\text{-}308}$, which contained only the catalytic domain and a particularly basic region of the NTD (residues 82–110), was capable of condensing chromatin with nearly the same efficiency as histone H1.1. Given that the catalytic domain alone poorly oligomerized arrays ($TDG_{111\text{-}308}$), this suggests that TDG's ability to condense chromatin is localized to residues 82–110 of the NTD. Previous studies using short DNA duplexes have shown that these residues form high-affinity non-specific complexes with DNA [159], [234]. Thus, in the context of chromatin, it is possible that residues 82–110 may bind DNA between arrays to facilitate oligomerization. Our observation that TDG-mediated oligomerization is impaired by the presence of its full N- and C-terminal domains provides additional support for this mechanism, as those domains have been shown to destabilize non-specific interactions between residues 82–110 and DNA [234]. We cannot rule out, however, that TDG dimerization, which has been observed at very high TDG concentrations (> 1 μM) [147], contributes to array oligomerization. If so, our results suggest that the CTD may also destabilize this interaction (compare $TDG_{111\text{-}308}$ to $TDG_{111\text{-}410}$) (Figure 4.5c).

In order to gain further support for a mechanism involving TDG's NTD, we tethered residues 1–110 to a nucleosome-targeting peptide from the Kaposi's sarcoma-associated herpesvirus protein LANA (LANA-$TDG_{1\text{-}110}$) [113], [235]. Remarkably, incubating 12-mer arrays with the LANA-$TDG_{1\text{-}110}$ fusion protein induced oligomerization to a similar extent as full-length TDG (Figure 4.5d and C.7). This effect

was dependent on the attachment of TDG_{1-110} to LANA, as proteolytic cleavage of their linkage significantly impaired array oligomerization ($LANA+TDG_{1-110}$). These results not only confirm that TDG-mediated oligomerization is derived largely from its disordered NTD, but also suggests that TDG's folded catalytic domain (i.e. the "reader" domain) serves in this context to recruit the lysine-rich NTD to chromatin. These data are also in agreement with the previous finding that the NTD binds DNA regardless of whether or not it is attached to TDG, and through similar interactions [234]. As expected, a fusion protein comprising LANA and TDG's CTD ($LANA-TDG_{309-410}$) had no effect on array solubility.

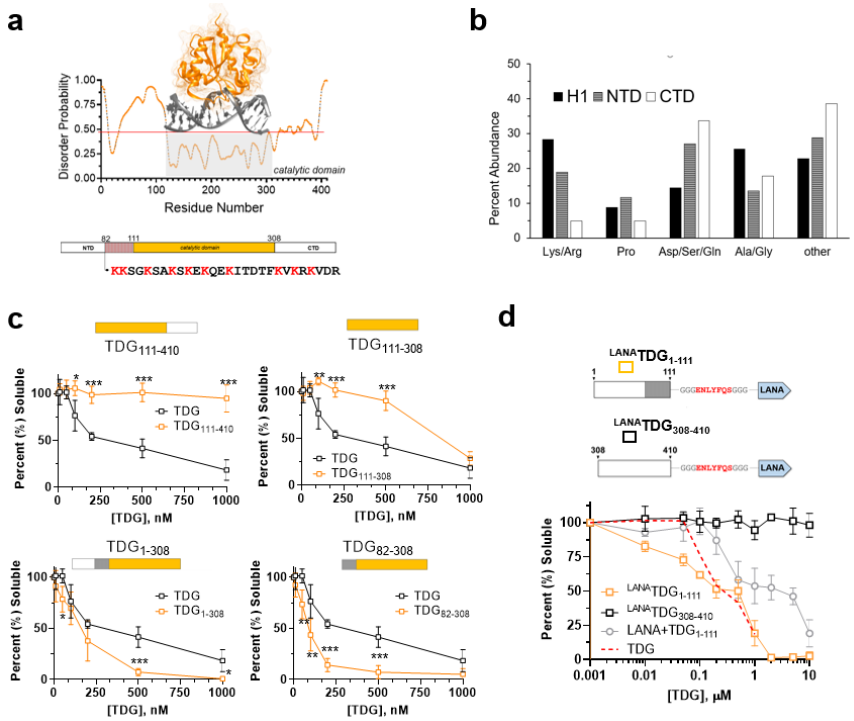


Figure 4.5: TDG-mediated chromatin oligomerization is regulated by its intrinsically disordered N- and C-terminal domains. (a) Disorder probability for all residues within TDG determined by the Protein DisOrder prediction System (PrDOS).[236] (b) Percent abundance of low complexity amino acids for Histone H1.1 and the indicated TDG tail domains. (c) Side-by-side comparison of the solubility plots for the indicated TDG deletion mutant and wild-type TDG. (d) The same solubility plots depicted in part (c), but for experiments conducted with TDG-LANA fusion constructs. Error bars represent standard deviation from at least three independent experiments.

4.1.3 TDG-mediated chromatin condensation is reversible

If chromatin fiber oligomerization is driven by non-specific inter-fiber interactions between TDG's NTD and DNA, it should be possible to disrupt, and thus re-solubilize, the resulting oligomers using competitor DNA. To test this, we treated insoluble TDG-chromatin oligomers with increasing concentrations of free 207 bp 601 DNA and measured the soluble fraction. Consistent with our hypothesis, excess free DNA was capable of reversing array oligomerization by TDG (Figures 4.6 and C.7). Importantly, the 12-mer arrays remained intact throughout the precipitation and redissolution cycle (Figure C.7), as does TDG's catalytic activity (Figure C.7c). Therefore, like oligomerization by divalent cations [58], TDG-mediated chromatin condensation is freely reversible. Insoluble H1.1-chromatin oligomers were also reversible by free DNA, but this process was much less gradual than for TDG (Figure 4.6 and C.7). This further highlights the different mechanisms used by these two proteins to condense chromatin. Surprisingly, free DNA was unable to re-solubilize arrays that had been precipitated by $TDG_{1\text{-}308}$ and $TDG_{82\text{-}308}$, indicating that reversibility is highly dependent on the presence of the CTD (Figure C.7). This again is consistent with the CTD acting to destabilize inter-array interactions between the NTD (presumably residues 82–110) and DNA, in this case being required to prevent irreversible oligomerization. Moreover, these data imply that reversal of array oligomerization by DNA does not involve TDG's catalytic domain.

4.1.3.1 Growth Arrest and DNA Damage Inducible Alpha (Gadd45)

The N- and C-terminal domains of TDG have been shown to mediate interactions with numerous protein partners [159], [160], [225], [235]. On the basis of our data above, we predict that these interactions might be capable of altering the formation and/or stability (i.e. reversibility) of TDG-mediated chromatin oligomers. We decided to explore this possibility using growth arrest and DNA damage inducible alpha (GADD45a). GADD45α is a low molecular weight (18 kDa) and highly acidic protein (pI = 4.36), present in every cell type, that has known implications in DNA repair, cell cycle maintenance, and epigenetic reprogramming [238]–[240]. Although a detailed mechanistic understanding of how GADD45α participates in these pathways is not clear, *in vivo* and *in vitro* data indicate this acidic protein increases DNA accessibility at the nucleosome and chromatin level through electrostatic interactions with the core histone tails (Figure C8) [239], [241]. In addition to its roles in cell growth control, genomic stability, and DNA repair, GADD45a has been shown to functionally and physically interact with TDG to promote removal of 5fC/5caC from genomic DNA [181], [237], [240]. Importantly, these interactions involve both the N- and C-terminal domains of TDG (residues 1–132 and 178–397, respectively) [237]. As before, we exposed insoluble TDG-chromatin oligomers to increasing concentrations of GADD45a and monitored the change in solubility (Figures 4.6 and C.8). We found that GADD45a readily reversed array oligomerization by full length TDG, with nearly half the precipitated arrays becoming re-solubilized in the presence of a stoichiometric amount of GADD45a relative to TDG.

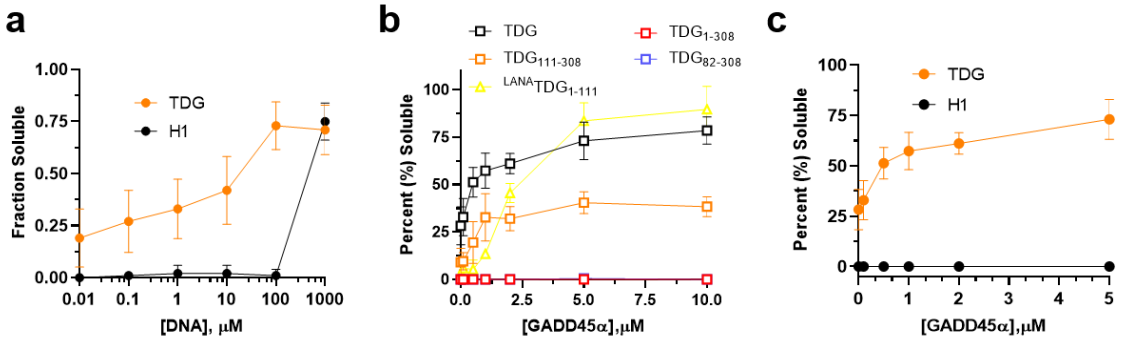


Figure 4.6: TDG-mediated chromatin oligomerization is reversible. Insoluble chromatin oligomers were incubated with the indicated concentration of 601 DNA (a) or GADD45a (b-c), and the change in solubility was monitored following centrifugation. Error bars represent standard deviation from at least three independent experiments.

Importantly, GADD45a does not re-solubilize H1.1-chromatin oligomers (Figure 4.6c and C.8), which further demonstrates the specificity of the interaction between GADD45a and the disordered tail domains of TDG. GADD45a was also capable of reversing array oligomerization by LANA-$TDG_{1\text{-}110}$, although with reduced efficiency, revealing that the functional protein-protein interaction involves at a minimum TDG's NTD. In agreement with our DNA competition experiments, re-solubilization of TDG-array oligomers by GADD45a was dependent on the presence of TDG's CTD in the context of the full-length protein (Figure 4.6b), further supporting a model wherein TDG's CTD potentiates the disruption of NTD-mediated inter-fiber interactions by external regulators. Collectively, these data demonstrate that TDG-mediated chromatin condensation can be regulated through protein-protein interactions involving its NTD (and presumably its CTD), and importantly, implicate GADD45a in controlling chromatin structural organization through its association with TDG. It is worth noting that multiple lysine residues within the N- and C- terminal domains of TDG undergo posttranslational

modification (e.g. acetylation, phosphorylation, and Sumoylation), which has been shown to influence TDG's interactions with DNA and other proteins [242]. By extension of our results above, we anticipate that these modifications will also impact the formation and/or stability of TDG-mediated chromatin oligomers, thereby meriting further investigation.

4.1.4 DNA methylation impairs chromatin condensation by TDG

Finally, we tested whether TDG-dependent chromatin oligomerization was sensitive to the methylation status of the underlying DNA using nucleosome arrays that had been hypermethylated by the CpG methyltransferase MssSI (Figure 4.7a and C.9). We found that DNA methylation significantly impaired array oligomerization by TDG, with the majority of methylated arrays (~70 %) remaining soluble following exposure to 1 μM TDG (Figure 4.7b). DNA methylation also inhibited chromatin condensation by $^{LANA}TDG_{1\text{-}111}$, albeit to a lesser extent than with the full protein.

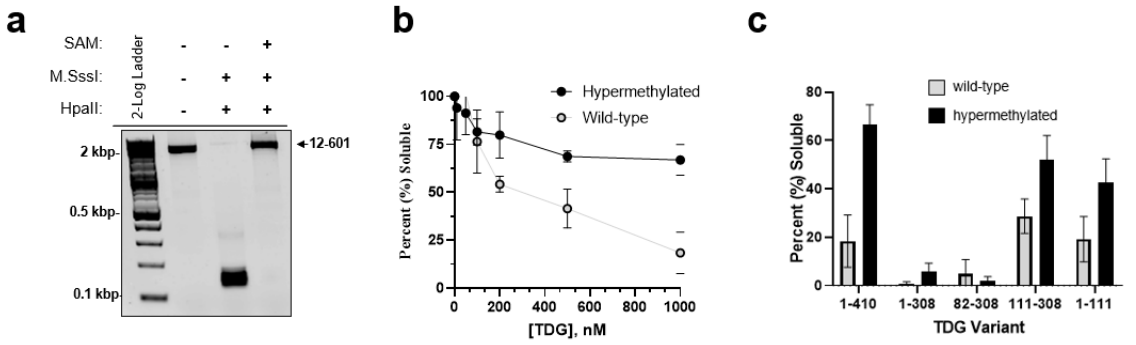


Figure 4.7: DNA methylation inhibits TDG-mediated chromatin condensation. (a) Hypermethylation of CpG dinucleotides within **12-601** DNA confirmed by resistance to HpaII digestion (0.7% agarose). (a) Precipitation assay to monitor nucleosome array oligomerization. (Un)methylated nucleosome arrays were incubated with the indicated concentration of TDG, oligomers were removed by centrifugation, and the percentage of arrays remaining in solution was determined by gel electrophoresis. (b) Soluble fraction following treatment of (un)methylated arrays with different TDG variants (1 μM). Error bars represent standard deviation from at least three independent experiments.

In contrast, TDG variants lacking their CTD were capable of fully aggregating methylated arrays (Figure 4.7c). Together, these data suggest that DNA methylation weakens inter-array interactions mediated by TDG's NTD, which is magnified by the destabilizing effects of the CTD. One possible explanation is the increased rigidity imparted on the DNA duplex by 5mC, which has been shown to alter nucleosome stability and dynamics (i.e. DNA accessibility) [76]. These changes could, for example, promote intra-array interactions of TDG's NTD at the expense of inter-array binding and condensation. Additionally, the reduced flexibility of methylated DNA could hinder DNA bending by TDG [143], which may play an important role during condensation. Future investigations are required to determine the exact mechanism. It will also be important to examine how other cytosine modifications impact the ability of TDG to condense chromatin, particularly 5fC, which has been shown to greatly enhance DNA flexibility [76]. Most excitingly, these data support a potential regulatory mechanism wherein 5mC prevents the formation of TDG-dependent chromatin structures at methylated (or inactive) genomic regions *in vivo*.

4.1.5 TDG-chromatin condensates are liquid-liquid droplets

4.1.5.1 Regulation of chromatin by liquid-liquid phase separation (LLPS)

Recent biochemical studies have revealed that some salt- (i.e. Mg^{2+}) and protein-induced chromatin condensates are actually phase separated liquid-liquid droplets [73], [113], [243], [244]. In biology, liquid-liquid phase separation (LLPS) underlies the

formation of membraneless compartments/organelles such as the nucleolus, P-bodies, and stress granules [245], [246]. This process achieves high local concentrations of proteins and nucleic acids relative to the bulk solution and creates compartments with specific functions; in the context of chromatin these compartments vary from repressed heterochromatin to transcriptionally hyper-active chromatin domains [53], [54], [230]. Proteins that induce chromatin LLPS have intrinsically disordered regions (IDRs) capable of forming weak, multivalent, interactions with other proteins and chromatin molecules that, collectively, facilitate the exclusion of water and the formation of macro-molecular assemblies (e.g. multi-phase systems). Based on the evidence presented above, which directly links the extensive IDRs of TDG to chromatin condensation, we expect TDG-chromatin condensates are actually LLPS droplets.

LLPS has been tossed around as a mechanism for regulating macroscopic chromatin organization for a while and is supported by computational models [63] and the high level of disorder present in native chromatin [51], [52]. Only recently, however, has biochemical evidence for chromatin LLPS been reported in the literature. Still, LLPS is a relatively new concept in the chromatin field and a standard set of *in vitro* and *in vivo* experiments to validate the liquid-liquid character of biomolecular condensates has yet to be established. The study of LLPS systems requires careful consideration and should be evaluated based on the criteria depicted in Table 4.1 (below). One challenge with confirming the presence of LLPS droplets is their shared properties with other macroscopic assemblies like gels, fibrils, and solid/glass-like aggregates, which themselves can result from LLPS droplet aging or maturation. The primary method for

detecting and studying LLPS systems is direct visualization by microscopy. Preliminary evidence to support a LLPS-driven model for chromatin regulation by TDG is provided in the following section.

Table 4.1: List of four basic criteria for classifying chromatin condensates as LLPS droplets.

Physiological conditions
1. Round (spherical) droplets
2. Concentration-dependent assembly
3. Sensitivity to salts and aliphatic alcohols
4. Dynamic and capable of mixing/fusing

4.1.5.2 Characterization of TDG-chromatin LLPS droplets

The TDG precipitation experiments described above were pre-formed in the presence of 25 mM NaCl, conditions that are not suited for chromatin LLPS (Figure 4.8a, first panel) [113]. Interestingly, when we added K^+ and Mg^{2+} to intracellular concentrations (100 mM and 2 mM, respectively) and looked at the samples with a 40× objective we observed robust droplet formation by all TDG variants having at least one intact IDR, even with $TDG_{308\text{-}411}$, which failed to aggregate chromatin in precipitation experiments, but moderately promoted chromatin oligomerization in inter-fiber FRET experiments (Figure 4.8a and C.10b). TDG-chromatin (labelled with Cy3 or Cy5) condensates were observed in solution and on coverslips using an FL-Auto2 wide field imaging system or a confocal microscopy* (Figure 4.8). These preliminary findings demonstrate that TDG-chromatin droplets (1) form under physiological conditions and (2)

are round/spherical in nature. Future quantitation of droplet sphericity will confirm the roundness of these structures, which appear spherical to the naked eye.

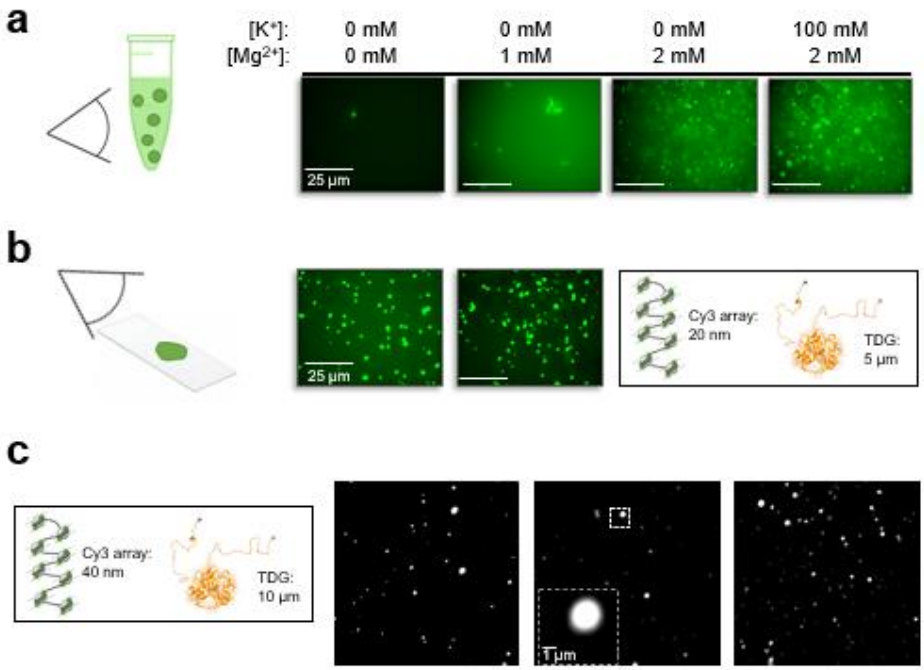


Figure 4.8: Visualization of TDG-chromatin LLPS droplets *in vitro*. (a) Wide-field fluorescence microscopy images of TDG-chromatin complexes in solution and (b) after being dripped onto a coverslip at the indicated concentrations. (c) High resolution confocal microscopy images of TDG-chromatin complexes.

The second criterion for LLPS is concentration-dependent assembly. To test if TDG promotes LLPS in such a manner we exposed 20 nM nucleosome arrays to increasing TDG. This revealed initial phase separation (i.e. condensation) occurs at concentrations ranging from 0.5 to 5 µM, depending on the variant used, and that droplet abundance, and size to some extent, appeared to increase in direct correlation to protein concentration (Figure C.10). As expected, trends in LLPS potency closely resembled solubility trends from precipitation experiments, where $TDG_{82\text{-}308}$ and $TDG_{1\text{-}308}$ were the most effective inducers of phase separation (Figure C.10). A common method for

demonstrating concentration-dependent assembly in the literature is through the construction of phase diagrams. Protein-nucleic acid phase diagrams are generated for two-component systems by systematically varying the concentration of one component while holding the other constant. Here, we generated Protein-DNA phase diagrams by visualizing mixtures of TDG, TDG_{82-308}, and H1.1 with Cy5-601 DNA (Figure 4.9).

Phase diagrams reveal a striking similarity between the phase behavior of TDG_{82-308} and H1.1, pointing to a similar mechanism of DNA condensation for the cationic residues within TDG_{82-111} and the H1.1 CTD. Interestingly, full-length TDG, which differs from the other proteins in that it has a large, hydrophobic, and uncharged IDR ($TDG_{309-411}$), was less effective at promoting LLPS. Nonetheless, all proteins tested impacted LLPS at low micromolar concentrations. Before assigning biological significance to *in vitro* LLPS it is important to consider if the phase separation phenomena occurs under conditions (i.e. concentrations) of nucleic acid and protein that can reasonably be achieved in the cell. Ultimately, it is difficult to assess this component because there are a multitude of variables that regulate protein recruitment and distribution throughout the nucleus, however, the concentration regimes in which we detect phase separation are similar to, or substantially lower than, conditions that have been reported for other transcriptional regulators that promote LLPS *in vitro*. For instance, HP1 phase separates chromatin at >1 µM [71], [241], BRD4 at concentrations >10 µM [113], [201], MED1 at >10 µM [201], VRN at > 2.5 µM [247], and H1.1 and PC4 at ratios of 1:1 with respect to individual nucleosome, to name a few [68], [246]. Based on these observations, and the intimate association of TDG with genomic elements suspected to undergo LLPS *in vivo* (e.g.

enhancers) [55], [200], [201], we believe chromatin condensation is a *bona fide* function of TDG.

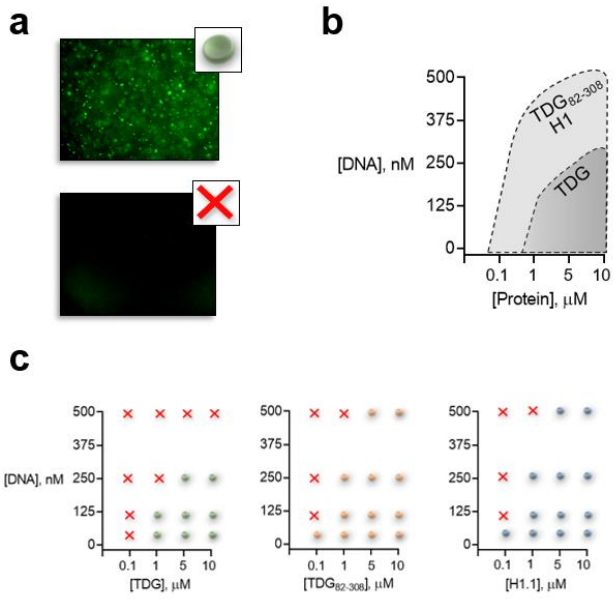


Figure 4.9: Protein-DNA phase diagrams. (a) Representative images depicting the method in which LLPS was determined. Colored spheres indicate LLPS, red X's demarcate the absence of LLPS (b) Over-layed phase diagrams for TDG, $TDG_{82\text{-}308}$, and H1.1. (c) Individual phase diagrams for the indicated proteins and a Cy5-labelled DNA (207 bp). Phase diagrams were produced as described [54], [245], [247], [249], [250].

A third criteria for determining if biomolecular condensates are in fact liquid-like is by exposing them to aliphatic alcohols, such as 1,6-hexanediol, and salts, which have been used to disrupt LLPS bodies involving protein IDRs, RNA, and chromatin *in vitro* and *in vivo*. Weak hydrophobic contacts are disrupted by 1,6-hexanediol and ionic interactions are destabilized by elevated salt concentrations, thus the sensitivity of biomolecular condensates to these factors can provide valuable information regarding the contribution of hydrophobic and electrostatic interactions to droplet assembly. To test for effects of 1,6-hexanediol and salt on TDG-chromatin condensates we generated LLPS

droplets using 20 nM Cy3-chromatin and 5 μM protein as before and exposed them to increasing concentrations of the desired antagonist (Figure 4.10).

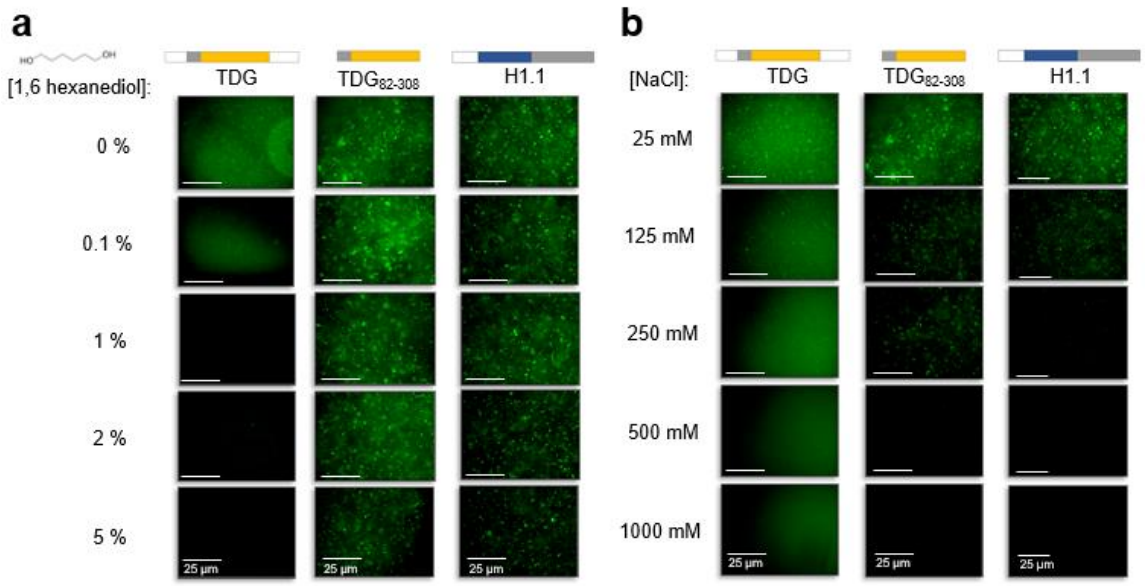


Figure 4.10: TDG-chromatin droplets are sensitive to 1,6-hexanediol and Na^+. (a) Fluorescent microscopy images of chromatin droplets, generated with the indicated protein (5 μM), exposed to increasing concentrations of 1,6-hexanediol. (b) Fluorescent microscopy images of the same droplets described in (a) exposed to the indicated concentration of NaCl.

Fluorescent microscopy images of chromatin droplets revealed that full-length TDG was extremely sensitive to 1,6-hexanediol, with droplet formation completely abrogated in the presence of just 0.1% alcohol (Figure 4.10a). In contrast, droplets formed by $TDG_{82\text{-}308}$ and H1.1 were much more resilient to disruption by 1,6-hexanediol, suggesting that residues 309-411 of TDG contribute a hydrophobic element to chromatin droplets, whereas $TDG_{82\text{-}308}$ and H1.1 are largely stabilized by ionic interactions. Consistent with this interpretation, we find chromatin droplets formed by $TDG_{82\text{-}308}$ and H1.1 were highly sensitive to Na^+ (Figure 4.10b). We note TDG catalysis is also highly sensitive to ionic conditions, whereas the ability of TDG to bind DNA is not (see Figure

C.13 for added discussion). Taken together, the preliminary data presented thus far in the form of (1) visible (sphere-like) droplets, (2) concentration-dependent threshold for assemble, and (3) sensitivity to disruption by hydrophobic and ionic small molecules, strongly point to a LLPS-model for chromatin regulation by TDG; however, additional quantitative investigations will ultimately be required both *in vitro* and *in vitro*.

The final criterion for classifying a LLPS system involves thorough interrogation of droplet dynamics. If the condensate is truly liquid-like then its constituent protein and nucleic acids should be capable of free diffusion and, therefore, should exhibit internal mixing profiles and viscosity properties comparable to other liquid-like systems. Additionally, liquid-droplets consisting of the same components should be capable of interacting and undergoing fusion, or some other form of binding, which leads to droplet growth and/or progression into fibril-, gel-, or solid-like states. In the future, fluorescence recovery after photobleaching (FRAP) and fluorescence lifetime measurements will be used to probe the phase properties of TDG-chromatin droplets. Interestingly, in conducting preliminary FRAP experiments we observed that droplets induced by the hydrophobic IDR of TDG ($TDG_{111-410}$), although quite small, remained in a liquid-like state for extended periods of time, whereas the droplets mediated by the cationic IDR of TDG (TDG_{82-308}) were only temporarily soluble, and amalgamated into gel-like aggregates that settled to the bottom of the sample vessel (Figure 4.11b). Consistent with these TDG-mediated condensates being gel- and not solid-like, we did observe minimal FRAP in pilot experiments (Figure 4.11).

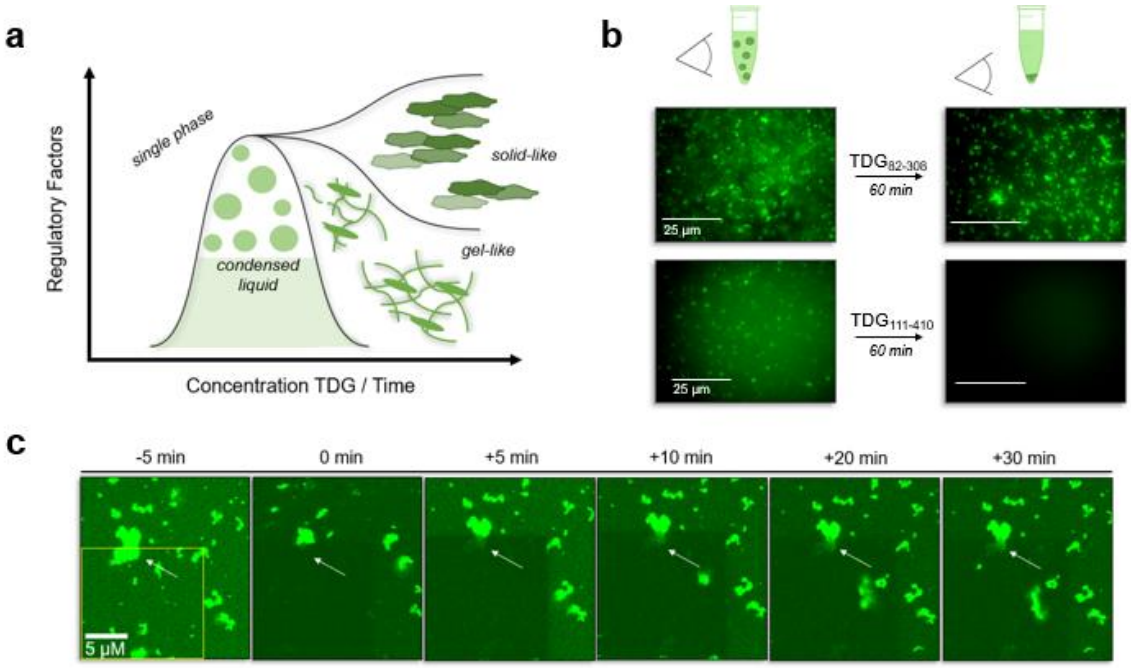
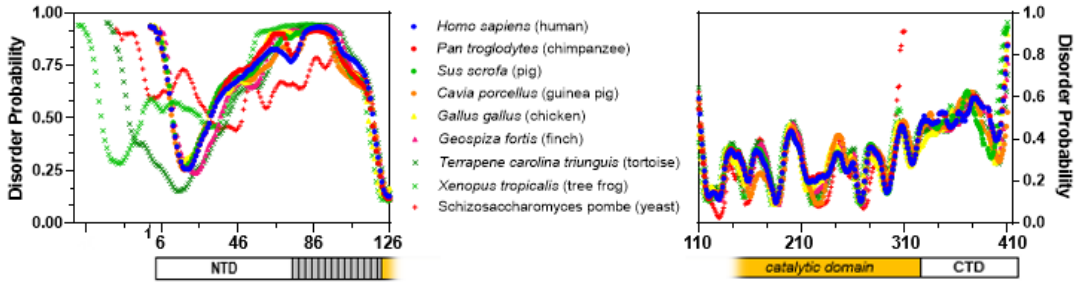


Figure 4.11: : Chromatin LLPS droplets mediated by TDG mature into gel-like aggregates in an N-terminal domain-dependent manner. (a) Schematic depicting the expected types of TDG-chromatin phases. (b) Droplet aging assay to detect aggregation in which the bottom of the sample vessel was imaged via fluorescence microscopy after a one hour incubation period. (c) Preliminary FRAP experiment conducted on a Cy3-labelled chromatin aggregate generated by treatment of 20 nM arrays with 5 µM full-length TDG for > 2 hours.

4.1.5.3 Evolutionary basis for IDRs in TDG and BER enzymes

More experiments are required to fully understand the liquid-like properties of TDG-chromatin condensates, both *in vitro* and in cells. Nevertheless, the importance of these disordered domains is further reflected in TDG phylogenic analysis, which indicates that although sequence conservation is low within the NTD compared to the rest of the protein, the overall IDR content of the NTD remains fairly constant across species (4.13). We also note the apparent trend in shortening and refinement of the NTD disorder profile as you progress from amphibian to mammal. From an evolutionary perspective, we speculate the change in size and IDR profile reflects an effort to 'fine-tune' the entropy of the NTD. Low complexity IDRs are associated with high conformational entropies, contain little information, and are theoretically disfavored by evolution- which molds proteins based on structure:function-driven relationships- however, their ability to self-interact through weak, multi-valent, interactions and form phase separated compartments reduces the entropic cost of other information related processes. Because IDR-dense phases have a large number of isoenergetic microstates (i.e. high entropy) they can form spontaneously and bring other proteins and nucleic acids along with them, creating apparent order and organization. In the case of TDG, perhaps IDR's contribute to this form of organization at genomic sites that require rapid and dynamic changes in DNA epigenetics, something that may have played an important role in the evolution of hormone-responsive gene control. While TDG is the first DNA repair protein we are aware of that has been reported to induce chromatin LLPS, exploratory studies also led to the

discovery that two other BER enzymes with IDRs, APE1 and PARP1, are also capable of inducing chromatin LLPS (Figure C.14).



Organism	Mean Disorder Probability	Amino Acid Abundance (NTD)				Percent Similarity	
	NTD	K/R	P	A/G	Q/N/S	NTD	Overall
6996-*Homo sapiens*	67.6%	16.8%	10.4%	13.6%	17.6%	-	-
452188-*Pan troglodytes*	67.7%	17.6%	11.2%	13.6%	17.6%	99.2%	99.3%
100155184-*Sus scrofa*	70.5%	17.5%	10.8%	14.2%	16.7%	85.8%	91.9%
100724604-*Cavia porcellus*	65.1%	18.1%	11.8%	22.8%	11.8%	78.6%	82.0%
395510-*Gallus gallus*	68.8%	20.0%	8.8%	15.2%	9.6%	55.7%	75.7%
102035597-*Geospiza fortis*	66.8%	20.3%	9.4%	14.8%	12.5%	55.9%	75.2%
112106684-*Terrapene Carolina triunguis*	61.5%	15.7%	10.7%	14.3%	15.0%	65.1%	80.6%
448432-*Xenopus tropicalis*	62.1%	11.2%	9.4%	14.1%	18.2%	43.8%	61.3%
2539432-*Schizosaccaromyces pombe*	60.7%	18.3%	2.0%	9.2%	20.9%	N/A	36.6%

Figure 4.12: Computed disorder probability and residue content for select vertebrate and yeast TDG homologs. Proteins were aligned based on their catalytic domain structures and the x-axis is labelled based of *h*TDG. Percent abundance for select residues was determined only for the NTD, and percent similarity was calculated via ncbi protein blast alignment tool. The Gene IDs for each protein is provided.

4.2 Conclusions

In summary, this work provides the first evidence that TDG directly alters chromatin structure through its physical interactions with DNA, and further expands TDG's functional repertoire to include chromatin condensation. The proposed model based off our findings is depicted in Figure 4.13 and our results are summarized in Table 4.2. While the biological significance of this remarkable property remains to be

determined, TDG's involvement in a number of gene regulatory pathways, such as DNA demethylation and transcription, suggests that chromatin remodeling by TDG will have important biological consequences. For example, TDG's intrinsic ability to bind and locally open compact chromatin fibers may play a role in its ability to recruit and/or promote the activity of downstream factors during transcriptional activation, such has been observed for pioneering TFs [251]. Importantly, these "pioneering" activities could be targeted to sites enriched with 5fC/5caC, with TDG's slow off-rate following excision allowing for stable recruitment of activating transcription factors and further chromatin opening [152], [179].

One particularly exciting possibility offered by our results is that TDG directly participates in the formation of long-range chromatin fiber interactions, for example, between gene enhancers and promoters during transcriptional activation (i.e. chromatin looping). In support of this hypothesis, genome-wide studies have shown that, in response to 17β-estradiol (E2), TDG localizes to sites that are involved in the interactions between promoters and enhancers of E2-responsive genes [64]. Importantly, the three-dimensional (3D) reorganization of E2-responsive genes upon E2 stimulation is abrogated upon TDG depletion, indicating that TDG plays a central role in 3D chromosomal rearrangements during E2-mediated transcriptional activation *in vivo*. Furthermore, our results support a phase-separation model of chromatin regulation by TDG, which is consistent with the interaction of TDG with E2-responsive genes known to undergo LLPS [201], [252]. To the best of our knowledge this is the first example of a DNA glycosylase capable of

regulating higher order chromatin structure, a function that has thus far been reserved for architectural proteins and transcription factors.

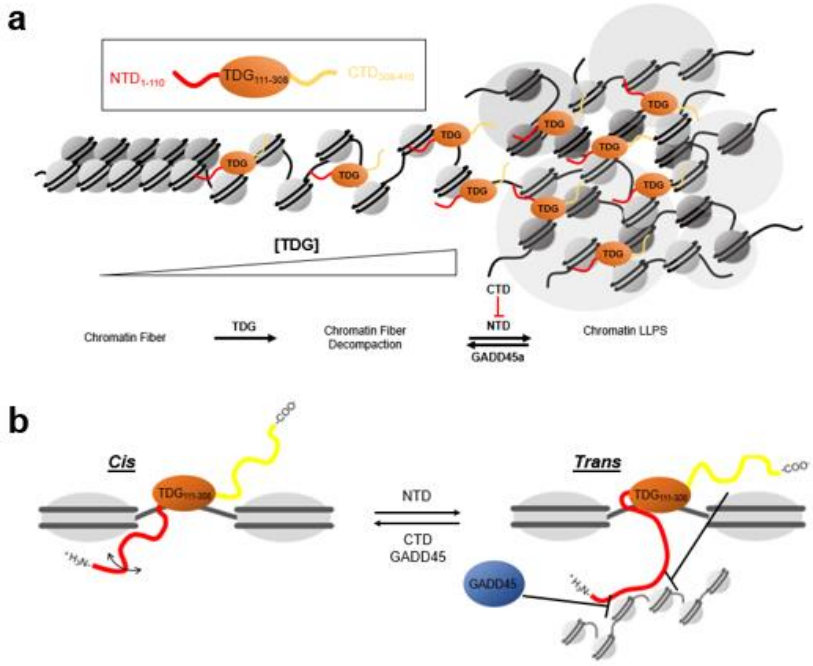


Figure 4.13: Proposed model for the regulation of chromatin structure by TDG. (a) Macroscopic mechanism. Upon recruitment, TDG binds to linker DNA between nucleosomes and de-compacts the local fiber structure. As TDG molecules accumulate within an individual fiber the overall architecture becomes conducive for oligomerization and then, when a critical density is reached, TDG-bound chromatin condenses into a liquid-like state. This process is driven by the NTD and antagonized by the CTD, GADD45, and potentially other regulators that interact with the NTD. (b) TDG-level mechanism. The ability of TDG's catalytic domain to bind DNA is enhanced by the positively charged NTD [159], [234]; interactions of the NTD with the same DNA as the catalytic domain is referred to as *cis* binding. Our findings suggest a new function for the NTD as an effector of chromatin condensation, which we propose is driven by *trans* DNA-binding events.

Table 4.2: **S**ummary of condensation and reversibility results. Chromatin condensation potency was classified based off protein A_{50} (i.e. the concentration regimes that yield >50% precipitation): an $A_{50} \leq 250$ nM is depicted as +++, an A_{50} between 250- 500 nM is depicted as ++, an $A_{50} > 500$ nM is depicted as +, and a lack of detectable precipitation is represented by a (-). Sensitivity to cytosine methylation (5mC) and reversibility is indicated as yes (Y) or no (N). ND = not determined.

Protein	Oligomerization		5mC-sensitive	Reversibility	
	Precipitation	LLPS		DNA	GADD45a
TDG	++	++	Y	Y	Y
H1.1	+++	+++	N	Y	N
TDG_{1-308}	+++	+++	N	N	N
TDG_{82-308}	+++	+++	N	N	N
$TDG_{111-308}$	+	ND	Y	Y	Y
$TDG_{308-410}$	-	+	ND	ND	ND
LANA-TDG_{1-110}	++	++	Y	ND	Y
LANA-$TDG_{309-410}$	-	ND	ND	ND	ND

4.3 Materials and Methods

4.3.1 General

Restriction enzymes (EcoRV, PflMI, BstXI, DraIII-HF, HaeII, DraI), UDG, hSMUG, NEB Next dsDNA Fragmentase, M.SssI CpG Methyltransferase, and micrococcal nuclease (MNase) were purchased from New England Biolabs (Ipswich, MA). Maleimide- Cy3 and Cy5 dyes (cat. nos. 21380, 23380) used in the labeling of $H2A_{N110C}$ were acquired from LμMiprobe Life Science Solutions (Hunt Valley, MD). Sequencing grade trypsin (cat. no. 90057) was purchased from Fisher Scientific (Waltham, MA). Synthetic oligonucleotides were either purchased from Integrated DNA Technologies (IDT) or prepared by solid-phase synthesis on an Expedite 8909 DNA/RNA synthesizer using standard methods. DNA synthesis reagents and nucleoside

phosphoramidites were purchased from Glen Research (Sterling, VA). Mixed human genomic DNA (cat. no. G3041) was purchased from Promega Corp. (Madison, WI).

4.3.2 Histone preparation and octamer refolding

Recombinant human histones (H2AN1110C, H2A.1, H2B.1, H3.1, and H4.1) were expressed and purified using established protocols.1–3 Tail-less histone proteins (H3 residues 38-135, H4 residues 17-99) were purchased from the Histone Source (Ft. Collins, CO). Histone H1.1 (cat. no. ab198676) was purchased from Abcam (Cambridge, MA). Histone $H2A_{N1110C}$ was fluorescently labelled with maleimide Cy3 and Cy5 dyes using an established protocol [101], and histone octamers were refolded and purified as previously described in Chapter 2. Purified histone octamers were stored at 4 °C in Octamer Buffer (2 M NaCl, 5 mM BME, 0.2 mM PMSF, 10 mM HEPES, pH 7.8) until further use.

4.3.3 Protein expression and purification

Full-length human TDG (410 amino acids) and truncated TDG variants were expressed in Escherichia coli and purified as described in Chapter 3. Expression plasmids for all truncated TDG variants were generated by deleting the corresponding nucleotides from plasmid **pET28a-hTDG** using inverse PCR followed by re-ligation of the linearized plasmid. All truncated TDG variants were confirmed to be catalytically active using a 5fC-containing DNA duplex as described in Chapter 3 (Figure C.6) The codon optimized gene fragments for expression of LANA-TDG_{1-110}, LANA-$TDG_{309-410}$, and full-length human GADD45a (UniProt identifier: P24522) were purchased as gBlock Gene Fragments from

IDT and assembled by PCR as recommended by the manufacturer. The assembled DNA was cloned into the pET28a expression vector (Novagen) between the HindIII and NdeI restriction sites, generating plasmids **pET28-LANA.1-110^{TDG}**, **pET28-LANA.309-410^{TDG}**, and **pET28a-GADD45**, respectively. Correct assembly of all plasmids was verified by DNA sequencing (Eton Bio, San Diego, CA).

LANA-TDG fusion proteins were expressed and purified as other TDG variants, except no activity screening was carried out. For GADD45a preparation, the plasmid assembled above (**pET28a-GADD45**) was transformed into BL21 (DE3) cells and the outgrowth (0.8 mL) was used to seed 1 × 100 mL mL cultures of 2YT media suuplemented with 50 μg/mL Kanamycin. After shaking overnight at 37 °C, 25 mL of overnight culture was used to innoculate 4 × 1 L of 2YT media supplemented with 50 ug/mL Kanamycin. The cells were grown to an OD600 ~0.800 at 37 °C with vigorous shaking and expression was induced with 0.2 mM IPTG at 37 °C for 5 hours. Next, the cells were pelleted by centrifuging at 3900 RPM for 60 minutes using a swinging bucket rotor (4 °C) and the resulting cell pellet was resuspended in Buffer H10 (50 mM HEPES, pH 8.0, 300 mM NaCl, 10% glycerol, 10 mM imidazole, 5 mM BME); 40 mL of Buffer H10 was used per 1 L of cell culture. To begin lysis, the resuspended cell pellet was frozen at -80 °C (until solid) and then thawed on water at 4 °C. The cells were further lysed by sonicating on ice for 5 minutes, and lysates were cleared by centrifugation (10,000 × g, 60 minutes) and subsequently filtered with a 0.2 μM syringe tip filter. The filtered lysate was applied to a 5 mL HisTrap FF colμMn (GE Healthcare Lifesciences) equilibrated with 5 colμMn volμMes (CV) of Buffer H10. The protein-bound resin was then washed with a 10 CV

Buffer H10, then GADD45 was eluted with a linear gradient (0 →100%) of buffer H1000 (buffer H10 supplemented with 1 M imidazole) over 10 CV. Fractions containing pure protein were combined and exchanged into Buffer HP50 (50 mM HEPES, pH 8, 50 mM NaCl 10% glycerol, 10 mM BME, 1 mM PMSF) using a 5mL HiTrap Desalting colμMn (GE Healthcare Lifesciences). Protein samples were stored at -80 °C until use.

4.3.4 Preparation of DNA templates

The DNA templates (**12-601** and **12-601-FRET**) used to reconstitute nucleosome arrays consisted of 12 copies of the "Widom 601" positioning sequence, each of which is separated by 30 bp of linker DNA (Figure C.1). These DNAs were assembled as previously described in Chapters 2 and 3. See Figure C.1 caption for more details.

4.3.5 Reconstitution of mononucleosomes and nucleosome arrays.

Reconstitution of both mononucleosomes and nucleosome arrays was carried out as described in Chapters 2 and 3 using histone octamers described above. Immediately following the reconstitution step, samples were centrifuged at 13,000 × RPM for 20 min and the resulting pellets were discarded. Soluble chromatin substrates were stored at 4 °C in buffer NB (25 mM NaCl, 0.1 mM PMSF, 10 mM HEPES, pH 7.8) for later use. Reconstituted arrays were analyzed by 0.6% agarose gel electrophoresis (Figures C.1b, C.1e, C.4b, and C.5b) and reconstituted mononucleosomes were analyzed by 5% native PAGE (59:1 acrylamide:bisacrylamide) (Figure C.2) to check for free DNA. Nucleosome arrays reconstituted from DNA template **12-601** and **12-601-FRET** (Figure C.1) are referred to as **12-NCP** and **12-NCP-FRET**, respectively.

In the case of human genomic DNA, reconstitution by salt dialysis was conducted as described above, however, the histone octamer:DNA ratios were varied more broadly (0.5:1 – 3.0:1) to identify a suitable ratio for producing soluble nucleosome arrays. In our hands, a ratio of 0.75:1 (octamer;DNA) most efficiently reconstituted fragmented human genomic DNA into chromatin (Figure C.4).

4.3.6 Nucleosome occupancy assay.

Nucleosome saturation of arrays was confirmed by digestion of ~150 ng (~ 120 fmol) of reconstituted arrays with 7.5 units PflMI and BstXI in buffer NB supplemented with 2 mM $MgCl_2$. Free **12-601** DNAs were also digested under the same conditions, and both sets of samples (naked DNA and arrays) were analyzed side-by-side with native PAGE (5%, 59:1, acrylamide: bisacrylamide) (Figure C.1c, C.1f, and C.5c). Prior to gel loading, the final glycerol concentration was adjusted to 5% using a solution consisting of nucleosome buffer supplemented with 30% glycerol. The presence of a nucleosome band as well as the absence of significant free DNA (<1%) demonstrates full nucleosome occupancies in these reconstituted arrays.

4.3.7 Micrococcal nuclease digestion of free and bound arrays

The presence of well positioned 147 bp-nucleosomes was confirmed for **12-NCP** arrays by complete micrococcal nuclease (MNase) digestion. Arrays (150 ng) were digested with 12 units of MNase, in a 20 μL reaction, for 10 minutes at 37 ˚C in buffer NB supplemented with 0.1 mM $MgCl_2$. Reactions were stopped with the addition of SDS

loading buffer (LB) to final concentrations of 0.1% SDS and 5% glycerol. The fully digested DNA was analyzed on a 1% agarose gel and visualized by ethidium bromide staining (Figure C.4b). Nucleosome occupancy following reconstitution of human genomic DNA was confirmed in a similar manner.

For the TDG protection assay (Figure 4.1d), 12-mer arrays (5 nM) were pre-incubated with 0.1, 0.5, or 1 μM TDG in a 25 μL reaction mixture containing buffer NB supplemented with 0.2 mM $MgCl_2$ for 15 minutes at 37 ˚C. At that point, 2.7 μL of MNase reaction buffer (1.5 U/μL MNase and 0.2 mM $MgCl_2$) was added and the reaction was allowed to proceed at 37 ˚C. Aliquots were taken at the indicated times and quenched with SDS LB as before. Digestion reactions were analyzed by 0.7% agarose (1 × Sodium Borate (SB) buffer, 195 V, 25 minutes) and visualized by ethidium bromide staining (Figure 4.d). Analysis of chromatin fiber compaction via intra-fiber FRET. Intra-fiber compaction experiments were carried out as previously reported5 using an identically labeled nucleosome array (**12-NCP-FRET**) (Figure C.1).5 Briefly, 10 nM **12-NCP-FRET** arrays were equilibrated in buffer NB supplemented with 2 mM $MgCl_2$ at 37 °C for 5 minutes. At that point, either 200 nM TDG or 500 nM FOXA1 was added, and the reaction mixture was incubated for an additional 20 minutes at 37 °C prior to being transferred to a Nunc 384-Well Optical (glass) Bottom Plate (Thermofisher). The plate was imaged and processed as described in Chapter 3. All FRET intensities reported herein were corrected for spectral overlap as previously described [124], [225].

4.3.8 Electrophoretic mobility shift assays (EMSA).

The binding affinity of TDG to naked DNA and mononucleosomes was determined by EMSA (Figure C.2). 5'-[^{32}P]-labelled 601 DNA or mononucleosomes (5 nM; Figure C.2) were incubated with the indicated concentrations of TDG (0–1 μM) in buffer NB supplemented with 0.2 mM $MgCl_2$ and 5% glycerol. The binding reactions were carried out at 37 °C for 20 minutes and were resolved by 5% native PAGE (59:1 acrylamide:bisacrylamide), which were run at 160 V for 30 minutes at 4 °C. The gel was visualized using a GE Typhoon gel imager and quantified using ImageQuant TL software imager (GE Healthcare Lifesciences) (Figure C.2).

4.3.9 Analysis of chromatin oligomerization via precipitation

Chromatin oligomerization was determined by precipitation as previously described [42], [78], [90], [253]. Briefly, nucleosome arrays (5 nM) were incubated in the presence of the indicated protein in a reaction mixture consisting of buffer NB for 10 minutes at 37 ˚C. Unless indicated otherwise (Figure 4.2), Mg^{2+} ($MgCl_2$) was not included in the reaction mixture. Following the incubation, samples were centrifuged (13,000 RPM) at 4 ˚C for 15 minutes and an aliquot of the supernatant was combined with SDS LB and analyzed by 0.7% agarose gel electrophoresis (1×SB Buffer, 195 V, 20 minutes).

4.3.10 Generation of histone tail deleted nucleosome arrays

Nucleosome arrays lacking individual tail domains (Figure 4.3) were reconstituted using **12-601** DNA and recombinant histone octamers, refolded from either H3 or H4 tail-

deleted proteins (globular domains only, H3 residues: 38-135 & H4 residues 17-99). Nucleosome arrays lacking all histone tail domains were generated via Trypsin digestion of intact **12-NCP** arrays. Briefly, lyophilized Trypsin was dissolved in 50 mM acetic acid (100 ng/µL) and diluted 1:1 with 250 mM TRSI (pH 7.9). Nucleosome arrays (25 nM) were then digested with 0.5 ng/µL Trypsin at room temperature for 30 minutes. Reactions were quenched by addition of TQ buffer (final concentration: 50 ng/µL aprotinin, 0.25 mg/mL BSA, 10 mM HEPES pH 7.8, 25 mM NaCl, 0.1 mM PMSF) and samples were stored on ice until time of use. Successful tail removal was confirmed by native and denaturing gel electrophoresis (Figure 4.3a).

4.3.11 Fragmentation of human genomic DNA

Human genomic DNA was thawed on ice, and then 20 µg was diluted into 180 µL of 1.1× NEB NEXT® dsDNA Fragmentase® buffer. Next, the Fragmentase stock solution was vortexed for 3 seconds and then 20 µL of enzyme was added directly to the DNA mixture, bringing the total volume to 200 µL. The final reaction mixture was vortexed for an additional 3 seconds, then incubated at 37 ˚C for 8 minutes, with gentle vortexing every 2 minutes. The reaction was quenched by addition of SDS to a final concentration of 0.1%, and successful generation of 0.5-3kbp DNA fragments was confirmed by agarose gel (0.7%) electrophoresis (Figure C.4a). The DNA was then desalted with an EconoSpin mini spin column (cat. no. 1920-050/250, Epoch Life Sciences) and eluted in milli-Q H2O.

4.3.12 Analysis of chromatin oligomerization via inter-fiber FRET

In order to monitor FRET between arrays, two separate arrays were reconstituted using histone octamers labeled with either Cy3 (donor arrays) or Cy5 (acceptor arrays) dyes via histone $H2A_{N1110C}$ (Figure C.5). To allow for accurate determination of the FRET intensity, three separate reaction mixtures were prepared for each concentration of $MgCl_2$ or TDG tested:

DA: 1:1 mixture of Donor to Acceptor arrays (FRET sample)

DO: 1:1 mixture of Donor to unlabeled arrays (Donor only)

AO: 1:1 mixture of Acceptor to unlabeled arrays (Acceptor only)

All reactions contained a final chromatin concentration of 5 nM in buffer NB. Following the addition of TDG (or $MgCl_2$), each sample was incubated at 37 °C for 15 minutes and the reaction mixture was transferred to a Nunc 384-Well Optical (glass) Bottom Plate (Thermofisher) for imaging. Prior to imaging, optical plates were siliconized with Sigmacote (Sigma Aldrich) according to the manufacturer's instructions. All reaction mixtures were imaged using a Typhoon FLA 9500 multimode imager (GE Healthcare Lifesciences) at 100 μm resolution using the three laser settings and emission filters in the order indicated below:

F: FRET channel (532 nm ex.; 665 nm em.; PMT voltage: 675)

A: Acceptor channel (635 nm ex.; 665 nm em.; PMT voltage: 675)

D: Donor channel (532 nm ex.; 575 nm em.; PMT voltage: 625)

As before, we determined the corrected FRET (Fcorr.) intensities for TDG- and Mg^{2+}-treated chromatin samples using Eq. 1; where F, A, and D represent the fluorescent

intensities from the FRET, Acceptor, and Donor channels, respectively, from the DA chromatin sample containing a 1:1 mixture of donor to acceptor arrays.

$$F_{corr} = F - (D \bullet X_d) - (A \bullet X_a)$$

In order to account for background signal, which result from donor bleed-through into the acceptor emission (X_a), as well as background from direct acceptor excitation at the donor's excitation wavelength (X_d), we determined X_a and X_d values (Eqs. 2 & 3) for each Magnesium or TDG concentration using the respective AO and DO samples described above. Xa is obtained from the ratio of F and A intensities from the AO sample and X_d is obtained from the ratio of F and D intensities with the DO sample, all of which contain identical Magnesium or TDG concentrations as their corresponding DA sample.

$$X_a = F/A \qquad (3)\ X_d = F/D$$

4.3.13 Chromatin aggregation reversibility assay

Pre-formed chromatin oligomers were generated by incubating 1 μM TDG (or the indicated TDG variant) with 5 nM 12-mer arrays in a reaction mixture (8 uL) containing buffer NB at 37 ˚C for 15 minutes. At this point, 4 uL of a solution containing either 601 DNA or GADD45a, both at 3-times the final desired concentration (Figure 4.6, C.7, and C.8), was added to the mixture. After incubating for 10 minutes at 37 ˚C, samples were centrifuged at (13,000 RPM) at 4 ˚C for 15 minutes and an aliquot of the supernatant was combined with SDS LB and analyzed by agarose gel (0.7%) electrophoresis (Figure C.7, and C.8).

4.3.14 M.SssI methylation of nucleosome arrays

For methylation reactions, 12-601 DNA (60 nM) was incubated with 0.35 U/μL M.SssI in CutSmart buffer supplemented with 0.4 mM SAM at 37 ˚C for 4 hours. To confirm CpG sites were fully methylated, a 75 fmol aliquot was digested with 10 units HpaII in a 10 μL reaction containing 1 × CutSmart buffer at 37 ˚C for 45 minutes. Following the digest, glycerol was added (5%, v:v) and the reactions were analyzed via agarose gel (0.7%) electrophoresis. HpaII-resistant 12-601 DNA was used in subsequent nucleosome reconstitutions and confirmed to form chromatin via native agarose gel electrophoresis.

4.3.15 Statistical Analysis

All FRET data were presented as means and standard deviations. Statistical analysis of intra- and inter-fiber FRET studies was conducted using GraphPad Prism (v8.4.2). For comparison of the corrected FRET intensities (Figure 1b and 2X) between samples containing either free 12-NCP arrays or 12-NCP arrays that had been oligomerized by TDG or Mg2+, all data sets were first compared by unpaired one-way analysis of variance (ANOVA), then significant differences were determined between each condition tested using a Tukey's multiple comparisons test ($\alpha = 0.05$). An identical analysis (ANOVA and Tukey's multiple comparisons) was used to compare TDG-induced precipitation of wild-type and hyper-methylated arrays (Figure 5).

4.3.16 TDG activity assay

TDG activity following a cycle of oligomerization was determined by adding 325 fmol of a 5fC-containing 601 DNA (**1-601-5fC49**, described in Chapter 2) to pre-formed TDG-chromatin aggregates before or after they were incubated at 37 ˚C for 15 or 30 minutes. Abasic sites generated by TDG were cleaved by including APE1 (35 nM) and 0.1 mM $MgCl_2$ in the final reaction mixture. The reactions were incubated at 37 ˚C for 45 minutes, then quenched with Native loading buffer (5% glycerol, 0.1% SDS, 10 mM HEPES, pH 7.8), and nicked/intact **1-601-5fC49** were separated with 10% (29:1) Native PAGE (Figure C7c).

4.3.17 Analysis of sequence and disorder among TDG homologs

The disorder probability for TDG homologs were computed using Protein Disorder prediction System (PrDOS) and aligned using the calculated disorder probabilities from the catalytic domain. Mean disorder probability was calculated from the averaged predicted values for the TDG N-terminal domain, which considered every amino acid upstream of the PrDOS minima bordering the catalytic domain from each TDG species; this site corresponds to residue D_{126} in *h*TDG. Sequence similarities were determined for the same N-terminal residues, and the full-length proteins, with respect to *h*TDG via ncbi blast protein alignment. Gene IDs are as follows: *S. pombe* (yeast): 2539432, *T. carolina triunguis* (tortoise): 112106684, *X. tropicalis* (tree frog): 448432 *G. fortis* (finch): 102035597, *G. gallus* (chicken): 395510, *C. porcellus* (guinea pig):

100724604, *S. scrofa* (pig):100155184, *P. troglodytes* (chimpanzee):452188, *H. sapiens* (human) 6996.

4.3.18 Fluorescence microscopy of chromatin LLPS droplets

If not otherwise stated, LLPS was induced by mixing protein (10 μM) and nucleic acid (40 nM) samples at 1:1 ratios. Proteins were prepared in buffer HP50 (10% glycerol, 10 mM HEPES, pH 7.8, 50 mM NaCl) and combined with nucleic acid sample in 2× LLPS buffer (1×: 100 mM KCl, 2 mM $MgCl_2$, 10 mM HEPES, pH 7.8, 25 mM NaCl). The final concentration of glycerol was 5%. Mixtures were incubated at 37˚C for 10-15 minutes, then transferred to either a 384-well (Nunc) glass-bottom plate or a glass coverslip. A 10 μL volume is required for glass-plate imaging and only 1-2 μL was used to wet the coverslip when imaging on slides. Fluorescence microscopy images, unless otherwise stated, were obtained using an Invitrogen FL-Auto2 wide field imaging system (40×) objective. Images and videos were processed in ImageJ.

CHAPTER 5

SUMMARY OF THE POLYCOMB REPRESSIVE COMPLEX 2 BINDS RNA IRRESPECTIVE OF STEREOCHEMISTRY

5.1 Background

5.1.1 Long non-coding RNAs

In the central dogma of molecular biology RNA is typically depicted as simple mediator between genotype and phenotype but, in reality, RNA participates in almost every facet of gene regulation. To put in perspective the importance of RNA's non-messenger roles: it is estimated that ~90% percent of the genome is transcribed whereas only 1.5-2% these total genes encode for proteins. These non-protein coding genes account for rRNA, tRNA, miRNA, as well as lncRNAs, which have recently received increased attention due to their involvement in a wide range of genomic processes [254]–[256]. Indeed, transcriptomic data suggests that nature may have relied heavily on the evolution of lncRNA-mediated pathways to achieve biological complexity which, as mentioned in Chapter 1, is not correlated to the number of protein coding genes [14], [255], [257], [258]. In these pathways,lncRNAs are expected to function as signals, guides, decoys, scaffolds, or structural elements [256].

*Parts of this chapter have been reprinted with permission from Deckard III, C. E.; Sczepanski, J.T., Polycomb Repressive Complex 2 Binds RNA Irrespective of Stereochemistry, *Chemical Communications*, **2018**, *54*, 12061-12064.

As *structural* components, lncRNA molecules can function as transcriptional repressors by directly binding to chromatin, facilitating its condensation into a densely compacted state. In humans this has important functions in silencing autosomal genes that should only be expressed from one of the diploid chromosomes. The lncRNA Xist, for example, is transcribed on the inactive X chromosome (X_i) in females and coats the entire chromosome to facilitate its condensation into an aggregated Barr body [259]; other transcripts like Kcnqtot1 and Air also silence target genes by binding to chromatin *in cis* [255], [255], [260]. In addition to binding-induced silencing, lncRNAs to serve as molecular *scaffolds* and *guides* to recruit and maintain transcriptional repressors at chromatin near target genes, such as PRC2 (see *5.1.2*), [255], [256], [261]. In the case of PRC2 and related chromatin modifiers the outcome is gene silencing but, as discussed in section *1.2.3.2* and Chapter 4, eRNAs (a type of lncRNA) can provide scaffolding functions for enzymes involved in DNA demethylation and transcription [64], [68], [181], and their interaction with chromatin can lead to condensed states with high levels of expression [55], [64], [200]. Finally, lncRNAs can function as decoys by essentially distracting proteins from binding chromatin or other target molecules [256].

5.1.2 Polycomb Repressive Complex 2

The Polycomb Repressive Complex 2 (PRC2) is perhaps the most represented lncRNA-associated protein(s) in the literature. This multimeric protein assembly contains the H3K27-specific histone methyltransferase: Enhancer of Zeste Homologue 2 (EZH2). EZH2 catalyzes the mono-/di-/tri-methylation of H3K27(me1/2/3), which leads to

transcriptional silencing [262]. Fully functional PRC2 complexes also require the Early Embryonic Development (EED) factor, the Supressor of Zeste variant 12 (SUZ12) protein, and the Retinoblastoma binding protein 4/8 (Rbbp4/8); and participation from other subunits such as AEBP2 and JARID2 can lead to more active PRC2 complexes [262]–[264]. The timely recruitment of PRC2 to specific loci by lncRNAs is critical for embryonic development, X-inactivation, and the proper maintenance of lineage specific genes [254], [255], [257], [263], [3]; while dysregulation of PRC2 is the cause of numerous cancers [263], [264]. Not surprisingly, the activity of PRC2 is tightly controlled.

5.1.3 PRC2 in cancer

PRC2 dysregulation occurs through numerous mechanisms and each is associated with a unique diseased phenotype. For example, PRC2 components are overexpressed in breast cancer, prostate cancer and melanomas; gain of function mutations in EZH2 are implicated in multiple lymphomas; and overexpression of lncRNAs associated with PRC2 recruitment are linked to metastatic tumors in breast, ovarian, and esophageal cancers (lncRNA: HOTAIR) and neuronal melanomas (lncRNA: ANRIL), to a name a few [267]–[272]. As a result, substantial efforts have been devoted to developing PRC2 inhibitors. These include active site inhibitors, allosteric inhibitors, and lncRNA anti-sense oligonucleotides (ASOs). Despite the presence of multiple strategies for inhibiting PRC2, off-target effects from small molecule HMT inhibitors and chemistry-dependent toxicities associated with modified ASOs still pose a challenge from a therapeutic perspective.

Based on recent evidence that PRC2 interacts with RNA promiscuously *in vitro* and *in vivo* [273]–[276], that these interactions have important gene regulatory functions in cancer biology [255], [257], [265], and that this interaction is antagonistic to PRC2 catalysis [276], [277], we predicted that synthetic oligonucleotides capable of mirroring the RNA motifs that PRC2 bound most tightly could be potential therapeutic targets. Although the molecular basis for the interactions between PRC2 and lncRNAs remain unclear, emerging evidence now suggests that the presence of guanine (G)-rich RNA motifs are a key determinant for binding by PRC2. For example, Kaneko *et al*. showed that poly(G), but not poly(A), was bound by PRC2 *in vitro* [278]. Moreover, Wang *et al.* recently reported that PRC2 binds G > C,U >> A in single stranded RNA and has a preference for binding folded G-quadruplex (G4) RNA structures [275], [277]. These i*n vitro* data are consistent with the preferential binding of PRC2 to RNAs containing G-tracts *in vivo* [273], [274], [279].

We believe an alternative strategy for targeting PRC2, one based on L-RNA inhibitors, could offer several advantages over other methods. L-nucleic acids (L-RNA/DNA) are the synthetic enantiomer of naturally occurring D-nucleic acids [280]. Although they share identical physical and chemical properties, oligonucleotides of one chirality are unable to interact with complimentary nucleic acids of the opposite chirality [280], [281]. Moreover, the complete exclusion of L-nucleic acids from nature have rendered them “invisible” to enantioselective nucleases, and other cellular components, that have evolved to only recognize the natural D-form [280], [281]. In light of these facts,

we were motivated to ask whether the promiscuity of PRC2 towards G-rich RNAs could be extended to mirror image L-RNA and, specifically, L-G4 RNA structures.

5.2 Results

5.2.1 PRC2 binds G4 RNA independent of stereochemistry

To initially test for potential interactions between PRC2 and L-G4 RNAs, we synthesized both D- and L-RNA versions of $(GGAA)_{10}$, a G4- forming RNA previously shown to bind PRC2 with high affinity (K_d = 7.7 ± 2.4 nM) [274], [275]. For consistency, both enantiomers of $(GGAA)_{10}$ were Cyanine 3 (Cy3)-labeled at their 5′ ends (Figure D.1). Formation of G4 structures was confirmed by circular dichroism (CD) spectroscopy for both enantiomers of $(GGAA)_{10}$ [209], which exhibited the expected mirror symmetry (Figure 5.1a) [281]. Folding of these RNAs was further verified by gel electrophoresis (Figure D.1). We then evaluated the ability of each enantiomer of $(GGAA)_{10}$ to bind PRC2 using an electrophoretic mobility shift assay (EMSA) (Figure 5.1b) [275]. Remarkably, we found that PRC2 bound with similar affinity to both D- and L- $(GGAA)_{10}$ (K_d = 39 ± 5 and 20 ± 4 nM, respectively). Moreover, the Hill coefficients were nearly identical (~4), suggesting that PRC2 factors bound both enantiomers of $(GGAA)_{10}$ using a common mode of cooperativity.

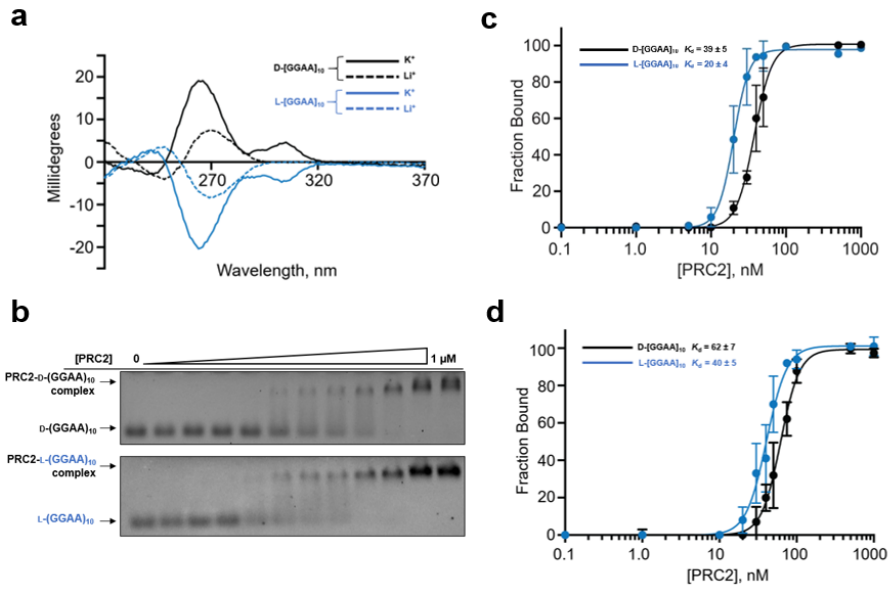


Figure 5.1: PRC2 binds G4 RNA irrespective of stereochemistry. (a) CD spectra of D- and L-$(GGAA)_{10}$ RNAs in the presence of either 100 mM KCl (K^+) or 100 mM LiCl (Li^+). (b) Representative EMSA gels (1% agarose) of $(GGAA)_{10}$ binding to PRC2 (0–1 μM). (c) Saturation plot for binding of either D-$(GGAA)_{10}$ or L-$(GGAA)_{10}$ to PRC2 in the presence of 100 mM KCl or (d) 100 mM LiCl. Error bars show SD (n = 3).

PRC2's broad, multi-subunit RNA binding site [275], [282], as well as its preference for long RNAs (>300) over short RNAs [273], suggests that PRC2 could accommodate multiple bound $(GGAA)_{10}$ RNA motifs. However, the exact binding stoichiometry of PRC2:RNA is unknown and is likely to vary depending on the RNA target. Finally, to demonstrate that binding of L-G4 RNA by PRC2 was not unique to $(GGAA)_{10}$, we prepared a second G4-forming RNA, $(G3A4)_4$, and found that it too bound PRC2 irrespective of stereochemistry (Kd = 57 ± 5 and 52 ± 4 nM for D- and L-$(G3A4)_4$, respectively) (Figure 5.2). To the best of our knowledge, this is the first reported example of a native RNA-binding protein (or protein of any type) recognizing L-RNA.

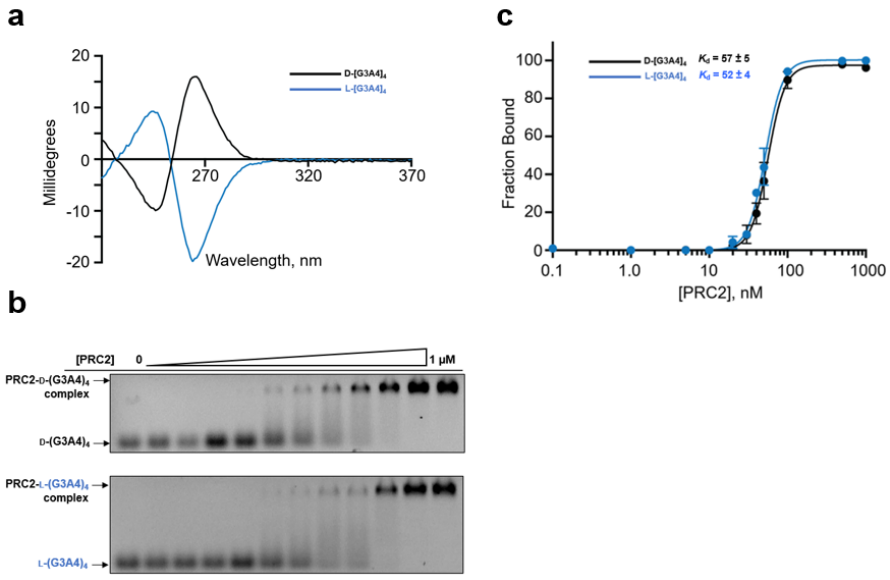


Figure 5.2: PRC2 binds similarly to both D- and L-$(G3A4)_4$ G4 RNAs. (a) CD spectra of D- and L-$(G3A4)_4$. (b) Representative EMSA gels (1% agarose, 0.2 TBE supplemented with 10 mM KOAc) of $(GGAA)_{10}$ binding to PRC2 (0–1 μM). Binding conditions were the same as described in Figure 1b (main text). (c) Saturation plot for binding of either D- or L-$(G3A4)_4$ to PRC2. Error bars show SD (n= 3).

To assess the importance of a folded G4 structure on the binding of L-RNA by PRC2, we replaced the K^+ cations in the EMSA binding buffer with Li^+, which results in destabilization of the G4 structure. We found that the affinity of PRC2 for both D- and L-$(GGAA)_{10}$ was reduced by ~2-fold under these conditions (Figure 5.1d). This modest loss of affinity in the absence of a folded G4 is in agreement with Wang *et al*. and further confirms that folded L-G4 structures are not the sole determinant for PRC2 binding to G-rich RNAs [275]. Consistently, PRC2 was able to bind tightly to both D and L versions of the non-G4- forming RNA $(GA)_{20}$ (K_d = 66 ± 7 and 59 ± 6 nM, respectively) (Figure 5.3). In fact, the affinity of PRC2 for $(GA)_{20}$ was similar to that of $(GGAA)_{10}$ in the absence of K^+ (i.e. no G4s). In contrast, PRC2 bound very weakly to both enantiomers of $(A)_{40}$

(estimated K_d > 800 nM; Figure 5.4), which is consistent with its preference for G-rich RNA motifs. Interestingly, we found that the deoxyribose version of D- $(GGAA)_{10}$, D-(dGGAA), also bound very weakly to PRC2 (estimated K_d > 1000 nM; Figure 5.4). This is despite evidence that D-$(dGGAA)_{10}$ forms G4 structures. This result was somewhat unexpected given the ability of PRC2 to bind tightly to diverse G-rich sequences, even in the absence of G4 structure formation [275]. Therefore, while the chirality of the sugar moiety is not a determinant for binding of G-rich oligonucleotides by PRC2, the identity of the sugar (ribose versus deoxyribose) is critical.

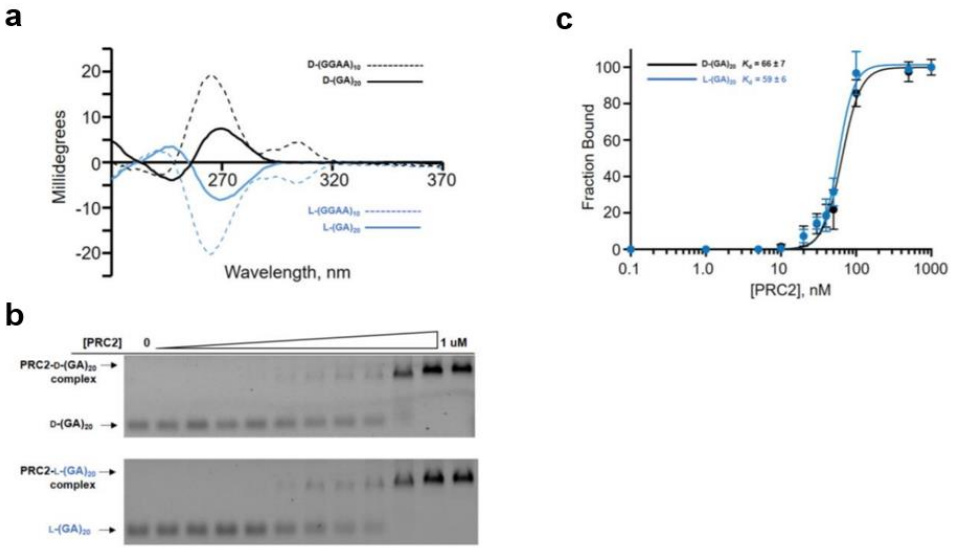


Figure 5.3: PRC2 binds similarly to both D- and L-$(GA)_{20}$ RNAs. (a) CD spectra of D- and L-$(GA)_{20}$. The spectral data for both D- and L-$(GGAA)_{10}$ RNAs is overlaid for comparison. (b) Representative EMSA gels (1% agarose, 0.2 TBE supplemented with 10 mM KOAc) of D-(GA)20 and L-$(GA)_{20}$ binding to PRC2 (0–1 μM). Binding conditions were the same as described in Figure 1b (main text). (c) Saturation plot for binding of either D- or L- $(GA)_{20}$. Error bars show SD (n= 3).

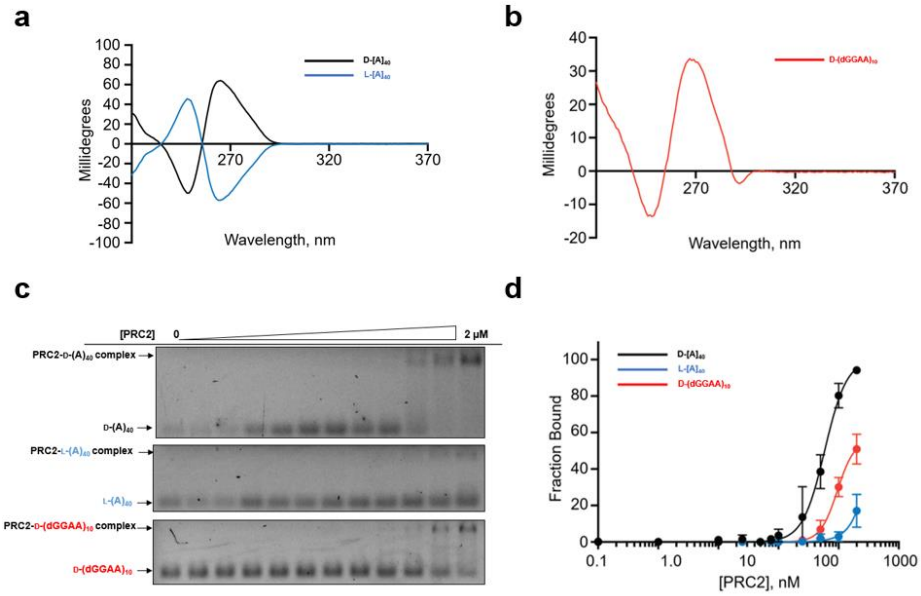


Figure 5.4: PRC2 binds weakly to $(A)_{40}$ and D-$(dGGAA)_{10}$. (a,b) CD spectra of both enantiomers of $(A)_{40}$ and the D-$(dGGAA)_{10}$, respectively. c) Representative EMSA gels (1% agarose, 0.2 TBE supplemented with 10 mM KOAc) of D-$(A)_{40}$, L-$(A)_{40}$, and D- $(dGGAA)_{10}$ binding to PRC2 (0–2 μM). Binding conditions were the same as described in Figure 1b (main text). (d) Saturation plot for binding of D-$(A)_{40}$, L-$(A)_{40}$, and D-$(dGGAA)_{10}$ to PRC2. Error bars show SD (n= 3).

5.2.2 D- and L- RNA G4s share the same PRC2 binding

Because D- and L-$(GGAA)_{10}$ are simply mirror images of each other, we reasoned that they bind to the same site on PRC2. This is important because it implies that L-RNA could potentially inhibit PRC2 from binding endogenous D-RNA targets. To test this hypothesis, we carried out competition assays in which increasing concentrations of unlabeled L-$(GGAA)_{10}$ was added to pre-formed complexes of PRC2 and Cy3-labeled D-$(GGAA)_{10}$, and vice versa (Figure 5.5). These data revealed that both D- and L-$(GGAA)_{10}$ could outcompete their respective enantiomers for binding to PRC2. However, the L-$(GGAA)_{10}$-PRC2 complex was somewhat more resistant to higher concentrations of competitor than its D-RNA counterpart, which may reflect the slightly higher affinity of L- $(GGAA)_{10}$ for PRC2 as compared to D-$(GGAA)_{10}$ (Figure 5.1). As expected, L-$(A)_{40}$

failed to compete against both enantiomers of $(GGAA)_{10}$, even when present in 1,000-fold excess (Figure 5.5). Taken together, these data strongly suggest that the same RNA-binding site on PRC2 recognizes both D- and L-$(GGAA)_{10}$ RNA and, based on their similar binding properties, we conclude that it does so independent of nucleic acid chirality.

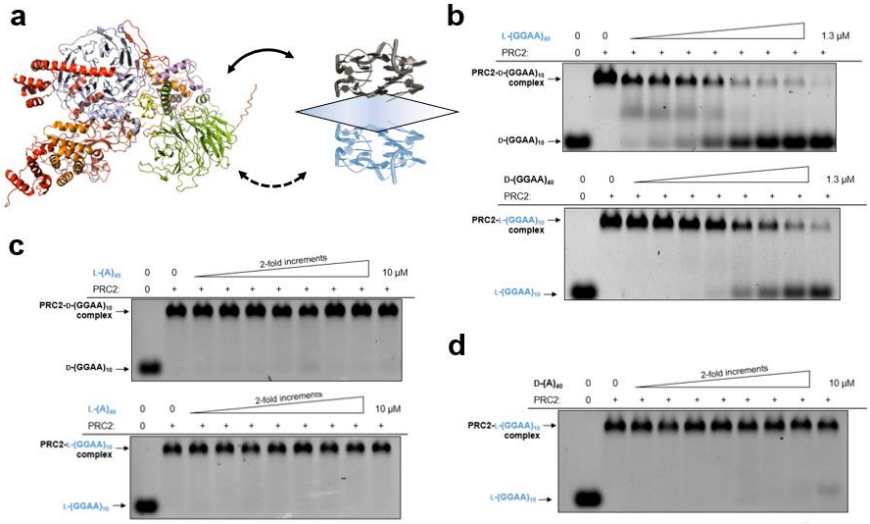


Figure 5.5: Both L- and D- versions of $(GGAA)_{10}$ bind to the same site on PRC2 (a). Pre-formed PRC2-D-$(GGAA)_{10}$ complexes are disrupted by L-$(GGAA)_{10}$ and vice versa (b). (c) Competitive binding experiments for L-$(A)_{40}$ versus pre-formed PRC2- $(GGAA)_{10}$ complexes.

5.2.3 L- RNA G4s antagonize PRC2-substrate complex formation

Motivated by the above results, we next asked whether L- $(GGAA)_{10}$ could inhibit PRC2 from binding the long noncoding (lnc)RNA HOTAIR, a bona fide in *vivo* target required for PRC2 occupancy and H3K27 trimethylation of the HOXD loci and many other genomic sites [254], [255], [257], [283]. HOTAIR is also overexpressed in numerous human cancers and has been shown to promote breast cancer invasiveness and metastasis in a manner that is dependent on PRC2 [284]–[286]. Thus, disrupting the PRC2-HOTAIR interaction represents a promising approach for developing effective cancer therapy. For these studies, we employed the first 300 nucleotides from the 5′ end of HOTAIR (HOTAIR-300), which was previously shown to bind PRC2 *in vitro* (Figures 5.6 and D.4) [255], [273], [279], [287]. As before, we titrated unlabeled L-$(GGAA)_{10}$ with pre-formed

complexes of PRC2 and Cy5-labeled HOTAIR-300 (Figure 5.6). At a stoichiometric concentration of L-$(GGAA)_{10}$ relative to PRC2 (250 nM), we observed almost complete dissociation of HOTAIR-300 from PRC2. Similar results were obtained using D-$(GGAA)_{10}$, whereas L-poly$(A)_{40}$ failed to compete (Figure 5.6). These results demonstrate that L-$(GGAA)_{10}$ is an effective inhibitor of lncRNA-PRC2 interactions and further support a common binding site for both D- and L-RNA.

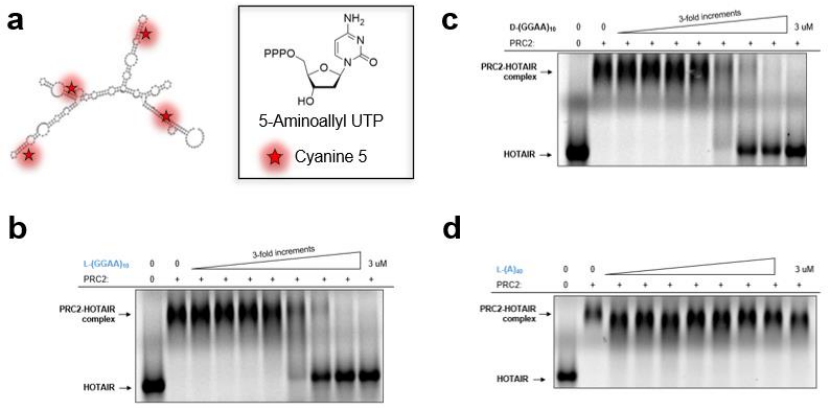


Figure 5.6: L-$(GGAA)_{10}$ outcompetes native substrates for binding to PRC2. (a) Schematic depiction of Cy5-labelled domain I (nucleotides 1-300) of the lncRNA HOTAIR. Internal Cy5 dyes were installed via NHS ester coupling to 5-aminoallyl UTP, which was stochastically introduced during transcription. (b) Pre-formed complexes between PRC2 (250 nM) and HOTAIR-300 (25 nM) are disrupted by L-$(GGAA)_{10}$ and (c) D-$(GGAA)_{10}$, but not (d) L-$(A)_{40}$.

Recent studies have shown that native D-RNA, including D-$(GGAA)_{10}$, is able to disrupt the association of PRC2 with both naked DNA and nucleosomes *in vitro*, suggesting that RNA and chromatin share the same or mutually exclusive binding sites on PRC2 [275], [287]. In line with these studies, we tested whether L-$(GGAA)_{10}$ could also prevent PRC2 from binding to chromatin. For these experiments, we employed a Cy5-labeled 12-mer oligonucleosome array reconstituted in vitro using recombinant human histones (Figures 5.7, D.3, and D.4). We found that L-$(GGAA)_{10}$ was able to disrupt pre-

formed complexes between PRC2 and the oligonucleosome array in a concentration-dependent manner, with no discernable PRC2-chromatin complexes remaining upon the addition of a stoichiometric concentration of L-$(GGAA)_{10}$ relative to PRC2 (1 μM) (Figure 5.8). Again, these results closely mirrored those obtained using D-$(GGAA)_{10}$ (Figure 5.8). Thus, we conclude that, like native D-RNA, the interaction of PRC2 with L-RNA and chromatin is mutually antagonistic.

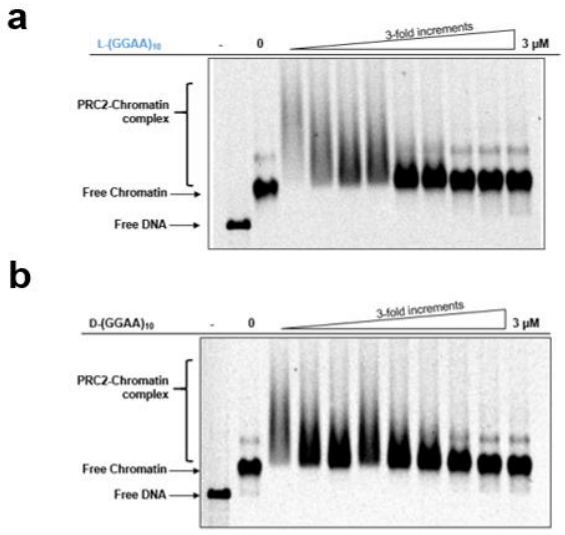


Figure 5.7: L- and D- RNA G4s antagonize chromatin binding by PRC2. (a) Pre-formed complexes between PRC2 (1000 nM) and a 12-mer oligonucleosome array (8 nM) are disrupted by L-$(GGAA)_{10}$ and (b) D-$(GGAA)_{10}$.

5.3 Conclusions, Outlook, & Broader Impact

In summary, we demonstrated that PRC2's promiscuous binding to RNA extends to mirror image L-RNA, thereby providing the first evidence that native proteins are capable of recognizing L-oligonucleotides. Remarkably, we found that PRC2 bound similarly to both enantiomers of G4-forming RNAs, suggesting a chirality-independent mode of recognition. This unexpected and wholly novel finding dramatically broadens the definition of "promiscuous" RNA binding, which now must be expanded to include nucleic acid chirality. Previous studies have shown that native D-RNA is capable of inhibiting PRC2's methyltransferase activity by preventing it from binding its nucleosome substrates [275]. Thus, our discovery that D- and L-RNA bind competitively to the same site on PRC2 opens the door for therapeutic targeting of PRC2 using nuclease-resistant L-G4 RNAs [280], [281], [288], [289]. An important next step towards achieving this goal will be to demonstrate that L-G4 RNAs inhibit PRC2 methyltransferase activities in vitro and in human cells.

Importantly, our discovery that a native RNA-binding protein recognizes L-RNA challenges the prevailing assumption that L-oligonucleotides are "invisible" to the stereospecific environment of the cell and implies that protein interactions should be taken into consideration when designing L-oligonucleotides for intracellular applications. Given the large number of proteins that have been shown to interact with nucleic acids in a nonspecific or "promiscuous" manner [290], it is reasonable to predict that the stereochemical promiscuity observed herein is not unique to PRC2. Therefore, it will be important to undertake future efforts aimed at identifying additional proteins that are

capable of interacting with L-RNA (and L-DNA), which if successful will contribute to the future development of intracellular L-oligonucleotide technologies and may ultimately lead to new therapeutic opportunities. The application of structured, guanine-rich, L-RNA oligonucleotides as inhibitors of PRC2 was also successfully patented through our efforts.

5.4 Materials and Methods

5.4.1 General

The DNA and RNA oligonucleotides were either purchased from Integrated DNA Technologies (IDT, Coralville, IA) or prepared using an Expedite 8909 DNA/RNA synthesizer. Oligonucleotide synthesis reagents, D-nucleoside phosphoramidites, and Cyanine 3 (Cy3) phosphoramidite were purchased from Glen Research (Sterling, VA), and L-nucleoside phosphoramidites were purchased from ChemGenes (Wilmington, Ma). All oligonucleotides were purified by polyacrylamide gel electrophoresis (PAGE) and desalted by ethanol precipitation; the purity and integrity of each oligonucleotide was confirmed with mass spectrometry (Novatia, Newton, PA). Polycomb Repressive Complex 2 was purchased from Active Motif (Carlsbad, CA). N-Hydroxysuccinimide (NHS) ester of Cyanine 5 (Cy5) used in the labeling of HOTAIR was acquired from LμMiprobe Life Science Solutions (Hallendale Beach, FL).

5.4.2 Electrophoretic Mobility Shift Assays (EMSA)

Prior to use, Cy3-labeled oligonucleotides (Appendix D) were diluted to 100 nM in TE Buffer (10 mM TRIS pH 7.5, 1 mM EDTA) and denatured at 95 °C for 10 minutes before being snap cooled on ice for 5 minutes. The oligonucleotides were then diluted to 50 nM in PRC2 binding buffer (50 mM Tris pH 7.4, 100 mM KCl, 2.5 mM $MgCl_2$, 0.1 mM $ZnCl_2$, 2 mM BME, 0.1 mg/mL BSA, and 5% glycerol) and allowed to fold at 37 °C for 30 minutes. In some instances, the KCl was replaced with LiCl (see Figure 5.1). The oligonucleotides were then diluted to 2 nM into individual binding reactions (10 µL) containing PRC2 binding buffer and increasing concentrations of PRC2 (0.1–1000 nM). Binding reactions were carried out at 30 °C for 30 minutes and bound and unbound fractions were subsequently separated by 1% agarose gel electrophoresis (0.2 X TBE supplemented with 10 mM KOAc or LiOAc as indicated). Agarose gels were run at 4 °C for 75 minutes at 44V. The gels were visualized using GE Typhoon gel imager using the Cy3-emmision filter (excitation: 532 nm; PMT: 950 V) and quantified using ImageQuant TL software.

We found that the proximity of the Cy3 dye to the terminal guanosines within our G4 RNAs resulted in fluorescent quenching (~2.5-fold). However, upon PRC2-binding, an increased Cy3 emission was observed that we attributed to exclusion of the dye from proximal guanosine residues. This phenomenon has been observed previously for G-rich sequences.3,4 To account for this phenomenon in our calculations, we corrected all unbound fluorescent intensities by a factor equal to the maximµM Cy3-signal as measured in the presence of saturating PRC2 divided by the fluorescence of unbound Cy3-RNA.

5.4.3 Circular dichroism (CD) spectroscopy

For CD experiments, oligonucleotides (9.8 μM) were folded as described above in a buffer containing 2 mM sodiμM phosphate (pH 7.0), 0.1 mM EDTA, and 100 mM of either KCl or LiCl as indicated. Data were obtained from a 450 μL sample in a quartz cuvette using an Applied Photophysics Chirascan spectrophotometer (Leatherhead, England) at 1 nm intervals from 220 to 370 nm. All data were collected at a constant temperature of 23 °C.

5.4.4 $(GGAA)_{10}$ competition assay.

Complexes of PRC2 (100 nM) and Cy3-labeled $(GGAA)_{10}$ (10 nM) were pre-formed in PRC2 binding buffer as described for EMSAs (30 minutes at 30 °C). Competitive binding experiments were carried out by adding variable concentrations (10–1300 nM) of unlabeled D-$(GGAA)_{10}$ competitor to the pre-formed PRC2-Cy3-L-$(GGAA)_{10}$ complexes (or vise versa), and the reaction was allowed to proceed for 30 minutes at 30 °C. Bound versus unbound fractions were subsequently separated by 1% agarose gel electrophoresis (0.2 X TBE supplemented with 10 mM KOA) and quantified as described above.

5.4.5 HOTAIR-binding and Competition assay

A DNA fragment representing the first 300 nt of HOTAIR (HOTAIR-300) was prepared via PCR assembly using gBlocks Gene Fragments (IDT; Coralville, IA). The resulting DNA was added directly into a 100 μL transcription reaction containing 10 U/μL

T7 RNA polymerase, 0.001 U/μL Inorganic pyrophosphatase (IPP), 25 mM $MgCl_2$, 2 mM spermidine, 10 mM DTT, 40 mM Tris (pH 7.9), and 5 mM of each of the four NTPs, where 5-aminoallyl-UTP (Thermo Fisher Scientific, Waltham, MA) was supplemented in the transcription reaction at 0.5 mM. The reaction mixture was incubated at 37 °C for 2 hours, then enzymes, DNA, and unincorporated NTPs were removed using a Quick-RNA Mini Prep Plus Kit (Zymo Research, Irvine, California) and pure HOTAIR RNA was obtained in 1 X TE buffer. We then used the internally positioned amine functional groups (on the 5-aminoallyl-UTP) to couple a Cy5 NHS-ester (LμMiprobe Life Science Solutions, Hallendale Beach, FL) using the provided procedure. For the competition experiments, HOTAIR-PRC2 complexes (25 and 250 nM, respectively) were pre-formed in PRC2 binding buffer as described for EMSAs, and unlabeled $(GGAA)_{10}$ or $(A)_{40}$ competitor RNA was added in 3-fold increments from 1 nM to 3 μM. (Figure 5.6). Bound versus unbound fractions were subsequently separated by 1% agarose gel electrophoresis (0.2 TBE supplemented with 10 mM KOA) and quantified as described above.

5.4.6 Assembly and Characterization of Cy5-labelled 12×601 arrays

Human histone proteins were expressed and purified as described in Chapter 2 and the Cy5-labelled nucleosome array was assembled using our recently published "plug and play" approach.2 Briefly, we utilized two internally positioned nicking endonuclease sites (Nt. BstNBI) within the fifth 601 unit (N5) of the 12 601 array (Figure 5.7) to generate two single-stranded breaks flanking a region of 28 nucleotides (nt). The dual-nicked 12-601 array DNA was then mixed with 20-fold excess of a Cy5-labelled (internally)

oligonucleotide insert consisting of a sequence identical to the 28 nt fragment generated by the Nt. BstNBI nicking endonuclease. The mixture was then heated at 80°C for 20 minutes before being cooled to room temperature at -1°C/min. Following the annealing step (~1 hour), T4 DNA ligase and ATP (2 mM final concentration) were added to the mixture to reseal the nicks and generate an intact DNA strand. The efficiency of the exchange process (nicking, insertion, and ligation) was carefully monitored in order to ensure complete insertion of the modified oligonucleotide (see Figure 5.7). Oligonucleosome reconstitutions were carried out via salt dialysis and the arrays were purified by selective Mg2+ -induced precipitation. Nucleosome saturation was confirmed by selective restriction enzyme digestion (Figure 5.7).

5.4.7 Chromatin-binding and Competition assay

In order to confirm that PRC2 was capable of binding the Cy5-labled oligonucleosome array, we performed an EMSA using the same conditions described for the $(GGAA)_{10}$ binding experiments (Figure D.4). Using 8 nM arrays, we found that 1:1 PRC2-chromatin complexes were initiated at PRC2 concentrations 500 nM), we observed non-stoichiometric binding by PRC2, resulting higher molecular weight complexes that migrated significantly slower than unbound chromatin when analyzed by agarose gel electrophoresis (0.7%). For the competition assay, we chose a concentration of PRC2 that resulted in a clearly visible interaction by gel electrophoresis (1000 nM PRC2) and generated PRC2-chromatin complexes by incubating PRC2 with the Cy5- labeled array (8 nM) at 30 °C for 30 minutes in PRC2 binding buffer. We then added unlabeled competitor

$(GGAA)_{10}$ RNA in 3-fold increments from 1 nM up to 3 μM and analyzed the results by 0.7% agarose gel electrophoresis (0.2 TBE, 10 mM KOAc, 44 V, 5 hours) (Figure 5.8).

REFERENCES

(1) Watson, J. D.; Crick, F. H. C. The Structure of Dna. *Cold Spring Harb Symp Quant Biol* **1953**, *18*, 123–131. https://doi.org/10.1101/SQB.1953.018.01.020.

(2) Marx, A. Chemical Biology of DNA Replication: Probing DNA Polymerase Selectivity Mechanisms with Modified Nucleotides. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 63–71. https://doi.org/10.1002/9780470664001.ch3.

(3) Oberhuber, M. Nucleic Acid-Templated Chemistry. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 73–101. https://doi.org/10.1002/9780470664001.ch4.

(4) Jäschke, A. RNA as a Catalyst: The Diels–Alderase Ribozyme. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 339–354. https://doi.org/10.1002/9780470664001.ch16.

(5) Fritsch, V.; Westhof, E. The Architectural Motifs of Folded RNAs. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 141–174. https://doi.org/10.1002/9780470664001.ch7.

(6) Wang, Q.; Unrau, P. J. Evolving an Understanding of RNA Function by In Vitro Approaches. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 355–376. https://doi.org/10.1002/9780470664001.ch17.

(7) Ames, T. D.; Breaker, R. R. Bacterial Riboswitch Discovery and Analysis. In *The Chemical Biology of Nucleic Acids*; John Wiley & Sons, Ltd, **2010**; pp 433–454. https://doi.org/10.1002/9780470664001.ch20.

(8) DNA and RNA. *Bihar PSC Exam Notes*, **2017**; https://bihar.pscnotes.com/life-science/dna-and-rna/

(9) Olson, W. K.; Colasanti, A. V.; Lu, X.-J.; Zhurkin, V. B. Watson–Crick Base Pairs: Character and Recognition. In *Wiley Encyclopedia of Chemical Biology*; American Cancer Society, **2008**; pp 1–11. https://doi.org/10.1002/9780470048672.wecb452.

(10) What is a gene?; *Genetics Home Reference*; **2020**; https://medlineplus.gov/genetics/understanding/basics/gene/

(11) Venter, J. C. *et, al*; The Sequence of the Human Genome. *Science* **2001**, *291* (5507), 1304–1351. https://doi.org/10.1126/science.1058040.

(12) Brent, M. R. How Does Eukaryotic Gene Prediction Work? *Nat Biotechnol* **2007**, *25* (8), 883–885. https://doi.org/10.1038/nbt0807-883.

(13) The FANTOM Consortium. The Transcriptional Landscape of the Mammalian Genome. *Science* **2005**, *309* (5740), 1559–1563. https://doi.org/10.1126/science.1112014.

(14) Cooper, G. M. The Complexity of Eukaryotic Genomes. *The Cell: A Molecular Approach. 2nd edition* **2000**.

(15) Triyoso, D. H.; Good, T. A. Pulsatile Shear Stress Leads to DNA Fragmentation in Human SH-SY5Y Neuroblastoma Cell Line. *J Physiol* **1999**, *515* (Pt 2), 355–365. https://doi.org/10.1111/j.1469-7793.1999.355ac.x.

(16) Scully, R.; Panday, A.; Elango, R.; Willis, N. A. DNA Double-Strand Break Repair-Pathway Choice in Somatic Mammalian Cells. *Nature Reviews Molecular Cell Biology* **2019**, *20* (11), 698–714. https://doi.org/10.1038/s41580-019-0152-0.

(17) Chapman, J. R.; Taylor, M. R.; Boulton, S. J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell Review* **2012**, *47*, 497–510. https://doi.org/10.1016/j.molcel.2012.07.029.

(18) Hardison, Ross; B-Form, A-Form, and Z-Form of DNA; *Biology LibreTexts*; **2020**; https://bio.libretexts.org/Bookshelves/Genetics/Book.

(19) Paweletz, N. Walther Flemming: Pioneer of Mitosis Research. *Nature Perspectives* **2001**, *2*, 72–75.

(20) Hernandez, V. “Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III” (1944) by Oswald Avery, Colin MacLeod and Maclyn McCarty.

(21) Hershey, A.; Chase, M. The Hershey-Chase Experiments (1952) https://embryo.asu.edu/pages/hershey-chase-experiments-1952-alfred-hershey-and-martha-chase.

(22) Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **1974**, *184* (4139), 868–871. https://doi.org/10.1126/science.184.4139.868.

(23) Huberman, J. A.; Attardi, G. Isolation of Metaphase Chromosomes from HeLa Cells. *The Journal of Cell Biology* **1966**, *31* (1), 95–105. https://doi.org/10.1083/jcb.31.1.95.

(24) Olins, A. L.; Olins, D. E. Spheroid Chromatin Units (Ngr Bodies). *Science* **1974**, *183* (4122), 330–332. https://doi.org/10.1126/science.183.4122.330.

(25) Pooley, A. S.; Pardon, J. F.; Richards, B. M. The Relation between the Unit Thread of Chromosomes and Isolated Nucleohistone. *Journal of Molecular Biology* **1974**, *85* (4), 533–549. https://doi.org/10.1016/0022-2836(74)90314-3.

(26) Luger, K.; Dechassa, M. L.; Tremethick, D. J. New Insights into Nucleosome and Chromatin Structure: An Ordered State or a Disordered Affair? *Nat. Rev. Mol. Cell Biol.* **2012**, *13* (7), 436–447. https://doi.org/10.1038/nrm3382.

(27) Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F.; Richmond, T. J. Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution. *Nature* **1997**, *389* (6648), 251–260. https://doi.org/10.1038/38444.

(28) McGhee, J. D.; Felsenfeld, G. Nucleosome Structure. *Annual Review of Biochemistry* **1980**, *49* (1), 1115–1156. https://doi.org/10.1146/annurev.bi.49.070180.005343.

(29) Chromatin Structure as Probed by Nucleases and Proteases: Evidence for the Central Role of Hitones H3 and H4. *Cell* **1976**, *9* (1), 179–193. https://doi.org/10.1016/0092-8674(76)90063-5.

(30) Finch, J. T.; Lutter, L. C.; Rhodes, D.; Brown, R. S.; Rushton, B.; Levitt, M.; Klug, A. Structure of Nucleosome Core Particles of Chromatin. *Nature* **1977**, *269* (5623), 29–36. https://doi.org/10.1038/269029a0.

(31) Hayes, J. J.; Tullius, T. D.; Wolffe, A. P. The Structure of DNA in a Nucleosome. *Proceedings of the National Academy of Sciences* **1990**, *87* (19), 7405–7409. https://doi.org/10.1073/pnas.87.19.7405.

(32) Mirzabekov, A. D.; Shick, V. V.; Belyavsky, A. V.; Bavykin, S. G. Primary Organization of Nucleosome Core Particle of Chromatin: Sequence of Histone Arrangement along DNA. *Proceedings of the National Academy of Sciences* **1978**, *75* (9), 4184–4188. https://doi.org/10.1073/pnas.75.9.4184.

(33) Dorigo, B.; Schalch, T.; Bystricky, K.; Richmond, T. J. Chromatin Fiber Folding: Requirement for the Histone H4 N-Terminal Tail. *Journal of Molecular Biology* **2003**, *327* (1), 85–96. https://doi.org/10.1016/S0022-2836(03)00025-1.

(34) Pepenella, S.; Murphy, K. J.; Hayes, J. J. Intra- and Inter-Nucleosome Interactions of the Core Histone Tail Domains in Higher-Order Chromatin Structure. *Chromosoma* **2014**, *123* (0), 3–13. https://doi.org/10.1007/s00412-013-0435-8.

(35) Anderson, J. D.; Widom, J. Sequence and Position-Dependence of the Equilibrium Accessibility of Nucleosomal DNA Target Sites. *J. Mol. Biol.* **2000**, *296* (4), 979–987. https://doi.org/10.1006/jmbi.2000.3531.

(36) Travers, A. A. DNA Bending and Nucleosome Positioning. *Trends in Biochemical Sciences* **1987**, *12*, 108–112. https://doi.org/10.1016/0968-0004(87)90050-8.

(37) A genomic code for nucleosome positioning | Nature https://www.nature.com/articles/nature04979 (accessed Aug 31, 2020).

(38) Thåström, A.; Bingham, L. M.; Widom, J. Nucleosomal Locations of Dominant DNA Sequence Motifs for Histone–DNA Interactions and Nucleosome Positioning. *Journal of Molecular Biology* **2004**, *338* (4), 695–709. https://doi.org/10.1016/j.jmb.2004.03.032.

(39) Thåström, A.; Lowary, P. T.; Widlund, H. R.; Cao, H.; Kubista, M.; Widom, J. Sequence Motifs and Free Energies of Selected Natural and Non-Natural Nucleosome Positioning DNA Sequences. *J. Mol. Biol.* **1999**, *288* (2), 213–229. https://doi.org/10.1006/jmbi.1999.2686.

(40) Allahverdi, A.; Chen, Q.; Korolev, N.; Nordenskiöld, L. Chromatin Compaction under Mixed Salt Conditions: Opposite Effects of Sodium and Potassium Ions on Nucleosome Array Folding. *Scientific Reports* **2015**, *5* (1), 8512. https://doi.org/10.1038/srep08512.

(41) Korolev, N.; Allahverdi, A.; Yang, Y.; Fan, Y.; Lyubartsev, A. P.; Nordenskiöld, L. Electrostatic Origin of Salt-Induced Nucleosome Array Compaction. *Biophys J* **2010**, *99* (6), 1896–1905. https://doi.org/10.1016/j.bpj.2010.07.017.

(42) Kan, P.-Y.; Caterino, T. L.; Hayes, J. J. The H4 Tail Domain Participates in Intra- and Internucleosome Interactions with Protein and DNA during Folding and Oligomerization of Nucleosome Arrays. *Molecular and Cellular Biology* **2009**, *29* (2), 538–546. https://doi.org/10.1128/MCB.01343-08.

(43) Kalashnikova, A. A.; Porter-Goff, M. E.; Muthurajan, U. M.; Luger, K.; Hansen, J. C. The Role of the Nucleosome Acidic Patch in Modulating Higher Order Chromatin Structure. *J R Soc Interface* **2013**, *10* (82). https://doi.org/10.1098/rsif.2012.1022.

(44) Robinson, P. J. J.; Fairall, L.; Huynh, V. A. T.; Rhodes, D. EM Measurements Define the Dimensions of the "30-Nm" Chromatin Fiber: Evidence for a Compact, Interdigitated Structure. *PNAS* **2006**, *103* (17), 6506–6511. https://doi.org/10.1073/pnas.0601212103.

(45) Song, F.; Chen, P.; Sun, D.; Wang, M.; Dong, L.; Liang, D.; Xu, R.-M.; Zhu, P.; Li, G. Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **2014**, *344* (6182), 376–380. https://doi.org/10.1126/science.1251413.

(46) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. X-Ray Structure of a Tetranucleosome and Its Implications for the Chromatin Fibre. *Nature* **2005**, *436* (7047), 138–141. https://doi.org/10.1038/nature03686.

(47) Li, G.; Reinberg, D. Chromatin Higher-Order Structures and Gene Regulation. *Curr Opin Genet Dev* **2011**, *21* (2), 175–186. https://doi.org/10.1016/j.gde.2011.01.022.

(48) Grigoryev, S. A.; Arya, G.; Correll, S.; Woodcock, C. L.; Schlick, T. Evidence for Heteromorphic Chromatin Fibers from Analysis of Nucleosome Interactions. *PNAS* **2009**, *106* (32), 13317–13322. https://doi.org/10.1073/pnas.0903280106.

(49) Scheffer, M. P.; Eltsov, M.; Frangakis, A. S. Evidence for Short-Range Helical Order in the 30-Nm Chromatin Fibers of Erythrocyte Nuclei. *Proc Natl Acad Sci U S A* **2011**, *108* (41), 16992–16997. https://doi.org/10.1073/pnas.1108268108.

(50) Bajpai, G.; Jain, I.; Inamdar, M.; Das, D.; Padinhateeri, R. Binding of DNA-Bending Non-Histone Proteins Destabilizes Regular 30-Nm Chromatin Structure. *PLOS Computational Biology* **2017**, 1–19.

(51) Ou, H.; Phan, S.; Deerinck, T.; Thor, A.; Ellisman, M.; O'Shea, C. ChromEMT: Visualizing 3D Chromatin Structure and Compaction in Interphase and Mitotic Cells. *Science* **2017**, *357* (6349), 1–13.

(52) Ricci, M. A.; Manzo, C.; Garcia-Parajo, M. F.; Lakadamyali, M. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **2015**, *160*, 1145–1158. https://doi.org/10.1016/j.cell.2015.01.054.

(53) Bonev, B.; Cavalli, G. Organization and Function of the 3D Genome. *Nat Rev Genet* **2016**, *17* (11), 661–678. https://doi.org/10.1038/nrg.2016.112.

(54) Strom, A. R.; Emelyanov, A. V.; Mir, M.; Fyodorov, D. V.; Darzacq, X.; Karpen, G. H. Phase Separation Drives Heterochromatin Domain Formation. *Nature* **2017**, *547* (7662), 241–245. https://doi.org/10.1038/nature22989.

(55) Hnisz, D.; Shrinivas, K.; Young, R. A.; Chakraborty, A. K.; Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **2017**, *169* (1), 13–23. https://doi.org/10.1016/j.cell.2017.02.007.

(56) Li, Y.; Haarhuis, J. H. I.; Sedeño Cacciatore, Á.; Oldenkamp, R.; van Ruiten, M. S.; Willems, L.; Teunissen, H.; Muir, K. W.; de Wit, E.; Rowland, B. D.; Panne, D. The Structural Basis for Cohesin–CTCF-Anchored Loops. *Nature* **2020**, *578* (7795), 472–476. https://doi.org/10.1038/s41586-019-1910-z.

(57) Hansen, A. S.; Pustova, I.; Cattoglio, C.; Tjian, R.; Darzacq, X. CTCF and Cohesin Regulate Chromatin Loop Stability with Distinct Dynamics. *eLife* **2017**, *6*, e25776. https://doi.org/10.7554/eLife.25776.

(58) Schwarz, P. M.; Felthauser, A.; Fletcher, T. M.; Hansen, J. C. Reversible Oligonucleosome Self-Association: Dependence on Divalent Cations and Core Histone Tail Domains. *Biochemistry* **1996**, *35* (13), 4009–4015. https://doi.org/10.1021/bi9525684.

(59) Lu, X.; Hamkalo, B.; Parseghian, M. H.; Hansen, J. C. Chromatin Condensing Functions of the Linker Histone C-Terminal Domain Are Mediated by Specific Amino Acid Composition and Intrinsic Protein Disorder. *Biochemistry* **2009**, *48* (1), 164–172. https://doi.org/10.1021/bi801636y.

(60) Risca, V. I.; Denny, S. K.; Straight, A. F.; Greenleaf, W. J. Variable Chromatin Structure Revealed by in Situ Spatially Correlated DNA Cleavage Mapping. *Nature* **2017**, *541* (7636), 237–241. https://doi.org/10.1038/nature20781.

(61) Nizovtseva, E. V.; Clauvelin, N.; Todolli, S.; Polikanov, Y. S.; Kulaeva, O. I.; Wengrzynek, S.; Olson, W. K.; Studitsky, V. M. Nucleosome-Free DNA Regions Differentially Affect Distant Communication in Chromatin. *Nucleic Acids Res* **2017**, *45* (6), 3059–3067. https://doi.org/10.1093/nar/gkw1240.

(62) Kharerin, H.; Bhat, P. J.; Padinhateeri, R. Role of Nucleosome Positioning in 3D Chromatin Organization and Loop Formation. *J Biosci* **2020**, *45* (1), 14. https://doi.org/10.1007/s12038-019-9976-1.

(63) Jost, D.; Carrivain, P.; Cavalli, G.; Vaillant, C. Modeling Epigenome Folding: Formation and Dynamics of Topologically Associated Chromatin Domains. *Nucleic Acids Res* **2014**, *42* (15), 9553–9561. https://doi.org/10.1093/nar/gku698.

(64) Kolendowski, B.; Hassan, H.; Krstic, M.; Isovic, M.; Thillainadesan, G.; Chambers, A. F.; Tuck, A. B.; Torchia, J. Genome-Wide Analysis Reveals a Role for TDG in Estrogen Receptor-Mediated Enhancer RNA Transcription and 3-Dimensional Reorganization. *Epigenetics & Chromatin* **2018**, *11* (1). https://doi.org/10.1186/s13072-018-0176-2.

(65) Buenrostro, J.; Wu, B.; Chang, H.; Greenleaf, W. ATAC-Seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **2015**, *109*, 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109.

(66) Song, C.-X.; Szulwach, K. E.; Dai, Q.; Fu, Y.; Mao, S.-Q.; Lin, L.; Street, C.; Li, Y.; Poidevin, M.; Wu, H.; Gao, J.; Liu, P.; Li, L.; Xu, G.-L.; Jin, P.; He, C. Genome-Wide Profiling of 5-Formylcytosine Reveals Its Roles in Epigenetic Priming. *Cell* **2013**, *153* (3), 678–691. https://doi.org/10.1016/j.cell.2013.04.001.

(67) Jin, C. (1); Felsenfeld, G. (1); Zang, C. (2); Peng, W. (2); Wei, G. (3); Cui, K. (3); Zhao, K. (3). H3.3/H2A.Z Double Variant-Containing Nucleosomes Mark "nucleosome-Free Regions" of Active Promoters and Other Regulatory Regions. *Nature Genetics* **2009**, *41* (8), 941–945. https://doi.org/10.1038/ng.409.

(68) Hassan, H. M.; Kolendowski, B.; Isovic, M.; Bose, K.; Dranse, H. J.; Sampaio, A. V.; Underhill, T. M.; Torchia, J. Regulation of Active DNA Demethylation through RAR-Mediated Recruitment of a TET/TDG Complex. *Cell Reports* **2017**, *19* (8), 1685–1697. https://doi.org/10.1016/j.celrep.2017.05.007.

(69) Hagège, H.; Klous, P.; Braem, C.; Splinter, E.; Dekker, J.; Cathala, G.; de Laat, W.; Forné, T. Quantitative Analysis of Chromosome Conformation Capture Assays (3C-QPCR). *Nat Protoc* **2007**, *2* (7), 1722–1733. https://doi.org/10.1038/nprot.2007.243.

(70) Han, J.; Zhang, Z.; Wang, K. 3C and 3C-Based Techniques: The Powerful Tools for Spatial Genome Organization Deciphering. *Mol Cytogenet* **2018**, *11*. https://doi.org/10.1186/s13039-018-0368-2.

(71) Hug, C. B.; Vaquerizas, J. M. The Birth of the 3D Genome during Early Embryonic Development. *Trends in Genetics* **2018**, *34* (12), 903–914. https://doi.org/10.1016/j.tig.2018.09.002.

(72) Narlikar, G. J.; Sundaramoorthy, R.; Owen-Hughes, T. Mechanisms and Functions of ATP-Dependent Chromatin-Remodeling Enzymes. *Cell* **2013**, *154* (3), 490–503. https://doi.org/10.1016/j.cell.2013.07.011.

(73) Sanulli, S.; Trnka, M. J.; Dharmarajan, V.; Tibble, R. W.; Pascal, B. D.; Burlingame, A. L.; Griffin, P. R.; Gross, J. D.; Narlikar, G. J. HP1 Reshapes Nucleosome Core to Promote Phase Separation of Heterochromatin. *Nature* **2019**, *575* (7782), 390–394. https://doi.org/10.1038/s41586-019-1669-2.

(74) Cirillo, L. A.; Lin, F. R.; Cuesta, I.; Friedman, D.; Jarnik, M.; Zaret, K. S. Opening of Compacted Chromatin by Early Developmental Transcription Factors HNF3 (FoxA) and GATA-4. *Molecular Cell* **2002**, *9* (2), 279–289. https://doi.org/10.1016/S1097-2765(02)00459-8.

(75) Leger, H.; Smet-Nocca, C.; Attmane-Elakeb, A.; Morley-Fletcher, S.; Benecke, A. G.; Eilebrecht, S. A TDG/CBP/RARα Ternary Complex Mediates the Retinoic Acid-Dependent Expression of DNA Methylation-Sensitive Genes. *Genomics Proteomics Bioinformatics* **2014**, *12*, 8–18.

(76) Ngo, T. T. M.; Yoo, J.; Dai, Q.; Zhang, Q.; He, C.; Aksimentiev, A.; Ha, T. Effects of Cytosine Modifications on DNA Flexibility and Nucleosome Mechanical Stability. *Nat Commun* **2016**, *7* (1), 1–9. https://doi.org/10.1038/ncomms10813.

(77) Pennings, S. DNA Methylation, Nucleosome Formation and Positioning. *Briefings in Functional Genomics and Proteomics* **2005**, *3* (4), 351–361. https://doi.org/10.1093/bfgp/3.4.351.

(78) Jimenez-Useche, I.; Nurse, N. P.; Tian, Y.; Kansara, B. S.; Shim, D.; Yuan, C. DNA Methylation Effects on Tetra-Nucleosome Compaction and Aggregation. *Biophysical Journal* **2014**, *107* (7), 1629–1636. https://doi.org/10.1016/j.bpj.2014.05.055.

(79) Bogdanović, O.; Veenstra, G. J. C. DNA Methylation and Methyl-CpG Binding Proteins: Developmental Requirements and Function. *Chromosoma* **2009**, *118* (5), 549–565. https://doi.org/10.1007/s00412-009-0221-9.

(80) Mellén, M.; Ayata, P.; Heintz, N. 5-Hydroxymethylcytosine Accumulation in Postmitotic Neurons Results in Functional Demethylation of Expressed Genes. *Proc Natl Acad Sci USA* **2017**, *114* (37), E7812–E7821. https://doi.org/10.1073/pnas.1708044114.

(81) Raiber, E. A.; Portella, G.; Cuesta, S. M.; Hardisty, R.; Murat, P.; Li, Z.; Iularo, M.; Dean, W.; Spindel, J.; Beraldi, D.; Liu, Z.; Dawson, M. A.; Reik, W.; Balasubramanian, S. 5-Formylcytosine Organizes Nucleosomes and Forms Schiff Base Interactions with Histones in Mouse Embryonic Stem Cells. *Nature Chemistry* **2018**, *10*, 1258–1266.

(82) Kellinger, M. W.; Song, C.-X.; Chong, J.; Lu, X.-Y.; He, C.; Wang, D. 5-Formylcytosine and 5-Carboxylcytosine Reduce the Rate and Substrate Specificity of RNA Polymerase II Transcription. *Nat. Struct. Mol. Biol.* **2012**, *19* (8), 831–833. https://doi.org/10.1038/nsmb.2346.

(83) Kitsera, N.; Allgayer, J.; Parsa, E.; Nadine, G.; Rossa, M.; Carell, T.; Khobta, A. Functional Impacts of 5-Hydroxymethylcytosine, 5-Formylcytosine, and 5-Carboxycytosine at a Single Hemi-Modified CpG Dinucleotide in a Gene Promoter. *Nucleic Acids Research* **2017**, *45* (19).

(84) Gansen, A.; Tóth, K.; Schwarz, N.; Langowski, J. Opposing Roles of H3- and H4-Acetylation in the Regulation of Nucleosome Structure—a FRET Study. *Nucleic Acids Res* **2015**, *43* (3), 1433–1443. https://doi.org/10.1093/nar/gku1354.

(85) Garcia-Ramirez, M.; Rocchini, C.; Ausio, J. Modulation of Chromatin Folding by Histone Acetylation. *J. Biol. Chem.* **1995**, *270* (30), 17923–17928. https://doi.org/10.1074/jbc.270.30.17923.

(86) Banerjee, T.; Chakravarti, D. A Peek into the Complex Realm of Histone Phosphorylation ▿. *Mol Cell Biol* **2011**, *31* (24), 4858–4873. https://doi.org/10.1128/MCB.05631-11.

(87) Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, Erasing and Reading Histone Lysine Methylations. *Experimental & Molecular Medicine* **2017**, *49* (4), e324–e324. https://doi.org/10.1038/emm.2017.11.

(88) Dhall, A.; Weller, C. E.; Chu, A.; Shelton, P. M. M.; Chatterjee, C. Chemically Sumoylated Histone H4 Stimulates Intranucleosomal Demethylation by the LSD1-CoREST Complex. *ACS Chem. Biol.* **2017**, *12* (9), 2275–2280. https://doi.org/10.1021/acschembio.7b00716.

(89) Debelouchina, G. T.; Gerecht, K.; Muir, T. W. Ubiquitin Utilizes an Acidic Surface Patch to Alter Chromatin Structure. *Nat. Chem. Biol.* **2017**, *13* (1), 105–110. https://doi.org/10.1038/nchembio.2235.

(90) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B Ubiquitylation Disrupts Local and Higher-Order Chromatin Compaction. *Nature Chemical Biology* **2011**, *7*, 113–110.

(91) Caldecott, K. W. Protein ADP-Ribosylation and the Cellular Response to DNA Strand Breaks. *DNA Repair* **2014**, *19*, 108–113. https://doi.org/10.1016/j.dnarep.2014.03.021.

(92) Fenley, A. T.; Anandakrishnan, R.; Kidane, Y. H.; Onufriev, A. V. Modulation of Nucleosomal DNA Accessibility via Charge-Altering Post-Translational Modifications in Histone Core. *Epigenetics & Chromatin* **2018**, *11* (1). https://doi.org/10.1186/s13072-018-0181-5.

(93) Pradeepa, M. M.; Grimes, G. R.; Kumar, Y.; Olley, G.; Taylor, G. C. A.; Schneider, R.; Bickmore, W. A. Histone H3 Globular Domain Acetylation Identifies a New Class of Enhancers. *Nature Genetics* **2016**, *48* (6), 681–686. https://doi.org/10.1038/ng.3550.

(94) Kebede, A. F.; Nieborak, A.; Shahidian, L. Z.; Le Gras, S.; Richter, F.; Gómez, D. A.; Baltissen, M. P.; Meszaros, G.; Magliarelli, H. de F.; Taudt, A.; Margueron, R.; Colomé-Tatché, M.; Ricci, R.; Daujat, S.; Vermeulen, M.; Mittler, G.; Schneider, R. Histone Propionylation Is a Mark of Active Chromatin. *Nature Structural & Molecular Biology* **2017**, *24* (12), 1048–1056. https://doi.org/10.1038/nsmb.3490.

(95) Sabari, B. R.; Tang, Z.; Huang, H.; Yong-Gonzalez, V.; Molina, H.; Kong, H. E.; Dai, L.; Shimada, M.; Cross, J. R.; Zhao, Y.; Roeder, R. G.; Allis, C. D. Intracellular Crotonyl-CoA Stimulates Transcription through P300-Catalyzed Histone Crotonylation. *Molecular Cell* **2015**, *58* (2), 203–215. https://doi.org/10.1016/j.molcel.2015.02.029.

(96) Jing, Y.; Liu, Z.; Tian, G.; Bao, X.; Ishibashi, T.; Li, X. D. Site-Specific Installation of Succinyl Lysine Analog into Histones Reveals the Effect of H2BK34 Succinylation on Nucleosome Dynamics. *Cell Chemical Biology* **2018**, *25* (2), 166-174.e7. https://doi.org/10.1016/j.chembiol.2017.11.005.

(97) Mishra, L. N.; Pepenella, S.; Rogge, R.; Hansen, J. C.; Hayes, J. J. Acetylation Mimics Within a Single Nucleosome Alter Local DNA Accessibility In Compacted Nucleosome Arrays. *Scientific Reports* **2016**, *6* (1), 1–11. https://doi.org/10.1038/srep34808.

(98) Xie, Z.; Dai, J.; Dai, L.; Tan, M.; Cheng, Z.; Wu, Y.; Boeke, J. D.; Zhao, Y. Lysine Succinylation and Lysine Malonylation in Histones. *Molecular & Cellular Proteomics* **2012**, *11* (5), 100–107. https://doi.org/10.1074/mcp.M111.015875.

(99) Brehove, M.; Wang, T.; North, J.; Luo, Y.; Dreher, S. J.; Shimko, J. C.; Ottesen, J. J.; Luger, K.; Poirier, M. G. Histone Core Phosphorylation Regulates DNA Accessibility. *J. Biol. Chem.* **2015**, *290* (37), 22612–22621. https://doi.org/10.1074/jbc.M115.661363.

(100) Sueoka, T.; Hayashi, G.; Okamoto, A. Regulation of the Stability of the Histone H2A–H2B Dimer by H2A Tyr57 Phosphorylation. *Biochemistry* **2017**, *56* (36), 4767–4772. https://doi.org/10.1021/acs.biochem.7b00504.

(101) Fierz, B.; Chatterjee, C.; McGinty, R. K.; Bar-Dagan, M.; Raleigh, D. P.; Muir, T. W. Histone H2B Ubiquitylation Disrupts Local and Higher-Order Chromatin Compaction. *Nat. Chem. Biol.* **2011**, *7* (2), 113–119. https://doi.org/10.1038/nchembio.501.

(102) Poirier, M. G.; Bussiek, M.; Langowski, J.; Widom, J. Spontaneous Access to DNA Target Sites in Folded Chromatin Fibers. *Journal of Molecular Biology 379* (4), 772–786. https://doi.org/10.1016/j.jmb.2008.04.025.

(103) Li, G.; Levitus, M.; Bustamante, C.; Widom, J. Rapid Spontaneous Accessibility of Nucleosomal DNA. *Nature Structural & Molecular Biology* **2004**, *12*, 46–53.

(104) Polach, K. J.; Widom, J. Mechanism of Protein Access to Specific DNA Sequences in Chromatin: A Dynamic Equilibrium Model for Gene Regulation. *Journal of Molecular Biology* **1995**, *254* (2), 130–149. https://doi.org/10.1006/jmbi.1995.0606.

(105) Clark, R. J.; Felsenfeld, G. Chemical Probes of Chromatin Structure. *Biochemistry* **1974**, *13* (17), 3622–3628. https://doi.org/10.1021/bi00714a034.

(106) Gansen, A.; Valeri, A.; Hauger, F.; Felekyan, S.; Kalinin, S.; Tóth, K.; Langowski, J.; Seidel, C. A. M. Nucleosome Disassembly Intermediates Characterized by Single-Molecule FRET. *PNAS* **2009**, *106* (36), 15308–15313. https://doi.org/10.1073/pnas.0903005106.

(107) Shlyakhtenko, L. S.; Lushnikov, A. Y.; Lyubchenko, Y. L. Dynamics of Nucleosomes Revealed by Time-Lapse Atomic Force Microscopy | Biochemistry. *Biochemistry* **2009**, *48* (33), 7842–7848. https://doi.org/10.1021/bi900977t.

(108) Wang, Z. A.; Liu, W. R. Proteins with Site-Specific Lysine Methylation. *Chemistry* **2017**, *23* (49), 11732–11737. https://doi.org/10.1002/chem.201701655.

(109) Wang, W. W.; Zeng, Y.; Wu, B.; Deiters, A.; Liu, W. R. A Chemical Biology Approach to Reveal Sirt6-Targeted Histone H3 Sites in Nucleosomes. *ACS Chem. Biol.* **2016**, *11* (7), 1973–1981. https://doi.org/10.1021/acschembio.6b00243.

(110) Chatterjee, C.; McGinty, R. K.; Fierz, B.; Muir, T. W. Disulfide-Directed Histone Ubiquitylation Reveals Plasticity in HDot1L Activation. *Nat. Chem. Biol.* **2010**, *6* (4), 267–269. https://doi.org/10.1038/nchembio.315.

(111) Fierz, B.; Muir, T. W. Chromatin as an Expansive Canvas for Chemical Biology. *Nat. Chem. Biol.* **2012**, *8* (5), 417–427. https://doi.org/10.1038/nchembio.938.

(112) Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M. The Site-Specific Installation of Methyl-Lysine Analogs into Recombinant Histones. *Cell* **2007**, *128* (5), 1003–1012. https://doi.org/10.1016/j.cell.2006.12.041.

(113) Gibson, B. A.; Doolittle, L. K.; Schneider, M. W. G.; Jensen, L. E.; Gamarra, N.; Henry, L.; Gerlich, D. W.; Redding, S.; Rosen, M. K. Organization of Chromatin by Intrinsic and Regulated Phase Separation. *Cell* **2019**, *179* (2), 470-484.e21. https://doi.org/10.1016/j.cell.2019.08.037.

(114) Wan, W.; Tharp, J. M.; Liu, W. R. Pyrrolysyl-TRNA Synthetase: An Ordinary Enzyme but an Outstanding Genetic Code Expansion Tool. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **2014**, *1844* (6), 1059–1070. https://doi.org/10.1016/j.bbapap.2014.03.002.

(115) Wang, Z. U.; Wang, Y.-S.; Pai, P.-J.; Russell, W. K.; Russell, D. H.; Liu, W. R. A Facile Method to Synthesize Histones with Posttranslational Modification Mimics. *Biochemistry* **2012**, *51* (26), 5232–5234. https://doi.org/10.1021/bi300535a.

(116) Banerjee, D. R.; Deckard III, C. E.; Zeng, Y.; Sczepanski, J. T. Acetylation of the Histone H3 Tail Domain Regulates Base Excision Repair on Higher-Order Chromatin Structures. *Scientific Reports* **2019**, *9*, 1–11. https://doi.org/10.1038/s41598-019-52340-0.

(117) Poirier, M.; Oh, E.; Tims, H.; Widom, J. Dynamics and Function of Compact Nucleosome Arrays. *Nature Structural & Molecular Biology* **2009**, *16* (9), 938–944. https://doi.org/10.1038/nsmb.1650.

(118) Kilic, S.; Felekyan, S.; Doroshenko, O.; Boichenko, I.; Dimura, M.; Vardanyan, H.; Bryan, L. C.; Arya, G.; Seidel, C. A. M.; Fierz, B. Single-Molecule FRET Reveals Multiscale Chromatin Dynamics Modulated by HP1α. *Nature Communications* **2018**, *9* (235), 1–14. https://doi.org/10.1038/s41467-017-02619-5.

(119) Menoni, H.; Shukla, M. S.; Gerson, V.; Dimitrov, S.; Angelov, D. Base Excision Repair of 8-OxoG in Dinucleosomes. *Nucleic Acids Res* **2012**, *40* (2), 692–700. https://doi.org/10.1093/nar/gkr761.

(120) Caruthers, M. H. A Brief Review of DNA and RNA Chemical Synthesis. *Biochem Soc Trans* **2011**, *39* (2), 575–580. https://doi.org/10.1042/BST0390575.

(121) Sczepanski, J. T.; Wong, R. S.; McKnight, J. N.; Bowman, G. D.; Greenberg, M. M. Rapid DNA–Protein Cross-Linking and Strand Scission by an Abasic Site in a Nucleosome Core Particle. *PNAS* **2010**, *107* (52), 22475–22480. https://doi.org/10.1073/pnas.1012860108.

(122) Rodriguez, Y.; Smerdon, M. J. The Structural Location of DNA Lesions in Nucleosome Core Particles Determines Accessibility by Base Excision Repair Enzymes. *J Biol Chem* **2013**, *288* (19), 13863–13875. https://doi.org/10.1074/jbc.M112.441444.

(123) Li, F.; Zhang, Y.; Bai, J.; Greenberg, M. M.; Xi, Z.; Zhou, C. 5-Formylcytosine Yields DNA–Protein Cross-Links in Nucleosome Core Particles https://pubs.acs.org/doi/pdf/10.1021/jacs.7b05495 (accessed Aug 9, 2020). https://doi.org/10.1021/jacs.7b05495.

(124) Chromatin Structure and the Pioneering Transcription Factor FOXA1 Regulate TDG-Mediated Removal of 5-Formylcytosine from DNA | Journal of the American Chemical Society https://pubs.acs.org/doi/full/10.1021/jacs.9b07576 (accessed Sep 19, 2019).

(125) Kukwikila, M.; Gale, N.; El-Sagheer, A. H.; Brown, T.; Tavassoli, A. Assembly of a Biocompatible Triazole-Linked Gene by One-Pot Click-DNA Ligation. *Nature Chem* **2017**, *9* (11), 1089–1098. https://doi.org/10.1038/nchem.2850.

(126) Kumar, R.; El-Sagheer, A.; Tumpane, J.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T. Template-Directed Oligonucleotide Strand Ligation, Covalent Intramolecular DNA Circularization and Catenation Using Click Chemistry. *J. Am. Chem. Soc.* **2007**, *129* (21), 6859–6864. https://doi.org/10.1021/ja070273v.

(127) Li, G.-M. Mechanisms and Functions of DNA Mismatch Repair. *Cell Research* **2008**, *18* (1), 85–98. https://doi.org/10.1038/cr.2007.115.

(128) Kunkel, T. A.; Erie, D. A. Dna Mismatch Repair. *Annual Review of Biochemistry* **2005**, *74* (1), 681–710. https://doi.org/10.1146/annurev.biochem.74.082803.133243.

(129) Li, X.; Heyer, W.-D. Homologous Recombination in DNA Repair and DNA Damage Tolerance. *Cell Res* **2008**, *18* (1), 99–113. https://doi.org/10.1038/cr.2008.1.

(130) Ranjha, L.; Howard, S. M.; Cejka, P. Main Steps in DNA Double-Strand Break Repair: An Introduction to Homologous Recombination and Related Processes. *Chromosoma* **2018**, *127* (2), 187–214. https://doi.org/10.1007/s00412-017-0658-1.

(131) Kisker, C.; Kuper, J.; Van Houten, B. Prokaryotic Nucleotide Excision Repair. *Cold Spring Harb Perspect Biol* **2013**, *5* (3). https://doi.org/10.1101/cshperspect.a012591.

(132) Schärer, O. D. Nucleotide Excision Repair in Eukaryotes. *Cold Spring Harb Perspect Biol* **2013**, *5* (10). https://doi.org/10.1101/cshperspect.a012609.

(133) Truglio, J. J.; Croteau, D. L.; Van Houten, B.; Kisker, C. Prokaryotic Nucleotide Excision Repair: The UvrABC System. *Chem. Rev.* **2006**, *106* (2), 233–252. https://doi.org/10.1021/cr040471u.

(134) Balliano, A. J.; Hayes, J. J. Base Excision Repair in Chromatin: Insights from Reconstituted Systems. *DNA Repair (Amst)* **2015**, *36*, 77–85. https://doi.org/10.1016/j.dnarep.2015.09.009.

(135) Schermerhorn, K. M.; Delaney, S. A Chemical and Kinetic Perspective on Base Excision Repair of DNA. *Acc Chem Res* **2014**, *47* (4), 1238–1246. https://doi.org/10.1021/ar400275a.

(136) Meas, R.; Wyrick, J. J.; Smerdon, M. J. Nucleosomes Regulate Base Excision Repair in Chromatin. *Mutation Research/Reviews in Mutation Research* **2019**, *780*, 29–36. https://doi.org/10.1016/j.mrrev.2017.10.002.

(137) Fleming, A. M.; Burrows, C. J. 8-Oxo-7,8-Dihydroguanine, Friend and Foe: Epigenetic-like Regulator versus Initiator of Mutagenesis. *DNA Repair (Amst.)* **2017**, *56*, 75–83. https://doi.org/10.1016/j.dnarep.2017.06.009.

(138) Nilsen, H.; Rosewell, I.; Robins, P.; Skjelbred, C. F.; Andersen, S.; Slupphaug, G.; Daly, G.; Krokan, H. E.; Lindahl, T.; Barnes, D. E. Uracil-DNA

Glycosylase (UNG)-Deficient Mice Reveal a Primary Role of the Enzyme during DNA Replication. *Molecular Cell* **2000**, *5* (6), 1059–1065. https://doi.org/10.1016/S1097-2765(00)80271-3.

(139) Neddermann, P.; Jiricny, J. Efficient Removal of Uracil from G.U Mispairs by the Mismatch-Specific Thymine DNA Glycosylase from HeLa Cells. *Proc Natl Acad Sci U S A* **1994**, *91* (5), 1642–1646.

(140) Jacobs, A. L.; Schär, P. DNA Glycosylases: In DNA Repair and Beyond. *Chromosoma* **2012**, *121* (1), 1–20. https://doi.org/10.1007/s00412-011-0347-4.

(141) Howard, M. J.; Rodriguez, Y.; Wilson, S. H. DNA Polymerase β Uses Its Lyase Domain in a Processive Search for DNA Damage. *Nucleic Acids Res* **2017**, *45* (7), 3822–3832. https://doi.org/10.1093/nar/gkx047.

(142) Chafin, D. R.; Vitolo, J. M.; Henricksen, L. A.; Bambara, R. A.; Hayes, J. J. Human DNA Ligase I Efficiently Seals Nicks in Nucleosomes. *EMBO J* **2000**, *19* (20), 5492–5501. https://doi.org/10.1093/emboj/19.20.5492.

(143) Buechner, C. N.; Maiti, A.; Drohat, A. C.; Tessmer, I. Lesion Search and Recognition by Thymine DNA Glycosylase Revealed by Single Molecule Imaging. *Nucleic Acids Research* **2015**, *43* (5), 2716–2729. https://doi.org/10.1093/nar/gkv139.

(144) Esadze, A.; Rodriguez, G.; Weiser, B. P.; Cole, P. A.; Stivers, J. T. Measurement of Nanoscale DNA Translocation by Uracil DNA Glycosylase in Human Cells. *Nucleic Acids Research* **2017**, *45* (21), 12413–12424. https://doi.org/10.1093/nar/gkx848.

(145) Schonhoft, J. D.; Kosowicz, J. G.; Stivers, J. T. DNA Translocation by Human Uracil DNA Glycosylase: Role of DNA Phosphate Charge. *Biochemistry* **2013**, *52* (15), 2526–2535. https://doi.org/10.1021/bi301561d.

(146) Schonhoft, J. D.; Stivers, J. T. Timing Facilitated Site Transfer of an Enzyme on DNA. *Nature Chemical Biology* **2012**, *8* (2), 205–210. https://doi.org/10.1038/nchembio.764.

(147) Maiti, A.; Morgan, M. T.; Pozharski, E.; Drohat, A. C. Crystal Structure of Human Thymine DNA Glycosylase Bound to DNA Elucidates Sequence-Specific Mismatch Recognition. 6.

(148) Jiang, Y. L.; Stivers, J. T. Mutational Analysis of the Base-Flipping Mechanism of Uracil DNA Glycosylase †. *Biochemistry* **2002**, *41* (37), 11236–11247. https://doi.org/10.1021/bi026226r.

(149) Matsubara, M.; Tanaka, T.; Terato, H.; Ohmae, E.; Izumi, S.; Katayanagi, K.; Ide, H. Mutational Analysis of the Damage-Recognition and Catalytic Mechanism of Human SMUG1 DNA Glycosylase. *Nucleic Acids Res* **2004**, *32* (17), 5291–5302. https://doi.org/10.1093/nar/gkh859.

(150) Werner, R. M.; Stivers, J. T. Kinetic Isotope Effect Studies of the Reaction Catalyzed by Uracil DNA Glycosylase: Evidence for an Oxocarbenium Ion−Uracil Anion Intermediate. *Biochemistry* **2000**, *39* (46), 14054–14064. https://doi.org/10.1021/bi0018178.

(151) Esadze, A.; Rodriguez, G.; Cravens, S. L.; Stivers, J. T. APE1 Accelerates Turnover of HOGG1 by Preventing Retrograde Binding to the Abasic Site Product. *Biochemistry* **2017**, *56* (14), 1974–1986. https://doi.org/10.1021/acs.biochem.7b00017.

(152) Waters, T. R.; Gallinari, P.; Jiricny, J.; Swann, P. F. Human Thymine DNA Glycosylase Binds to Apurinic Sites in DNA but Is Displaced by Human Apurinic Endonuclease 1. *J. Biol. Chem.* **1999**, *274* (1), 67–74. https://doi.org/10.1074/jbc.274.1.67.

(153) Tarantino, M. E.; Dow, B. J.; Drohat, A. C.; Delaney, S. Nucleosomes and the Three Glycosylases: High, Medium, and Low Levels of Excision by the Uracil DNA Glycosylase Superfamily. *DNA Repair (Amst)* **2018**, *72*, 56–63. https://doi.org/10.1016/j.dnarep.2018.09.008.

(154) Beard, B. C.; Wilson, S. H.; Smerdon, M. J. Suppressed Catalytic Activity of Base Excision Repair Enzymes on Rotationally Positioned Uracil in Nucleosomes. *PNAS* **2003**, *100* (13), 7465–7470. https://doi.org/10.1073/pnas.1330328100.

(155) Cole, H. A.; Tabor-Godwin, J. M.; Hayes, J. J. Uracil DNA Glycosylase Activity on Nucleosomal DNA Depends on Rotational Orientation of Targets. *J. Biol. Chem.* **2010**, *285* (4), 2876–2885. https://doi.org/10.1074/jbc.M109.073544.

(156) Hinz, J. M. Impact of Abasic Site Orientation within Nucleosomes on Human APE1 Endonuclease Activity. *Mutat Res* **2014**, *0*, 19–24. https://doi.org/10.1016/j.mrfmmm.2014.05.008.

(157) Ed, O.; S, D. Differential Ability of Five DNA Glycosylases to Recognize and Repair Damage on Nucleosomal DNA. *ACS Chem Biol* **2017**, *12* (3), 692–701. https://doi.org/10.1021/acschembio.6b00921.

(158) Banerjee, D. R.; Deckard, C. E.; Elinski, M. B.; Buzbee, M. L.; Wang, W. W.; Batteas, J. D.; Sczepanski, J. T. Plug-and-Play Approach for Preparing Chromatin Containing Site-Specific DNA Modifications: The Influence of Chromatin Structure on Base Excision Repair. *Journal of the American Chemical Society* **2018**, *140* (26), 8260–8267. https://doi.org/10.1021/jacs.8b04063.

(159) Coey, C. T.; Malik, S. S.; Pidugu, L. S.; Varney, K. M.; Pozharski, E.; Drohat, A. C. Structural Basis of Damage Recognition by Thymine DNA Glycosylase: Key Roles for N-Terminal Residues. *Nucleic Acids Res* **2016**, *44* (21), 10248–10258. https://doi.org/10.1093/nar/gkw768.

(160) Rodriguez, Y.; Hinz, J. M.; Laughery, M. F.; Wyrick, J. J.; Smerdon, M. J. Site-Specific Acetylation of Histone H3 Decreases Polymerase β Activity on Nucleosome Core Particles in Vitro. *Journal of Biological Chemistry* **2016**, *291* (21), 11434–11445. https://doi.org/10.1074/jbc.M116.725788.

(161) Cortázar, D.; Kunz, C.; Saito, Y.; Steinacher, R.; Schär, P. The Enigmatic Thymine DNA Glycosylase. *DNA Repair* **2007**, *6* (4), 489–504. https://doi.org/10.1016/j.dnarep.2006.10.013.

(162) Tini, M.; Benecke, A.; Um, S.-J.; Torchia, J.; Evans, R. M.; Chambon, P. Association of CBP/P300 Acetylase and Thymine DNA Glycosylase Links DNA Repair and Transcription. *Molecular Cell* **2002**, *9* (2), 265–277. https://doi.org/10.1016/S1097-2765(02)00453-7.

(163) Santos, F.; Hendrich, B.; Reik, W.; Dean, W. Dynamic Reprogramming of DNA Methylation in the Early Mouse Embryo. *Developmental Biology* **2002**, *241* (1), 172–182. https://doi.org/10.1006/dbio.2001.0501.

(164) Smith, Z. D.; Chan, M. M.; Mikkelsen, T. S.; Gu, H.; Gnirke, A.; Regev, A.; Meissner, A. A Unique Regulatory Phase of DNA Methylation in the Early Mammalian Embryo. *Nature* **2012**, *484* (7394), 339–344. https://doi.org/10.1038/nature10960.

(165) Serandour, A. A.; Avner, S.; Percevault, F.; Demay, F.; Bizot, M.; Lucchetti-Miganeh, C.; Barloy-Hubler, F.; Brown, M.; Lupien, M.; Metivier, R.; Salbert, G.; Eeckhoute, J. Epigenetic Switch Involved in Activation of Pioneer Factor FOXA1-Dependent Enhancers. *Genome Research* **2011**, *21* (4), 555–565. https://doi.org/10.1101/gr.111534.110.

(166) Metivier, R.; Gallais, R.; Tiffoche, C.; Le Peron, C.; Jurkowska, R. Z.; Carmouche, R. P.; Ibberson, D.; Barath, P.; Demay, F.; Reid, G.; Benes, V.; Jeltsch, A.; Gannon, F.; Salbert, G. Cyclical DNA Methylation of a Transcriptionally Active Promoter. *Nature* **2008**, *452*, 45–50. https://doi.org/10.1038/nature06544.

(167) Wu, S. C.; Zhang, Y. Active DNA Demethylation: Many Roads Lead to Rome. *Nat Rev Mol Cell Biol* **2010**, *11* (9), 607–620. https://doi.org/10.1038/nrm2950.

(168) Weber, A. R.; Krawczyk, C.; Robertson, A. B.; Kuśnierczyk, A.; Vågbø, C. B.; Schuermann, D.; Klungland, A.; Schär, P. Biochemical Reconstitution of TET1–TDG–BER-Dependent Active DNA Demethylation Reveals a Highly Coordinated Mechanism. *Nature Communications* **2016**, *7*, 10806. https://doi.org/10.1038/ncomms10806.

(169) X, W.; Y, Z. TET-Mediated Active DNA Demethylation: Mechanism, Function and Beyond. *Nat Rev Genet* **2017**, *18* (9), 517–534. https://doi.org/10.1038/nrg.2017.33.

(170) Schuermann, D.; Weber, A. R.; Schär, P. Active DNA Demethylation by DNA Repair: Facts and Uncertainties. *DNA Repair* **2016**, *44*, 92–102. https://doi.org/10.1016/j.dnarep.2016.05.013.

(171) Zhu, J.-K. Active DNA Demethylation Mediated by DNA Glycosylases. *Annual Review of Genetics* **2009**, *43* (1), 143–166. https://doi.org/10.1146/annurev-genet-102108-134205.

(172) Kohli, R. M.; Zhang, Y. TET Enzymes, TDG and the Dynamics of DNA Demethylation. *Nature* **2013**, *502* (7472), 472–479. https://doi.org/10.1038/nature12750.

(173) Hashimoto, H.; Hong, S.; Bhagwat, A. S.; Zhang, X.; Cheng, X. Excision of 5-Hydroxymethyluracil and 5-Carboxylcytosine by the Thymine DNA Glycosylase Domain: Its Structural Basis and Implications for Active DNA Demethylation. *Nucleic Acids Research* **2012**, *40* (20), 10203–10214. https://doi.org/10.1093/nar/gks845.

(174) Kunz, C.; Focke, F.; Saito, Y.; Schuermann, D.; Lettieri, T.; Selfridge, J.; Schär, P. Base Excision by Thymine DNA Glycosylase Mediates DNA-Directed Cytotoxicity of 5-Fluorouracil. *PLoS Biol* **2009**, *7* (4). https://doi.org/10.1371/journal.pbio.1000091.

(175) Morgan, M. T.; Bennett, M. T.; Drohat, A. C. Excision of 5-Halogenated Uracils by Human Thymine DNA Glycosylase: Robust Activity for DNA Contexts Other than CpG. *J Biol Chem* **2007**, *282* (38), 27578. https://doi.org/10.1074/jbc.M704253200.

(176) Raiber, E.-A.; Murat, P.; Chirgadze, D. Y.; Beraldi, D.; Luisi, B. F.; Balasubramanian, S. 5-Formylcytosine Alters the Structure of the DNA Double Helix. *Nat. Struct. Mol. Biol.* **2015**, *22* (1), 44–49. https://doi.org/10.1038/nsmb.2936.

(177) Szulik, M. W.; Pallan, P. S.; Nocek, B.; Voehler, M.; Banerjee, S.; Brooks, S.; Joachimiak, A.; Egli, M.; Eichman, B. F.; Stone, M. P. Differential Stabilities and Sequence-Dependent Base Pair Opening Dynamics of Watson-Crick Base Pairs with 5-Hydroxymethylcytosine, 5-Formylcytosine, or 5-Carboxylcytosine. *Biochemistry* **2015**, *54* (5), 1294–1305. https://doi.org/10.1021/bi501534x.

(178) Zhang, L.; Lu, X. (1); Liang, H. (1); Dai, Q. (1); He, C. (1); Lu, J. (2); Luo, C. (2); Jiang, H. (2); Xu, G.-L. (3). Thymine DNA Glycosylase Specifically Recognizes 5-Carboxylcytosine-Modified DNA. *Nature Chemical Biology* **2012**, *8* (4), 328–330. https://doi.org/10.1038/nchembio.914.

(179) Morgan, M. T.; Maiti, A.; Fitzgerald, M. E.; Drohat, A. C. Stoichiometry and Affinity for Thymine DNA Glycosylase Binding to Specific and Nonspecific DNA. *Nucleic Acids Research* **2011**, *39* (6), 2319–2329. https://doi.org/10.1093/nar/gkq1164.

(180) Maiti, A.; Drohat, A. C. Thymine DNA Glycosylase Can Rapidly Excise 5-Formylcytosine and 5-Carboxylcytosine: Potential Implications for Active Demethylation of CpG Sites. *Journal of Biological Chemistry* **2011**, *286* (41), 35334–35338. https://doi.org/10.1074/jbc.C111.284620.

(181) Arab, K.; Park, Y. J.; Lindroth, A. M.; Schäfer, A.; Oakes, C.; Weichenhan, D.; Lukanova, A.; Lundin, E.; Risch, A.; Meister, M.; Dienemann, H.; Dyckhoff, G.; Herold-Mende, C.; Grummt, I.; Niehrs, C.; Plass, C. Long Noncoding RNA TARID Directs Demethylation and Activation of the Tumor Suppressor TCF21 via GADD45A. *Molecular Cell* **2014**, *55* (4), 604–614. https://doi.org/10.1016/j.molcel.2014.06.031.

(182) Cortázar, D.; Kunz, C.; Selfridge, J.; Lettieri, T.; Saito, Y.; MacDougall, E.; Wirz, A.; Schuermann, D.; Jacobs, A. L.; Siegrist, F.; Steinacher, R.; Jiricny, J.; Bird, A.; Schär, P. Embryonic Lethal Phenotype Reveals a Function of TDG in Maintaining Epigenetic Stability. *Nature* **2011**, *470* (7334), 419–423. https://doi.org/10.1038/nature09672.

(183) Wu, H.; Wu, X.; Shen, L.; Zhang, Y. Single-Base Resolution Analysis of Active DNA Demethylation Using Methylase-Assisted Bisulfite Sequencing. *Nature Biotechnology* **2014**, *32* (12), 1231–1240. https://doi.org/10.1038/nbt.3073.

(184) Cortellino, S.; Xu, J.; Sannai, M.; Moore, R.; Caretti, E.; Cigliano, A.; Le Coz, M.; Devarajan, K.; Wessels, A.; Soprano, D.; Abramowitz, L. K.; Bartolomei, M. S.; Rambow, F.; Bassi, M. R.; Bruno, T.; Fanciulli, M.; Renner, C.; Klein-Szanto, A. J.; Matsumoto, Y.; Kobi, D.; Davidson, I.; Alberti, C.; Larue, L.; Bellacosa, A. Thymine DNA Glycosylase Is Essential for Active DNA Demethylation by Linked Deamination-Base Excision Repair. *Cell* **2011**, *146* (1), 67–79. https://doi.org/10.1016/j.cell.2011.06.020.

(185) Lu, X.; Han, D.; Zhao, B. S.; Song, C.-X.; Zhang, L.-S.; Doré, L. C.; He, C. Base-Resolution Maps of 5-Formylcytosine and 5-Carboxylcytosine Reveal Genome-Wide DNA Demethylation Dynamics. *Cell Res* **2015**, *25* (3), 386–389. https://doi.org/10.1038/cr.2015.5.

(186) Booth, M. J.; Raiber, E.-A.; Balasubramanian, S. Chemical Methods for Decoding Cytosine Modifications in DNA. *Chem. Rev.* **2015**, *115* (6), 2240–2254. https://doi.org/10.1021/cr5002904.

(187) Zhu, C.; Gao, Y.; Guo, H.; Xia, B.; Song, J.; Wu, X.; Zeng, H.; Kee, K.; Tang, F.; Yi, C. Single-Cell 5-Formylcytosine Landscapes of Mammalian Early Embryos and ESCs at Single-Base Resolution. *Cell Stem Cell* **2017**, *20* (5), 720-731.e5. https://doi.org/10.1016/j.stem.2017.02.013.

(188) Jin, C.; Felsenfeld, G. Nucleosome Stability Mediated by Histone Variants H3.3 and H2A.Z. *Genes Dev.* **2007**, *21* (12), 1519–1529. https://doi.org/10.1101/gad.1547707.

(189) Li, C.; Delaney, S. Histone H2A Variants Enhance the Initiation of Base Excision Repair in Nucleosomes. *ACS Chemical Biology* **2019**. https://doi.org/10.1021/acschembio.9b00229.

(190) Horikoshi, N.; Arimura, Y.; Taguchi, H.; Kurumizaka, H. Crystal Structures of Heterotypic Nucleosomes Containing Histones H2A.Z and H2A. *Open Biol* **2016**, *6* (6). https://doi.org/10.1098/rsob.160127.

(191) Talbert, P. B.; Henikoff, S. Histone Variants — Ancient Wrap Artists of the Epigenome. *Nature Reviews Molecular Cell Biology* **2010**, *11* (4), 264–275. https://doi.org/10.1038/nrm2861.

(192) Chen, P.; Zhao, J.; Wang, Y.; Wang, M.; Long, H.; Liang, D.; Huang, L.; Wen, Z.; Li, W.; Li, X.; Feng, H.; Zhao, H.; Zhu, P.; Li, M.; Wang, Q.; Li, G. H3.3 Actively Marks Enhancers and Primes Gene Transcription via Opening Higher-Ordered Chromatin. *Genes Dev* **2013**, *27* (19), 2109–2124. https://doi.org/10.1101/gad.222174.113.

(193) Pradham, S. K.; Su, T.; Yen, L.; Jacquet, K.; Huang, C.; Cote, J.; Kurdistani, S. K.; Carey, M. F. EP400 Deposits H3.3 into Promoters and Enhancers during Gene Activation. *Molecular Cell* **2016**, *61* (1), 27–38. https://doi.org/10.1016/j.molcel.2015.10.039.

(194) Gévry, N.; Hardy, S.; Jacques, P.-É.; Laflamme, L.; Svotelis, A.; Robert, F.; Gaudreau, L. Histone H2A.Z Is Essential for Estrogen Receptor Signaling. *Genes Dev* **2009**, *23* (13), 1522–1533. https://doi.org/10.1101/gad.1787109.

(195) Fan, J. Y.; Gordon, F.; Luger, K.; Hansen, J. C.; Tremethick, D. J. The Essential Histone Variant H2A.Z Regulates the Equilibrium between Different Chromatin Conformational States. *Nature Structural Biology* **2002**, *9* (3), 172.

(196) Valdés-Mora, F.; Gould, C. M.; Colino-Sanguino, Y.; Qu, W.; Song, J. Z.; Taylor, K. M.; Buske, F. A.; Statham, A. L.; Nair, S. S.; Armstrong, N. J.; Kench, J. G.; Lee, K. M. L.; Horvath, L. G.; Qiu, M.; Ilinykh, A.; Yeo-Teh, N. S.; Gallego-Ortega, D.; Stirzaker, C.; Clark, S. J. Acetylated Histone Variant H2A.Z Is Involved in the Activation of Neo-Enhancers in Prostate Cancer. *Nature Communications* **2017**, *8* (1), 1346. https://doi.org/10.1038/s41467-017-01393-8.

(197) Neri, F.; Incarnato, D.; Krepelova, A.; Rapelli, S.; Anselmi, F.; Parlato, C.; Medana, C.; Dal Bello, F.; Olivero, S. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Reports* **2015**, *10* (5), 674–683. https://doi.org/10.1016/j.celrep.2015.01.008.

(198) Henry, R. A.; Mancuso, P.; Kuo, Y.-M.; Tricarico, R.; Tini, M.; Cole, P. A.; Bellacosa, A.; Andrews, A. J. Interaction with the DNA Repair Protein Thymine DNA Glycosylase Regulates Histone Acetylation by P300. *Biochemistry* **2016**, *55* (49), 6766–6775. https://doi.org/10.1021/acs.biochem.6b00841.

(199) Mohan, R. D.; Litchfield, D. W.; Torchia, J.; Tini, M. Opposing Regulatory Roles of Phosphorylation and Acetylation in DNA Mispair Processing by Thymine DNA Glycosylase. *Nucleic Acids Research* **2010**, *38* (4), 1135–1148. https://doi.org/10.1093/nar/gkp1097.

(200) Nair, S. J.; Yang, L.; Meluzzi, D.; Oh, S.; Yang, F.; Friedman, M. J.; Wang, S.; Suter, T.; Alshareedah, I.; Gamliel, A.; Ma, Q.; Zhang, J.; Hu, Y.; Tan, Y.; Ohgi, K. A.; Jayani, R. S.; Banerjee, P. R.; Aggarwal, A. K.; Rosenfeld, M. G. Phase Separation of Ligand-Activated Enhancers Licenses Cooperative Chromosomal Enhancer Assembly. *Nature Structural & Molecular Biology* **2019**, *26* (3), 193–203. https://doi.org/10.1038/s41594-019-0190-5.

(201) Sabari, B. R.; Dall'Agnese, A.; Boija, A.; Klein, I. A.; Coffey, E. L.; Shrinivas, K.; Abraham, B. J.; Hannett, N. M.; Zamudio, A. V.; Manteiga, J. C.; Li,

C. H.; Guo, Y. E.; Day, D. S.; Schuijers, J.; Vasile, E.; Malik, S.; Hnisz, D.; Lee, T. I.; Cisse, I. I.; Roeder, R. G.; Sharp, P. A.; Chakraborty, A. K.; Young, R. A. Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control. *Science* **2018**, *361* (6400). https://doi.org/10.1126/science.aar3958.

(202) Zhang, Y.; Zhang, D.; Li, Q.; Liang, J.; Sun, L.; Yi, X.; Chen, Z.; Yan, R.; Xie, G.; Li, W.; Liu, S.; Xu, B.; Li, L.; Yang, J.; He, L.; Shang, Y. Nucleation of DNA Repair Factors by FOXA1 Links DNA Demethylation to Transcriptional Pioneering. *Nature Genetics* **2016**, *48* (9), 1003–1013. https://doi.org/10.1038/ng.3635.

(203) Hurtado, A.; Holmes, K. A.; Ross-Innes, C. S.; Schmidt, D.; Carroll, J. S. FOXA1 Is a Key Determinant of Estrogen Receptor Function and Endocrine Response. *Nature Genetics* **2011**, *43* (1), 27–33. https://doi.org/10.1038/ng.730.

(204) Müller, M. M.; Fierz, B.; Bittova, L.; Liszczak, G.; Muir, T. W. A Two-State Activation Mechanism Controls the Histone Methyltransferase Suv39h1. *Nature Chemical Biology* **2016**, *12* (3), 188–193. https://doi.org/10.1038/nchembio.2008.

(205) Azzaz, A. M.; Vitalini, M. W.; Thomas, A. S.; Price, J. P.; Blacketer, M. J.; Cryderman, D. E.; Zirbel, L. N.; Woodcock, C. L.; Elcock, A. H.; Wallrath, L. L.; Shogren-Knaak, M. A. Human Heterochromatin Protein 1α Promotes Nucleosome Associations That Drive Chromatin Condensation. *J Biol Chem* **2014**, *289* (10), 6850–6861. https://doi.org/10.1074/jbc.M113.512137.

(206) Lindahl, T.; Barnes, D. E. Repair of Endogenous DNA Damage. *Cold Spring Harb Symp Quant Biol* **2000**, *65*, 127–134. https://doi.org/10.1101/sqb.2000.65.127.

(207) Luzzietti, N.; Brutzer, H.; Klaue, D.; Schwarz, F. W.; Staroske, W.; Clausing, S.; Seidel, R. Efficient Preparation of Internally Modified Single-Molecule Constructs Using Nicking Enzymes. *Nucleic Acids Res* **2011**, *39* (3), e15–e15. https://doi.org/10.1093/nar/gkq1004.

(208) Luhnsdorf, B.; Kitsera, N.; Warken, D.; Lingg, T.; Epe, B.; Khobta, A. Generation of Reporter Plasmids Containing Defined Base Modifications in the DNA Strand of Choice. *ResearchGate* **2012**, *425* (1), 47–53. https://doi.org/10.1016/j.ab.2012.03.001.

(209) Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57* (1), 64–75. https://doi.org/10.1016/j.ymeth.2012.03.011.

(210) Biyani, M.; Nishigaki, K. Structural Characterization of Ultra-Stable Higher-Ordered Aggregates Generated by Novel Guanine-Rich DNA Sequences. *Gene* **2005**, *364*, 130–138. https://doi.org/10.1016/j.gene.2005.05.041.

(211) Hansen, J. C. Conformational Dynamics of the Chromatin Fiber in Solution: Determinants, Mechanisms, and Functions. *Annu. Rev. Biophys. Biomol. Struct.* **2002**, *31* (1), 361–392. https://doi.org/10.1146/annurev.biophys.31.101101.140858.

(212) Nakanishi, S.; Prasad, R.; Wilson, S. H.; Smerdon, M. Different Structural States in Oligonucleosomes Are Required for Early versus Late Steps of Base Excision Repair. *Nucleic Acids Res* **2007**, *35* (13), 4313–4321. https://doi.org/10.1093/nar/gkm436.

(213) Muthurajan, U.; Mattiroli, F.; Bergeron, S.; Zhou, K.; Gu, Y.; Chakravarthy, S.; Dyer, P.; Irving, T.; Luger, K. In Vitro Chromatin Assembly: Strategies and Quality Control. *Meth. Enzymol.* **2016**, *573*, 3–41. https://doi.org/10.1016/bs.mie.2016.01.002.

(214) Wiebauer, K.; Jiricny, J. Mismatch-Specific Thymine DNA Glycosylase and DNA Polymerase Beta Mediate the Correction of G.T Mispairs in Nuclear Extracts from Human Cells. *Proc Natl Acad Sci U S A* **1990**, *87* (15), 5842–5845.

(215) Drohat, A. C.; Coey, C. T. Role of Base Excision "Repair" Enzymes in Erasing Epigenetic Marks from DNA. *Chem Rev* **2016**, *116* (20), 12711–12729. https://doi.org/10.1021/acs.chemrev.6b00191.

(216) Bernardo, G. M.; Keri, R. A. FOXA1: A Transcription Factor with Parallel Functions in Development and Cancer. *Biosci Rep* **2012**, *32* (2), 113–130. https://doi.org/10.1042/BSR20110046.

(217) Laganière, J.; Deblois, G.; Lefebvre, C.; Bataille, A. R.; Robert, F.; Giguère, V. Location Analysis of Estrogen Receptor α Target Promoters Reveals That

FOXA1 Defines a Domain of the Estrogen Response. *PNAS* **2005**, *102* (33), 11651–11656. https://doi.org/10.1073/pnas.0505575102.

(218) Lupien, M.; Eeckhoute, J.; Meyer, C. A.; Wang, Q.; Zhang, Y.; Li, W.; Carroll, J. S.; Liu, X. S.; Brown, M. FoxA1 Translates Epigenetic Signatures into Enhancer-Driven Lineage-Specific Transcription. *Cell* **2008**, *132* (6), 958–970. https://doi.org/10.1016/j.cell.2008.01.018.

(219) Adams, E. J.; Karthaus, W. R.; Hoover, E.; Liu, D.; Gruet, A.; Zhang, Z.; Hyunuwoo, C.; DiLoreto, R.; Chhangawala, S.; Liu, Y.; Rohit, B.; Christina, S. L.; Sawyers, C. L. FOXA1 Mutations Alter Pioneering Activity, Differentiation and Prostate Cancer Phenotypes | Nature. *Nature* **2019**, *571*, 408–412.

(220) Pehrson, J. R. Probing the Conformation of Nucleosome Linker DNA in Situ with Pyrimidine Dimer Formation. *J. Biol. Chem.* **1995**, *270* (38), 22440–22444.

(221) Calo, E.; Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell* **2013**, *49* (5), 825–837. https://doi.org/10.1016/j.molcel.2013.01.038.

(222) Li, Z.; Gadue, P.; Chen, K.; Jiao, Y.; Tuteja, G.; Schug, J.; Li, W.; Kaestner, K. H. Foxa2 and H2A.Z Mediate Nucleosome Depletion during Embryonic Stem Cell Differentiation. *Cell* **2012**, *151* (7), 1608–1616. https://doi.org/10.1016/j.cell.2012.11.018.

(223) Muthurajan, U.; Mattiroli, F.; Bergeron, S.; Zhou, K.; Gu, Y.; Chakravarthy, S.; Dyer, P.; Irving, T.; Luger, K. Chapter One - In Vitro Chromatin Assembly: Strategies and Quality Control. In *Methods in Enzymology*; Marmorstein, R., Ed.; Enzymes of Epigenetics, Part A; Academic Press, 2016; Vol. 573, pp 3–41. https://doi.org/10.1016/bs.mie.2016.01.002.

(224) Zaret, K. S.; Stevens, K. Expression of a Highly Unstable and Insoluble Transcription Factor in Escherichia Coli: Purification and Characterization of the Fork Head Homolog HNF3 Alpha. *Protein Expr. Purif.* **1995**, *6* (6), 821–825. https://doi.org/10.1006/prep.1995.0014.

(225) Hieb, A. R.; D'Arcy, S.; Kramer, M. A.; White, A. E.; Luger, K. Fluorescence Strategies for High-Throughput Quantification of Protein Interactions. *Nucleic Acids Research* **2011**, *40* (5), 1–13. https://doi.org/10.1093/nar/gkr1045.

(226) Eagen, K. P. Principles of Chromosome Architecture Revealed by Hi-C. *Trends Biochem Sci* **2018**, *43* (6), 469–478. https://doi.org/10.1016/j.tibs.2018.03.006.

(227) Sjolund, A.; Senejani, A.; Sweasy, J. MBD4 and TDG: Multifaceted DNA Glycosylases With Ever Expanding Biological Roles. *Mutat Res* **2013**, *0*, 12–25. https://doi.org/10.1016/j.mrfmmm.2012.11.001.

(228) The C-terminal Domain Is the Primary Determinant of Histone H1 Binding to Chromatin in Vivo https://www.jbc.org/content/279/19/20028.long (accessed Aug 18, 2020).

(229) Allan, J.; Mitchell, T.; Harborne, N.; Bohm, L.; Crane-Robinson, C. Roles of H1 Domains in Determining Higher Order Chromatin Structure and H1 Location. *Journal of Molecular Biology* **1986**, *187* (4), 591–601. https://doi.org/10.1016/0022-2836(86)90337-2.

(230) Izzo, A.; Kamieniarz-Gdula, K.; Ramírez, F.; Noureen, N.; Kind, J.; Manke, T.; van Steensel, B.; Schneider, R. The Genomic Landscape of the Somatic Linker Histone Subtypes H1.1 to H1.5 in Human Cells. *Cell Reports* **2013**, *3* (6), 2142–2154. https://doi.org/10.1016/j.celrep.2013.05.003.

(231) Yang, S.-M.; Kim, B. J.; Toro, L. N.; Skoultchi, A. I. H1 Linker Histone Promotes Epigenetic Silencing by Regulating Both DNA Methylation and Histone H3 Methylation. *PNAS* **2013**, *110* (5), 1708–1713. https://doi.org/10.1073/pnas.1213266110.

(232) Maeshima, K.; Rogge, R.; Tamura, S.; Joti, Y.; Hikima, T.; Szerlong, H.; Krause, C.; Herman, J.; Seidel, E.; DeLuca, J.; Ishikawa, T.; Hansen, J. C. Nucleosomal Arrays Self-Assemble into Supramolecular Globular Structures Lacking 30-Nm Fibers. *EMBO J.* **2016**, *35* (10), 1115–1132. https://doi.org/10.15252/embj.201592660.

(233) Blacketer, M. J.; Feely, S. J.; Shogren-Knaak, M. A. Nucleosome Interactions and Stability in an Ordered Nucleosome Array Model System. *J. Biol. Chem.* **2010**, *285* (45), 34597–34607. https://doi.org/10.1074/jbc.M110.140061.

(234) Smet-Nocca, C.; Wieruszeski, J.-M.; Chaar, V.; Leroy, A.; Benecke, A. The Thymine-DNA Glycosylase Regulatory Domain: Residual Structure and DNA Binding. *Biochemistry* **2008**, *47* (25), 6519–6530. https://doi.org/10.1021/bi7022283.

(235) Barbera, A. J.; Chodaparambil, J. V.; Kelley-Clarke, B.; Joukov, V.; Walter, J. C.; Luger, K.; Kaye, K. M. The Nucleosomal Surface as a Docking Station for Kaposi's Sarcoma Herpesvirus LANA. *Science* **2006**, *311* (5762), 856–861. https://doi.org/10.1126/science.1120541.

(236) Ishida, T.; Kinoshita, K. PrDOS: Prediction of Disordered Protein Regions from Amino Acid Sequence. *Nucleic Acids Res* **2007**, *35* (suppl_2), W460–W464. https://doi.org/10.1093/nar/gkm363.

(237) Li, Z.; Gu, T.-P.; Weber, A. R.; Shen, J.-Z.; Li, B.-Z.; Xie, Z.-G.; Yin, R.; Guo, F.; Liu, X.; Tang, F.; Wang, H.; Schär, P.; Xu, G.-L. Gadd45a Promotes DNA Demethylation through TDG. *Nucleic Acids Research* **2015**, *43* (8), 3986–3997. https://doi.org/10.1093/nar/gkv283.

(238) Barreto, G. (1); Schäfer, A. (1); Swaminathan, S. k. (1); Handa, V. (1); Döderlein, G. (1); Maltry, N. (1); Wu, W. (1, 3); Niehrs, C. (1); Marhold, J. (2); Stach, D. (2); Lyko, F. (2). Gadd45a Promotes Epigenetic Gene Activation by Repair-Mediated DNA Demethylation. *Nature* **2007**, *445* (7128), 671–675. https://doi.org/10.1038/nature05515.

(239) Li, L.; Chen, K.; Wu, Y.; Long, Q.; Zhao, D.; Ma, B.; Pei, D.; Liu, X. Gadd45a Opens up the Promoter Regions of MiR-295 Facilitating Pluripotency Induction. *Cell Death Dis* **2017**, *8* (10), e3107–e3107. https://doi.org/10.1038/cddis.2017.497.

(240) Niehrs, C.; Schäfer, A. Active DNA Demethylation by Gadd45 and DNA Repair. *Trends in Cell Biology* **2012**, *22* (4), 220–227. https://doi.org/10.1016/j.tcb.2012.01.002.

(241) Carrier, F.; Georgel, P. T.; Pourquier, P.; Blake, M.; Kontny, H. U.; Antinore, M. J.; Gariboldi, M.; Myers, T. G.; Weinstein, J. N.; Pommier, Y.; Fornace, A. J. Gadd45, a P53-Responsive Stress Protein, Modifies DNA Accessibility on Damaged Chromatin. *Molecular and Cellular Biology* **1999**, *19* (3), 1673–1685. https://doi.org/10.1128/MCB.19.3.1673.

(242) Carter, R. J.; Parsons, J. L. Base Excision Repair, a Pathway Regulated by Posttranslational Modifications. *Mol Cell Biol* **2016**, *36* (10), 1426–1437. https://doi.org/10.1128/MCB.00030-16.

(243) Larson, A. G.; Elnatan, D.; Keenen, M. M.; Trnka, M. J.; Johnston, J. B.; Burlingame, A. L.; Agard, D. A.; Redding, S.; Narlikar, G. J. Liquid Droplet Formation by HP1α Suggests a Role for Phase Separation in Heterochromatin. *Nature* **2017**, *547* (7662), 236–240. https://doi.org/10.1038/nature22822.

(244) Crabtree, M.; Nott, T. These Organelles Have No Membranes. *The Scientist Magazine*®. November 30, 2018.

(245) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176* (3), 419–434. https://doi.org/10.1016/j.cell.2018.12.035.

(246) Falahati, H.; Haji-Akbari, A. Thermodynamically Driven Assemblies and Liquid-Liquid Phase Separations in Biology. *Soft Matter* **2019**, *15* (6), 1135–1154. https://doi.org/10.1039/c8sm02285b.

(247) Zhou, H.; Song, Z.; Zhong, S.; Zuo, L.; Qi, Z.; Qu, L.-J.; Lai, L. Mechanism of DNA-Induced Phase Separation for Transcriptional Repressor VRN1. *Angewandte Chemie* **2019**, *131* (15), 4912–4916. https://doi.org/10.1002/ange.201810373.

(248) Das, C.; Hizume, K.; Batta, K.; Kumar, B. R. P.; Gadad, S. S.; Ganguly, S.; Lorain, S.; Verreault, A.; Sadhale, P. P.; Takeyasu, K.; Kundu, T. K. Transcriptional Coactivator PC4, a Chromatin-Associated Protein, Induces Chromatin Condensation. *MCB* **2006**, *26* (22), 8303–8315. https://doi.org/10.1128/MCB.00887-06.

(249) Wang, L.; Hu, M.; Zuo, M.-Q.; Zhao, J.; Wu, D.; Huang, L.; Wen, Y.; Li, Y.; Chen, P.; Bao, X.; Dong, M.-Q.; Li, G.; Li, P. Rett Syndrome-Causing Mutations Compromise MeCP2-Mediated Liquid–Liquid Phase Separation of Chromatin. *Cell Research* **2020**, *30* (5), 393–407. https://doi.org/10.1038/s41422-020-0288-7.

(250) Le Treut, G.; Kepes, F.; Orland, H. Phase Behavior of DNA in the Presence of DNA-Binding Proteins. *Biophysical Journal* **2016**, *110* (1), 51–62. https://doi.org/10.1016/j.bpj.2015.10.027.

(251) Zaret, K. S.; Carroll, J. S. Pioneer Transcription Factors: Establishing Competence for Gene Expression. *Genes Dev* **2011**, *25* (21), 2227–2241. https://doi.org/10.1101/gad.176826.111.

(252) Nair, S. J.; Yang, L.; Meluzzi, D.; Oh, S.; Yang, F.; Friedman, M. J.; Wang, S.; Suter, T.; Alshareedah, I.; Gamliel, A.; Ma, Q.; Zhang, J.; Hu, Y.; Tan, Y.; Ohgi, K. A.; Jayani, R. S.; Banerjee, P. R.; Aggarwal, A. K.; Rosenfeld, M. G. Phase Separation of Ligand-Activated Enhancers Licenses Cooperative Chromosomal Enhancer Assembly. *Nat Struct Mol Biol* **2019**, *26* (3), 193–203. https://doi.org/10.1038/s41594-019-0190-5.

(253) Carruthers, L. M.; Hansen, J. C. The Core Histone N Termini Function Independently of Linker Histones during Chromatin Condensation. *J. Biol. Chem.* **2000**, *275* (47), 37285–37290. https://doi.org/10.1074/jbc.M006801200.

(254) Chu, C.; Qu, K.; Zhong, F. L.; Artandi, S. E.; Chang, H. Y. Genomic Maps of Long Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. *Mol. Cell* **2011**, *44* (4), 667–678. https://doi.org/10.1016/j.molcel.2011.08.027.

(255) Tsai, M.-C.; Manor, O.; Wan, Y.; Mosammaparast, N.; Wang, J. K.; Lan, F.; Shi, Y.; Segal, E.; Chang, H. Y. Long Noncoding RNA as Modular Scaffold of Histone Modification Complexes. *Science* **2010**, *329* (5992), 689–693. https://doi.org/10.1126/science.1192002.

(256) Wang, K. C.; Chang, H. Y. Molecular Mechanisms of Long Noncoding RNAs. *Mol Cell* **2011**, *43* (6), 904–914. https://doi.org/10.1016/j.molcel.2011.08.018.

(257) Rinn, J. L.; Chang, H. Y. Genome Regulation by Long Noncoding RNAs. *Annu Rev Biochem* **2012**, *81*. https://doi.org/10.1146/annurev-biochem-051410-092902.

(258) Pertea, M. The Human Transcriptome: An Unfinished Story. *Genes (Basel)* **2012**, *3* (3), 344–360. https://doi.org/10.3390/genes3030344.

(259) Engreitz, J. M.; Pandya-Jones, A.; McDonel, P.; Shishkin, A.; Sirokman, K.; Surka, C.; Kadri, S.; Xing, J.; Goren, A.; Lander, E. S.; Plath, K.; Guttman, M. The Xist LncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **2013**, *341* (6147), 1237973–1237973. https://doi.org/10.1126/science.1237973.

(260) Andresini, O.; Rossi, M. N.; Matteini, F.; Petrai, S.; Santini, T.; Maione, R. The Long Non-Coding RNA Kcnq1ot1 Controls Maternal P57 Expression in Muscle Cells by Promoting H3K27me3 Accumulation to an Intragenic MyoD-Binding Region. *Epigenetics & Chromatin* **2019**, *12* (1), 8. https://doi.org/10.1186/s13072-019-0253-1.

(261) Li, Y. RNA-DNA Triplex Formation by Long Noncoding RNAs. *Cell Chemical Biology* 9.

(262) Margueron, R.; Reinberg, D. The Polycomb Complex PRC2 and Its Mark in Life. *Nature* **2011**, *469* (7330), 343–349. https://doi.org/10.1038/nature09784.

(263) Son, J.; Shen, S. S.; Margueron, R.; Reinberg, D. Nucleosome-Binding Activities within JARID2 and EZH1 Regulate the Function of PRC2 on Chromatin. *Genes Dev.* **2013**, *27* (24), 2663–2677. https://doi.org/10.1101/gad.225888.113.

(264) Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Drury III, W. J.; Voigt, P.; Martin, S. R.; Taylor, W. R.; De Marco, V.; Pirrotta, V.; Reinberg, D.; Gamblin, S. J. Role of the Polycomb Protein EED in the Propagation of Repressive Histone Marks. *Nature* **2009**, *461* (7265), 762–767. https://doi.org/10.1038/nature08398.

(265) Kadoch, C.; Copeland, R. A.; Keilhack, H. PRC2 and SWI/SNF Chromatin Remodeling Complexes in Health and Disease. *Biochemistry* **2016**, *55* (11), 1600–1614. https://doi.org/10.1021/acs.biochem.5b01191.

(266) Sato, T.; Kaneda, A.; Tsuji, S.; Isagawa, T.; Yamamoto, S.; Fujita, T.; Yamanaka, R.; Tanaka, Y.; Nukiwa, T.; Marquez, V. E.; Ishikawa, Y.; Ichinose, M.; Aburatani, H. PRC2 Overexpression and PRC2-Target Gene Repression Relating to Poorer Prognosis in Small Cell Lung Cancer. *Scientific Reports* **2013**, *3* (1). https://doi.org/10.1038/srep01911.

(267) Fatemi, R. P.; Velmeshev, D.; Faghihi, M. A. De-Repressing LncRNA-Targeted Genes to Upregulate Gene Expression: Focus on Small Molecule Therapeutics. *Molecular Therapy - Nucleic Acids* **2014**, *3*, e196. https://doi.org/10.1038/mtna.2014.45.

(268) Kim, K. H.; Roberts, C. W. M. Targeting EZH2 in Cancer. *Nature Medicine* **2016**, *22* (2), 128–134. https://doi.org/10.1038/nm.4036.

(269) Bachmann, I. M.; Halvorsen, O. J.; Collett, K.; Stefansson, I. M.; Straume, O.; Haukaas, S. A.; Salvesen, H. B.; Otte, A. P.; Akslen, L. A. EZH2 Expression Is Associated With High Proliferation Rate and Aggressive Tumor Subgroups in Cutaneous Melanoma and Cancers of the Endometrium, Prostate, and Breast. *JCO* **2006**, *24* (2), 268–273. https://doi.org/10.1200/JCO.2005.01.5180.

(270) Varambally, S.; Dhanasekaran, S. M.; Zhou, M.; Barrette, T. R.; Kumar-Sinha, C.; Sanda, M. G.; Ghosh, D.; Pienta, K. J.; Sewalt, R. G. A. B.; Otte, A. P.; Rubin, M. A.; Chinnaiyan, A. M. The Polycomb Group Protein EZH2 Is Involved in Progression of Prostate Cancer. *Nature* **2002**, *419* (6907), 624–629. https://doi.org/10.1038/nature01075.

(271) Bracken Adrian P.; Helin, K. Polycomb Group Proteins: Navigators of Lineage Pathways Led Astray in Cancer. *Nature Reviews Cancer* **2009**, *9*, 773–784. https://doi.org/10.1038/nrc2736.

(272) T, E.; Aj, C.; J, S.; Ce, H.-C.; C, B.; Av, J.; K, W.; K, Z.; Fm, R.; A, R.; A, H.; Hg, D.; A, D.; F, C.; D, O.; J, B.; Fh, G.; Nc, C. Inactivating Mutations of the Histone Methyltransferase Gene EZH2 in Myeloid Disorders. *Nat Genet* **2010**, *42* (8), 722–726. https://doi.org/10.1038/ng.621.

(273) Davidovich, C.; Zheng, L.; Goodrich, K. J.; Cech, T. R. Promiscuous RNA Binding by Polycomb Repressive Complex 2. *Nat Struct Mol Biol* **2013**, *20* (11), 1250–1257. https://doi.org/10.1038/nsmb.2679.

(274) Davidovich, C.; Wang, X.; Cifuentes-Rojas, C.; Goodrich, K. J.; Gooding, A. R.; Lee, J. T.; Cech, T. R. Toward a Consensus on the Binding Specificity and Promiscuity of PRC2 for RNA. *Molecular Cell* **2015**, *57* (3), 552–558. https://doi.org/10.1016/j.molcel.2014.12.017.

(275) Wang, X.; Goodrich, K. J.; Gooding, A. R.; Naeem, H.; Archer, S.; Paucek, R. D.; Youmans, D. T.; Cech, T. R.; Davidovich, C. Targeting of Polycomb Repressive Complex 2 to RNA by Short Repeats of Consecutive Guanines. *Molecular Cell* **2017**, *65* (6), 1056-1067.e5. https://doi.org/10.1016/j.molcel.2017.02.003.

(276) Beltran, M.; Yates, C. M.; Skalska, L.; Dawson, M.; Reis, F. P.; Viiri, K.; Fisher, C. L.; Sibley, C. R.; Foster, B. M.; Bartke, T.; Ule, J.; Jenner, R. G. The Interaction of PRC2 with RNA or Chromatin Is Mutually Antagonistic. *Genome Res.* **2016**, *26* (7), 896–907. https://doi.org/10.1101/gr.197632.115.

(277) Wang, X.; Paucek, R. D.; Gooding, A. R.; Brown, Z. Z.; Ge, E. J.; Muir, T. W.; Cech, T. R. Molecular Analysis of PRC2 Recruitment to DNA in Chromatin and Its Inhibition by RNA. *Nat Struct Mol Biol* **2017**, *24* (12), 1028–1038. https://doi.org/10.1038/nsmb.3487.

(278) Kaneko, S.; Son, J.; Bonasio, R.; Shen, S. S.; Reinberg, D. Nascent RNA Interaction Keeps PRC2 Activity Poised and in Check. *Genes Dev.* **2014**, *28* (18), 1983–1988. https://doi.org/10.1101/gad.247940.114.

(279) Wu, L.; Murat, P.; Matak-Vinkovic, D.; Murrell, A.; Balasubramanian, S. Binding Interactions between Long Noncoding RNA HOTAIR and PRC2 Proteins. *Biochemistry* **2013**, *52* (52), 9519–9527. https://doi.org/10.1021/bi401085h.

(280) Vater, A.; Klussmann, S. Turning Mirror-Image Oligonucleotides into Drugs: The Evolution of Spiegelmer® Therapeutics. *Drug Discovery Today* **2015**, *20* (1), 147–155. https://doi.org/10.1016/j.drudis.2014.09.004.

(281) Klußmann, S.; Nolte, A.; Bald, R.; Erdmann, V. A.; Fürste, J. P. Mirror-Image RNA That Binds D-Adenosine. *Nature Biotechnology* **1996**, *14* (9), 1112–1115. https://doi.org/10.1038/nbt0996-1112.

(282) Long, Y.; Bolanos, B.; Gong, L.; Liu, W.; Goodrich, K. J.; Yang, X.; Chen, S.; Gooding, A. R.; Maegley, K. A.; Gajiwala, K. S.; Brooun, A.; Cech, T. R.; Liu, X.

Conserved RNA-Binding Specificity of Polycomb Repressive Complex 2 Is Achieved by Dispersed Amino Acid Patches in EZH2. *eLife* *6*. https://doi.org/10.7554/eLife.31558.

(283) Rinn, J. L.; Kertesz, M.; Wang, J. K.; Squazzo, S. L.; Xu, X.; Brugmann, S. A.; Goodnough, L. H.; Helms, J. A.; Farnham, P. J.; Segal, E.; Chang, H. Y. Functional Demarcation of Active and Silent Chromatin Domains in Human HOX Loci by Noncoding RNAs. *Cell* **2007**, *129* (7), 1311–1323. https://doi.org/10.1016/j.cell.2007.05.022.

(284) Gupta, R. A.; Shah, N.; Wang, K. C.; Kim, J.; Horlings, H. M.; Wong, D. J.; Tsai, M.-C.; Hung, T.; Argani, P.; Rinn, J. L.; Wang, Y.; Brzoska, P.; Kong, B.; Li, R.; West, R. B.; van de Vijver, M. J.; Sukumar, S.; Chang, H. Y. Long Non-Coding RNA HOTAIR Reprograms Chromatin State to Promote Cancer Metastasis. *Nature* **2010**, *464* (7291), 1071–1076. https://doi.org/10.1038/nature08975.

(285) Gutschner, T.; Diederichs, S. The Hallmarks of Cancer: A Long Non-Coding RNA Point of View. *RNA Biology* **2012**, *9* (6), 703–719. https://doi.org/10.4161/rna.20481.

(286) Schmitt, A. M.; Chang, H. Y. Long Noncoding RNAs in Cancer Pathways. *Cancer Cell* **2016**, *29*, 452–463. https://doi.org/10.1016/j.ccell.2016.03.010.

(287) Cifuentes-Rojas, C.; Hernandez, A. J.; Sarma, K.; Lee, J. T. Regulatory Interactions between RNA and Polycomb Repressive Complex 2. *Mol. Cell* **2014**, *55* (2), 171–185. https://doi.org/10.1016/j.molcel.2014.05.009.

(288) Hauser, N. C.; Martinez, R.; Jacob, A.; Rupp, S.; Hoheisel, J. D.; Matysiak, S. Utilising the Left-Helical Conformation of L-DNA for Analysing Different Marker Types on a Single Universal Microarray Platform. *Nucleic Acids Res* **2006**, *34* (18), 5101–5111. https://doi.org/10.1093/nar/gkl671.

(289) Heterochiral DNA Strand-Displacement Circuits | Journal of the American Chemical Society https://pubs.acs.org/doi/10.1021/jacs.7b10038 (accessed Aug 17, 2020).

(290) Jankowsky, E.; Harris, M. E. Specificity and Nonspecificity in RNA-Protein Interactions. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (9), 533–544. https://doi.org/10.1038/nrm4032.

(291) Meas, R.; Meas, J.; Smerdon, M. J. Nucleosomes Regulate Base Excision Repair in Chromatin. *Mutation Research/Reviews in Mutation Research* **2017**, *780*, 29–36. https://doi.org/10.1016/j.mrrev.2017.10.002.

(292) Berquist, B. R.; McNeill, D. R.; Wilson III, D. M. Characterization of Abasic Endonuclease Activity of Human Ape1 on Alternative Substrates, as Well as Effects of ATP and Sequence Context on AP Site Incision. *Journal of Molecular Biology* **2009**, *379* (1), 17–27. https://doi.org/10.1016/j.jmb.2008.03.053.

(293) Lee, A. J.; Warshaw, D. M.; Wallace, S. S. Insights into the Glycosylase Search for Damage from Single-Molecule Fluorescence Microscopy. *DNA Repair* **2014**, *20*, 23–31. https://doi.org/10.1016/j.dnarep.2014.01.007.

(294) Friedman, J. I.; Stivers, J. T. Detection of Damaged DNA Bases by DNA Glycosylase Enzymes. *Biochemistry* **2010**, *49* (24), 4957–4967. https://doi.org/10.1021/bi100593a.

(295) Venkatesh, S.; Workman, J. L. Histone Exchange, Chromatin Structure and the Regulation of Transcription. *Nat. Rev. Mol. Cell Biol.* **2015**, *16* (3), 178–189. https://doi.org/10.1038/nrm3941.

(296) Maiti, A.; Drohat, A. C. Dependence of Substrate Binding and Catalysis on PH, Ionic Strength, and Temperature for Thymine DNA Glycosylase: Insights into Recognition and Processing of G·T Mispairs. *DNA Repair* **2011**, *10* (5), 545–553. https://doi.org/10.1016/j.dnarep.2011.03.004.

A) APPENDIX

CHAPTER 2 SUPPLEMENTARY DATA

A.1 Supplementary Text

In order to gain insight into mechanisms of BER on chromatin, several previous studies have measured the combined activities of UDG and APE1 on mononucleosomes containing precisely positioned dU residues [122], [154]. During BER, UDG specifically recognizes and removes uracil bases in DNA resulting in formation of an abasic site, which is subsequently cleaved by APE1 in order to generate a nicked substrate for downstream processing [135]. In general, the catalytic activities of both enzymes are reduced on nucleosome substrates as compared to naked DNA [134], [160], [291]. This is attributed to reduced flexibility of the DNA when associated with histones, as well as steric hindrance by the histones globular and tail domains when the DNA is folded into nucleosomes [160]. Furthermore, the extent of inhibition is dependent on the rotational setting of the lesion, such that dU residues having their phosphate backbones oriented away from the histone octamer surface (referred to as outward facing) are converted to product more rapidly than those facing inward.

Despite these prior investigations, however, a detailed analysis of the impact of higher order chromatin folding on BER has not been reported using site-specifically damaged systems. This is because the vast majority of studies have been carried out on damaged mononucleosomes, which are incapable of folding into native chromatin structures. We are aware of only a single in vitro study addressing BER within the context

of folded chromatin (Nakanishi *et al.*; [212]). Despite the significance of this study, the authors utilized stochastically damaged nucleosome arrays as substrates, thereby precluding precise quantification of BER efficiency at specific damage sites. Therefore, whether conclusions based on these previous systems accurately reflect BER in vivo requires further consideration. Thus, having assembled site-specifically damaged nucleosome arrays, we set out to characterize the influence of the local chromatin environment, and in particular, the degree of chromatin compaction, on the initial steps of BER.

Using the two dU containing nucleosome arrays prepared above (**12-NCP-dU49** and **12-NCP-dU88**, Figure 2.1), we assessed the efficiency of uracil removal by the combined activities of UDG and APE1 within the context of multiple distinct chromatin states. In order to facilitate analysis of the UDG/APE1 digestion products by gel electrophoresis, the corresponding dU containing 12×601 DNA templates (**12- 601-dU49** and **12-601-dU88**) were assembled from **12-601-Nt** and **12-601-Nb**, respectively, using 5′-[^{32}P]-labeled oligonucleotide inserts (Table A.3) and subsequently reconstituted into oligonucleosome arrays as described above. In our assays, both naked DNA and oligonucleosome arrays (4 nM) containing the positioned lesion were treated with UDG and APE1 (1 nM each), resulting in the formation of a singlestrand break at the damage site. Following the reaction, the dU-containing 5′-[^{32}P]-labeled insert was specifically excised from the larger 12×601 array using the corresponding nicking endonuclease, and the products were resolved by denaturing gel electrophoresis (Figure A.13 and A.15). Distilling the large chromatin complex down to a short (< 40 nt) radiolabeled fragment

greatly simplifies the analysis of the UDG/APE1 digestion reaction and highlights another advantage of our plug-and-play approach. Complete removal of the damaged insert from the 12×601 DNA template was confirmed by the absence of high molecular weight bands in the gel (Figure A.15). In order to determine the impact of chromatin compaction on the activities of UDG/APE1, our precisely damaged nucleosome arrays (**12-NCP-dU49** and **12-NCP-dU88**) were digested in the presence of either 0.2 or 2 mM Mg^{2+}. Nucleosome arrays take on an extended beads-on-a-string conformation in the presence of 0.2 mM Mg^{2+}, representing an accessible chromatin state (i.e., euchromatin). However, in the presence of 2 mM Mg^{2+}, nucleosome arrays fold into maximally compact fibers similar to condensed chromatin *in vivo* (i.e., the 30 nm chromatin fiber) [34], [211], [232]. Indeed, AFM images of the damaged 12-mer arrays confirm that they are highly compacted at 2 mM Mg^{2+} (Figure A.11). In addition to the damaged arrays, a similar set of digestion reactions were carried out on mononucleosome substrates (**1-NCP-dU49** and **1-NCP-dU88**) containing identically positioned dU modifications (Figure A.3). By comparing UDG/APE1 digestion efficiencies between these two types of chromatin substrates (i.e., mononucleosomes and arrays), the impact of chromatin folding was readily separated from the intrinsic inhibitory effects of the underlying histone octamer. We first examined the series of substrates containing dU49, which is positioned within the nucleosome core and is outwardly oriented in relation to the histone octamer surface. In the presence of 0.2 mM Mg^{2+}, the combined action of UDG/APE1 cleaved both mononucleosomes (**1-NCP-dU49**) and nucleosome arrays (**12-NCP-dU49**) containing dU49 at nearly equivalent rates (2.8 ± 1.4 and 1.9 ± 0.5 min^{-1} , respectively), which was ∼2−3-fold slower than the rate

observed for digestion of the corresponding naked DNA (**12-601-dU49**; Figure A.14). The total extent of digestion was also similar for both chromatin substrates (~60%). This result indicates that accessibility of dU49 within the extended nucleosome array is dictated almost exclusively by the underlying histone octamer (i.e., nucleosome 5), whereas neighboring nucleosomes have little effect on the rate of BER. In contrast, when the concentration of Mg^{2+} was increased to 2 mM, which induces compaction of the extended array, cleavage of dU49 by UDG/APE1 was dramatically inhibited in the nucleosome array as compared naked DNA (Figure A.14). Indeed, the observed initial rates for these digestion curves (0.6 ± 0.1 and 14.1 ± 3.8 min^{-1}, respectively) indicate that dU49 was digested at least 20-fold slower within compacted chromatin than within the naked DNA substrate. This is despite an overall increase in UDG/APE1 activity (~2-fold) as a result of increased Mg^{2+} concentration [292]. The reduced UDG/APE1 efficiency is not due to array aggregation, because a chromatin oligomerization assay[89] failed to detect significant precipitation of a oligonucleosome array under the assay conditions (Figure A.9). Importantly, digestion of dU49 within the compact array (**12-NCP-dU49**) proceeded ~8-fold slower than on the corresponding mononucleosome substrate (**1-NCP-dU49**), suggesting that inter-nucleosome interactions are the major contributor of the inhibition observed at this site. This observation is significant because it implies that chromatin conformational dynamics, and regulators there of (e.g., histone PTMs), are likely to contribute to the rate of BER in vivo. In order to confirm that the observed inhibition was a direct result of inter-nucleosome interactions and chromatin folding, the UDG/APE1 digestion experiments above were repeated using nucleosome arrays reconstituted from

histone octamers lacking the H4 N-terminal tail (residues 1−19). The H4 tail domain mediates chromatin compaction through contacts with the H2A acidic patch on an adjacent nucleosome [26], [213]. Therefore, deletion of the H4 tail significantly impairs Mg^{2+}-induced compaction of nucleosome arrays without influencing the stability of individual nucleosome core particles [33]. In the presence of 2.0 mM Mg^{2+}, the tail-less array was digested ~14- fold faster by UDG/APE1 than the array containing full-length histone H4 (Figure A.14), indicating that UDG/APE1 activity could be rescued by disrupting chromatin compaction. In contrast, deletion of the H4 tail had a modest effect on the digestion efficiency of dU49 within mononucleosomes (80%) within the 12-mer array is sequestered within nucleosomes, which effectively reduces the amount of DNA that UDG and APE1 must sample prior to finding the damaged site. Alternatively, compaction of the damaged DNA into a 30 nm fiber, which is expected to increase molecular crowding and effective DNA concentration around the lesion [45], may promote a facilitated diffusion search mechanism by UDG/APE1, thereby increasing the efficiency of lesion identification and repair [144], [293], [294]. Regardless of the mechanism, these data presented here suggest that densely packed chromatin environments (i.e., heterochromatin) may promote the initial steps of BER. Given that BER within linker DNA is not well characterized, this is the first study examining BER of a site-specifically positioned DNA lesion within an oligonucleosome array, it will be important to investigate how the position of a DNA lesion within the linker region, as well as the "linker histone" H1 [119], [155], affects the activities of BER proteins in the context of different chromatin architectures. Using stochastically damaged 12-mer nucleosome arrays as BER substrates,

Nakanishi *et al* [212]. previously reported that the combined activities of UDG and APE1 are only modestly inhibited (∼2−3-fold) within compact chromatin. This observation led the authors to conclude that DNA base damage within heterochromatin remains mostly accessible to UDG/APE1, and thus the initial steps of BER do not require significant disruption of chromatin structure. However, the data presented above using site-specifically modified arrays (Figure 4), which provides a much higher resolution analysis of BER, now suggests otherwise. Indeed, we showed that the rate of uracil removal by UDG/APE1 within compacted nucleosome arrays varied by as much as 70-fold between the two positions tested (Figure A.14). Thus, we conclude that damaged DNA within heterochromatin does not remain mostly accessible to BER initiation, but rather the accessibility varies significantly in a position-dependent manner.

A.2 Supplementary Methods

A.2.1 Imaging of reconstituted arrays using Atomic Force Microscopy (AFM)

Structural characterization of the nucleosome arrays were performed via topographic atomic force microscopy (AFM) imaging in an Agilent 5500. Tapping mode was operated in a dry nitrogen environment (relative humidity 0.1%) to eliminate capillary effects. The nominal tip radius was 10 nm, the spring constant was typically ~50 N/m (Aspire conical tips, CT170R), and the setpoint for imaging was 4 V – 6 V. Image analysis was done in the software program Scanning Probe Imaging Processing (SPIP). Sample plates were newly prepared before each imaging session. Poly-L-lysine (PL) was

deposited on freshly cleaved mica by dropcasting 30 µL of a 10 µg/mL aqueous solution for a 30 second incubation, then rinsing from the surface with nanopore water (18.2 MΩ•cm, Barnstead) and drying under flowing nitrogen. Three types of samples were prepared for structural characterization: 1) the extended state chromatin, 2) compacted state chromatin, and 3) the chromatin-enzyme (APE1) complex. For each case, the final concentration of sample was 0.25 nM in low salt buffer (25 mM NaCl, 0.1 mM PMSF, 10 mM HEPES pH 7.8) and the sample was cross-linked with 0.1% (v/v) glutaraldehyde (EM grade, Electron Microscopy Science) overnight on ice. For extended state imaging, 30 µL of sample was drop-casted over poly-L-Lysine coated mica for a 1 min incubation time, then gently washed with nanopure water and dried under flowing nitrogen. For compact state imaging, the sample was incubated with 2 mM $MgCl^2$ for 30 minutes on ice before cross-linking, the rest of 98% saturation. This further validates the biochemical results obtained with the **12-NCP-dU49** chromatin.

AFM characterization of **12-NCP-dU88** identified 42 well-resolved strands in four, 1 µm x 1 µm scans (Figure A.10-A.12). In the higher resolution images (250 nm x 250 nm scan) all 12 individual nucleosomes were visible along with the linker DNA (Figure A.10). The heights of the nucleosomes were measured to be 2.2 nm – 3.5 nm and the height of the DNA was determined to be ~0.8 nm (Figure A.10). Out of the 42 well-resolved strands, 36 were fully saturated, while 3 of the arrays were missing a single nucleosome. An additional 3 substrates lacked 2 nucleosomes. Thus, of the 504 nucleosome positioning sequences present in the 42 arrays characterized, 495 were

wrapped into a nucleosomes, yielding >98% nucleosome occupancy. This further validates the biochemical results obtained for the **12-NCP-dU88** chromatin.

AFM imaging of both the **12-NCP-dU49** and **12-NCP-dU88** chromatin substrates in the presence of 2 mM $MgCl_2$ revealed that both arrays adopted a more 'compacted' conformation, representative of heterochromatin (Figure A.11). In the compacted individual nucleosomes and the linker DNA were not well-resolved. However, these images confirmed that when the samples were prepared in the presence of $MgCl_2$ the strands were folded into a compact structure representative of the 30 nm fiber. These observations are in accordance with previous reports that considered the compaction of native and *in vitro* reconstituted chromatin [45], [46], [295].

AFM imaging of APE1-treated **12-NCP-dU49** resulted in the identification of two extended arrays bound by APE1 on the N5 nucleosome (Figure A.12). Our ability to identify additional complexes was likely complicated by our use of Mg^{2+} precipitation to remove excess enzyme from the sample prior to fixing and imaging. Magnesium promotes APE1 turn-over and dissociation from AP sites , making it challenging to trap the complex. Nevertheless, to the best of our knowledge, this is the first time BER on chromatin (i.e. oligonucleosome arrays) has been imaged at single-molecule resolution.

The first APE1-bound array was completely saturated and shows APE1 localized to N5 (250 nm x 250 nm high resolution scan, Figure A.15). The stacking arrangement of APE1 and the underlying nucleosome was very distinct, as APE1 was laterally offset on top of N5, with a combined height of ~5 nm. The stacking observed in APE1-bound samples was unique compared to AFM images of unbound arrays, where the stacking of

multiple nucleosomes results in a vertical alignment. In the second APE1-chromatin complex (250 nm x 250 nm high resolution scan, Figure A.12), the chromatin substrate had eleven nucleosomes with APE1 bound at the N5 nucleosome with similar laterally offset stacking. Importantly, even though the second APE1-bound chromatin was not completely saturated, the nucleosome containing the AP lesion (N5) was occupied by histone octamer. The images reported here of chromatin bound with APE1 at N5 provides further evidence for the authenticity of our sample.

No APE1-chromatin complexes were identified in AFM images obtained using undamaged oligonucleosome arrays (**12-NCP-Nt**) (Figure A.12). In four 1 μm x 1 μm scans (Figure A.12), 16 well-resolved strands were identified. In the 250 nm x 250 nm higher resolution scan (Figure A.12), 12 individual nucleosomes were visible, with heights measured to be 2.3-3.1 nm. Out of the 16 well-resolved strands, 11 were fully saturated. Most importantly, in areas where clusters of nucleosomes appeared vertically stacked, there was no observed lateral offset that would be characteristics of the nucleosome-APE1 complex.

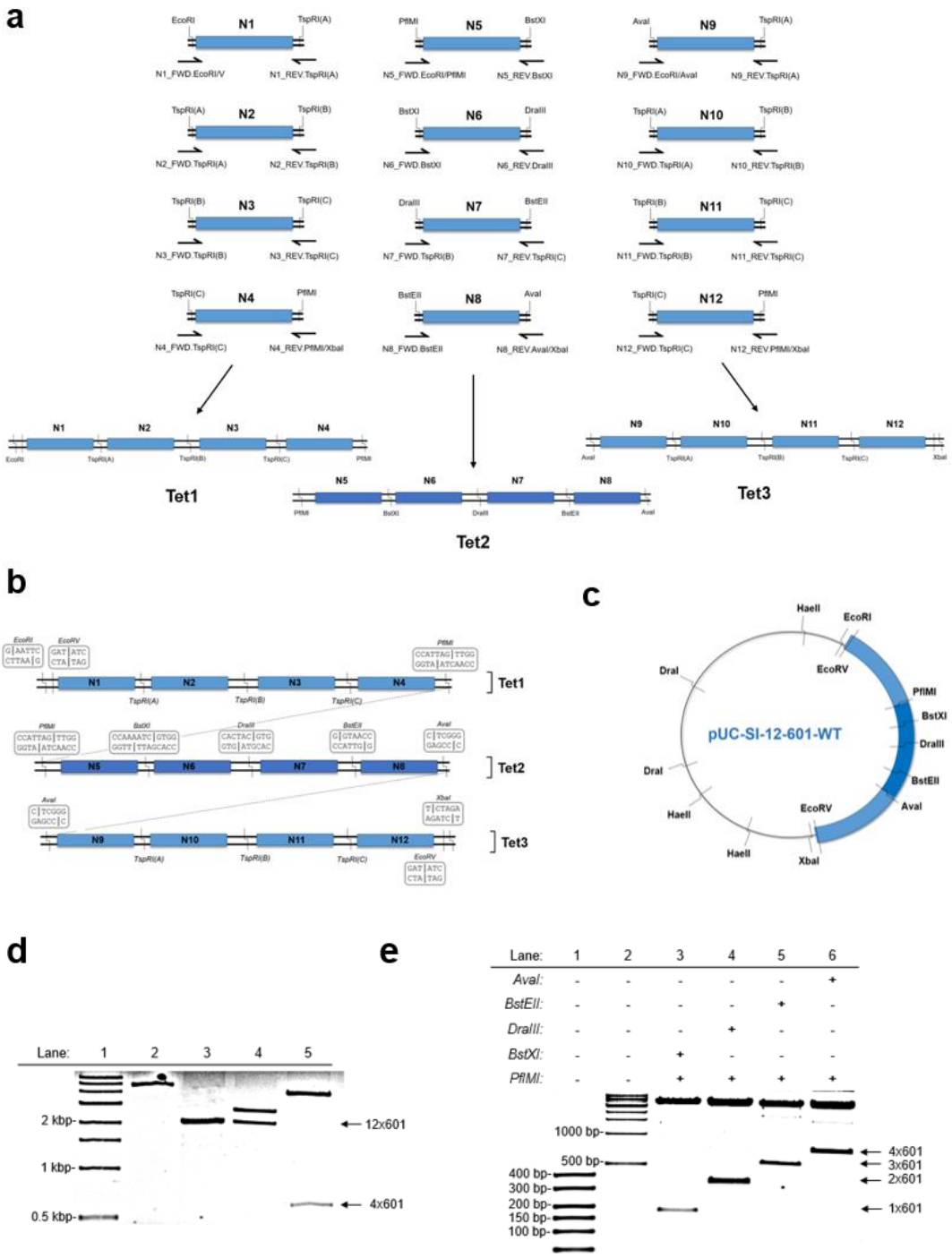
A.2.2 UDG/APE1 digestion

Reaction mixtures contained 4 nM of the appropriate substrate, 25 mM NaCl, 0.1 mM PMSF, 10 mM HEPES, pH 7.8, and either 0.2 or 2 mM $MgCl_2$. Reactions were initiated by adding UDG and APE1 (1 nM each) and incubated at 37 ºC for 0 to 60 minutes. For mononucleosome samples (**1-NCP-dU49** and **1-NCP-dU88**), reactions were supplemented with 44 nM (11 equivalents) of undamaged mononucleosomes such that the

ratio of damaged to undamaged mononucleosomes, as well as total number of 601 units, was identical to the 12- mer arrays. Aliquots were removed at different times and the reaction stopped with the addition of 1% SDS (v/v), $NaBH_4$ (10 mM), and 5 units proteinase K. The mixture was then extracted with phenol:chloroform:isoamyl alcohol (25:24:1, v/v, ThermoFisher Scientific) and the DNA desalted by ethanol precipitation. The resulting DNA was digested with 5 units of either Nt. BstNBI or Nb.BbvCI (depending on the substrate) in order to excise the dU containing 5′-[^{32}P]- labeled oligonucleotide insert from the large 12×601 DNA array (Figure A.13). The digested DNA was then resolved by 10% denaturing PAGE (19:1 acrylamide:bisacrylamide) and visualized using a Typhoon FLA 9500 gel imager (GE Healthcare) (Figure A.15). The fraction cleaved was quantified using ImageQuant TL software (GE Healthcare, v8.1). At least three replicates were carried out for each substrate, and initial rates were calculated by fitting the digestion data in to a single phase exponential and multiplying each rate constant by the Y_{max} as previously described [111], [122], [212].

A.3 Supplementary Figures

Chapter II Figure A.1 : Schematic of **SI-12-601-WT** array. (a) Design of the 12 individual 601 fragments (N1-12) used in the assembly of SI-12-601-WT. The primers used to generate each 601 are listed below their respective 601 units (see Table S1 for sequences). Ligation of N1-N2-N3-N4, N5-N6-N7-N8 and N9-N10-N11-N12 yielded the three tetramers (Tet1, Tet2 and Tet3) shown in (b). The blue rectangles correspond to the 601 sequence and the black lines represent linker DNA. (b) Diagram of the full **SI-12-601-WT** array with each tetramer (Tet1- top, Tet2-middle, Tet3-bottom), their individual 601 components (N1-N12), and restriction enzyme sites indicated. (c) Restriction map of the pUC-SI-12-601-WT plasmid construct. (d) Restriction digestion analysis of **pUC-SI-12-601-WT** (0.5% Agarose, 1x SodiμM Borate buffer). Lane 1, 0.5 kb–10 kb ladder; Lane 2, undigested plasmid; Lane 3, control 12×601 array; Lane 4, EcoRV digest of **pUC-SI-12-601-WT** to release the 12×601 DNA; Lane 5, PflMI and AvaI double digest of pUC-SI-12-601-WT to release Tet2 (middle 4x601). (e) Excision of specific 1-, 2-, 3-, or 4x601 fragments from within Tet2 by digestion with the indicated restriction enzymes (1x601→PflMI/BstXI; 2x601→PflMI/DraIII; 3x601→PflMI/BstEII; 4x601→PflMI/AvaI).


a
EcoRI
N1
TspRI(A)
N1_FWD.EcoRI/V
N1_REV.TspRI(A)
PflMI
N5
BstXI
N5_FWD.EcoRI/PflMI
N5_REV.BstXI
AvaI
N9
TspRI(A)
N9_FWD.EcoRI/AvaI
N9_REV.TspRI(A)
TspRI(A)
N2
TspRI(B)
N2_FWD.TspRI(A)
N2_REV.TspRI(B)
BstXI
N6
DraIII
N6_FWD.BstXI
N6_REV.DraIII
TspRI(A)
N10
TspRI(B)
N10_FWD.TspRI(A)
N10_REV.TspRI(B)
TspRI(B)
N3
TspRI(C)
N3_FWD.TspRI(B)
N3_REV.TspRI(C)
DraIII
N7
BstEII
N7_FWD.TspRI(B)
N7_REV.TspRI(C)
TspRI(B)
N11
TspRI(C)
N11_FWD.TspRI(B)
N11_REV.TspRI(C)
TspRI(C)
N4
PflMI
N4_FWD.TspRI(C)
N4_REV.PflMI/XbaI
BstEII
N8
AvaI
N8_FWD.BstEII
N8_REV.AvaI/XbaI
TspRI(C)
N12
PflMI
N12_FWD.TspRI(C)
N12_REV.PflMI/XbaI
N1
N2
N3
N4
EcoRI
TspRI(A)
TspRI(B)
TspRI(C)
PflMI
Tet1
N5
N6
N7
N8
PflMI
BstXI
DraIII
BstEII
AvaI
Tet2
N9
N10
N11
N12
AvaI
TspRI(A)
TspRI(B)
TspRI(C)
XbaI
Tet3
b
EcoRI
EcoRV
PflMI
N1
N2
N3
N4
TspRI(A)
TspRI(B)
TspRI(C)
Tet1
PflMI
BstXI
DraIII
BstEII
AvaI
N5
N6
N7
N8
Tet2
AvaI
XbaI
N9
N10
N11
N12
TspRI(A)
TspRI(B)
TspRI(C)
EcoRV
Tet3
c
HaeII
EcoRI
DraI
EcoRV
PflMI
BstXI
DraIII
BstEII
AvaI
pUC-SI-12-601-WT
DraI
HaeII
EcoRV
HaeII
XbaI
d
Lane: 1 2 3 4 5
2 kbp-
1 kbp-
0.5 kbp-
12x601
4x601
e
Lane: 1 2 3 4 5 6
AvaI: - - - - - +
BstEII: - - - - + -
DraIII: - - - + - -
BstXI: - - + - - -
PflMI: - - + + + +
1000 bp-
500 bp-
400 bp-
300 bp-
200 bp-
150 bp-
100 bp-
4x601
3x601
2x601
1x601

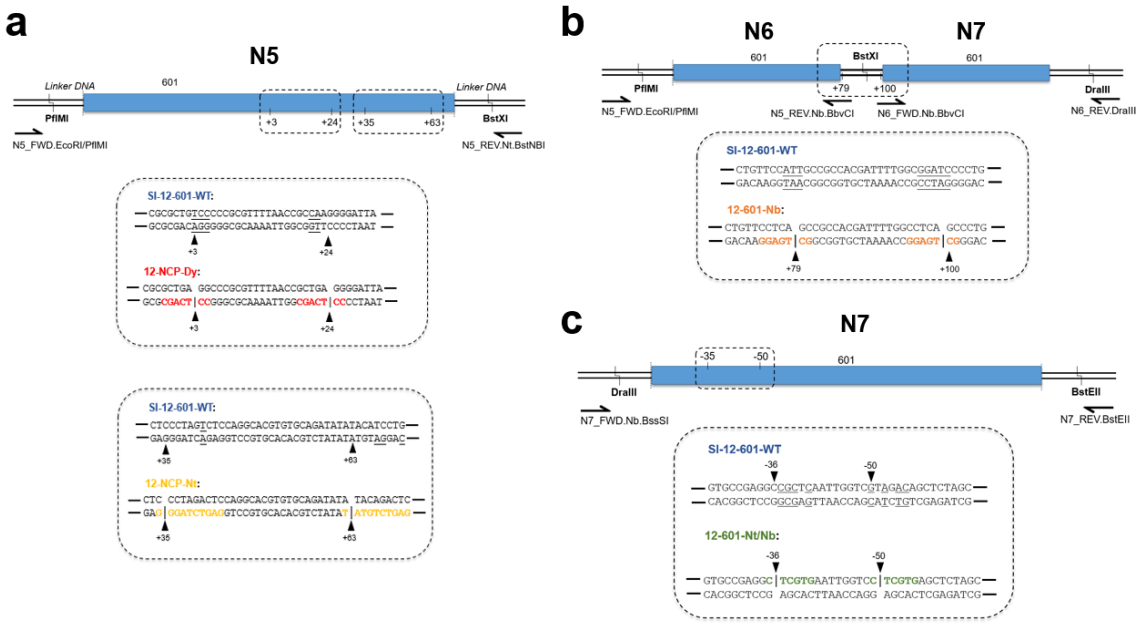


Figure A.2: Position and sequence details for all 'nickable' 12×601 templates prepared in this work compared to with WT 601 template (**SI-12-601-WT**). (a) Sequence-level view of N5 within **SI-12-601-WT** and changes made to generate template **12-601-Dy** and **12-601-Nt**. Red sequence represents the Nt.BbvCI and yellow sequence represents Nt.BstNBI recognition sites, respectively. (b) Sequence-level view of linker DNA between N5 and N6 within **SI-12-601-WT** and changes made to generate template **12-601-Nb**. Orange sequence represents the Nb.BbvCI recognition sites. Together, the constructs depicted in (a) and (b) offer nearly complete coverage over all possible translational conformations of the desired insert (see Figure 2.1). (c) Sequence-level view of N7 within **SI-12-601-WT** and changes made to generate template **12-601-Nt/SI**. Green sequence represents the Nb.BssSI recognition sites.

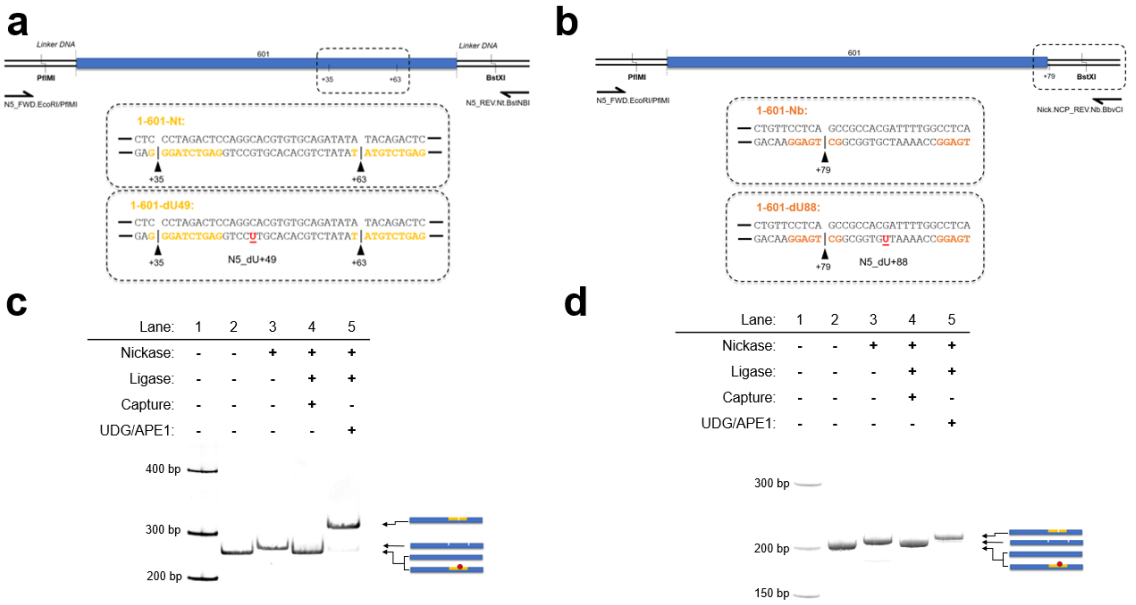


Figure A.3: Position and sequence details for dU containing mononucleosome substrates **1-NCP-dU49** and **1-NCP-dU88**. (a) Sequence-level view of the 1x601 DNA template (**1-601-Nt**) used to prepare mononucleosome **1-NCP-dU49**, which is identical to nucleosome 5 (N5) within array **12-NCP-dU49**. Yellow sequence represents the Nt.BstNBI recognition sites. (b) Sequence-level view of the 1x601 DNA template (**1-601-Nb**) used to prepare mononucleosome **1-NCP-dU88**, which is identical to nucleosome 5 (N5) and the adjacent linker DNA within array **12-NCP-dU88**. Orange sequence represents the Nb.BbvCI recognition sites. (c,d) Electrophoretic mobility shift assays demonstrating the insertion of dU containing oligonucleotides into **1-601-dU49** (a) and **1-601-dU88** (b). Lane 1, ladder; lane 2, unmodified **1-601-Nt** or **1-601-Nb** DNA DNA, respectively; lane 3, 1x601 DNA treated with nicking endonuclease results in doubly nicked DNA that migrates slightly higher than the intact DNA; lane 4, ligated 1x601 fragment containing a dU lesion; lane 5, UDG/APE1 treatment of ligated material (from lane 4) generates a single nick and thus allows for the dU incorporation efficiency to be determined based on the difference in intensities of the intact (lacking dU) and the nicked (containing dU) DNA bands.

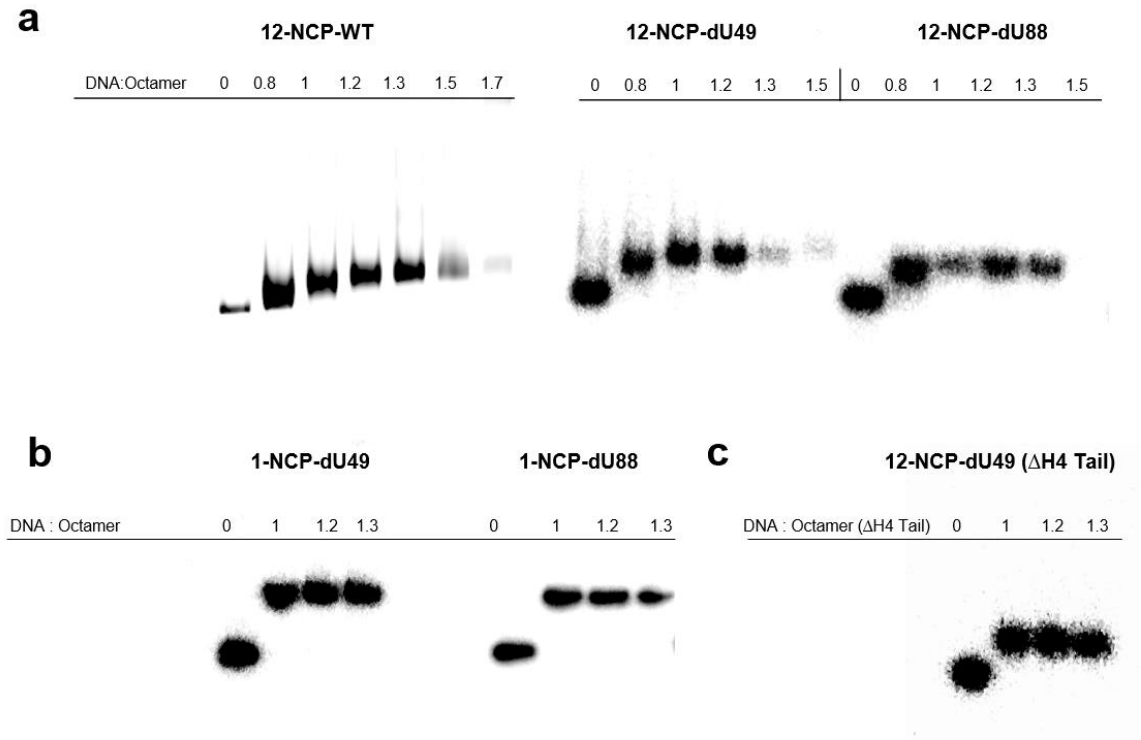


Figure A.4: Analysis of nucleosome and oligonucleosome array reconstitutions. (a) Agarose gel (0.6%, 0.2× TBE) analysis of 12×601 DNA templates reconstituted with increasing concentrations of histone octamer. The molar ratio of histone octamer to DNA template is indicated. (b) Native PAGE (5%, 59:1 acrylamide:bisacrylamide) analysis of 1×601 DNA templates reconstituted with increasing concentrations of histone octamer. The molar ratio of DNA to histone octamer is indicated. (c) Agarose gel (0.6%, 0.2× TBE) analysis of 12×601 DNA templates reconstituted with increasing concentrations of histone octamer lacking the histone H4 tail domain (ΔH4 tail). The molar ratio of 601 units to histone octamer is indicated.

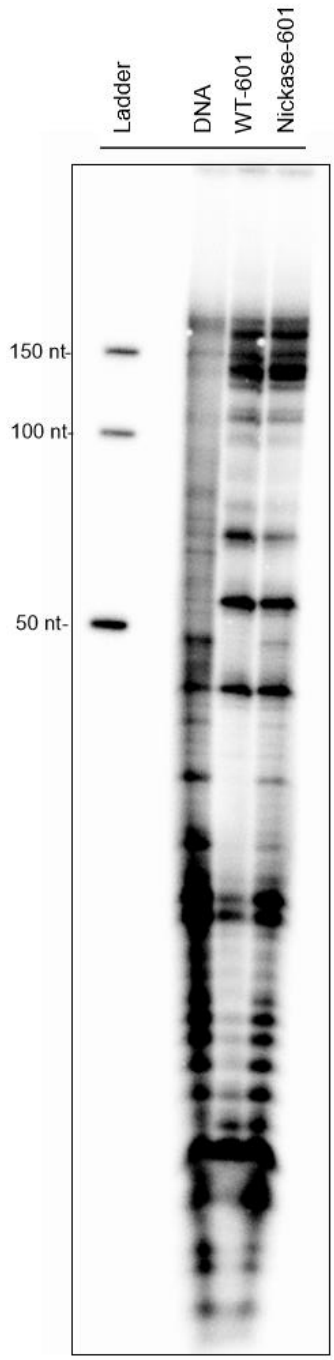


Figure A.5: DNase footprinting analysis (10% denaturing PAGE, 19:1 acrylamide:bisacrylamide) of free DNA and mononucleosomes reconstituted with the native 601 DNA sequence or one modified with Nt.BstNBI recognition sites. The Nt.BstNBI recognition sites are positioned as indicated in Figure 2.1 and A.2.

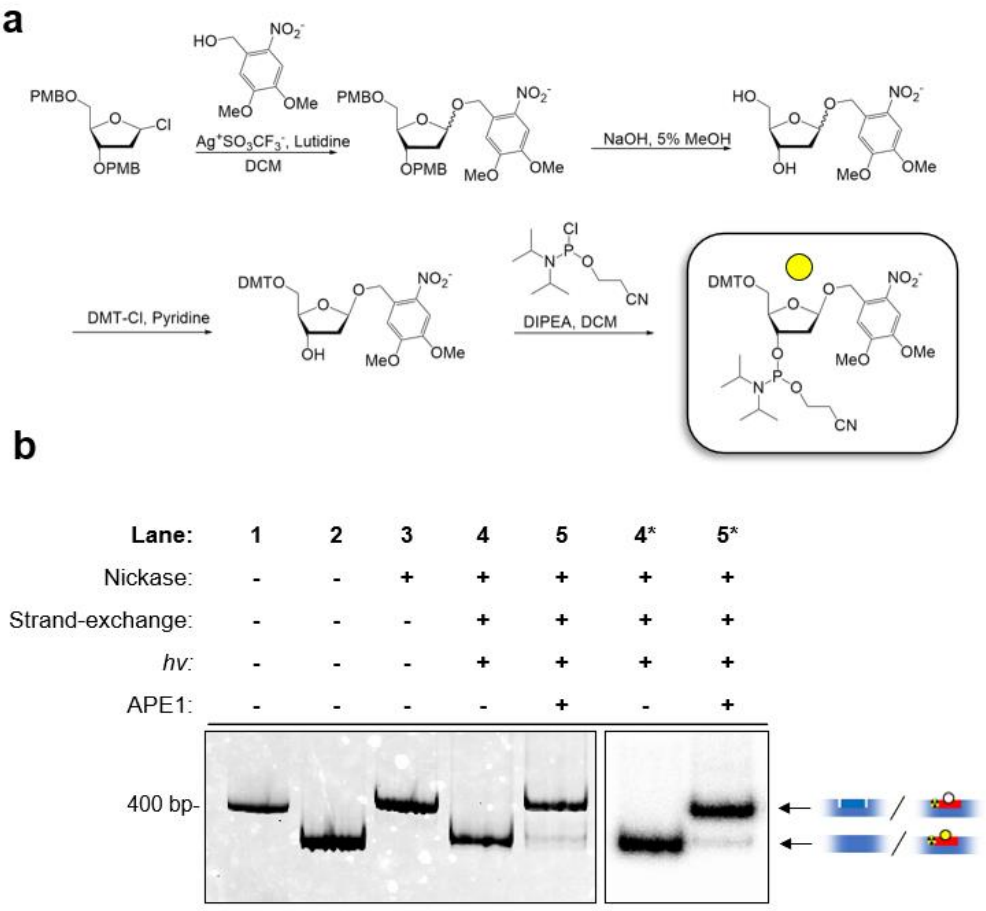


Figure A.6: Synthesis and incorporation of a protected AP lesion into a 12×601 template. (a) Synthetic scheme for preparing an *O*-nitrobenzyl protected phosphoramidite (compound S4). Compound S4 was prepared, characterized, and purified as described by Sczepanski *et. al* [121]. (b) Representative electrophoretic mobility shift assay demonstrating the insertion of an oligonucleotide containing the DNA modification depicted in (a) into **12-601-Nb**. All reactions were carried out on the corresponding plasmid DNA and the 601 DNA fragments (N6-7) containing the modified site was excised via PflMI and DraIII restriction digestion (Figure A.1) prior to analysis by native polyacrylamide gel electrophoresis.

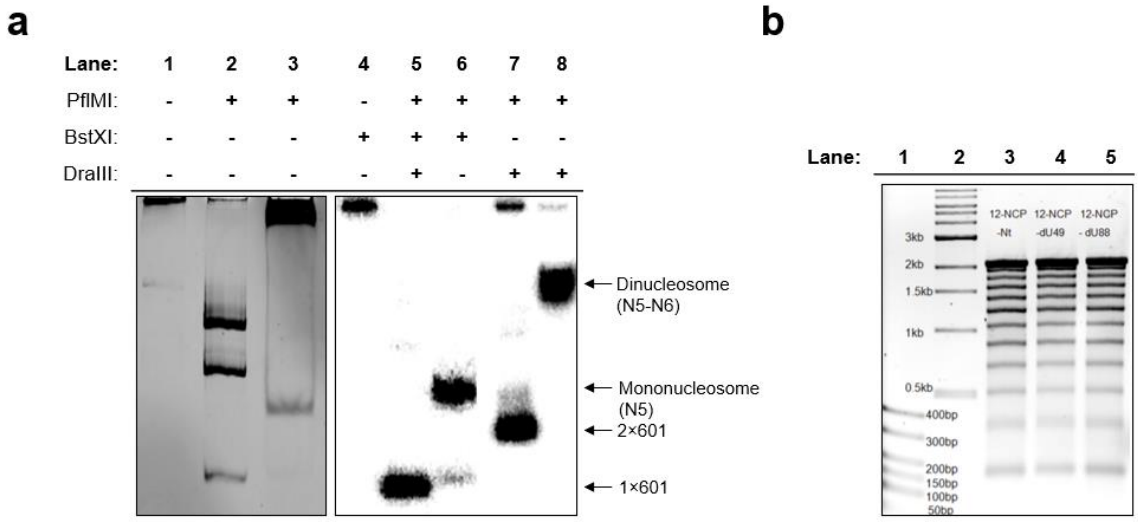


Figure A.7: Verification of full nucleosome occupancy in reconstituted oligonucleosome arrays. (a) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of reconstituted **12-NCP-Nt** (i.e. unmodified array; lanes 1-3) stained with ethidium bromide, **12-NCP-dU49** (lanes 4-6, autoradiogram) and **12-NCP-dU88** (lanes 7-8, autoradiogram). Both **12-NCP-Nt** (WT) and **12-NCP-dU49** arrays were digested with PflMI/BstXI (lanes 3 and 6 respectively) and **12-NCP-dU88** array was digested with PflMI/DraIII (lane 8). Naked DNAs (12-601-Nt, 12-601-dU49 and 12-601-dU88) were digested similarly (lanes 2, 5 and 7, respectively) and were run side-by-side with the digested arrays. Undigested **12-NCP-Nt** and undigested **12-NCP-dU49** (lanes 1 and 4, respectively) were used as controls. (b) Agarose gel (0.5%, 0.2× TBE) analysis of the partial micrococcal nuclease digestion of **12-NCP-Nt** (lane 3), **12-NCPdU49** (lane 4) and **12-NCP-dU88** (lane 5). Lanes 1 and 2 are DNA ladders (50–400 bp and 0.5– 10 kb, respectively).

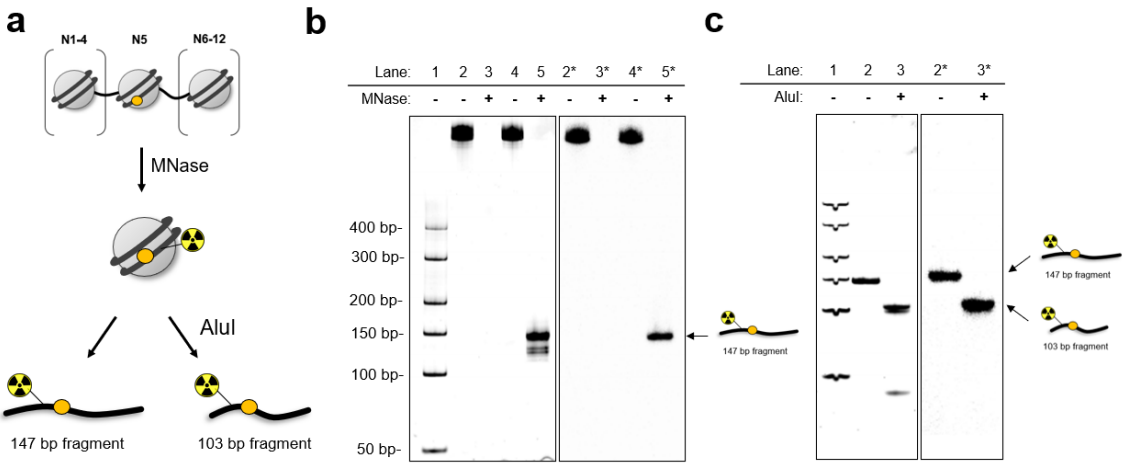


Figure A.8: Verification of the translational positioning of the dU49 lesion within **12-NCP-dU49**. (a) Schematic of the experiment described herein. (b) An aliquot of **12-601-dU49** and **12-NCP-dU49** were digested completely with Micrococcal nuclease (see *2.3.11*) and the full reaction mixture for each respective substrate (free DNA and chromatin) was analyzed via 10% Native PAGE (19:1, acrylamide: bisacrylamide). Representative images of an Ethidium Bromide stained digest gel (left) and an autoradiograph of the same gel (right). Prior to MNase treatment of the free DNA (**12-601-dU49**) a high molecular band is present (lane 2). Addition of MNase to the free DNA sample leads to degradation of **12-601-dU49** to small oligonucleotide fragments that do not remain in the gel (lane 3). Treatment of **12-NCP-dU49** with MNase converts the intact 12×601 DNA (lane 4) into a distinct band which migrates at ~150bp (lane 5). (c) The ~147 bp band in part (b) was cut and purified from the gel, then digested further with AluI. Treatment with AluI converts the ~147 bp fragment (lane 2) to a ~103 bp fragment (lane 3). The observed digestion pattern of the **12-NCP-dU49** indicates that the dU49 lesion is maintained within the nucleosome core particle and that the nucleosome (N5) containing the dU lesion is positioned as expected. Lane 1 in both parts (b) & (c) contains a DNA ladder consisting of 50, 100, 150, 200, 300 and 400 bp dsDNA fragments.

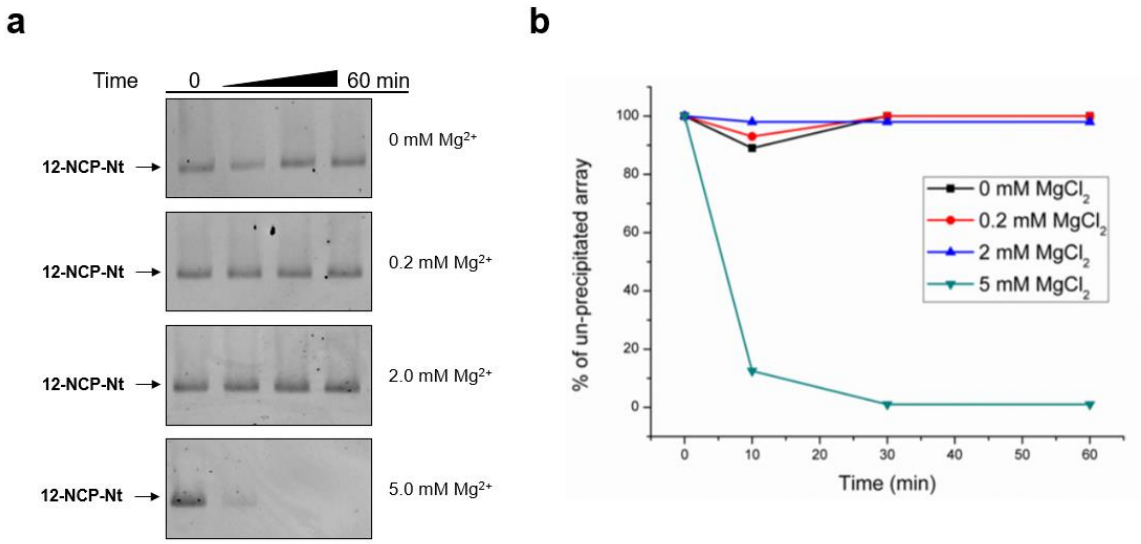
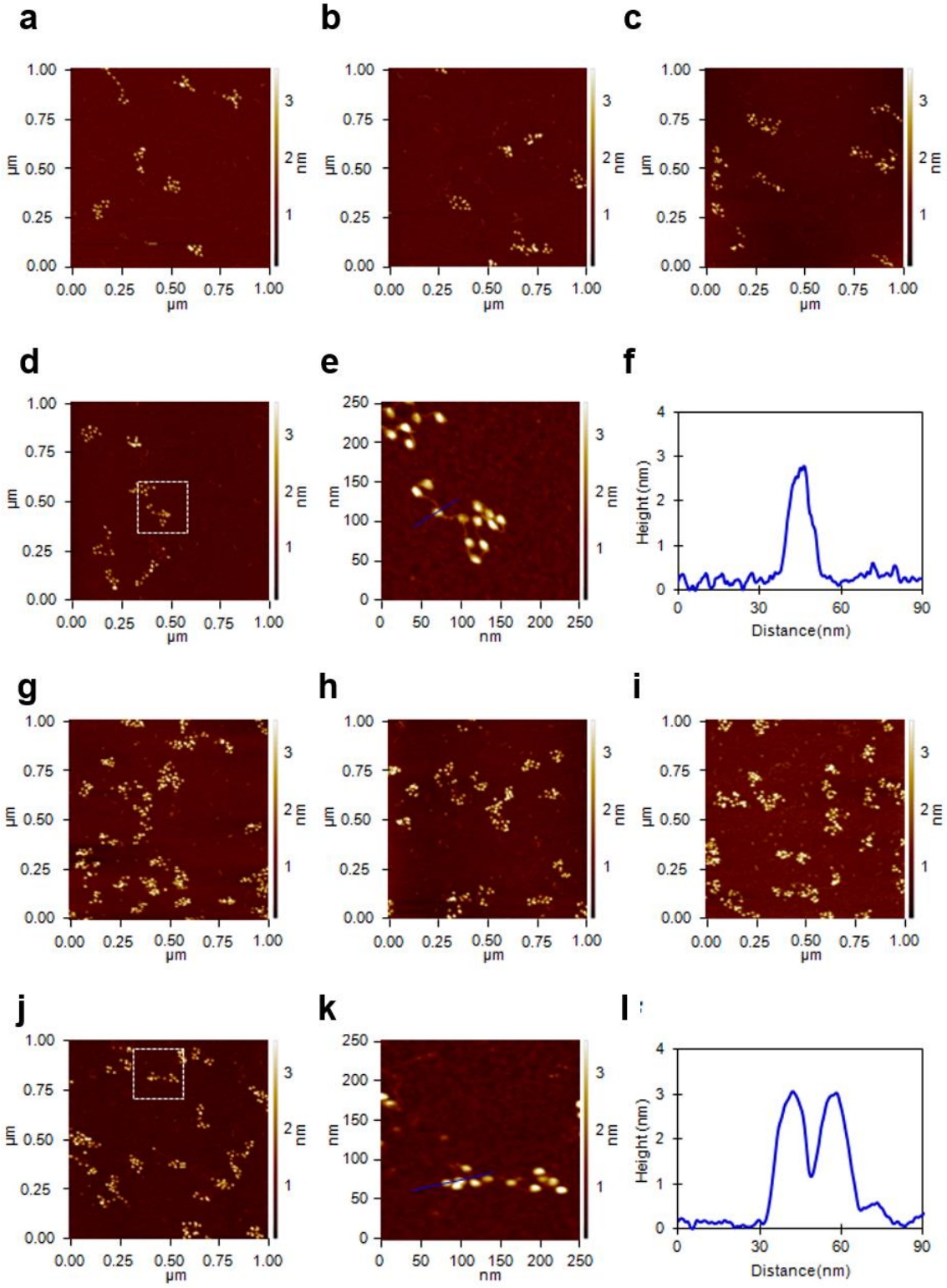


Figure A.9: Chromatin oligomerization assay. (a) Agarose gel (0.5%, 0.2× TBE) analysis of the supernatant fractions of **12-NCP-Nt** arrays as a function of Mg^{2+} concentrations. (b) Quantitative analysis of the band intensities shown in (a).

Figure A.10: AFM tapping mode topography images of the **12-NCP-dU49** and **12-NCP-dU88** chromatin in the extended states in absence of $MgCl_2$: (a-d) and (g-j), respectively. The white dashed box in (d) highlights the strand imaged under high resolution in (e). (e) High resolution 250 x 250 nm scan, all 12 nucleosomes and linking DNA are visible; (f) Height profile of scanned substrate**,** The blue line in e shows the corresponding cross section. (g-l) contain the same analyses, except for **12-NCP-dU88** chromatin.


a
b
c
d
e
f
g
h
i
j
k
l
µm
nm
Height (nm)
Distance (nm)

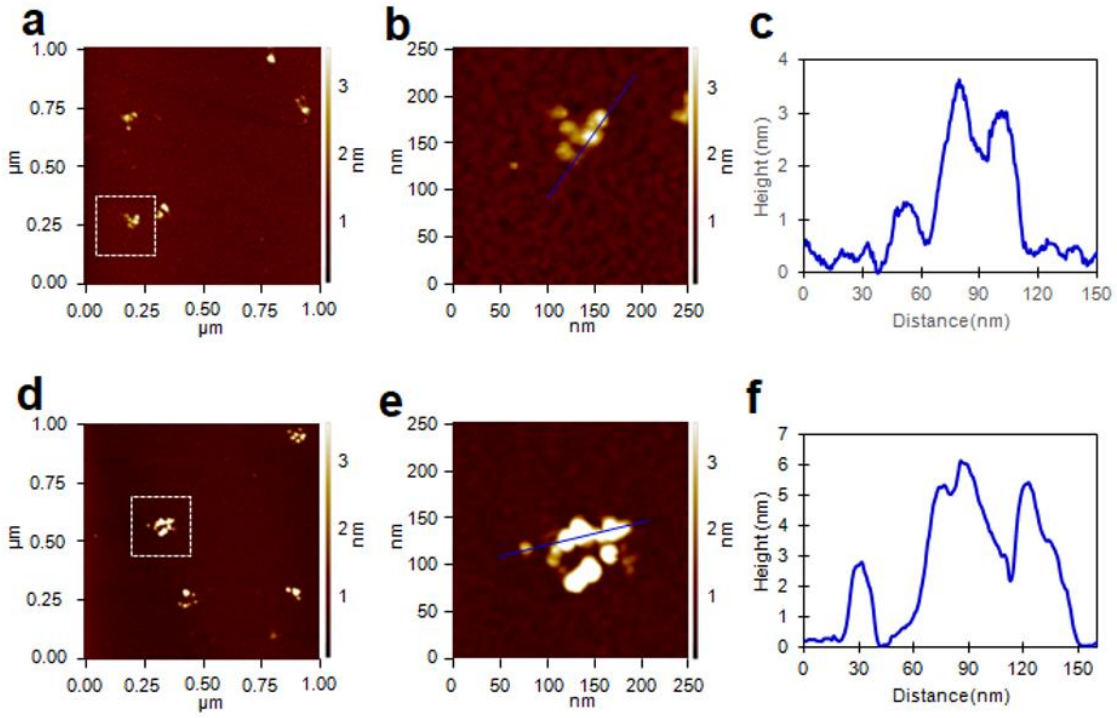


Figure A.11: AFM tapping mode topography images of the **12-NCP-dU49** and **12-NCP-dU88** in the compacted states in presence of 2 mM $MgCl_2$: (a) 1 x 1 µm image of compacted **12-NCP-dU49**, The white dashed box in **a** highlights the strand imaged under high resolution in (b); (b) High resolution 250 x 250 nM scan image of compacted **12-NCP-dU49**; (c) Height profile of scanned substrate**,** The blue line in (b) shows the corresponding cross section; (d) 1 x 1 µm image of compacted **12-NCP-dU88**, The white dashed box in **d** highlights the strand imaged under high resolution in (e); (e) High resolution 250 x 250 nM scan image of compacted **12-NCP-dU88**; (f) Height profile of scanned substrate**,** The blue line in **e** shows the corresponding cross section

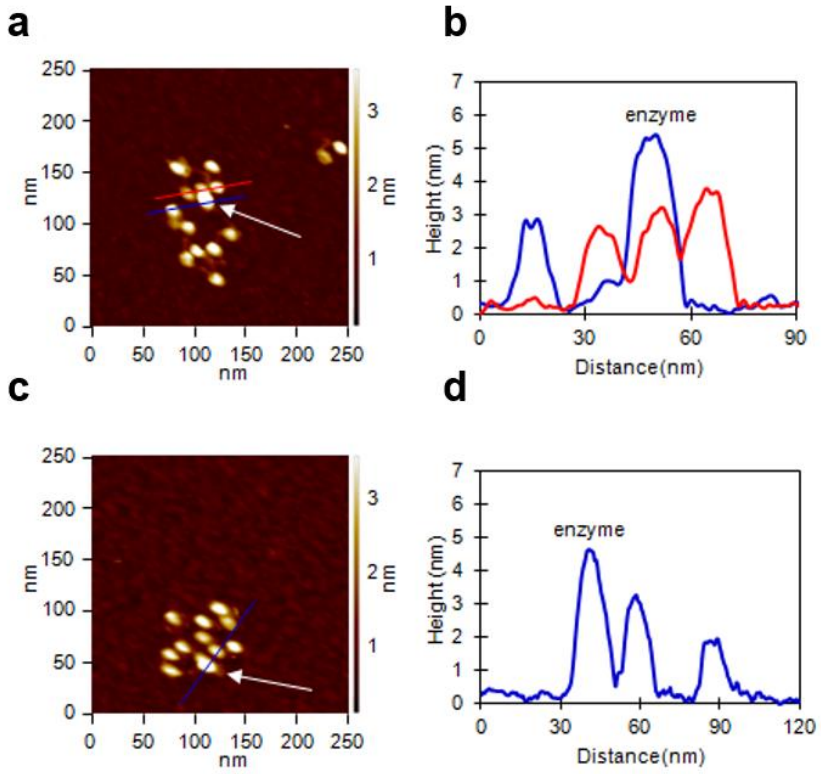
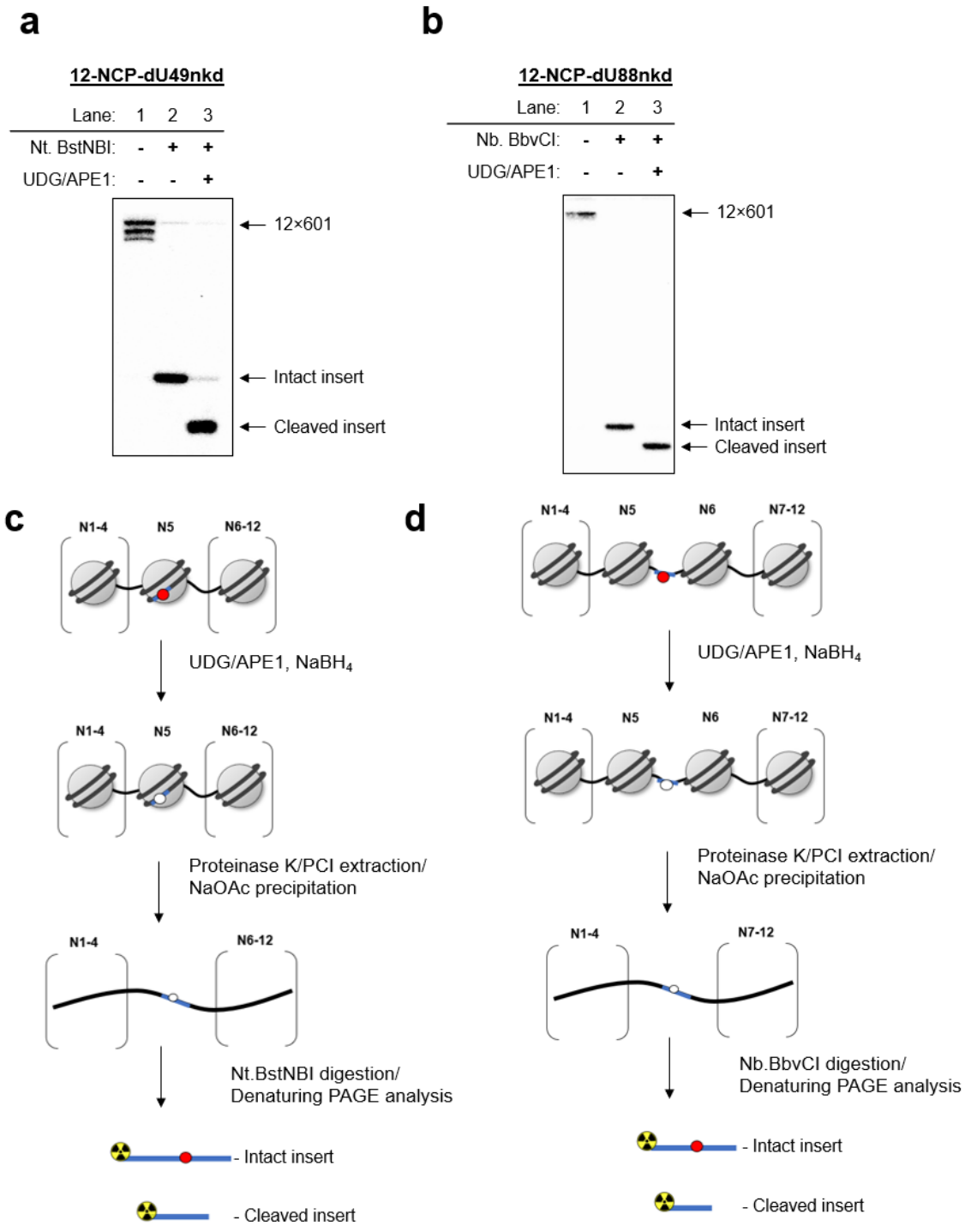
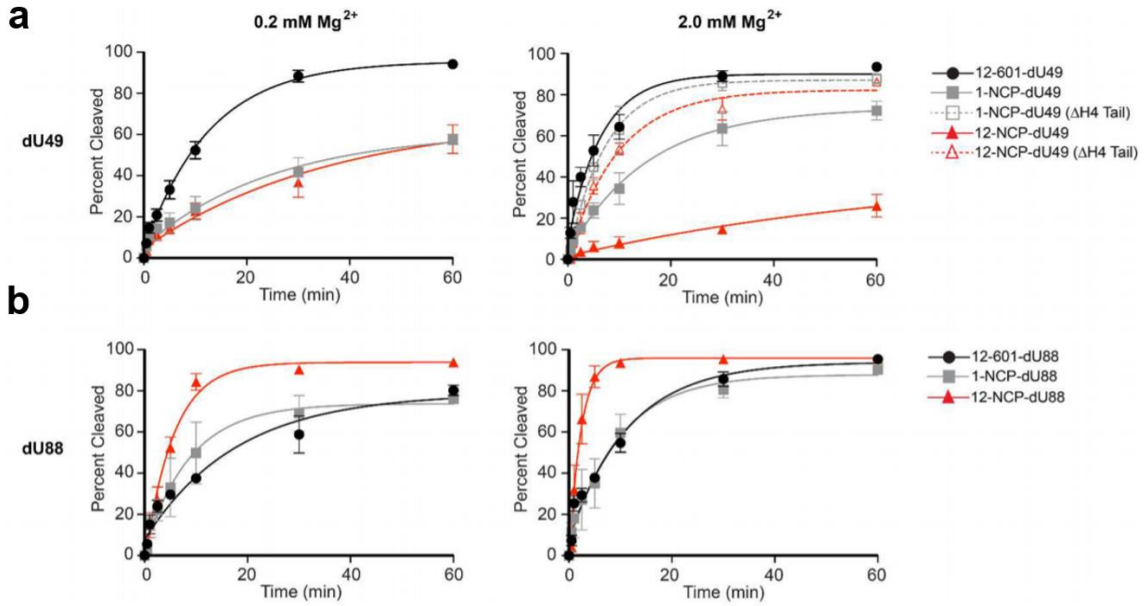


Figure A.12: AFM tapping mode topography images of the **12-NCP-dU49**/APE1 complex. (a) High resolution 250 x 250 nm image of the APE1 (denoted by the white arrow) complexed to the 5th nucleosome of a **12-NCP-dU49**. The red and blue lines show where the corresponding cross sections in (b) were taken. (b) The red cross section shows the height profiles of the nucleosome while the blue cross section shows the height of the enzyme supported by the 5th nucleosome. Similarly, (c) and (d) show a second **12-NCP-dU49**/APE1 complex image and cross section, respectively

Figure A.13: Schematic and gel images (10%, 19:1 denaturing PAGE, autoradiogram) showing the complete removal of damaged inserts from **12-601-dU49** (**a**) and **12-601-dU88** DNA (**b)**: Both 12-601 DNAs were digested with UDG/APE1 and specific nickase enzymes. After nickase digestion, >99% of inserts were removed from DNA templates (lane 2 in each gel); after UDG/APE1 digestion followed by nickase digestion, >99% of damaged insert were removed from 12-601 DNA templates. Schematic of UDG/APE1 combined kinetics against **12-NCP-dU49** (c) and **12-Link-dU88** (d) chromatin: Each kinetics assay was a combination of four steps- nicking of DNA through digestion of UDG/APE1 followed by $NaBH_4$ reduction of nicked 3' end, extraction of DNA through Proteinase K digestion followed by PCI extraction and precipiatation, specific nickase digestion to remove the damaged insert and separation in denaturing PAGE for analyzation. Respective naked DNAs (**12-601-dU49** and **12-601-dU88**) and mononucleosomes (**1-NCP-dU49** and **1-Link-dU88**) were also digested with UDG/APE1 in identical way.


a
12-NCP-dU49nkd
Lane: 1 2 3
Nt. BstNBI: - + +
UDG/APE1: - - +
12×601
Intact insert
Cleaved insert
b
12-NCP-dU88nkd
Lane: 1 2 3
Nb. BbvCI: - + +
UDG/APE1: - - +
12×601
Intact insert
Cleaved insert
c
N1-4
N5
N6-12
UDG/APE1, NaBH4
Proteinase K/PCI extraction/
NaOAc precipitation
Nt.BstNBI digestion/
Denaturing PAGE analysis
- Intact insert
- Cleaved insert
d
N1-4
N5
N6
N7-12
UDG/APE1, NaBH4
Proteinase K/PCI extraction/
NaOAc precipitation
Nb.BbvCI digestion/
Denaturing PAGE analysis
- Intact insert
- Cleaved insert



Lesion	[Mg^{2+}] (mM)	*k* (% cleaved per min)			
dU49	0.2	7.6 ± 0.9	2.8 ± 1.4	1.9 ± 0.5	—
	2.0	14.1 ± 3.8	4.8 ± 1.6 (11.9 ± 2.9)	—	0.6 ± 0.1 (8.5 ± 0.8)
dU88	0.2	4.3 ± 1.0	8.6 ± 1.2	17.0 ± 2.0	—
	2.0	8.1 ± 1.2	9.9 ± 4.8	—	44.9 ± 11.9

Figure A.14: BER of damaged DNA within various chromatin environments. (a,b) The indicated substrates (4 nM) were incubated with UDG and APE1 (1 nM each) in the presence of either 0.2 mM Mg^{2+} (left) or 0.2 mM Mg^{2+} (right). ΔH4 tail indicates substrates assembled using histone H4 lacking the N-terminal tail domain (residues 1−19). (c) Initial rates of cleavage of dU by UDG/APE1. Rates were calculated by fitting the digestion data in panels (a) and (b) to a single phase exponential and multiplying each rate constant by the Y_{max} as previously described [122], [155]. Values in parentheses (red) were obtained using substrates lacking the histone H4 tail domain.

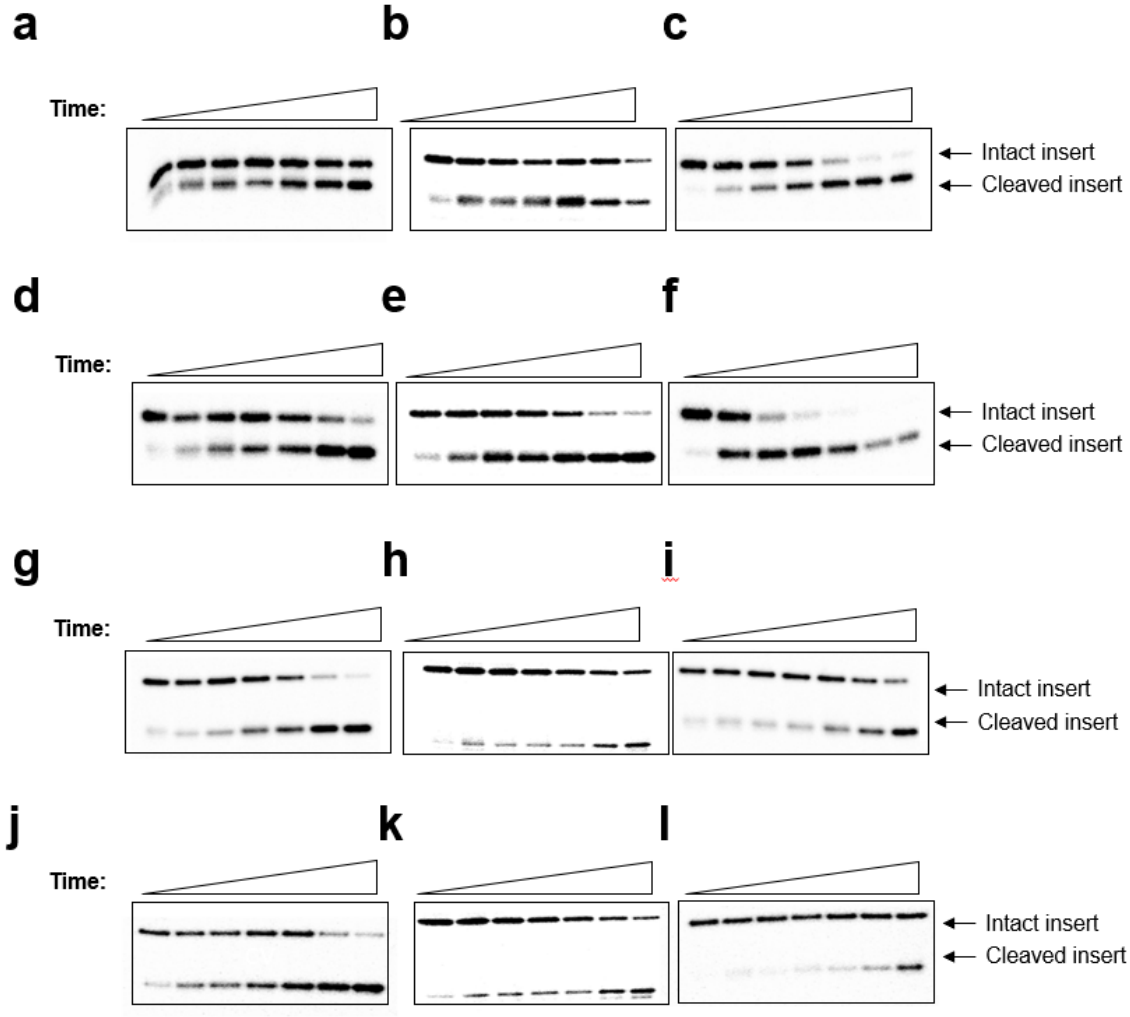


Figure A.15: Representative gel (10%, 19:1 denaturing PAGE) images (auto-radiograms) of UDG/APE1 combined kinetics against three different types of dU49 and dU88 chromatin substrates at two different $MgCl_2$ concentrations: (a) **12-601-dU49** at 0.2 mM $MgCl_2$ (b) **1-NCP-dU49** at 0.2 mM $MgCl_2$ (c) **12-NCP-dU49** at 0.2 mM $MgCl_2$, (d) **12-601-dU49** at 2 mM $MgCl_2$, (e) **1-NCP-dU49** at 2 mM $MgCl_2$ (f) **12-NCP-dU49** at 2 mM $MgCl_2$. (g) **12-601-dU88** at 0.2 mM $MgCl_2$ (h) **1-Link-dU88** at 0.2 mM $MgCl_2$ (i) **12-Link-dU88** at 0.2 mM $MgCl_2$, (j) **12-601-dU88** at 2 mM $MgCl_2$, (k) **1-Link-dU88** at 2 mM $MgCl_2$ (l) **12-Link-dU88** at 2 mM $MgCl_2$.

A.4 Supplementary Tables

Table A.1: Primer sequences used in the construction of the **SI-12-601-WT** DNA. Primer names were assigned in order to reflect the particular 601 unit they were used to generate (i.e. N1, N2, etc.), as well as the Restriction enzyme sites that were introduced (FWD= forward primer, REV= reverse primer). For clarity, the middle column (601 unit) depicts the specific 601, out of the 12 total, generated by the given primer.

Name	601	Sequence
N1_FWD.EcoRI/V	N1	GATCCGGAATTCAGCATCCGGATGATATCGAGAATCCCGGTGCCGAGG
N1_REV.TspRI(A)	N1	TCCTTATCCCCCCAGTGTCCGGCAATGGAACAGGATGTATATATCT GACACG
N2_FWD.TspRI(A)	N2	CATAAGGAGGACACTGGGATCCGGATCCCCTGGAGAATCCCGGTGCC
N2_REV.TspRI(B)	N2	TCCTTATCCCCTCACTGCCCGGCAATGGAACAGGATGTATATATCTGAC ACG
N3_FWD.TspRI(B)	N3	CATAAGGAGGGCAGTGAGATCCGGATCCCCTGGAGAATCCCGGTGCC
N3_REV.TspRI(C)	N3	CGGTATTGTCTAGATCCCGAGCGGCAATGGAACAGGATGTATATATCTGACA CG
N4_FWD.TspRI(C)	N4	ATAAGGAGAGCAGTGTTATCCGGATCCCCTGGAGAATCCCGGTGCC
N4_REV. PflMI/XbaI	N4	CGGTATTGTCTAGACCATTAGTTGGCGGCAATGGAACAGGATGTATATATCT GACACG
N5_FWD. EcoRI/PflMI	N5	GATCCGGAATTCCCAACTAATGGCCGGATCCCCTGGAGAATCCCGGTGCC
N5_REV.BstXI	N5	CGGTATTGCCAAAATCGTGGCGGCAATGGAACAGGATGTATATATCTGACAC G
N6_FWD.BstXI	N6	GATCCGCCACGATTTTGGCGGATCCCCTGGAGAATCCCGGTGCCGAGGCC
N6_REV.DraIII	N6	CGGTATTGCACGTAGTGCGGCAATGGAACAGGATGTATATATCTGACACG
N7_FWD.DraIII	N7	GATCCGCACTACGTGATCCGGATCCCCTGGAGAATCCCGGTGCCGAGGC
N7_REV.BstEII	N7	CGGTATTGGGTTACCCGGCAATGGAACAGGATGTATATATCTGACACG
N8_FWD.BstEII	N8	GATCCGGGTAACCGCATCCGGATCCCCTGGAGAATCCCGGTGCCGAGGCC
N8_REVAvaI/XbaI	N8	CGGTATTGTCTAGATCCCGAGCGGCAATGGAACAGGATGTATATATCTGA CACG
N9_FWD.EcoRI/AvaI	N9	CGACTTGAATTCAGCTCGGGCGCATCCGGATCCCCTGGAGAATCCCGGTG CCGAGGCC
N9_REV.TspRI(A)	N9	TCCTTATCCCCCCAGTGTCCGGCAATGGAACAGGATGTATATATCTGAC ACG
N10_FWD.TspRI(A)	N10	CATAAGGAGGACACTGGGATCCGGATCCCCTGGAGAATCCCGGTGCC
N10_REV.TspRI(B)	N10	TCCTTATCCCCTCACTGCCCGGCAATGGAACAGGATGTATATATCTGAC ACG
N11_FWD.TspRI(B)	N11	CATAAGGAGGGCAGTGAGATCCGGATCCCCTGGAGAATCCCGGTGCC
N11_REV.TspRI(C)	N11	CGGTATTGTCTAGATCCCGAGCGGCAATGGAACAGGATGTATATATCTGA CACG
N12_REV.TspRI(C)	N12	CATAAGGAGAGCAGTGTTATCCGGATCCCCTGGAGAATCCCGGTGCC
N12_REV.PflMI/XbaI	N12	CGGTATTGTCTAGATGATATCGGATGTATATATCTGACACGTGCC
WT.147_FWD	-	CTGGAGAATCCCGGTGCCGAGGCC
WT.147_REV	-	ACAGGATGTATATATCTGACACGTGCC

Table A.2: DNA sequences used in the preparation of the modified, nickase-site containing, 601s. Names were assigned just as described in Table A.1. Here, the nickase enzymes are reflected in the sequence names, even though the other restriction sites were maintained (i.e. **N5_REV.Nt.BstNBI** and **N5_REV.Nb.BbvCI** also contain the same BstXI site as **N5_REV.BstXI**, this was required for uploading of the nickase-601 into the original 12-601 array). "Temp.TOP" & "Temp.BOT" refer to oligonucleotides used to generate a template for PCRs via cross extension reactions. The middle column (601 Unit) emphasizes the specific 601 that was generated with the DNAs shown below. Oligonucleotides lacking a descriptor for "601 Unit" were used solely for generating 1-601s for mononucleosome studies.

Name	601	Sequence
N5_REV.Nt.BstNBI	N5	CGGTATTGCCAAAATCGTGGCGGCAATGGAAGAGTCTGT ATATATCTG ACACG
N5_Nick.Temp.TOP	N5	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTA GACAGCTCT AGCACCG CTTAAACGCACGTACGCGCTGTCCCCC
N5_Nick.Temp.BOT	N5	TGTATATATCTGACACGTGCCTGGAGTCTAGGGAGTAAT CCCCTTGGC GGTTAAAACG CGGGGGACAGCGCGTACGTGC
N5_Gap.Temp.TOP	N5	CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTA GACAGCACGAGCACCGCTTAAACGCACGAGCGCGCTGTC CCCC
N5_Gap.Temp.BOT	N5	TGAGGATGTATATATCTGACGCGCCGGTGGAGCTGAGGG AGTCATCCCCCTCGTGGTTAAAACGCGGGGGACAGCGCG CTCG
N5_REV.Nb.BbvCI	N5	CGGTATTGCCAAAATCGTGGCGGCTGAGGAACAGGATGT ATATATCTG ACACG
N5_FWD.Nb.BbvCI	N5	GATCCGCCACGATTTTGGCCTCAGCCCTGGAGAATCCCG GTGCCGAG GCC
N7_FWD.Nb.BssSI	N7	GATCCGCACTACGTGATCCGGATCCCCTGGAGAATCCCG GTGCCGAGG CTCGTG
N7_Nick.Temp.TOP	N7	CTGGAGAATCCCGGTGCCGAGGCTCGTGAATTGGTCCTC GTGAGCTCT AGCACCGCT TAAACGCACGTACGCGCTGTCCCCC
Nick.149_REV.FspI	-	TTGCCAAAATCGTGGCGGCAATGGATGCGCAGAGTCTGT ATATATCTG ACACG
Nick.NCP_REV.Nb.BbvCI	-	TGAGGCCAAAATCGTGGCGGCTGAGGAACAGGATGTATA TATCTGAC ACG
N5_FWD.Gap	N5	GATCCGGAATTCCCAACTAATGGCCGGATCCC CTGGAGAATCCCGGTGCC
N5_REV.Gap	N5	CGGTATTGCCAAAATCGTGGCGGCAATGGC TGAGGATGTATATATCTGACGCGCC

Table A.3: Sequences of DNAs and Lock nucleic acid (LNA) capture probes used in the current work. Oligonucleotides were named to reflect the specific 601 fragment they were capable of being inserted into (this is also represented by the middle column: "601 Unit"). The naming scheme also describes the particular modification being introduced and its position: dU+49= 2'-deoxyuridine residue at the +49 position with respect to the N5 dyad, Am.T= C6 Amino dT. The amine functionalities were used to conjugate NHS-Cy3 and NHS-Cy5 dyes. The capture probes: **LNA.Capt_NCP** and **LNA.Capt_Linker** were used during insertion reactions containing dU+49 or dU+88, respectively (the underlined residues correspond to positions of LNAs).

Name	601	Sequence
N5_dU+49	N5	Phos/TATATCTGACACGT/dU/CCTGGAGTCTAGG
N5_dU+88	N5	Phos/TGAGGCCAAAAT/dU/GTGGCGGC
N5_ONV+49	N5	Phos/TATATCTGACACGT/ONV/CCTGGAGTCTAGG
N5_ONV+88	N5	Phos/TGAGGCCAAAAT/ONV/GTGGCGGC
N5_dU+49_Am+39	N5	Phos/TATATCTGACACGT/dU/CCTGGAGTC/AmT/AGG
N7_Am.T-39	N7	Phos/TCGTGAATTGG/AmT/CC
LNA.Capt_NCP	-	CC**T**AGACTCC**A**GGC**A**CGTG**T**CAGA**T**ATA/Biotin
LNA.Capt_Linker	-	G**C**CGC**C**ACGATTTT**G**GCCT**C**A/Biotin
Capture Strand 1	-	GCACCGCTTAAACGCACGA
Capture Strand 2	-	GCGCCCTGTCCCCCGCGTTTTAACCACGA

B) APPENDX

CHAPTER 3 SUPPLEMENTARY DATA

B.1 Supplementary Figures

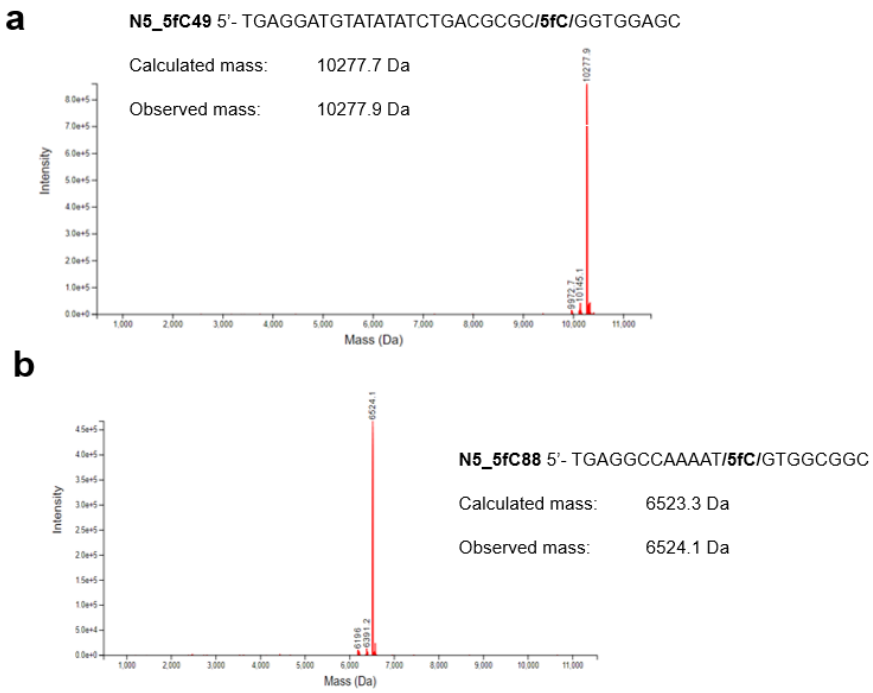


Figure B.1: ESI-MS spectra of 5fC-containing oligonucleotides N5_5fC49 (a) and 5N_5fC88 (b).

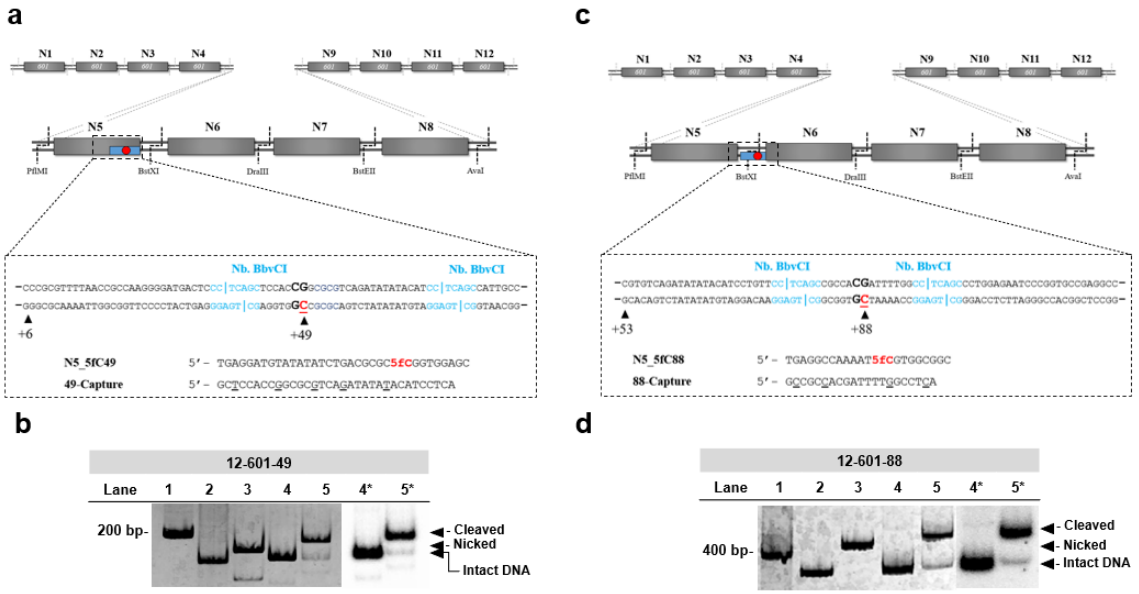


Figure B.2: Sequence and design details of the 5fC49- and 5fC88-containing 12×601 DNA templates (**12-601-49** and **12-601-88**, respectively). (a) Schematic representation of the position of 5fC49 within **12-601-49**, which was constructed from **pUC-12-601-49** via strand exchange using oligonucleotide **N5_dU49**. (b) Representative electrophoretic mobility shift assay (10% 29:1 acrylamide:bisacrylamide) demonstrating the insertion of a 5fC49 containing oligonucleotide (**N5_5fC49**) into plasmid **pUC-12-601-49** (i.e. **12-601-49**). All reactions were carried out on the corresponding plasmid DNA as described previously and the 601 DNA fragment (N5, 178bp) containing the modified site was excised via PflMI and BstXI restriction digestion prior to analysis by 10% native PAGE (19:1 acrylamide:bisacrylamide). Lane 1, 200 bp marker; lane 2, unmodified **pUC-12-601-49**; lane 3, **pUC-12-601-49** DNA treated with Nb.BbvCI; lane 4, ligated **pUC-12-601-49** following strand exchange with 5′-[^{32}P]-**N5_dU49**; lane 5, TDG treatment of ligated material from lane 4. Lanes 4* and 5* are radiograph images of lanes 4 and 5. (c) The same as (a) but for 5fC88 within **12-601-88**. (d) All reactions were carried out on the corresponding plasmid DNA as described previously and the 601 DNA fragment (N5–N6, 356bps) containing the modified site was excised via PflMI and DraIII restriction digestion prior to analysis by 8% native PAGE (29:1 acrylamide:bisacrylamide). Lane 1, 400 bp marker; lane 2, unmodified **pUC-12-601-88**; lane 3, **pUC-12-601-88** DNA treated with Nb.BbvCI; lane 4, ligated **pUC-12-601-88** following strand exchange with 5′-[^{32}P]-**N5_dU88**; lane 5, TDG treatment of ligated material from lane 4. Lanes 4* and 5* are radiograph images of lanes 4 and 5.

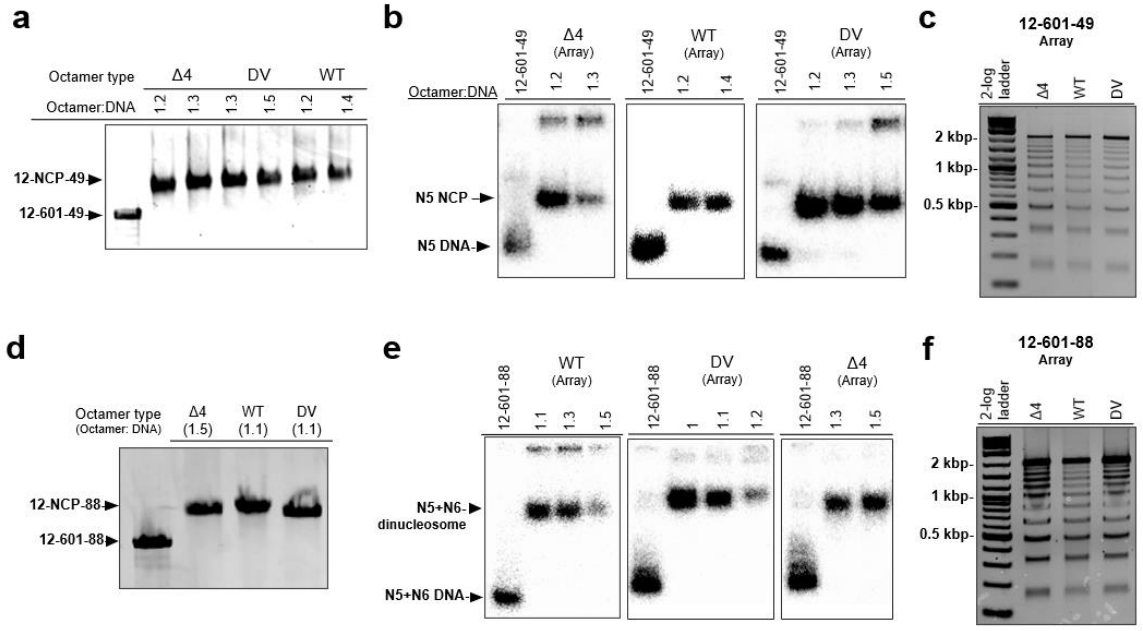


Figure B.3: Characterization of reconstituted nucleosome arrays. (a) Agarose gel electrophoresis (0.6%) analysis of **12-601-49** reconstituted with increasing concentrations of the indicated histone octamer. The molar ratio of histone octamer to DNA template is indicated. (b) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of **12-NCP-49** nucleosome arrays. Both naked DNA (**12-601-49**) and the indicated array were digested with PflMI/BstXI and were run side-by-side to determine nucleosome occupancy at N5. (c) Partial Micrococcal nuclease digestion analysis (1% agarose) of 5fC49-containing arrays. (d) Agarose gel electrophoresis (0.6%) analysis of **12-601-88** reconstituted with the indicated ratio of histone octamer to DNA template. (e) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of **12-NCP-88** nucleosome arrays. Both naked DNA (**12-601-88**) and the indicated array were digested with PflMI/DraIII and were run side-by-side to determine nucleosome occupancy at N5 and N6. (f) Partial Micrococcal nuclease digestion analysis (1% agarose) of 5fC88-containing arrays. WT: canonical octamers; Δ4: histone H4 tail deleted octamers; DV: histone H2A.Z/H3.3 dual-variant octamers.

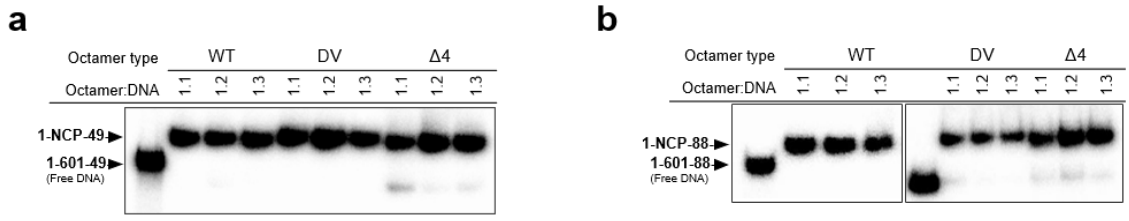


Figure B.4: Characterization of reconstituted mononucleosomes. Native PAGE (5%, 59:1 bisacrylamide:acrylamide) analysis of (a) **1-601-49** and (b) **1-601-88** reconstituted with increasing concentrations of the indicated histone octamer, yielding mononucleosomes **1-NCP-49** and **1-NCP-88,** respectively. The molar ratio of histone octamer to DNA template is indicated. WT: canonical octamers; Δ4: histone H4 tail deleted octamers; DV: histone H2A.Z/H3.3 dual-variant octamers.

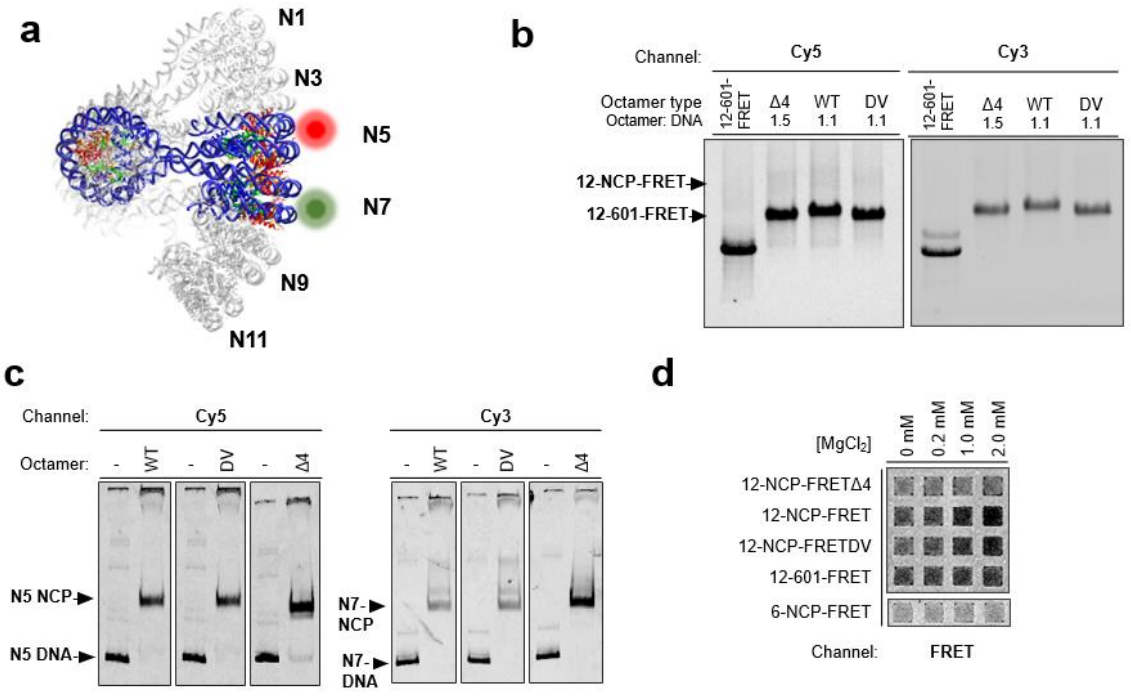


Figure B.5: Characterization of dye-labeled nucleosome arrays. (a) Structural representation of the Cy3/Cy5-labeled arrays (**12-NCP-FRET**) used for the FRET analysis. The Cy3/Cy5-labeled DNA (**12-601-FRET**) used to assemble these arrays was reported previously.[158] (b) Agarose gel electrophoresis (0.6%) analysis of **12-601-FRET** reconstituted with the indicated ratio of histone octamer to DNA template. Images from the Cy3 channel (532 nm ex.; 575 nm em.; right) and Cy5 channel (635 nm ex.; 665 nm em.; left) are shown. (c) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of **12-NCP-FRET** nucleosome arrays. Both naked DNA (**12-601-FRET)** and the indicated arrays were digested with either PflMI/BstXI (for release of N5) or DraIII/BstEII (for release of N7) and were run side-by-side to determine nucleosome occupancy at N5 and N7, respectively. WT: canonical octamers; Δ4: histone H4 tail deleted octamers; DV: histone H2A.Z/H3.3 dual-variant octamers. (d) Representative microplate images from the FRET experiments depicted in Figure 3.1c showing raw FRET data (F_{raw}) at different Mg^{2+} concentrations for the indicated arrays and DNA controls.

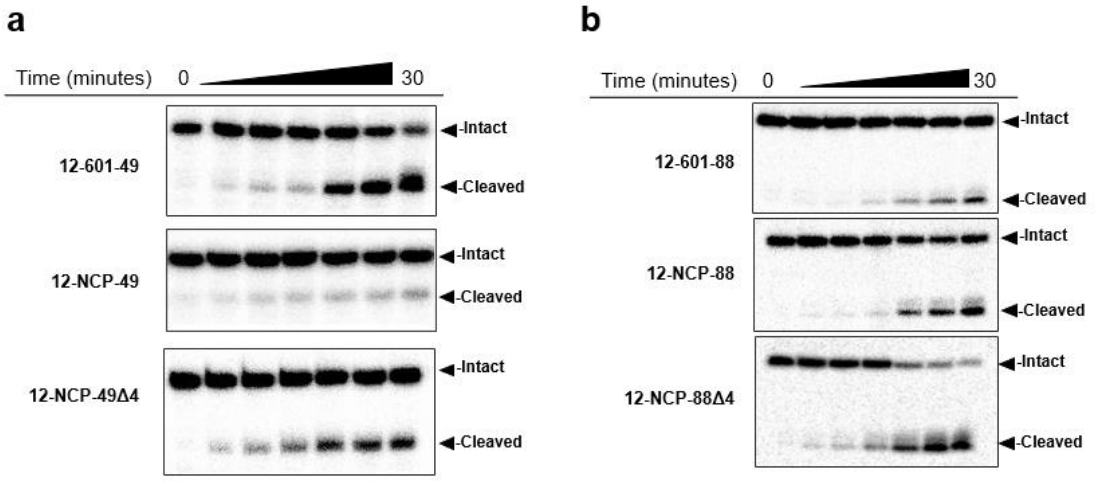


Figure B.6: Representative gel images of TDG digestion experiments on substrates containing either 5fC49 (a) or5fC88 (b). Reactions were carried out and analyzed as described in *3.3.10*.

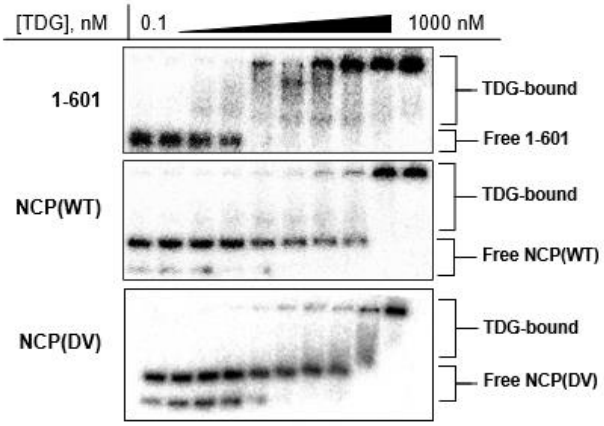


Figure B.7: Representative gel images from EMSA experiments between TDG (0.1 – 1000 nM)

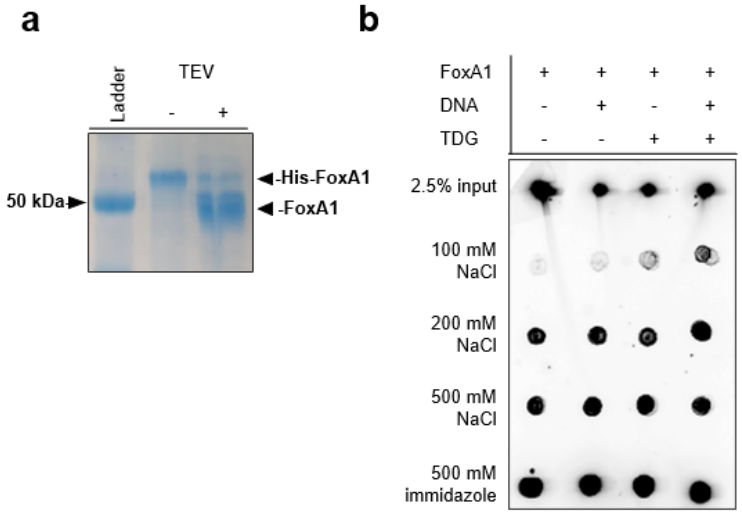


Figure B.8: TDG-FOXA1 pull-down data. (a) SDS PAGE analysis of His-tag removal from FOXA1 via TEV digestion. (b) Representative dot blot from the experiment described in *3.3.15*.

B.2 Supplementary Tables

Table B.1: Sequences of oligonucleotides described in this study. The underlined nucleotides correspond to locked nucleic acids (LNA).

Name	Sequence (5′ to 3′)
N5_5fC49	TGAGGATGTATATATCTGACGCGC/5fC/GGTGGAGC
N5_5fC88	GTGG/5fC/GGCTGAGGCCAAAATC
49-Capture	GCTCCACCGGCGCGTCAGATATATACATCCTCA
N5_5fC49-comp	GCTCCACCGGCGCGTCAGATATATACATCCTCA
N5_dU49	TGAGGATGTATATATCTGACGCGC/dU/GGTGGAGC
88-Capture	GCCGCCACGATTTTGGCCTCA
CpG-601-FWD	CTGGAGAATCCCGGTGCCGAGGCC
CpG-601-REV	TGAGGATGTATATATCTGACGCGCCGGTGGAGC
+51-Cy5 insert	TATATCTGACACGTUCCTGGAGTC/*C6 Amino-Mod dT*/AGG
FRET-capture	CCTAGACTCCAGGCACGTGTCAGATATA
-53-Cy3 insert	TCGTGAATTGG/*C6 Amino-Mod dT*/CC

C) APPENDIX

CHAPTER 4 SUPPLEMENTARY DATA

C.1 Supplementary Figures

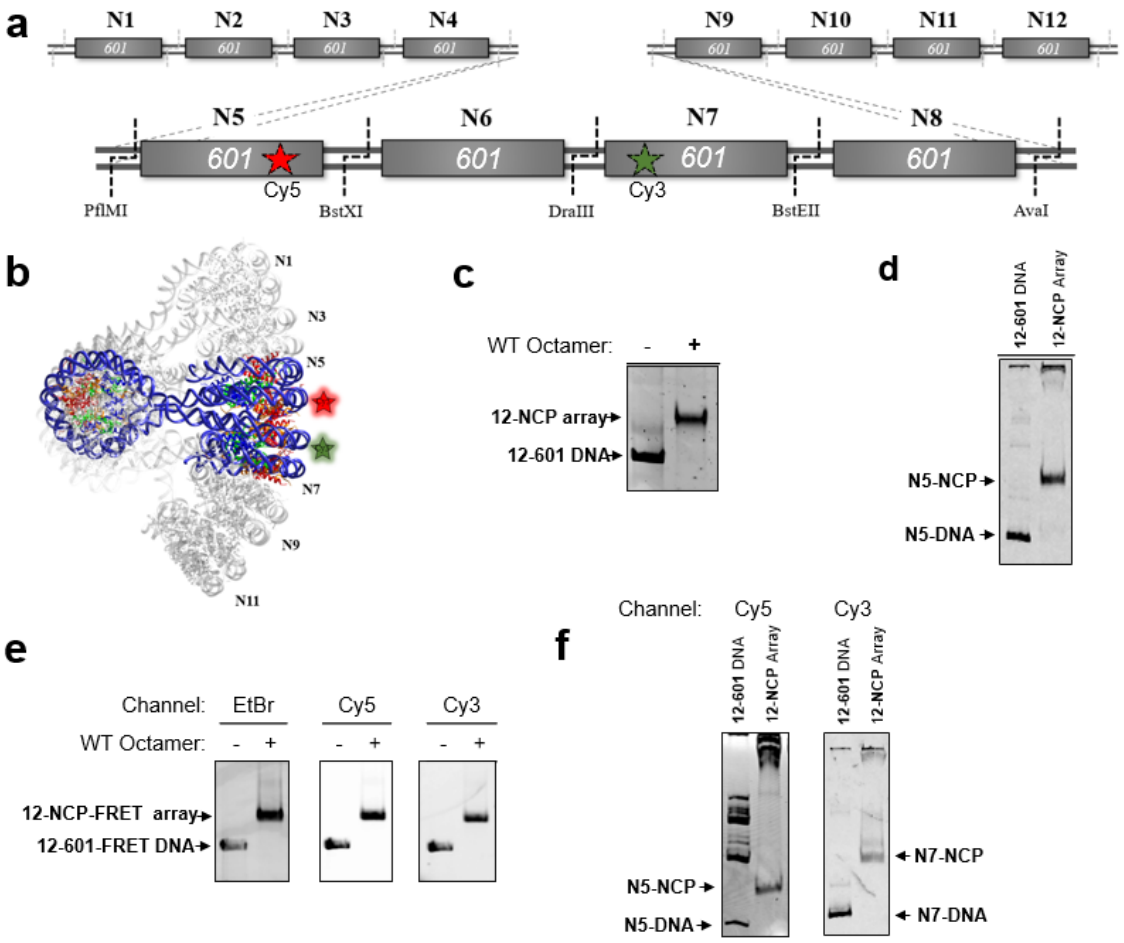


Figure C.1: The 12-mer nucleosome arrays used in this work. (a) Schematic of Cy5/Cy3-modified **12-601** DNA template used for FRET studies, referred to as **12-601-FRET** in Appendix C [117]. Assembly of this DNA is described in Chapters 2 and 3. (b) Structural representation of **12-NCP-FRET** arrays used for intra-fiber FRET analysis (pdb: 1zbb, emd: 2600). (c) Native gel analysis (0.6% agarose) of wild-type **12-NCP** and (e) **12-NCP-FRET** arrays reconstituted with wild-type (WT) histone octamer (molar ratios of 1.1 and 1.2, respectively). (d) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of **12-NCP** and (f) **12-NCP-FRET** arrays demonstrating complete nucleosome occupancy. For part (e), the same gel was imaged with the indicated fluorescent channel (EtBr, 532 nm ex.; Filter- LPG. Cy3, 532 nm ex.; Filter- LPG. Cy5, 635 nm ex.; Filter- LPR).

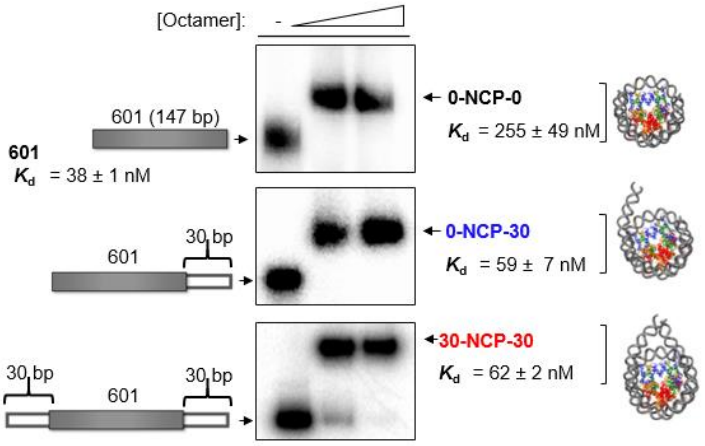


Figure C.2: Schematic representation (pdb: 1zbb) and characterization of the mononucleosomes used in this work. Complete reconstitution of all mononucleosome substrates was confirmed by native PAGE (5%, acrylamide:bisacrylamide) analysis. Dissociation constants for TDG binding are provided.

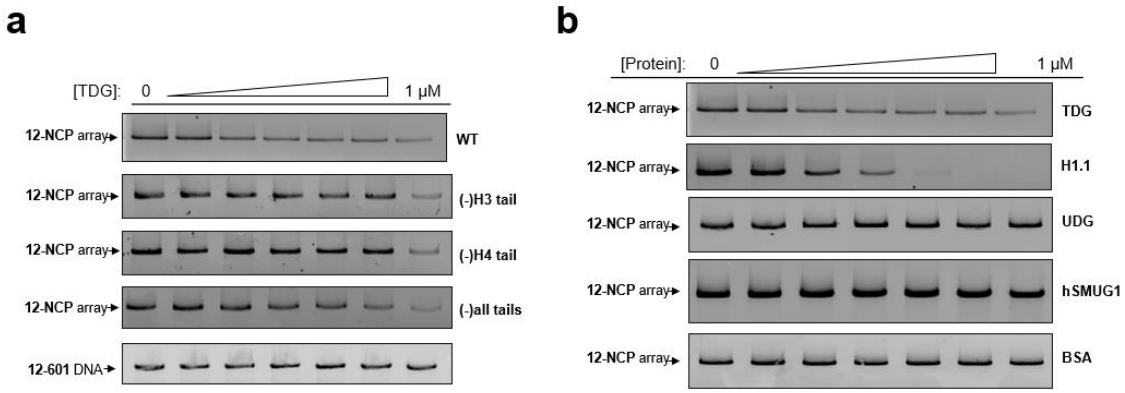


Figure C.3: Analysis of chromatin oligomerization via precipitation. (a) Representative gel images showing the soluble fraction of various tailless **12-NCP** arrays (or free **12-601** DNA) following treatment with increasing concentrations of TDG. Histone octamer composition is listed to the right of each gel image. (b) Representative gel images showing the soluble fraction of nucleosome arrays (**12-NCP**) following treatment with various proteins (indicated to the right) in the absence of Mg^{2+}. Reactions were carried out and analyzed as described in the section *4.3.9.*

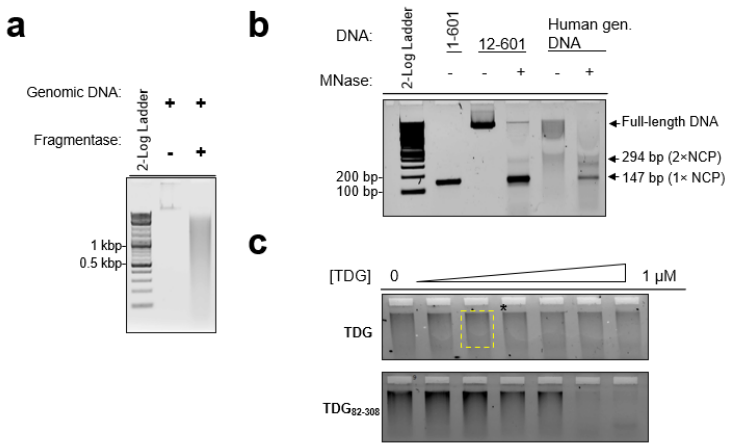


Figure C.4: Characterization and analysis of chromatin reconstituted from human genomic DNA. (a) Agarose gel (0.7%) gel analysis of human genomic DNA digested into 0.5-3 kbp fragments with NEB NEXT® dsDNA Fragmentase®, 2-log ladder refers to NEB cat. no. N3200. (b) Agarose gel analysis (0.6%) of chromatin reconstitution reactions using the fragmented genomic DNA depicted in part (a) and histone octamers containing Cy5-labeled $H2A_{N110C}$. A similarly labeled array reconstituted with **12-601** DNA was included for reference (**12-NCP-Cy5**). (c) Representative gel images showing the soluble fraction following treatment of chromatin reconstituted from genomic DNA with increasing concentrations of either TDG or TDG_{82-308}.

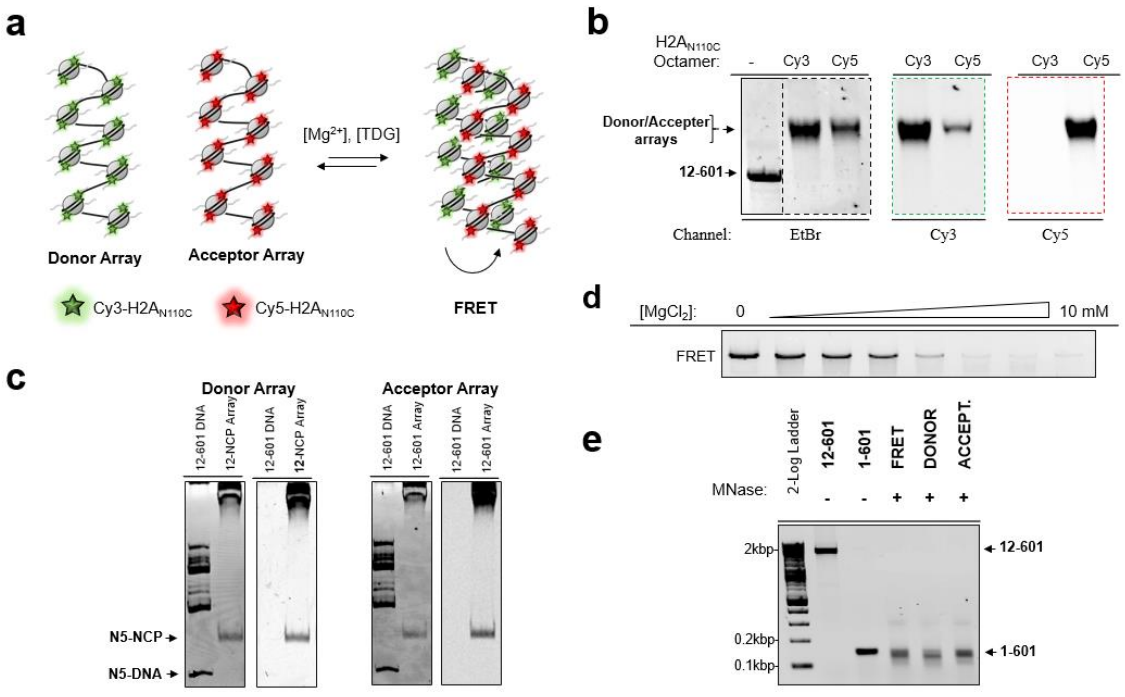


Figure C.5: Analysis of chromatin oligomerization via inter-fiber FRET. (a) Schematic depiction of the inter-fiber FRET assay. Two separate nucleosome arrays were labeled with either Cy3 (donor) or Cy5 acceptor) dyes via maleimide conjugation to histone $H2A_{N110C}$. Upon fiber oligomerization, which has been proposed to involve interdigitation of nucleosomes between different fibers[11], the dyes become close enough in space to allow efficient FRET. (b) Agarose gel (0.6%) analysis of donor and acceptor arrays reconstituted from **12-601** DNA template and histone octamers containing either Cy3- or Cy5-labeled histone $H2A_{N110C}$. The gel was visualized using the indicated fluorescent channel. (c) Restriction digestion analysis (5% native PAGE, 59:1 acrylamide:bisacrylamide) of donor and acceptor arrays demonstrating complete nucleosome occupancy. (d) Representative gel image showing the soluble fraction from a Mg^{2+} dependent (0 – 10 mM) oligomerization FRET experiment corresponding to the FRET analysis depicted in Figure 2c. (e) Agarose gel (1%) analysis confirming the digestion of **FRET-**, **DONOR-**, and **ACCEPTOR-** 12mer arrays to individual nucleosome particles via MNase. A fully intact **12-601** and **1-601 DNA** were included for reference.

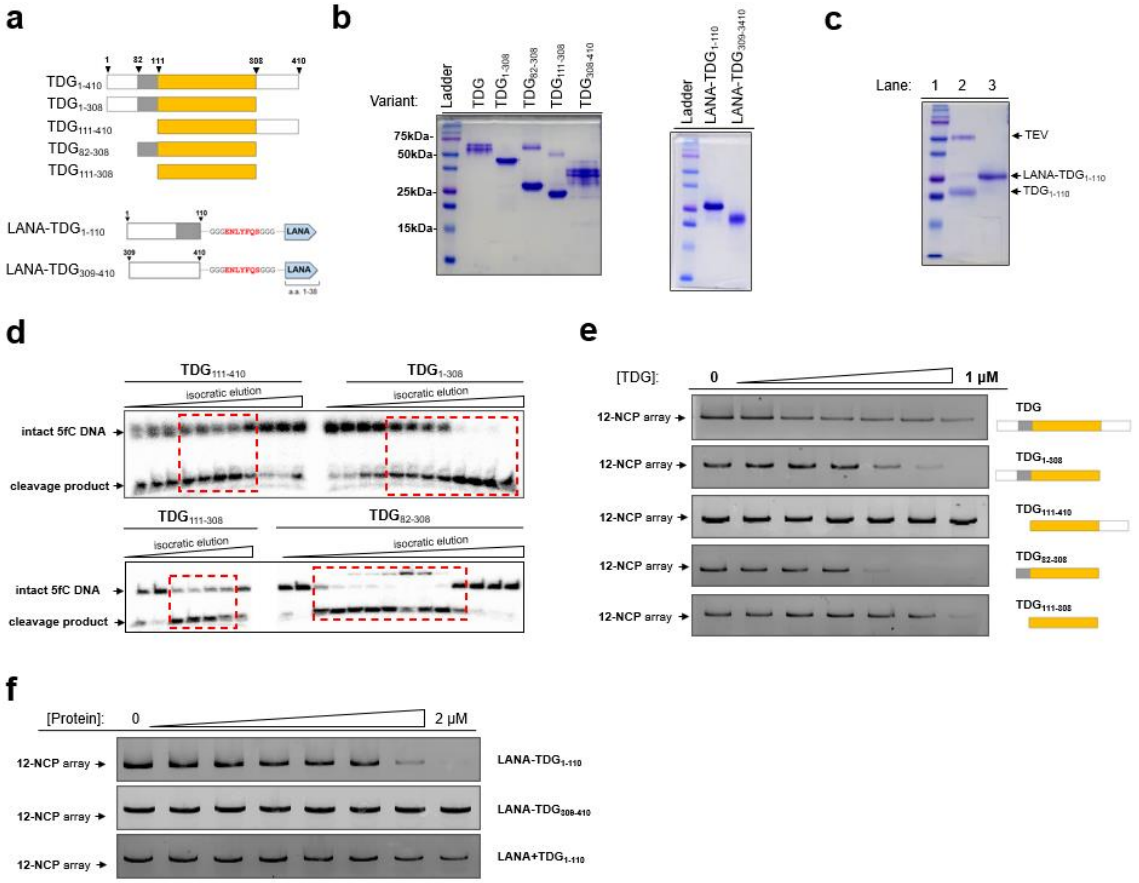


Figure C.6: TDG deletion and fusion proteins. (a) Schematic depiction of the TDG deletion mutants and LANA-TDG fusion proteins used in this work. LANA represents amino acids 1-38 of the latency-associated nuclear antigen 1 protein. (b) Denaturing SDS PAGE (15%, 29:1 acrylamide:bisacrylamide) analysis of purified TDG deletion variants and LANA-TDG proteins stained with Coomassie Brilliant Blue. "Ladder" is the Bio-Rad Precision Plus Pre-stained ladder (cat. no. 1610374). (c) Denaturing SDS PAGE (15%, 29:1 acrylamide:bisacrylamide) analysis of LANA-TDG$_{1\text{-}110}$ following treatment with TEV protease to separate the NTD residues 1-110 from the LANA peptide. Lane 1: Bio-Rad Precision Plus Pre-stained ladder; Lane 2: TEV digested LANA-TDG$_{1\text{-}110}$ reaction, Lane 3: LANA-TDG$_{1\text{-}110}$. (d) TDG activity screening assay conducted on fractions from the final chromatograph step of TDG purification. (e) Representative gel images showing the soluble fraction of **12-NCP** arrays following treatment with TDG deletion mutants depicted in (a). (f) Representative gel images showing the soluble fraction of **12-NCP** arrays following treatment with LANA-TDG proteins depicted in (a).

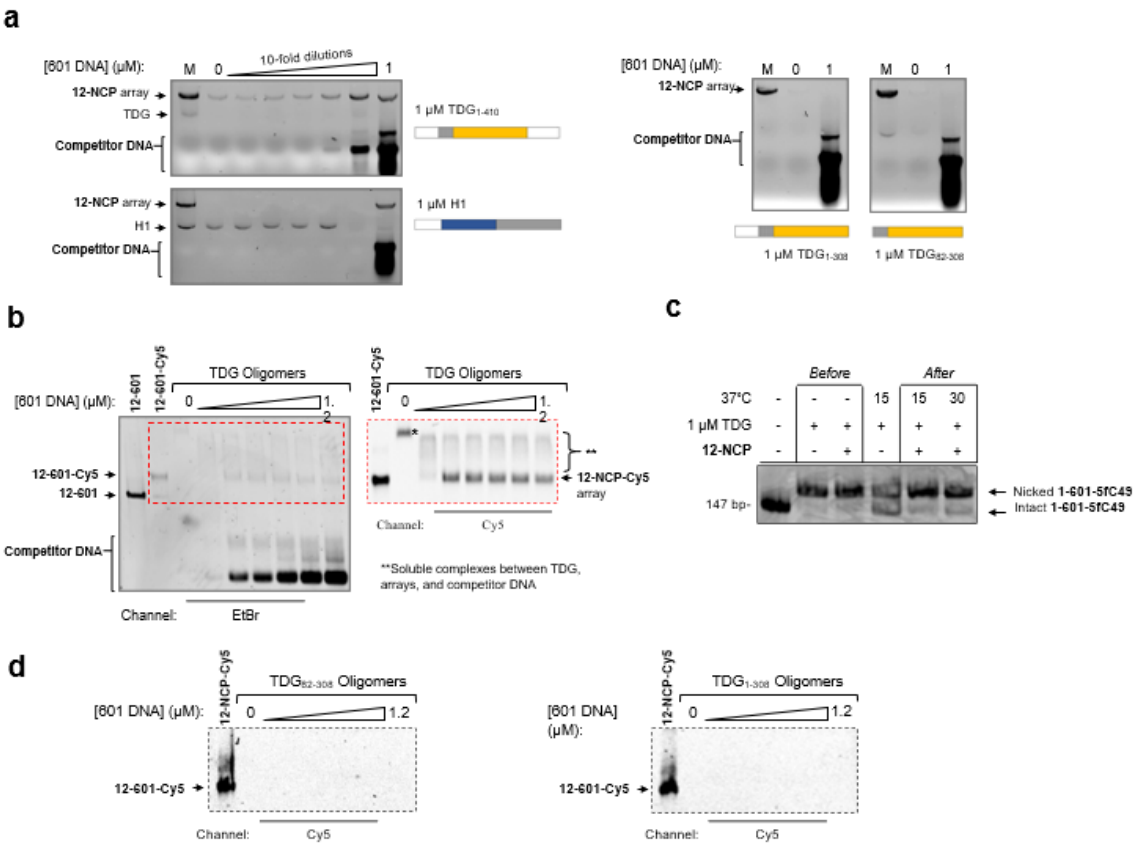


Figure C.7: Reversal of insoluble TDG-chromatin oligomers by DNA. (a) Agarose gel analysis of DNA reversibility experiments in which **12-NCP** aggregates, generated with the indicated protein, were exposed to 0-1 μM DNA prior to precipitation. (b) Representative gel images showing the soluble fraction from experiments in which **12-NCP-Cy5** arrays were first aggregated by full-length TDG and subsequently exposed to increased concentrations of 207 bp 601 DNA duplex. Gels were imaged either by EtBr staining (left) or Cy5 fluorescence (right). Asterisk indicates the fractions (~20%) of TDG-bound arrays that remain soluble following the addition of 1 μM full-length TDG. (c) Chromatin aggregates generated by $TDG_{1\text{-}308}$ and $TDG_{82\text{-}308}$ are not reversible by DNA. Same analysis as in part (b) but with only the Cy5 channel shown for simplicity. (d) TDG remains catalytically active following a cycle of chromatin condensation. Activity was determined by adding 325 fmol of a 5fC-containing 601 DNA (**1-601-5fC49**) to pre-formed TDG-chromatin aggregates *before* or *after* they were incubated at 37 ˚C for either 15 or 30 minutes. To facilitate cleavage of abasic sites generated by TDG, APE1 (35 nM) and 0.1 mM $MgCl_2$ was included in the final reaction mixture.

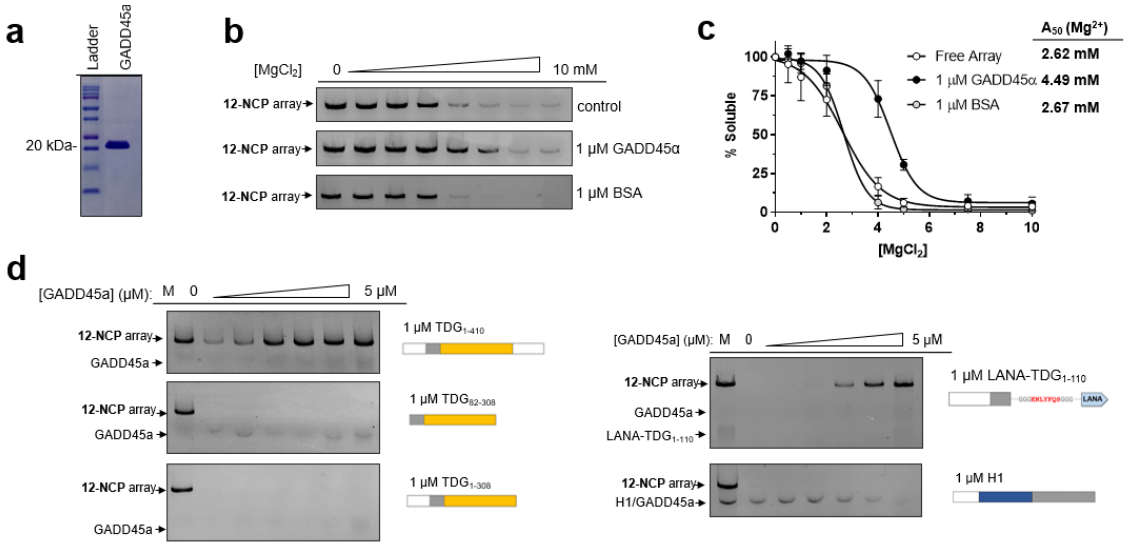


Figure C.8: GADD45a antagonizes TDG- and histone-dependent chromatin condensation. (a) Denaturing SDS PAGE (15%) analysis of recombinant human GADD45a protein stained with Coomassie Brilliant Blue. Ladder is the Bio-Rad Precision Plus Pre-stained ladder (cat. no. 1610374). (b) Representative gel images showing the soluble fraction from standard Mg^{2+}-precipitation assays in which 1 μM GADD45a or BSA was included. (c) Solubility plot from the experiments depicted in part (b). The average and standard deviation of n > 3 experiments is shown. (d) Representative gel images showing the soluble fraction from experiments in which **12-NCP** arrays were first aggregated by the indicated TDG variant or H1 (to the right) and subsequently exposed to increased concentrations of GADD45a. M: **12-NCP** array marker.

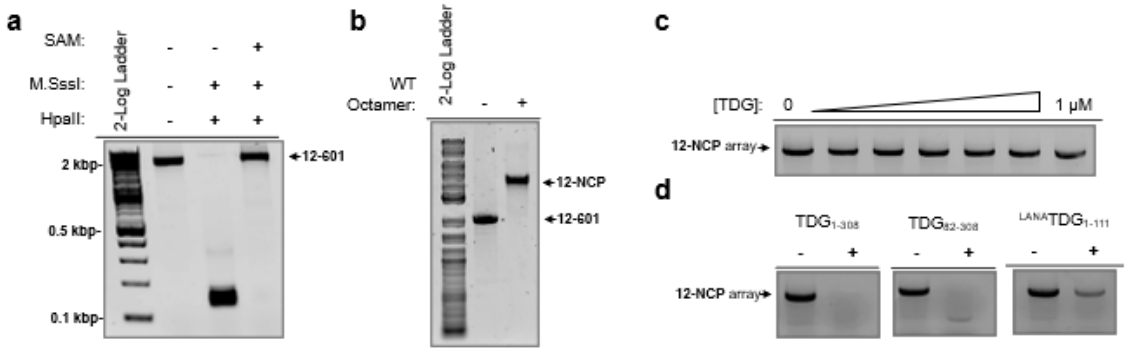


Figure C.9: Chromatin condensation by TDG is inhibited by DNA methylation. (a) Agarose gel electrophoresis (0.7%) analysis of CpG hypermethylated **12-601** DNA digested by HpaII. DNA treated with M.SssI and SAM was resistant to HpaII cleavage, while control reactions (lacking SAM) were digested completely by HpaII, confirming complete CpG methylation. (b) Agarose gel analysis (0.6%) of chromatin reconstitution reactions using the HpaII-resistant (hypermethylated) DNA depicted in part (a). (c,d) Representative gel images showing the soluble fraction of methylated **12-NCP** arrays following treatment with increasing concentrations of full-length TDG (c) or 1 μM truncated TDG variants (d).

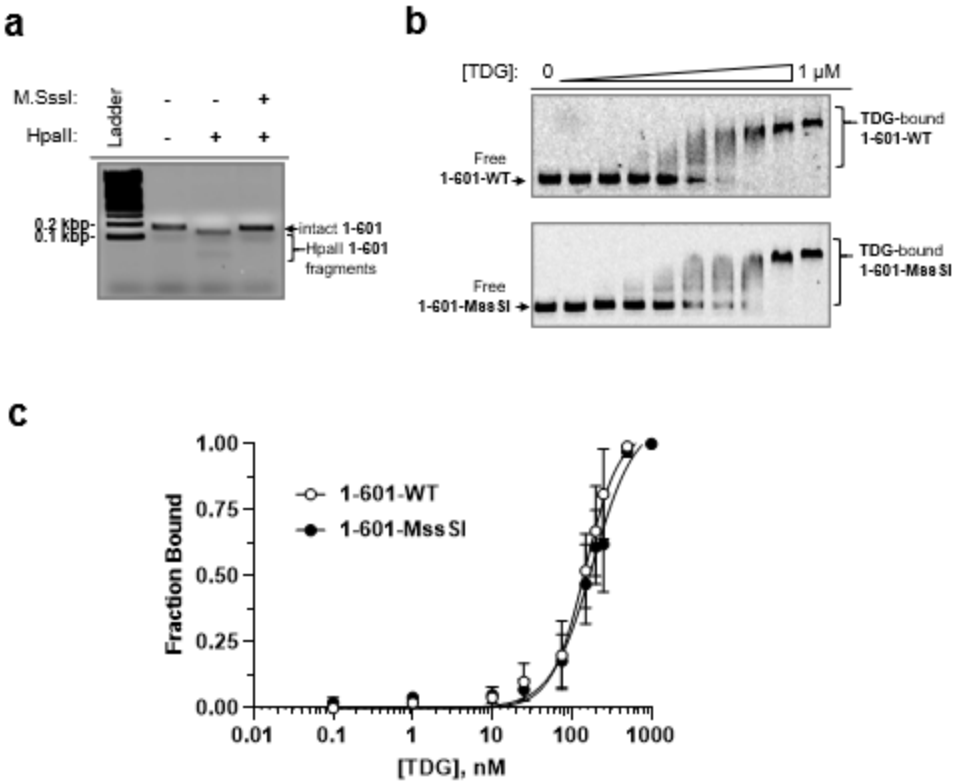


Figure C.10: TDG binds methylated 601 DNA similarly to unmethylated 601 DNA. (a) Agarose gel image confirming the resistance of MssSI-treated **1-601** (207 bp) to HpaII cleavage. (b) Representative agarose gel (0.8%) gel images of full-length TDG with wild-type and MssSI-treated **1-601** DNA. (c) Saturation plots of triplicate binding experiments like the ones depicted in part (b). TDG binds wild-type **1-601-WT** with a K_d of 148 nM and methylated **1-601-MssSI** with a K_d of 180 nM. At 1 µM TDG

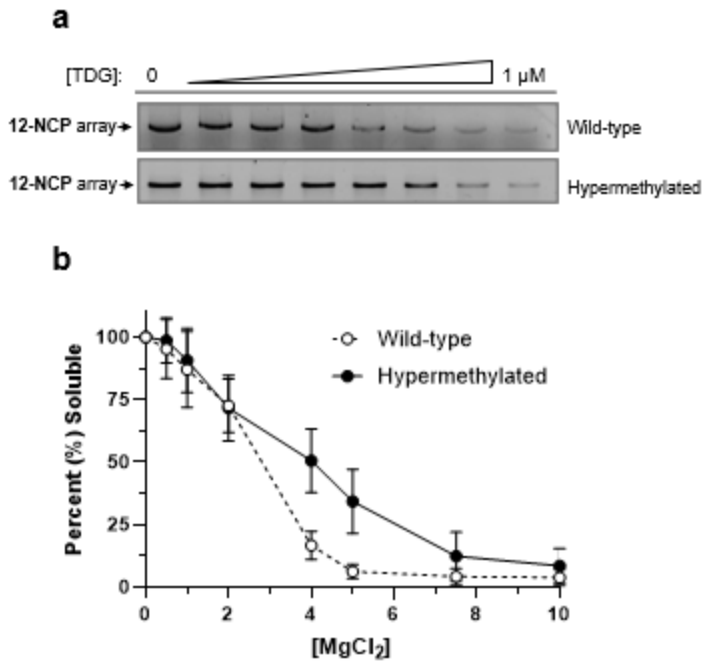


Figure C.11: DNA methylation reduces Mg^{2+}-induced precipitation of 12-mer nucleosome arrays. (a) Representative gel images showing the soluble fraction of wild-type and hypermethylated **12-NCP** arrays following treatment with full-length TDG. (b) Solubility plot of the experiments depicted in (a).

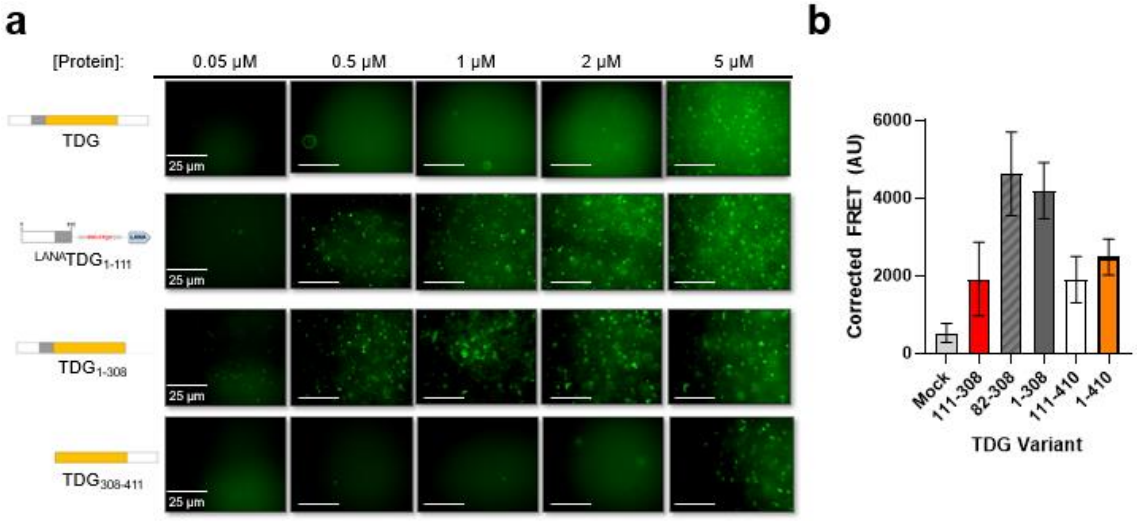


Figure C.12: TDG regulates chromatin LLPS in a concentration-dependent manner. (a) Fluorescence microscopy images of 20 nM Cy3-chromatin treated with the indicated TDG variants. (b) Inter-fiber FRET data demonstrating all TDG variants are capable of inducing chromatin oligomerization at 1 µM. FRET experiments as described in Figure 4.4 and *4.3.12*.

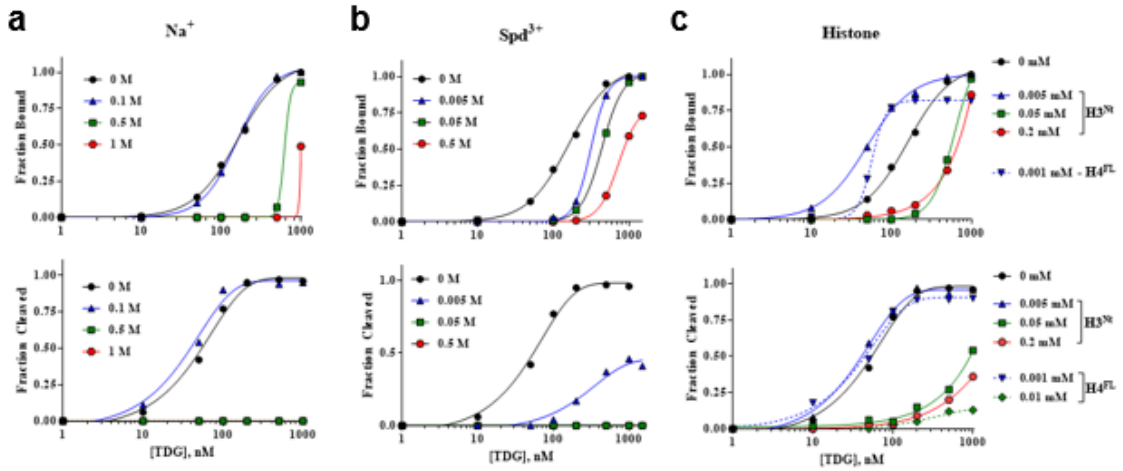


Figure C.13: Simultaneous look at the effects of mono- and polyvalent cations on TDG-binding and catalysis of 5fC. (a) Fraction bound (top) and fraction cleaved (bottom) following exposure of TDG-bound **1-601-5fC49** (0- 1µM TDG, 5 nM DNA) in the presence of the indicated concentration of NaCl. Data for 0 M control represents average of 4 replicates, all other data represent a single replicate. The fraction bound and fraction cleaved by analyzing aliquots from the same mixture via native or denaturing PAGE, respectively. For denaturing analysis, abasic sites were cleaved with NaOH (0.1 M) prior to gel-loading. (b) Same for Spermidine. (c) Same for the N-terminal tail of H3 (MW= 2054). (d) Same for full-length H4 protein; due to solubility not all H4 data points could be plotted. High ionic strength buffers diminish TDG catalysis by disrupting electrostatic interactions required for DNA-binding, arginine finger intercalation, and oxocarbenium stabilization [296]. We observed a complete loss in 5fC hydrolysis by TDG at $\geq$0.5 M Na^+ but, interestingly, found that TDG remained bound to **1-601-5fC49** even in the presence of 1 M NaCl (~K_d= 1000 nM). In general TDG exhibited similar trends in sensitivity to other cations, like: trivalent spermidine (Spd^{3+}), the N-terminal tail domain of histone H3 ($H3^{Nt}$), and the full-length H4 protein ($H4^{FL}$), each of which's presence inhibited TDG catalysis in a concentration-dependent manner. In all cases the DNA-binding capabilities of TDG were more resilient, most notably in the presence of polyamines. LLPS data indicates TDG-chromatin liquids have a high ionic character, and the findings presented above demonstrates TDG can remain stably bound to DNA under high ionic strength buffers, but possibly not in the capacity of an active glycosylase. Further investigation is required to determine if TDG is catalytically active in LLPS droplets, or if the ionic properties of NTD-mediated condensates inhibits catalysis.

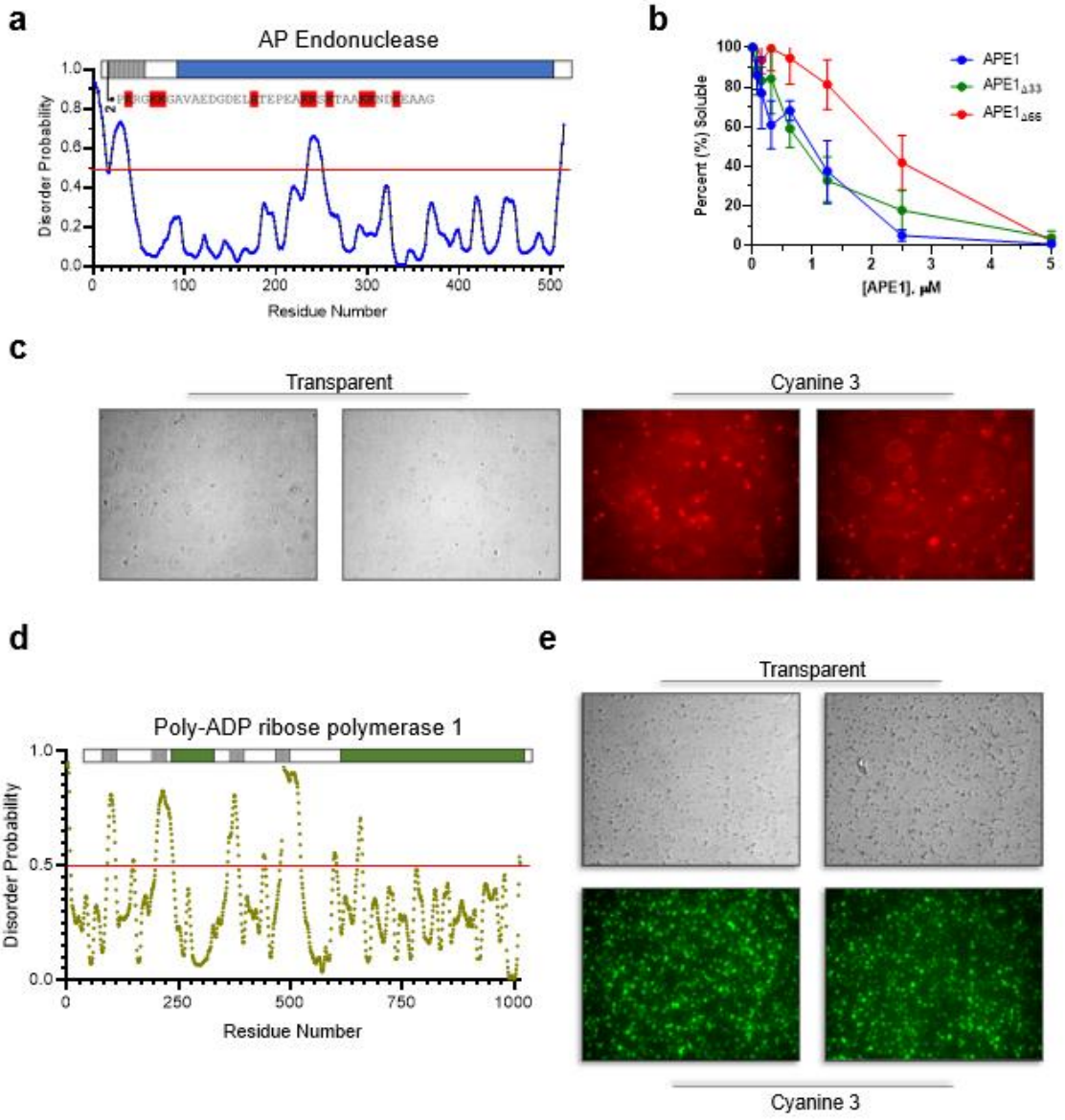


Figure C.14: DNA repair proteins induce chromatin LLPS. (a) Disorder probability for APE1 [236]. (b) Solubility plot for nucleosome arrays (5 nM) treated with the indicated concentration of APE1. (c) Fluorescent microscopy images of APE1-induced LLPS droplets generated with 5 μM APE1, 20 nM chromatin, and 2 mM Mg^{2+}in nucleosome buffer. (d) Disorder probability for PARP-1 [236]. (e) Fluorescent microscopy images of PARP-1-induced LLPS droplets generated with 2.5 μM protein, 20 nM chromatin, and 2 mM Mg^{2+}.

D) APPENDIX

CHAPTER 5 SUPPLEMENTARY DATA

D.1 Supplementary Figures

a

	Name	Sequence
D-RNA	D-$(GGAA)_{10}$	5′-Cy3-AA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA
D-RNA	D-$(A)_{40}$	5′-Cy3-AA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA
D-RNA	D-$(G3A4)_4$	5′-Cy3-AA-AAA-GGGAAAA-GGGAAAA-GGGAAAA-GGAAAA
DNA	D-$(dGGAA)_{10}$	5′-Cy3-dAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA-dGdGdAdA
L-RNA	L-$(GGAA)_{10}$	5′-Cy3-AA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA-GGAA
L-RNA	L-$(A)_{40}$	5′-Cy3-AA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA-AAAA
L-RNA	L-$(G3A4)_4$	5′-Cy3-AA-AAA-GGGAAAA-GGGAAAA-GGGAAAA-GGAAAA

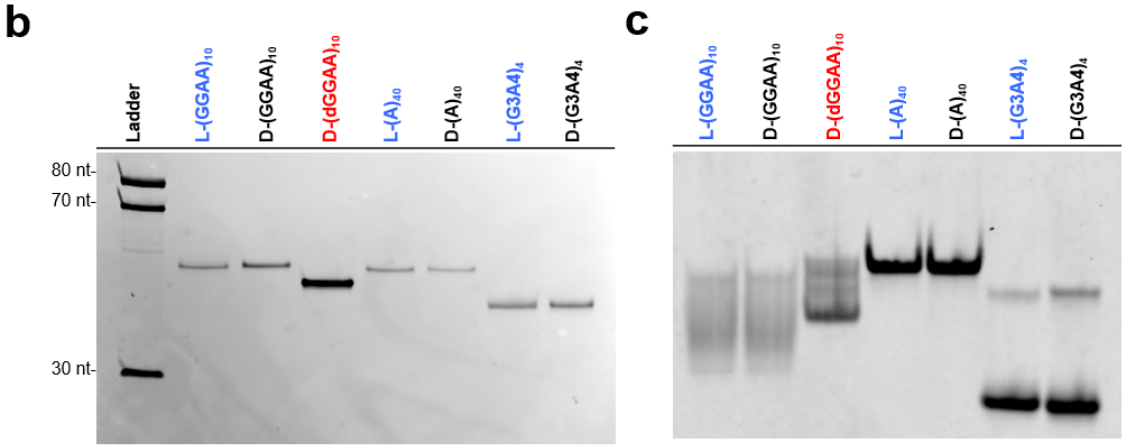
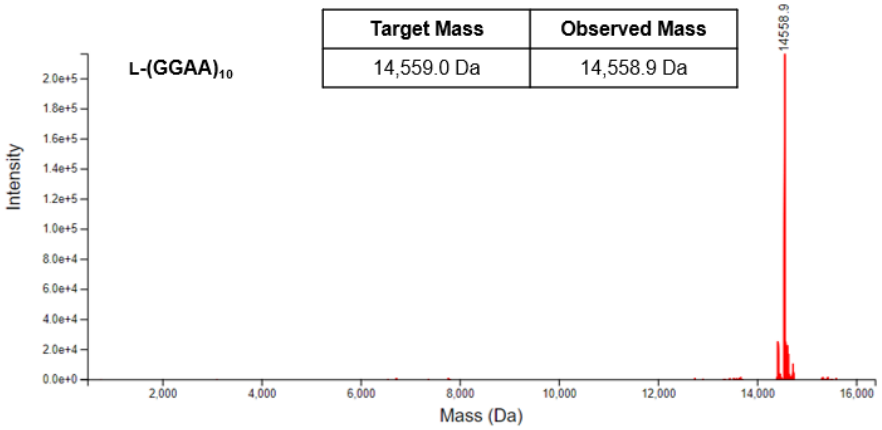
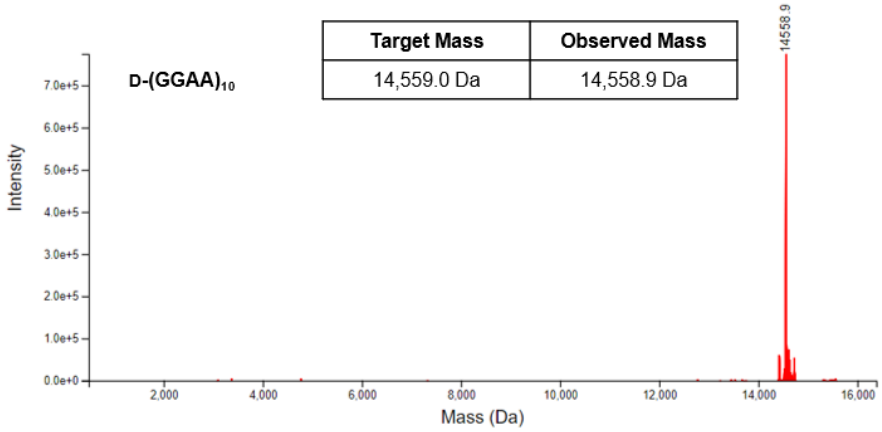
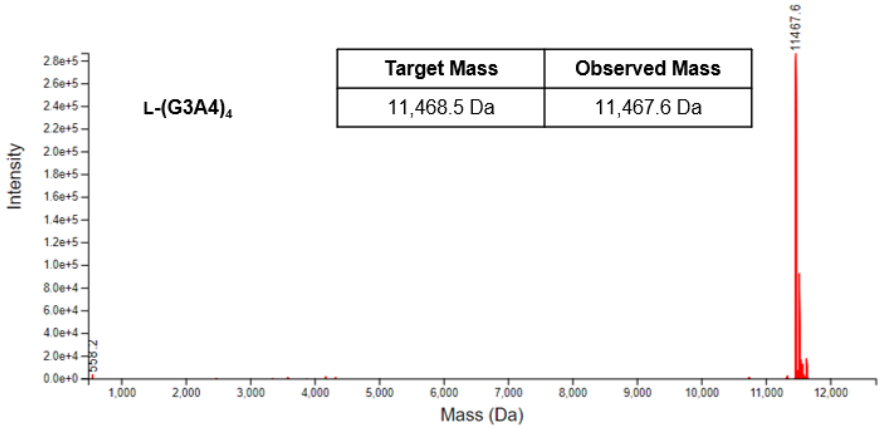
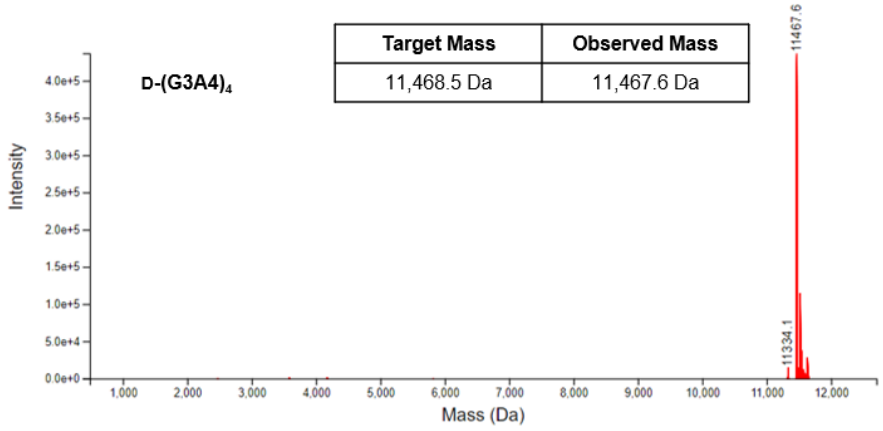


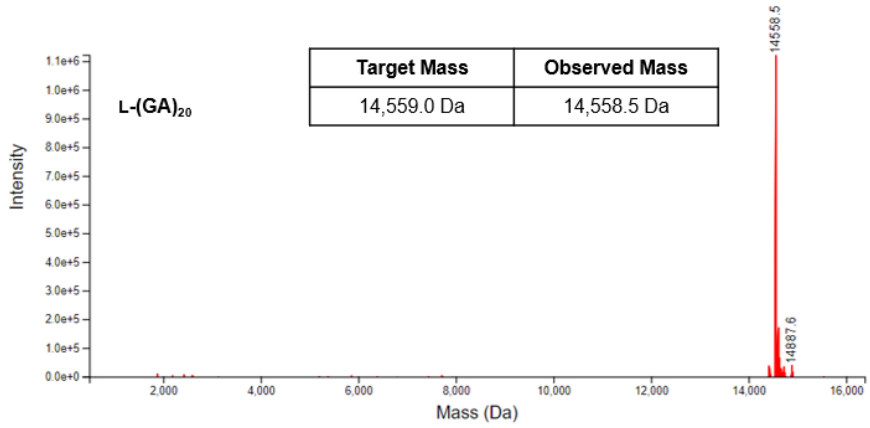
Figure D.1: Oligonucleotides used in this work. (a) Sequences of oligonucleotides used for binding EMSAs and competition experiments. Terminal D-deoxyribose residues (D-dA) on the RNA strands are underlined. (b) Denaturing PAGE analysis of the Cy3-labeled oligonucleotides presented in (a) (10%, 29:1 acrylamide:bisacrylamide). (c) Native PAGE analysis (10%, 29:1 acrylamide:bisacrylamide) of the same oligonucleotides in panel (b). The running buffer (1× TBE) was supplemented with 10 mM KOAc. The increased electrophoretic mobility of $(GGAA)_{10}$ and $(G3A4)_4$ relative to $(A)_{40}$ is indicative of the G4 structure formation by these oligonucleotides in the presence of K^+.

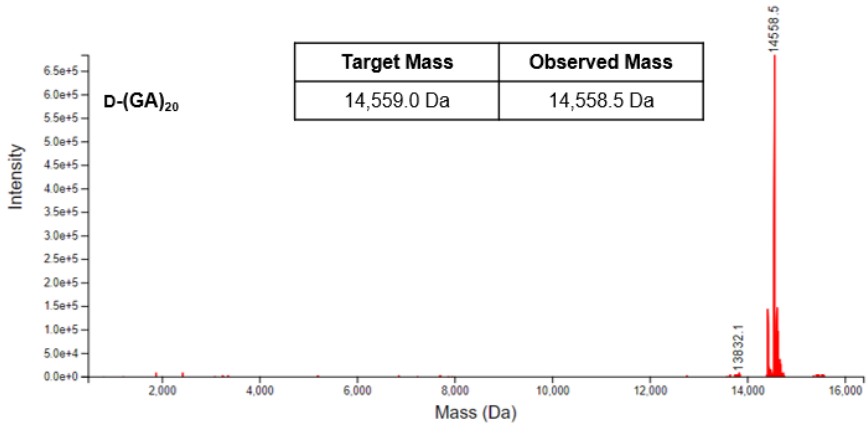

L-(GGAA)10
Target Mass
Observed Mass
14,559.0 Da
14,558.9 Da
14558.9
Intensity
Mass (Da)


D-(GGAA)10
Target Mass
Observed Mass
14,559.0 Da
14,558.9 Da
14558.9
Intensity
Mass (Da)


L-(G3A4)4
Target Mass
Observed Mass
11,468.5 Da
11,467.6 Da
11467.6
Intensity
Mass (Da)


D-(G3A4)4
Target Mass
Observed Mass
11,468.5 Da
11,467.6 Da
11467.6
11334.1
Intensity
Mass (Da)


L-(GA)20
Target Mass
Observed Mass
14,559.0 Da
14,558.5 Da
14558.5
14887.6
Intensity
Mass (Da)


D-(GA)20
Target Mass
Observed Mass
14,559.0 Da
14,558.5 Da
14558.5
13832.1
Intensity
Mass (Da)

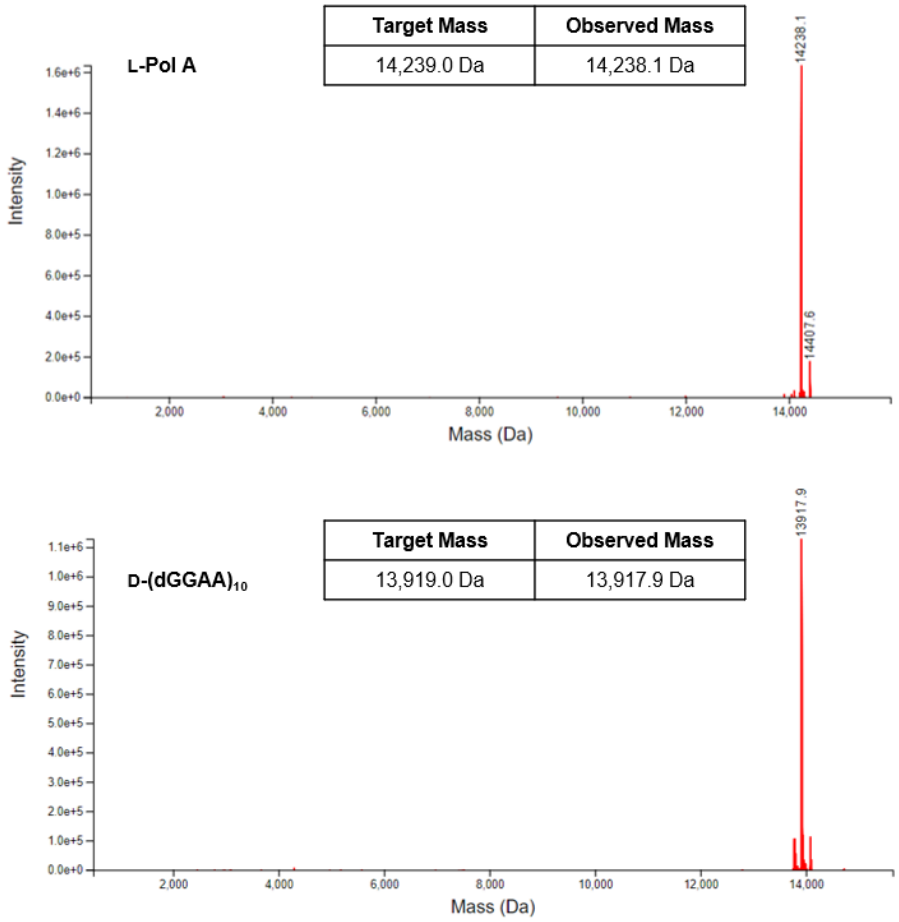


Figure D.2: ESI mass spectra of the indicated oligonucleotides.

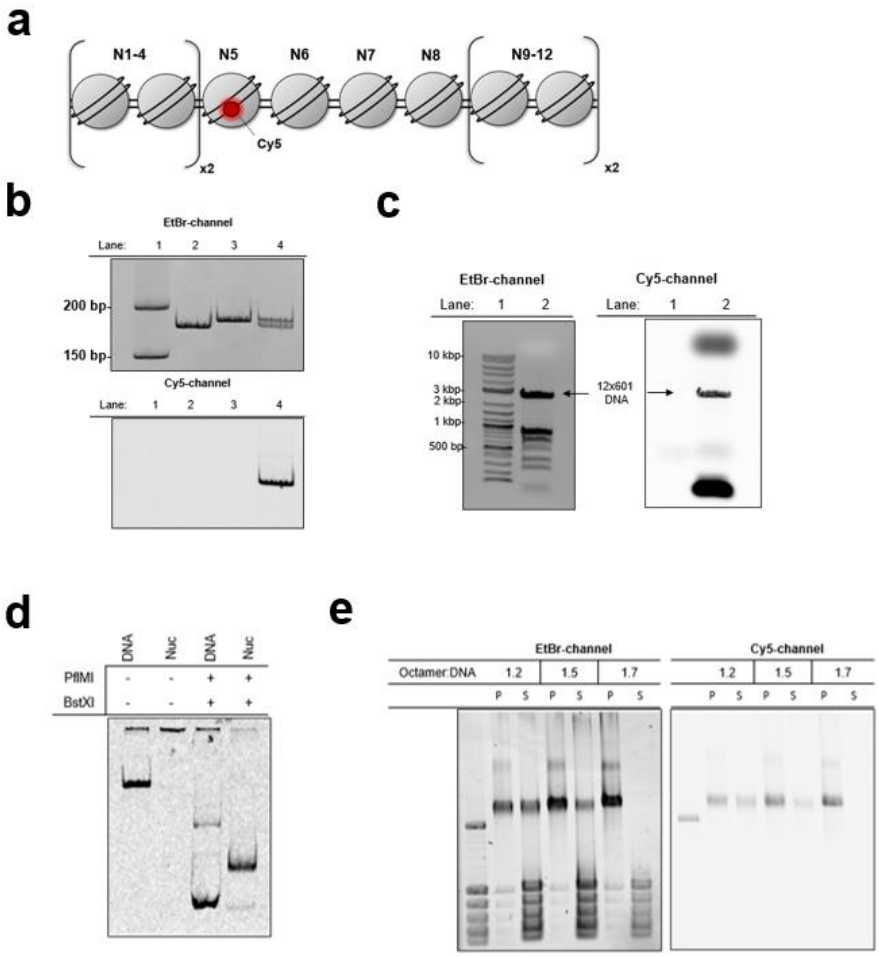


Figure D.3: Assembly of Cy5-labelled 12-mer oligonucleosome arrays. (a) Schematic of Cy5-labeled array employed in the PRC2 binding and competition assays. (b) Insertion of the Cy5 dye containing oligonucleotide into N5 was confirmed by 10% native PAGE (29:1, acrylamide:bisacrylamide). Lane 1, ladder; lane 2, unmodified N5 DNA fragment; lane 3, nicked N5 DNA fragment; lane 4, Cy5-labeled N5 DNA fragment following the strand exchange process. (c) Representative gel image depicting the removal of **Cy5-12-601** DNA from its parent plasmid. (d) Restriction enzyme digest analysis of the Cy5-containing N5 fragment (5%, 59:1 acrylamide:bisacrylamide) for the 1.5 octamer:DNA ratio depicted in part (e). Both naked (DNA) and reconstituted (Nuc) 12-mer arrays were digested similarly and their corresponding N5 fragments analyzed side-by-side. (e) Agarose gel depicting reconstituted Cy5-labeled oligonucleosome arrays using the indicated octamer:DNA ratio. Reconstitutions were precipitated with Mg^{2+} to remove free DNA and aliquots from the re-suspended nucleosome pellets (P) and the supernatant (S) were compared for array saturation.

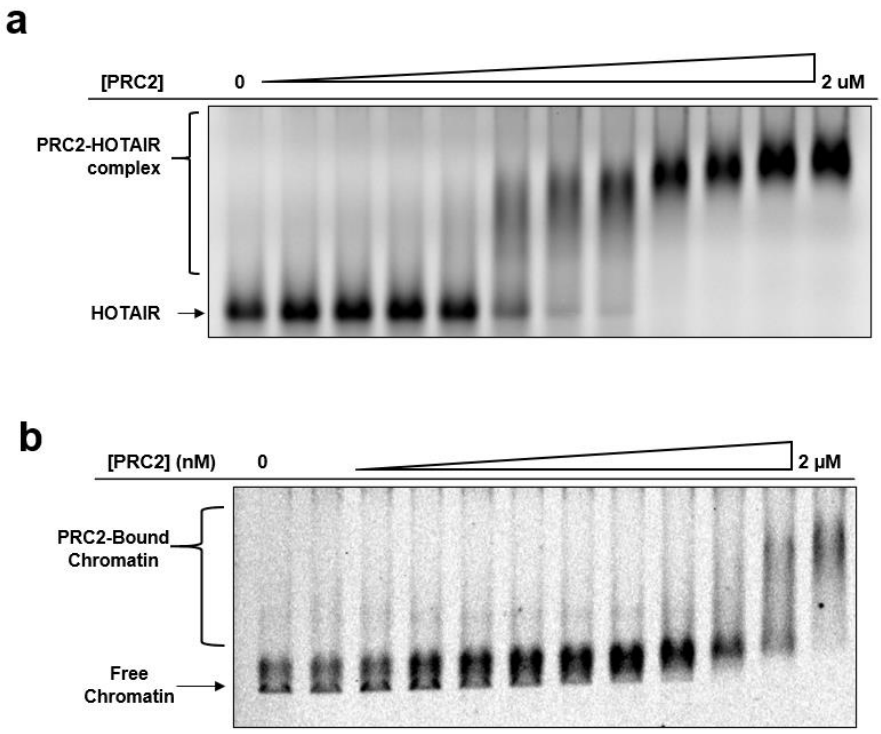


Figure D.4: Analysis of PRC2-binding to endogenous targets: (a) EMSA gel for PRC2 and Cy5-labelled HOTAIR residues 1-300 (see Figure 5.6 and section *5.4.5*). (b) EMSA gel for PRC2 and the Cy5-labelled chromatin depicted in Figure D.3. HOTAIR- and chromatin-binding was essentially conducted as all other RNA-binding experiments (see *5.4.5* and *5.4.7* for details), except reactions were analyzed with 0.5% agarose.

E) APPENDIX

POLY ADP-RIBOSE POLYMERASE-1 PRELIMARY DATA

E.1 Supplementary Text

E.1.1 Project Overview

Poly ADP-ribose polymerase (PARP-1) is a 1,014 amino acid, highly abundant, nuclear protein whose primary role is to maintain genome integrity through the detection of single- and double-strand DNA breaks; PARP-1 is also involved in the recruitment of Base Excision Repair (BER) proteins to apurinic/apyrimidinic (AP) sites. PARP-1 generates Poly ADP-ribose (PAR) polymers from NAD^+ in response to DNA damage; these PAR molecules either remain bound to PARP-1, or are transferred to nucleophilic resides on histone tails, and serve as molecular signals/scaffolds for repair proteins. The goal of this project is to study the impact of Poly-ADP ribose (PAR) on the repair of AP lesions in reconstituted chromatin.

An essential component of this work is the ability to generate AP lesions on command. AP sites are highly labile and can form unwanted DNA-protein cross-links during reconstitution and in chromatin (Figure E.1). With this in mind, an *O*-nitrobenzyl protected AP phosphoramidite was synthesized using an established protocol [121] and incorporated into short oligonucleotides compatible with PCR, as well as our plug-and-play chromatin, (Chapter 2 and E.1). We pursued several methods to achieve tethering of PAR chains in proximity to AP lesions in the context of NCPs and nucleosome arrays.

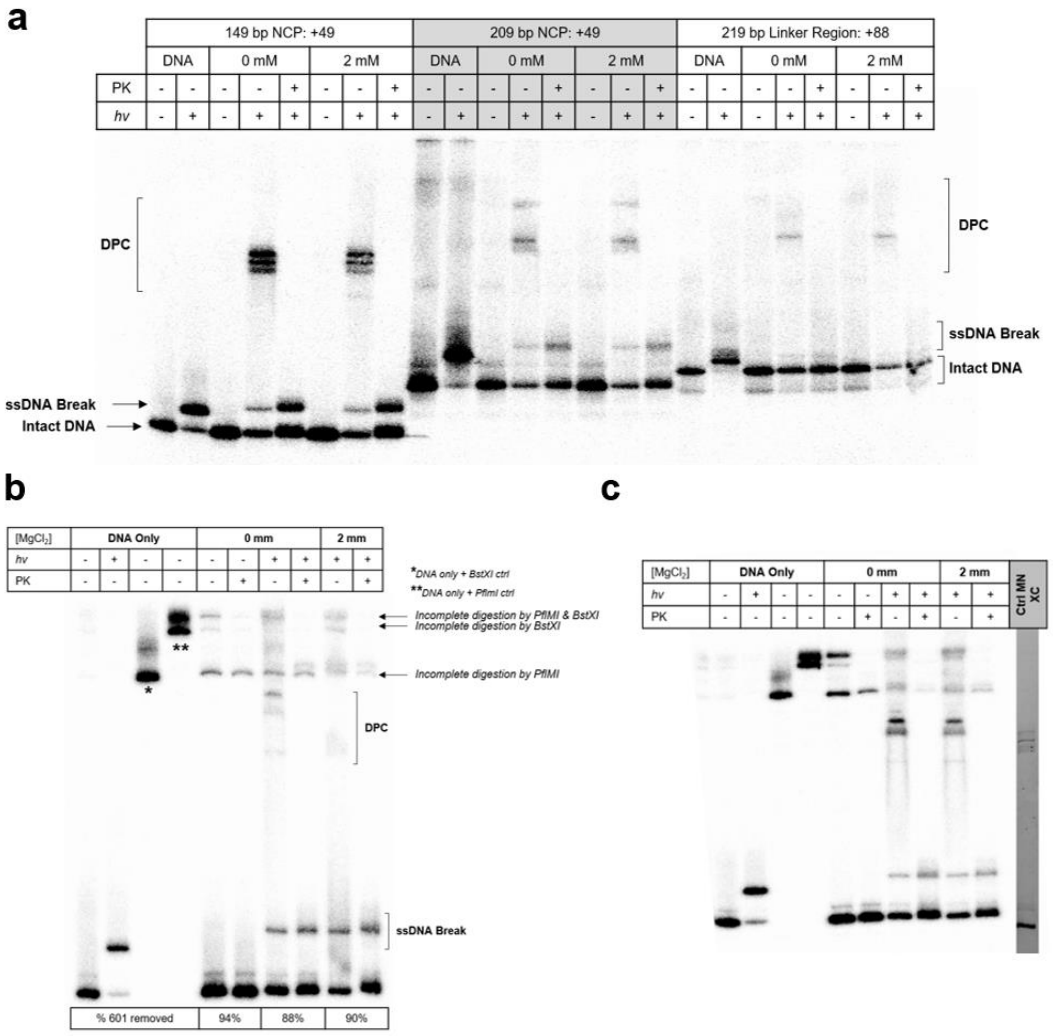


Figure E.1: DNA-histone cross-links monitored by denaturing PAGE. All experiments took advantage of different constructs derived from the *ONV*-protected AP precursor described in Figure A.6. (a) DNA-histone cross-links generated via photolysis of the indicated mononucleosome substrate. (b) DNA-histone cross-links generated via photolysis of chromatin having an AP lesion at the +49 position. (c) Replicate of experiment depicted in part (b). For cross-linking experiments, samples were prepared in 10 uL reactions containing 10 mM $NaCNBH_3$ and either 1 mM EDTA or 2 mM $MgCl_2$ in nucleosome buffer. The solutions were exposed to 350 nm light for 20 minutes at 4°C to generate an AP lesions and then incubated for 12 hours at 37°C. Reactions were quenched with 10 mM final $NaBH_4$ at room temperature for 30 minutes. NCPs reactions were split in half and digested with Proteinase K or quenched with loading buffer. Following $NaBH_4$ quenching, chromatin samples were heated at 60°C for 10 minutes in the presence of 0.1% SDS, NaOAc/EtOH precipitated, then resuspended in a 20 uL restriction digest reaction using the proper enzymes and 1x neb buffer for 5 hours at 37°C . Chromatin digests were then split in half and processed same as the mononucleosomes. All gels were 10% (29:1) SDS Tris/Glycine and were run either at 180 V for 14 hours or 400 V for 4 hours and exposed with a phosphor imaging screen and visualized using Typhoon imager.

E.1.2 PAR synthesis and purification

Poly ADP-ribose (PAR) polymers have been prepared enzymatically using PARP-$1_{1\text{-}1014}$ and PARP-$1_{379\text{-}1014}$, in combination, at ratios optimized to maximize yield. Both PARP-1 enzymes were expressed in DH5α *Escherichia coli* cells using established protocols. Initially, PAR-synthesis reactions were conducted with crude cell lysates, but this method proved to be ineffective at producing the desired quantity and size distribution of PAR chains. These experiements did, however, reveal that PARP-1 proficiency decreases when provided FAM-labelled-NAD^+ compared to its natural substrate (data not shown). His-tagged PARP-$1_{379\text{-}1014}$ was obtained in good yield (~32 mg/4 L expression) after purification with an Agarose Ni^{2+} gravity flow column. Full length, His-PARP-1 was obtained in significantly lower yield (5.86 mg) and required Ni^{2+}-enrichment and Heparin purification. PAR-synthesis reactoins were optimized *in vitro* by varying the concentrations of histones, NAD^+, activated calf thymus DNA, and the two pure/semi-pure PARP-1 enzymes. Conditions were evaluated qualitiatively by comparing the band intensities and mobility in 20%, 79:1 denaturing PAGE gel and quantitatively using LC-MS (Figure E.2)

PAR molecules of defined lengths were isolated from PARylation reactions by hydrolysis of the polymers from protein acceptors with 100mM CHES buffer (pH= 9), 10mM EDTA. Desired PAR-chain lengths were subsequently obtained via purification by denaturing PAGE (Appendix E.2b). The terminal aldehyde of PAR, which is required for condensation with histone proteins, is degraded when hydrolysis is performed with NaOH, but remains intact following CHES-hydrolysis.

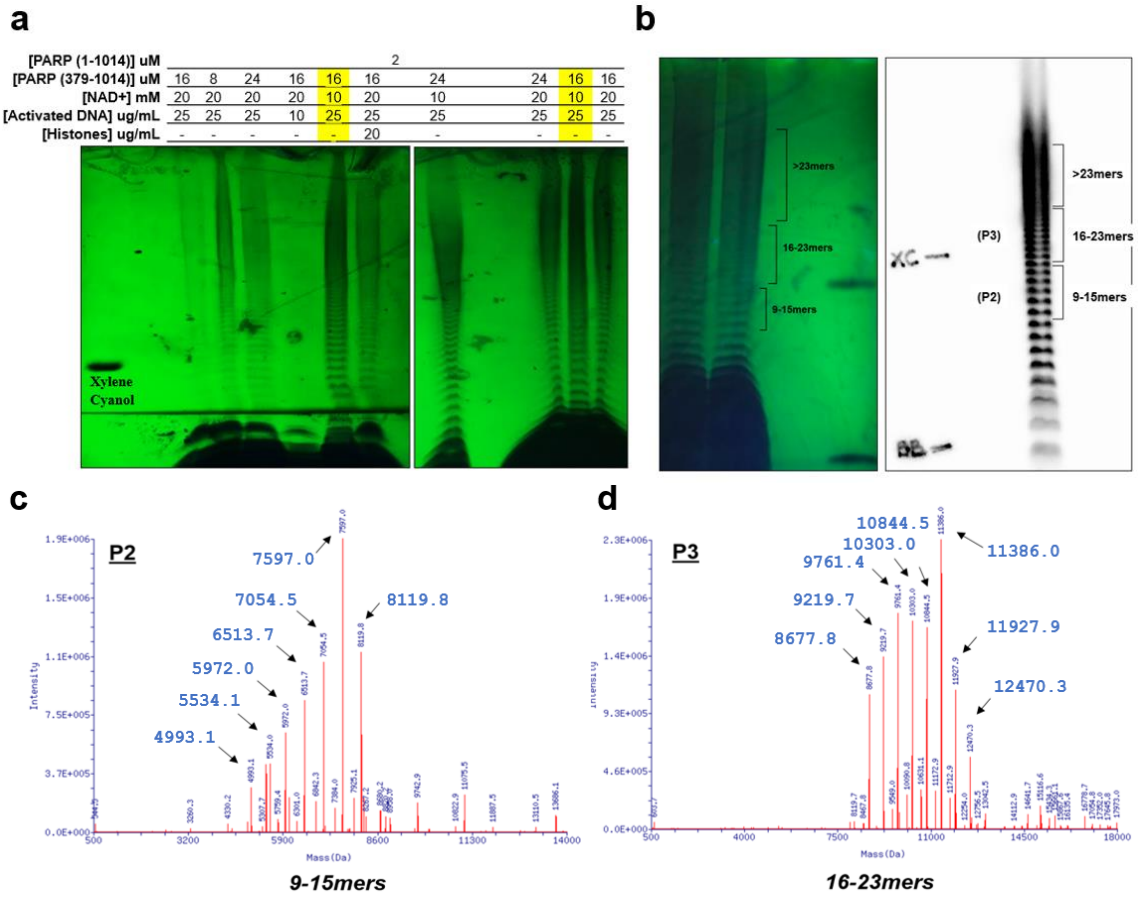


Figure E.2: PAR synthesis. (a) Reaction conditions were optimized to: 2uM PARP1 (1-1014), 16uM PARP1 (379-1014), 10mM NAD+, 25ug/mL DNA; 100mM Tris, 10mM $MgCl_2$, 1mM DTT. 20%, 79:1 Denaturing PAGE. (b) UV and radiograph images of PAR-purification gels with size ranges indicated. (c-d) Mass Spec Analysis of PAR (P2 & P3) and PARylated AO-601 primer. MS results for P2 PAR (length: 9-15mers) & P3 PAR (length: 16-23). Following *PAR synthesis reaction*s (see below), 400mM N-Cyclohexyl-2-aminoethanesulfonic acid (CHES, pH= 9.0) & 40mM EDTA were added to a final concentration of 100mM & 10mM respectively. Reactions were incubated at 37°C for two hours. After hydrolysis the pH was adjusted to 7-8 with 5M HCl and the reaction was washed with water in a 3kDa Amicon and concentrated to an appropriate volume to be combined with Formamide Loading Buffer. PAR chains were purified using 20% Denaturing PAGE (79:1). *PAR Synthesis Conditions*: 2uM PARP-1 (Full), 16uM PARP-1 (379-1014), 10mM NAD+, 25ug/mL Activated Calf Thymus DNA. 100mM Tris (pH 7.8), 10mM MgCl2, 1mM DTT. RT. 96 hours.

E.1.3 Covalent attachment of PAR to histones and DNA

To facilitate the synthetic attachment of PAR chains to histones and DNA a number of hetero linker strategies were considered, all of which exploited an aminooxy (AO) functionality for selectively targeting the terminal aldehyde of PAR. Current methods for *PAR-histone* conjugation employ a bi-functional linker having an AO and arylpropiolonitrile (APN= Thiol-reactive) interface. The goal was to, after AO-PAR condensation, react $H2A_{N110C}$ mutants with these thiol-reactive probes to generate covalent PAR-histone adducts. A few experiments suggest PAR may have been appended to $H2B_{E2C}$ via the aforementioned APN-AO linker however, aggregates appear during gel analysis in all experiments containing PAR and histones (w/& w/o linker) and attempts at removing the aggregates prior to gel loading show little sign of reducing aggregation (Figure E.3). I expect it will be best to perform the PAR-histone coupling after the nucleosome has been reconstituted.

In addition to introducing PAR into nucleosome(s) (arrays) through conjugation to histone tails, which would be the most biologically relevant attachment site, PAR-DNA attachment has also been explored as a means to prepare reconstituted systems containing PAR chains tethered near DNA lesions. We expect this will sufficiently replicate the influence of PAR proximity to lesion processing within a chromatin environment. PARP-1 has been reported to PARylate multiple DNA substrates *in vitro* and I have shown that PARP-1 can modify 5'-Phosphorylated overhangs and recessed ends to some extent (data not shown). Although there haven't been any reports of PARP-1 modifying DNA directly *in vivo*, PARylated DNA can still serve as a reasonable substrate for BER studies.

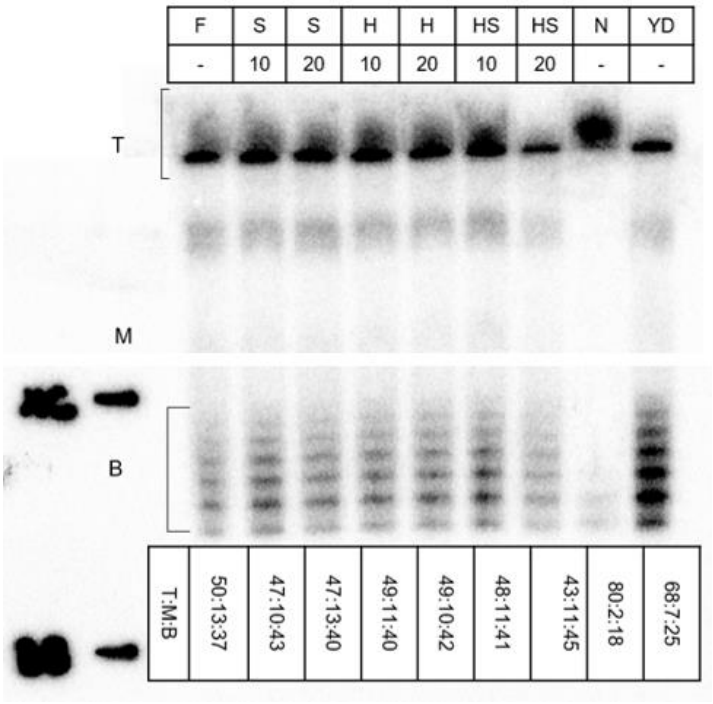


Figure E.3: PAR-histone coupling screening experiment. In effort to reduce PAR-Histone aggregation various methods were considered. PAR and $H2B_{E2C}$ were combined (PAR in excess) in 80mM PBS (pH= 7.4) and incubated for 3 hrs at RT, then subjected to the following prior to gel loading. F= Formamide Only, S= Spin @ 15k rpm, H= 37°C, N= Native LB, YD= Yellow Dot Filter, HS= Heat & Spin. None of the methods shown here were effective, the ratio of each band intensity is shown as Top:Middle:Bottom.

Currently, methods for conjugating PAR chains of heterogenous and defined lengths to short, AO-containing, oligonucleotides. These PAR-labelled DNAs will be used as primers to generate 601s and are compatible with our plug-and-play chromatin (see Chapter 2). Attempts at PARylating long DNA substrates via the AO-linker were unsuccessful (data not shown). However, smaller AO-DNAs were reactive towards purified PAR chains, and conditions have been optimized for AO-based PAR-DNA conjugations (Figure E.4) Alternatively, PAR chains can be directly reacted with the AO-containing molecule *in situ* by simply adding the aldehyde reactive compound to crude PARP-1 reaction mixtures (Figure E.5).

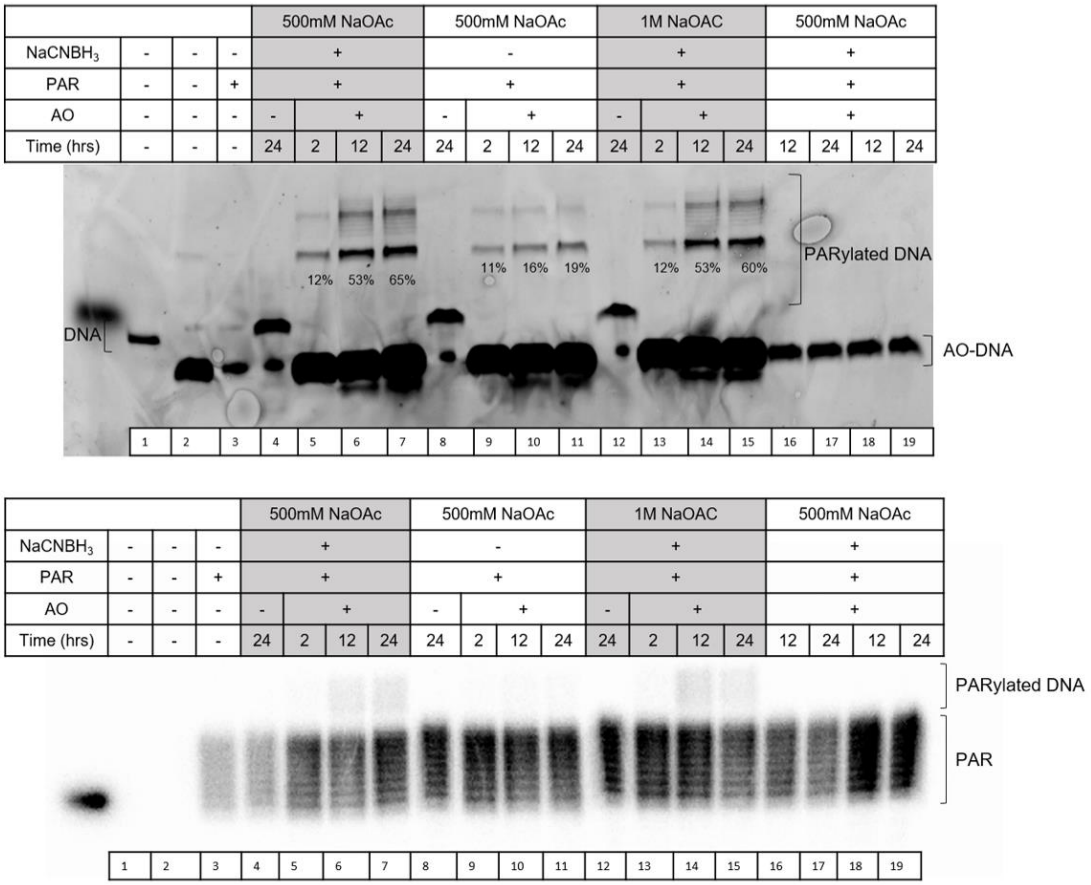


Figure E.4: Covalent modification of synthetic DNA oligonucleotides with pure PAR. PAR does not react with DNA lacking a 5'-AO moiety (lanes 4, 8 & 12). PAR+AO-DNA conjugation is most efficient at 500mM NaOAc (pH= 5.5), 10mM $NaCNBH_3$, 45°C, 24hrs (lane 7). A Radiograph from this experiment is shown below the EtBr stained image. $NaCNBH_3$ significantly improves yield (compare lanes 5-7 & 13-15 with lanes 9-11).

Unfortunately, degradation of PAR chains appeared to be taking place in most of these labelling reactions (Figure E.6). We expect this resulted from depurination of Adenosine during prolonged heating under acidic conditions. Although some promising data suggests PAR chains can survive thermal cycling and potentially be used in PCR reactions, alternative conditions will need to be optimized for the AO-PAR coupling (see discussion in *E.1.3*).

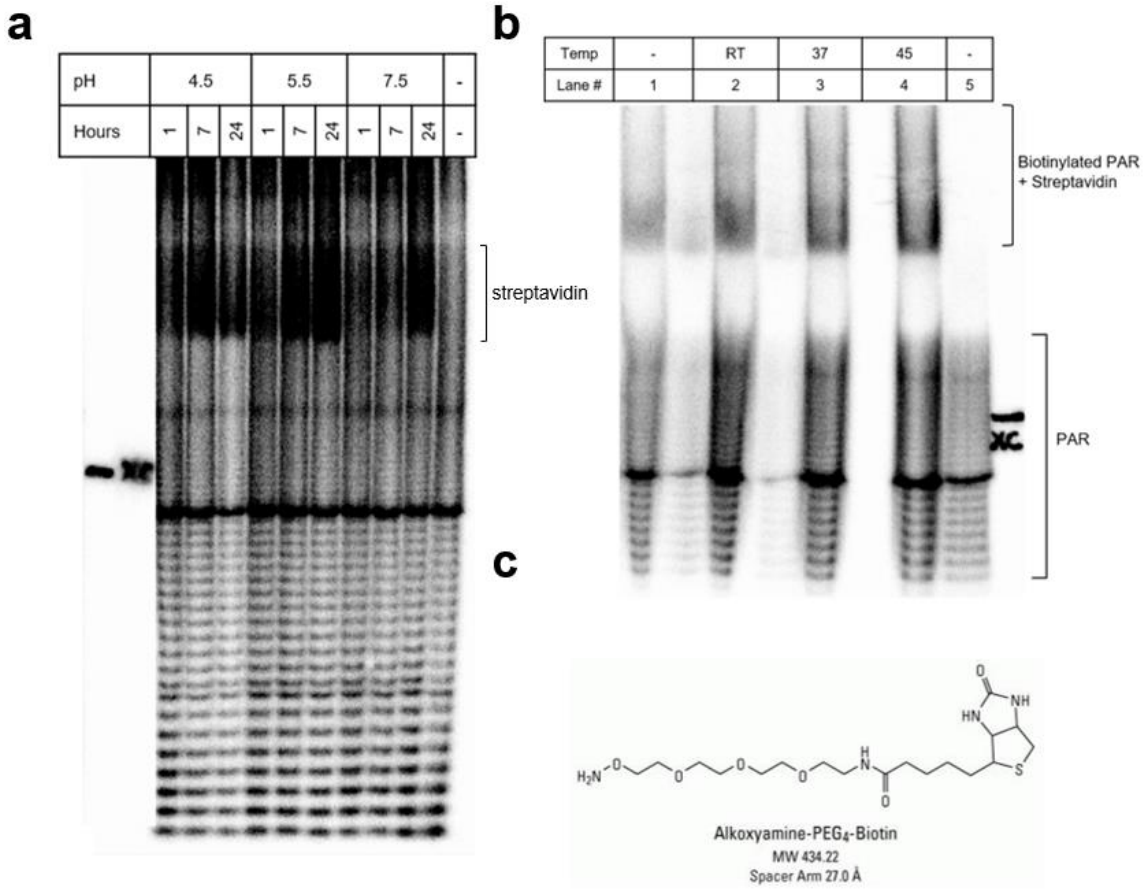
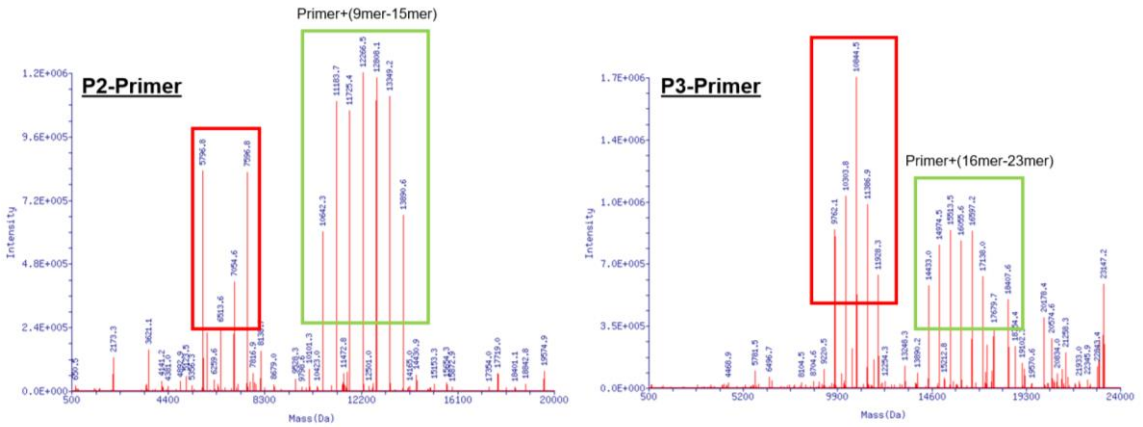


Figure E.5: Biotinylation of PAR *in situ.* (a) pH/time course experiment for PAR hydrolysis. Hydrolysis was conducted using the AO-Biotin reagent shown in (c) at the stated pH (37°C). The control lane (-) shows a ^{32}P-PAR reaction hydrolyzed using CHES conditions (AO-Biotin Absent). All reactions were incubated with Streptavidin and AO-PAR coupling efficiency as determined based on the intensity of radiation shifted to the streptavidin-labelled MW frame. pH 5.5, 24 hrs resulted in the highest yield (38%). (b) Temperature experiment for PAR hydrolysis using the conditions optimized in part (a); conditions listed below. (c) Structure of AO-biotin used in the above experiment (purchased from Thermo Fisher). 500mM NaOAc (pH= 5.4), 10-fold excess AO-linker, 45°C, 24 hrs. Biotinylated PAR was precipitated directly from the reaction by adding 3 volumes 100% EtOH. The pellets were resuspended directly in a Streptavidin solution containing 1 ug/uL Strep, 1M NaCl, 50mM Tris (pH= 7.4) and then loaded after incubation at 37 ˚C for 20 minutes.



	Mass
Aminooxy Primer	5786 Da

Figure E.6: ESI MS analysis of an AO-primer after coupling to the PAR of defined length outlined in figure E.1. The green boxed peaks represent

E.1.4 Future Directions and outlook

Overall, my work on this project established conditions for synthesizing and purifying PAR chains of defined length. The most effective method for targeting the aldehyde moiety of PAR is the AO functionality and, although degradation was observed under the experimental conditions described here, I expect performing the reaction at neutral pH in the presence of elevated NaCl could prevent degradation while still yielding PAR-AO conjugates. An important consideration for future researchers is the importance of maintaining the integrity of the terminal aldehyde. I expect during the purification process, and perhaps other unnoticed steps, Tris buffer was introduced to the PAR mixtures and could potentially have ‘capped’ the 3’-aldehyde which hampered my yields throughout this project.

F) APPENDIX

IN VITRO ASSEMBLY OF LNCRNA-CHROMATIN COMPLEXES

F.1 Supplementary Text

The lncRNA HOTAIR, like many other functional lncRNAs, is known to associate with chromatin remodeling complexes and is also expected to directly interact with chromatin itself. In general, lncRNAs regulate gene expression by either directly/indirectly influencing chromatin structure or recruiting proteins. In the case of HOTAIR, interactions with notable gene repressors: PRC2 and the lysine specific demethylase 1 (LSD1) proteins, result in recruitment of these factors to HOTAIR target genes and the establishment of a transcriptionally silent state. Several mechanisms have been proposed to explain the ability of lncRNAs like HOTAIR to reliably direct chromatin regulation, these include: RNA:DNA triplexes, R-loops, RNA-dependent ternary complexes, among others. The goal of this project is to study the interactions of HOTAIR, and other lncRNAs of interest, with chromatin and determine how chromatin structure is impacted by having large RNA transcripts bound in proximity nucleosome arrays. To facilitate the investigations outlined above, the initial goal of this project was to develop a platform for tethering HOTAIR to our laboratory's synthetic chromatin. To this end two approaches were pursued. Method (*1*) involves Cas9-mediated targeting of a HOTAIR-sgRNA fusion construct to our chromatin and method (*2*) involves direct hybridization of HOTAIR to our nucleosome arrays via a DNA linker.

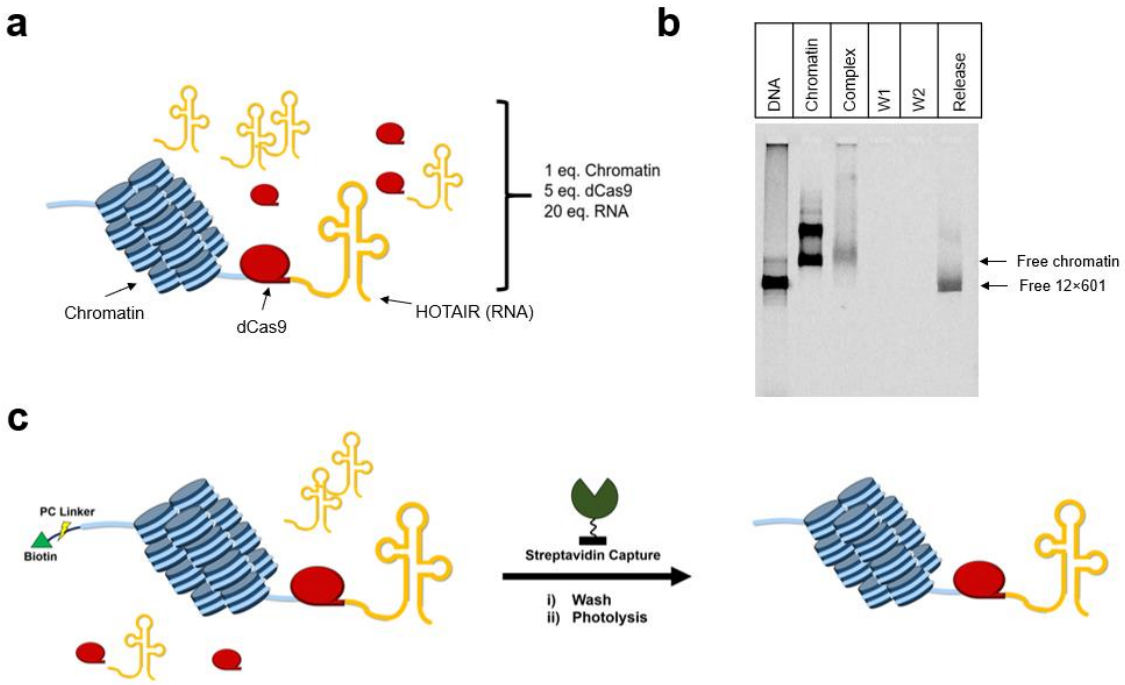


Figure F.1: HOTAIR-chromatin complexes are achieved with excess dCas9 and RNA, this excess material is removed using the following method: the chromatin is modified by the "plug and play" strategy to contain a 5'-biotin, attached to the chromatin via a photocleavable (PC) linker. Once HOTAIR-dCas9-chromatin complexes are achieved, the complex is retrieved using streptavidin coated beads and unbound RNA/dCas9 is washed away. Photolysis of the PC linker releases the desired complex.

(*1*) A catalytically dead Cas9 variant (dCas9) was used to guide HOTAIR to chromatin (Figure F.1). Dr. Yu Zeng examined various configurations of HOTAIR-sgRNA fusion constructs to identify the construct that was most efficiently capable of hybridizing to a complimentary sequence appended to the 601 DNA. To facilitate translation of this method to our synthetic chromatin platform I prepared a 12×601 sequence having the dCas9-sgRNA-HOTAIR recognition sequence located at the 3'-end of the array (Figure F.1). For ternary complexes with chromatin to form, 20-fold excess RNA and 5 fold excess dCas9 is required therefore, this crude mixture is unsuitable for

the desired biochemical analyses. To facilitate purification of the desired complex a plug-and-play based insertion site was introduced to the EcoRI end of the 12×601. This method effectively enables isolation of free 12×601 DNA following streptavidin pull-down in preliminary experiments, however, attempts at using this strategy on intact RNA-chromatin complex have given inconclusive results. One problem that might have to be overcome in using this method is the potential aggregation and disassembly of chromatin fibers when precipitated by streptavidin.

(*2*) The second method being pursued, which involves direct hybridization of DNA linkers to HOTAIR and chromatin offers two potential platforms for investigating the effect of RNA proximity on chromatin structure. These include either direct attachment via a two-sided linker (Figure F.2a) or immobilization of HOTAIR and chromatin on the same streptavidin molecule via biotinylated DNA linkers (Figure F.3a). Preliminary findings indicated that hybridization of the DNA linker to native HOTAIR is not effective. Under native conditions the DNA linker fails to hybridize effectively (~10%), indicating secondary structure within the 3'-end of HOTAIR could be precluding duplex formation, whereas when the mixture denatured and annealed, even though the hybridization efficiency increased 4-fold (~40%), a substantial amount of aggregation occurs. To overcome these two problems a new, extended, HOTAIR construct was prepared which consisted of domains 1-2 and an additional single stranded extension that was complimentary to our DNA linker (**HA-Ext**). Indeed, this **HA-Ext** RNA was much more reactive to DNA probes than wild-type HOTAIR (Figure F.2b).

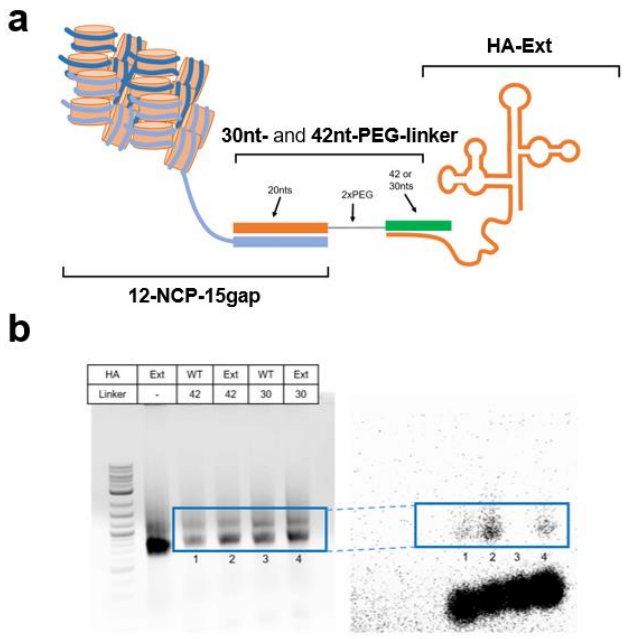


Figure F.2: Tethering of HOTAIR to chromatin via a DNA linker. (a) Schematic depicting method for attaching **HA-Ext** to **12-NCP-15gap** chromatin. (b) Preliminary data indicated the most efficient probe for hybridizing to **HA-Ext** is the **42nt-PEG-linker**. *Left*- EtBr stained Agarose (0.5%) shows HOTAIR extended RNA (bottom band) and the DNA template (top strand, DNase treatment not conducted)/ *Right*- autoadiograph of the same gel demonstrating **30nt-PEG-linker** does not hybridize with WT HOTAIR (lane 3) but does with the **HA-Ext** (lane 4). **42nt-PEG-linker** hybridizes to some extent with the WT HOTAIR (designed to be complimentary to terminal 12-nts of WT which are artifacts of the primer sequence) but still hybridizes to a greater extent with **HA-Ext**. Hybridization efficiencies: Lane 1: 4.5%, Lane 2: 36%, Lane 3: 0%, Lane 4: 3.5%.

(*2 continued*) In conjunction with **HA-Ext** development, a new 12×601 array was engineered (**12-601-15gap**) having an excisable fragment located at the 3'-end. When subjected to the strand exchange reaction described in Chapter 2, single stranded oligonucleotides having a **HA-Ext** complimentary linker or a 3'-biotin can be incorporated on the end of our chromatin template (Figure F.2). The DNA linkers that were synthesized for these purposes have been successfully hybridized to **HA-Ext** (Figure F.2b), with the 42-nucleotide linker (**42nt-PEG-linker** and **42nt-PEG-biotin**) working the best (>30%). This percentage most likely under-represents the true ratio of hybridized

probe because contamination from the **HA-Ext** DNA template. The ratio of DNA:RNA based on EtBr gel (Figure F.2b) is equal to 44:56, suggesting the efficiency of **42nt-PEG-linker** binding to HA-Ext is actually over 50%. In the future transcription reactions should be treated with DNase. Nevertheless, this method has facilitated HA-Ext could be tethered to a free 601 DNA (Figure F.4). Additionally, through the use of two separate 3'- biotin hybridization probes, both the **12-601-15gap** DNA and **HA-Ext** have been successfully immobilized. Almost 100% of the DNA and ≥85% of the RNA were able to be pulled out of solution using this method (data not shown).

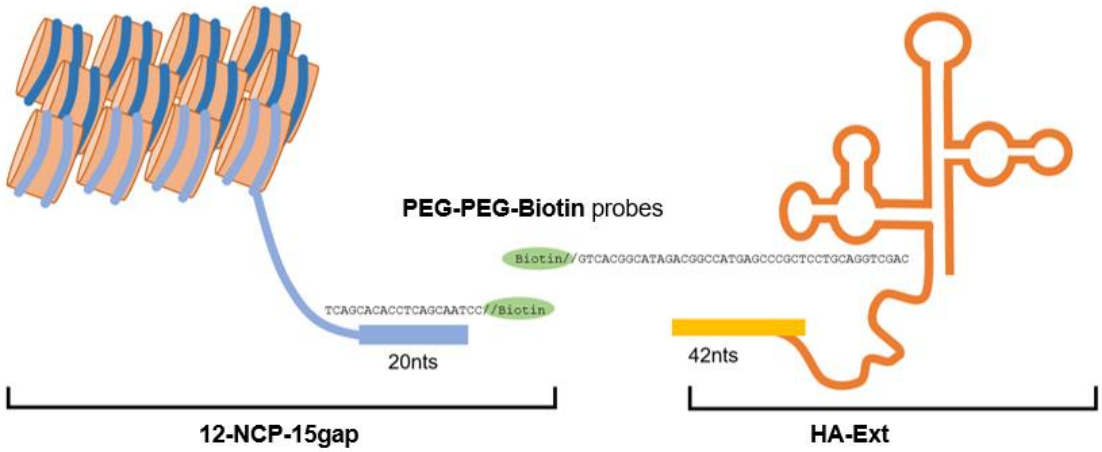


Figure F.3: Strategy for biotinylating **12-NCP-15gap** and **HA-Ext**. Once modified with a biotin, chromatin and RNA molecules can be immobilized on the same streptavidin molecule for study.

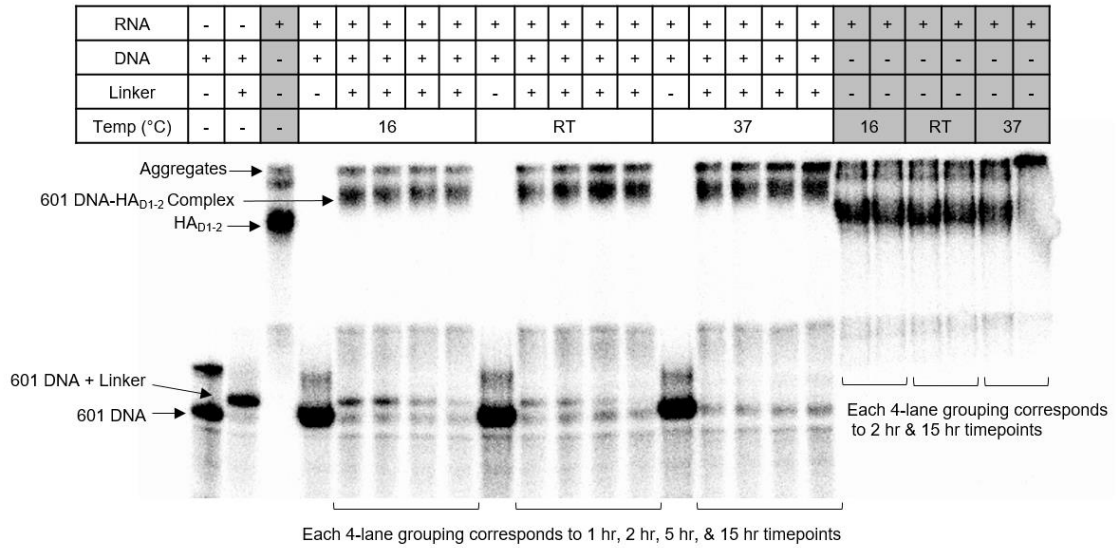


Figure F.4: Single 601 DNAs (**1-601-15gap**) were tethered to domains 1-2 of HOTAIR (**HA-Ext**) using a hybridizable DNA linker (**42nt-PEG-linker**). The linkers were first annealed with **1-601-15gap** via strand-exchange and then incubated with excess RNA for 15 hours at the indicated temperature. Timepoints were taken after 1, 2 and 5 hours. Lanes shaded grey indicate controls with ^{32}P-labelled **HA-Ext**; the DNA is the source of the radiation in all other lanes. The DNA is almost completely shifted (bound to HOTAIR) after 5 hours at 16°C, 2 hours at RT, and 1 hour at 37°C. Prolonged incubation at higher temperatures leads to increased aggregate formation. (5% Native PAGE, 0.5x TBE).